Lecture Notes in Energy 54

Sofia Ramos
Helena Veiga *Editors*

The Interrelationship Between Financial and Energy Markets

Springer

Lecture Notes in Energy

Volume 54

Lecture Notes in Energy (LNE) is a series that reports on new developments in the study of energy: from science and engineering to the analysis of energy policy. The series' scope includes but is not limited to, renewable and green energy, nuclear, fossil fuels and carbon capture, energy systems, energy storage and harvesting, batteries and fuel cells, power systems, energy efficiency, energy in buildings, energy policy, as well as energy-related topics in economics, management and transportation. Books published in LNE are original and timely and bridge between advanced textbooks and the forefront of research. Readers of LNE include postgraduate students and non-specialist researchers wishing to gain an accessible introduction to a field of research as well as professionals and researchers with a need for an up-to-date reference book on a well-defined topic. The series publishes single and multi-authored volumes as well as advanced textbooks.

More information about this series at http://www.springer.com/series/8874

Sofia Ramos · Helena Veiga
Editors

The Interrelationship Between Financial and Energy Markets

Springer

Editors
Sofia Ramos
Department of Finance
Instituto Universitário de Lisboa
(ISCTE-IUL)
Lisboa
Portugal

Helena Veiga
Department of Statistics
Universidad Carlos III de Madrid
Getafe
Spain

ISSN 2195-1284 ISSN 2195-1292 (electronic)
ISBN 978-3-642-55381-3 ISBN 978-3-642-55382-0 (eBook)
DOI 10.1007/978-3-642-55382-0

Library of Congress Control Number: 2014945256

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

In the last decade, energy markets have developed substantially due to the growing activity of financial investors. One consequence of this massive presence of investors is a stronger link between the hitherto segmented energy and financial markets. "The Interrelationship Between Financial and Energy Markets" is the title of this book, and it addresses some of the recent developments between financial and energy markets. It aims to further the understanding of the rich interplay between financial and energy markets by presenting several empirical studies that illustrate and discuss some of the main issues on this agenda.

Postgraduate students, researchers, and practitioners with a solid background in economic and finance theory are the target audience. Many chapters contain a strong component of quantitative methods applied to energy finance along with an up-to-date survey of the literature, thus allowing the reader to get up to speed on these topics.

A number of issues were omitted, including the regulatory aspects of the European energy markets and financial aspects on renewable and green energy, so as to avoid an overlap with the contents of other books by Springer, e.g. *Financial Aspects of Energy* and the *Handbook of Natural Resources and Energy Economics.*

As a whole, the 12 chapters of "The Interrelationship Between Financial and Energy Markets" aim to provide an overview of important aspects of the oil industry, the impact of oil shocks, electricity markets, and the analytical and quantitative tools applicable to energy finance.

This book is the result of input from many people. In particular, we would like to thank the authors of the chapters, as well as the reviewers for their helpful comments.

Acknowledgments

To our families

We are also grateful to the staff of Springer.

Last but not least we are grateful to our families and friends. Sofia is grateful to the support of Francisco, Miguel and Paulo, but also her parents António and Gracinda. Sofia is also grateful to the friendship of Helena and José Manuel in all these years of work together.

Helena is grateful to her husband, Marc Vorsatz, without him these last few years would be more difficult to go. Helena is also grateful to her parents, José and Conceição, for their unconditional support and to the friendship and enthusiasm of Sofia that makes the research together a pleasure.

Contents

Part III Electricity Markets

Introduction

Before the early 2000s, commodities and energy markets were partially segmented from outside financial markets and even from each other. During the 2000s, a series of academic articles publicized negative correlations between commodities and stock market and bond indexes, and the positive correlation with inflation. In investment language this means that investments in commodities, and in particular energy, provide diversification benefits as well as inflation hedging. These attractive features captured the attention of financial investors as they were coming to terms with the burst of the dot com bubble. The inflation hedging property was also welcomed by long-term investors for whom the inflation risk is a major concern.

World dependence on energy has grown steadily in the last decades. Figure 1 depicts the world consumption of primary energy and the striking increase is evident. Between 1965 and 2012, the consumption tripled. The steep upward trend of consumption in China and other emerging countries is also clear to see. Overall, the worldwide consumption of energy is dominated by China, the US, and the European Union, and together they accounted for 53 % of the world primary energy consumption in 2012.

According to the Energy Information Administration's International Energy Outlook 2013, world energy consumption will rise to 56 % in the next three decades driven by growth in developing countries such as China and India. Their projections also indicate that demand in China is expected to double that of the US by 2040. Therefore, there is a huge potential for growth in the energy markets.

The deregulation of the energy sector is another factor that has further boosted the links between markets. The 1980s marked the start of a wave of deregulation of several state-owned industries, including electricity, in many countries like the US and the UK. The deregulation process in the energy sector aimed to create a new institutional framework that benefited consumers and fostered welfare. The regulations were changed in order to attract a larger number of market participants to competitive sectors given the well-known difficulty in challenging incumbents. Prices were expected to fluctuate more as they were no longer being curbed, and this brought new risks for market participants. In addition, the Commodity Futures

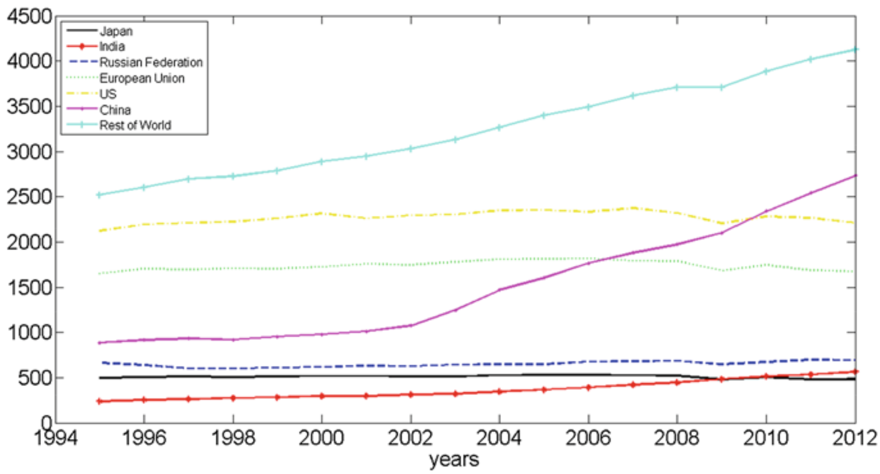


Fig. 1 World consumption of energy. *Data source* BP statistics. Primary energy comprises commercially traded fuels including modern renewables used to generate electricity

Modernization Act of 2000 ensured the deregulation of financial products known as over-the-counter derivatives on commodities.

The necessary conditions were in place: a sector with overwhelming growth (that is likely to continue) and many investment opportunities, and investors eager to find new asset classes and who have acknowledged the energy sector's growth potential. The trigger came with the wave of deregulation in the energy markets, which attracted a large number of new investors and strengthened the link between financial and energy markets.

This increased link between financial markets and energy has taken several forms:

The development of energy markets with a growing number of participants, and energy-based products are now being traded like financial assets. New spot and derivative energy markets have emerged and financial exchanges have extended business lines to energy products. Figure 2 shows the average open interest per year of futures contracts traded on the NYMEX since 1995 on crude oil and natural gas. The growing number of open interest in futures contracts clearly demonstrates the financial activity in the energy sector.

The new framework has some interesting implications. First, the interaction between supply and demand becomes more important for pricing. Second, markets share common investors, which leads to more commonalities between oil and stock markets; for instance, spillovers from one market to another are likely to become acute. Third, given that demand for energy is quite inelastic and supply tends to be rigid, price volatility is likely to increase, or prices will at least show spikes. Figures 3, 4 and 5 show the price of crude oil, natural gas and electricity; the three lines represent the maximum, average, and minimum price by year and they clearly show that the price spikes.

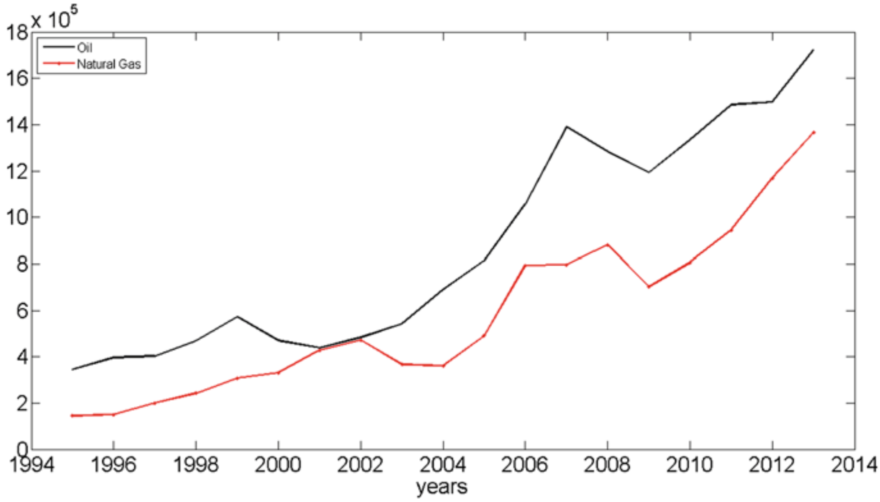


Fig. 2 Open interest: Yearly average-futures contract on the NYMEX. *Data source* Commitments of traders data from the US Commodity Futures Trading Commission

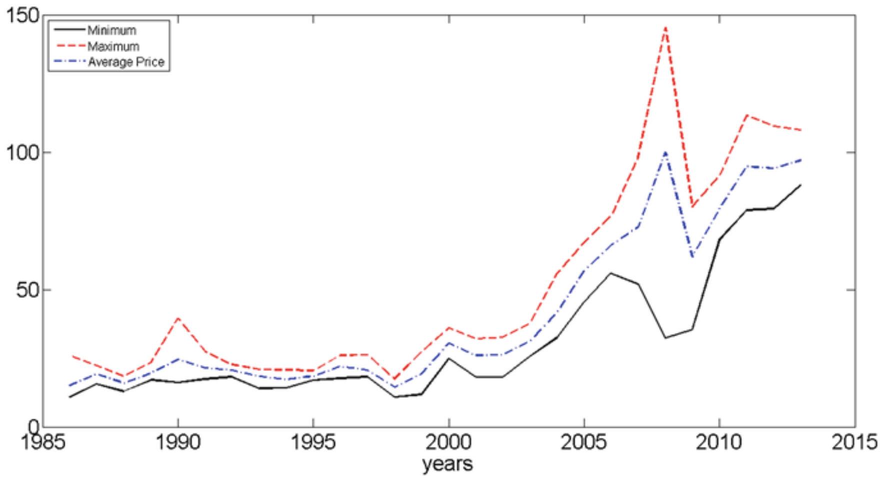


Fig. 3 Price of West Texas intermediate crude oil- ($/barrel). *Data source* Datastream

Funding of Energy Projects

Soaring demand pushed prices to new highs, triggering the development of new energy projects that had not previously been financially worthwhile. Figure 3 shows the sharp rise in the price of crude oil, interrupted briefly by the 2008 crisis before bouncing back. A side effect of the high cost of oil is that it stimulated

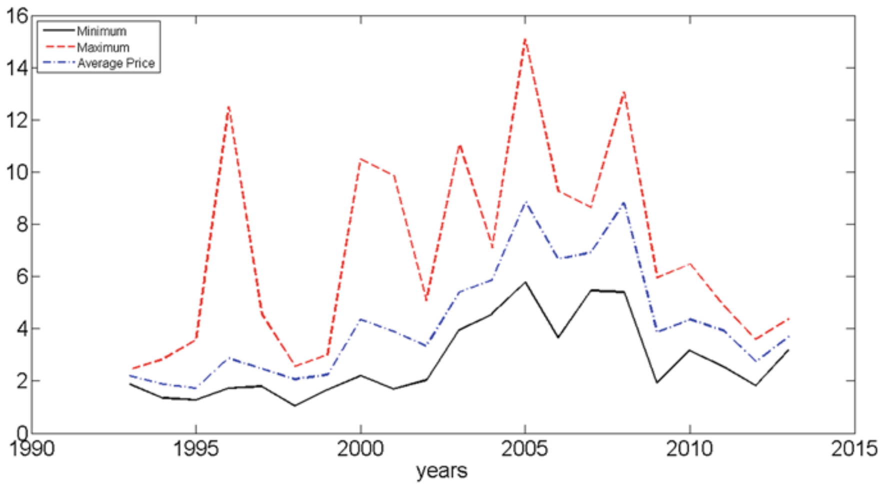


Fig. 4 Price of natural gas-Henry hub ($/MMBTU). *Data source* Datastream

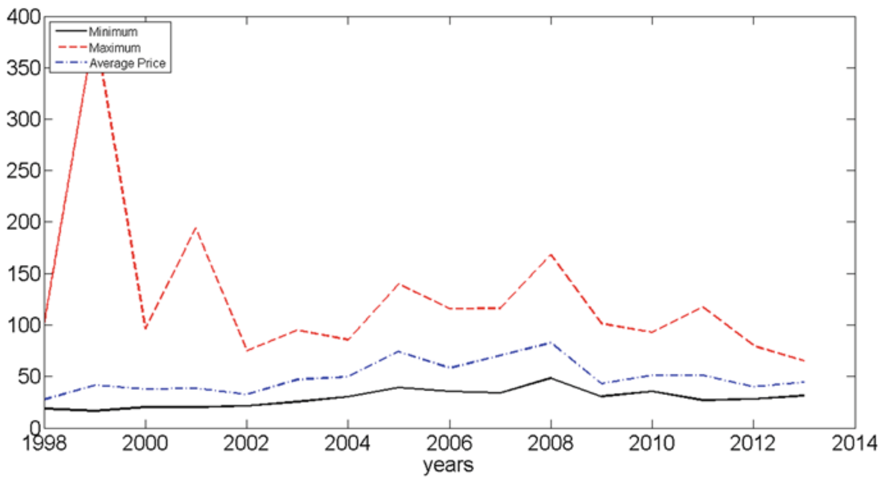


Fig. 5 DJPJM electricity—firm on PK—price index. *Data source* Datastream

investment in the production of "more expensive" oil like bearing shale formations, as well as of other sources of renewable energy.

The energy sector is characterized by large upfront investments in both research and technology (R&D) and also in infrastructures. Research and technology in technology is important because it enables technological advances that raise efficiency. For instance, technological developments in the electricity sector, e.g., in smart grids, storage capacity, and the integration of national or intra-national

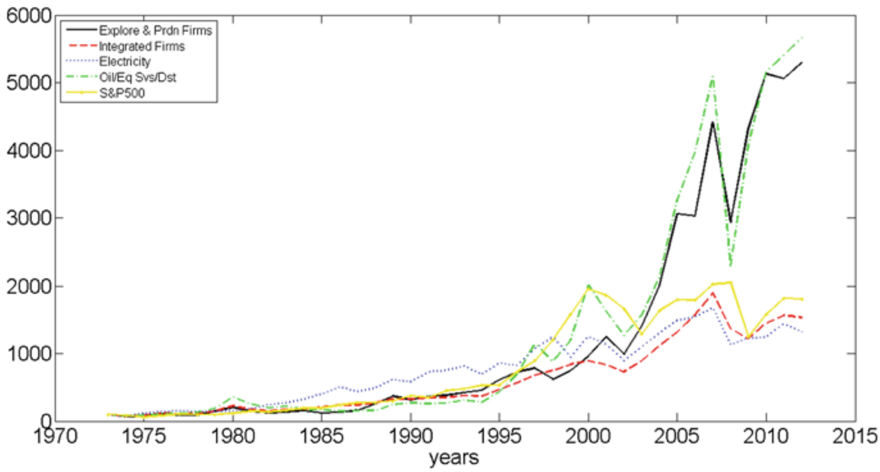


Fig. 6 US firms' market capitalization. *Data source* Datastream. All data series are standardized to the value 100 in 1973 to better allow comparisons

systems will increase efficiency and lead to substantial savings in costs. In the same way, technological development results in more efficiency due to cuts in costs when exploring renewable sources.

As a result of the growing consumption of energy, energy firms felt the need to make new investments, and these require funding. Huge amounts of money have been put into energy companies and financing energy projects. For instance, projects to increase the transmission capacity of electricity in developing markets due to industrialization and the growth in urbanization, and to integrate the different national electric grids.

Figure 6 shows the market value of indexes of U.S. oil and gas companies that go from exploration and production, integrated firms, to electricity and oil equipment and services and distribution. We also present the index Standard and Poor's 500 for comparison. All indexes are normalized to 100 for the first year to facilitate comparison. We can see the marked growth of exploration and production companies as well as oil equipment and services and distribution in relation to those of S&P 500.

Idiosyncrasies of Energy Markets

Despite greater integration with financial markets, the specificities of the energy markets are such that it is natural to analyze them separately from other commodities. An important distinctive feature of energy is that it is costly to transport and difficult to store (the extreme case is electricity), leading to rigidity on the supply side.

The inelasticity of demand and the rigid supply mean that any change in these two economic forces has a dramatic effect on prices, making energy prices quite volatile. Thus, modeling and forecasting prices and their volatility is important, as is an understanding of the impact of energy volatility on economic and financial variables such as stock market returns and Gross Domestic Product.

The Structure of the Book in Parts and Chapters

This book is divided into three parts. The first deals with the impact of oil price risk in firms, the second with the impact of oil shocks, and the third discusses electricity market issues. The recurrent theme of the book is the financial implications of energy prices and the importance of price risk from the perspective of both investors and firm managers. Special attention is given to econometric techniques that shed light on the dynamics of prices whether from spot or derivative markets, and forecasting the price of energy. The development of electricity markets is no less important, namely the design of markets, pricing, and the interaction of spot and derivatives prices.

Part I, devoted to the Oil Industry, consists of two chapters. The first section looks at the relation between the price of oil and the market value of companies. Oil has a central role in economies since it is a major input for most industries. The chapters look at the two groups of companies that are most directly dependent on oil: oil companies and airline companies.

Stockholders of energy companies need to know the risks facing these companies. The first chapter by Sofia Ramos, Helena Veiga, and Chi-Wei Wang, "Risk Factors in the Oil Industry: An Upstream and Downstream Analysis", looks at the drivers of the market value of oil firms upstream and downstream. Although there is of course a relation between the price of oil and the market value of the listed oil companies, oil is simultaneously the main output of the upstream industry and an input for the downstream segment. Thus, the chapter analyzes whether the impact is the same along the value chain of the oil industry. The results highlight the weak relationship with stock market returns, which means the industry is not pro-cyclical either upstream or downstream. However, complementary industry activities related with oil exploration are pro-cyclical industries. Second, the market returns of the industry show sensitivity to oil returns, and range from 1.45 % (downstream) to 4.19 % (machinery and equipment). This suggests that when the price of oil increases, this is likely to trigger higher oil production, and boost business activity in complementary industry segments. The oil industry is also shown to have some market power.

Oil price risk is a major issue in commercial airlines. The second chapter by Paul A. Laux, He Yan, and Chi Zhang, "Cost, Risk-Taking, and Value in the Airline Industry", analyzes the cost function of the commercial airline industry and analyzes whether it is worth hedging the oil price. The study develops empirical measurements of the cost functions of airline firms during the 1998–2009 period, using detailed data from jet fuel consumption and prices. The authors go on to analyze the potential value of hedging costs, and emphasize that the key to valuable risk management is the correlation between the risk source and the firm's investment range of opportunities. A positive correlation between them implies that hedging can be beneficial. The study finds that although airlines do hedge significantly, it is neither universal within the industry nor do any firms hedge fully. Further, the intensity of hedging varies substantially over time for many airlines. The authors find some explanation for this; the results show that unhedged fuel cost functions are concave on oil prices, which implies that not hedging the risk is beneficial to the value of airlines. Specifically, on average fuel costs tend toward concavity, suggesting that cost savings when oil prices drop exceed cost increases when oil prices spike. Thus airlines' cost structures are such that the gain from hedging is limited.

As oil is a critical input for almost all economic activities, oil price shocks slow economic activity. Part II, which consists of five chapters, describes several aspects related with The Impact of Oil Shocks. This is one of the most active areas of research in recent years, so here we find very heterogeneous chapters going from surveys on the economic effects of oil market developments to more technical chapters that evaluate econometric models and techniques for oil price forecasting and transmission of oil shocks and volatility to equity markets.

Chapter 3 by Ulrich Oberndorfer, "Oil Prices, Volatility, and Shocks: A Survey", summarizes the literature on the economic effects of oil market developments and analyzes the link between the oil market and economic outcomes. In particular, the chapter clarifies common definitions in oil markets such as price shock and price volatility measures, the theoretical and empirical findings on the impacts of oil on the macroeconomy, and the role played by the oil price in financial markets. The effects of oil on the economy are described mainly through two traditional mechanisms: the supply side and the income transfer mechanisms. In the first, the increase in oil price leads to a scarcity of this basic input, thus causing a decrease in output, real wage growth, and a rise in unemployment. The decrease in real wages also has additional negative effects like the increase in borrowing, interest rates, and inflation. The latter mechanism explains the effects of an increase in oil prices as a shift in purchasing power from oil-importing countries to oil-exporting countries.

Regarding the impact of oil price rises on financial markets, firms in the energy business often profit from increases in oil price because their profits and consequently their stock prices go up. The opposite occurs for firms in which the main input is oil. Nevertheless, it is oil price volatility that registers the biggest impact on financial markets due to the fact that it triggers higher expenditure and therefore induces hedging costs, decreases the production of the respective commodity, and impacts on the discounted expected future cash flows of firms, thus affecting their stock prices.

In the last decade, the literature on the nature of oil price shocks and their effects on equity markets has led to both greater understanding and new techniques to quantify these effects. Chapter 4 by Rania Jammazi, "Oil Shock Transmission to Stock Market Returns: Wavelet-Multivariate Markov Switching GARCH

Approach" addresses this topic and focuses on the transmission of oil shocks to stock markets using recent methodological developments that involve estimating a multivariate Markov switching GARCH to series of wavelet detail coefficients of stock market and oil returns. This methodology allows us to analyze the magnitude and the time-varying nature of the transmission. More specifically, it addresses two issues on the effect of crude oil shocks on stock market returns: the existence (or the inexistence) of oil shocks and/or oil volatility transmission to equity markets and, under the hypothesis of common increased volatility, whether this transmission is boosted with the international crisis. The main findings of this chapter are: the intensity of oil shocks is time-varying, there is a correlation between high oil price volatility and stock returns volatility, and international recessions and the responses of stock markets to the oil supply shocks vary in accordance with the geographical area. Therefore, for oil-importing countries, oil price shocks coming from non-European countries have a stronger effect than oil price shocks coming from European or Eurasian countries.

Oil price volatility has a detrimental effect on some important economic sectors such as automobile, chemical, oil and natural gas, and utilities. Some companies buy credit default swaps (CDS), namely financial products that provide protection from credit events, to insulate themselves from events that increase oil price volatility. Chapter 5 by Shawkat Hammoudeh and Ramazan Sari, "Forcing Variables in the Dynamics of Risk Spillovers in Oil-Related CDS Sectors, Equity, Bond and Oil Markets and Volatility Market Risks", examines migration, i.e., the deterioration in credit quality, and cascading of CDS risks, i.e., the spillover effects for the four oil related sectors.

The commodity market has seen an upward trend in financialization, i.e., a growing presence of financial investors without a commercial position on the commodity. There has been heated debate as to whether the presence of such investors, commonly dubbed financial speculators, disturbs and manipulates the price of commodities, including oil. Due to the physical limitations of investing in oil, the futures market has been the preferred venue for these investments. Giulio Cifarelli and Giovanna Paladino analyze the dynamic relation between spot and futures prices in Chapter 6 "Oil Futures Markets: A Dynamic Model of Hedging and Speculation". The authors present a model where spot and futures prices interact due to the presence of hedgers and speculators. The authors distinguish the interaction in two volatility regimes; they find evidence of the different behavior of hedgers and speculators and in particular that hedgers change their attitude in periods of high volatility when uncertainty is high. Overall, the authors conclude that the interaction between hedgers and speculators across volatility regimes has some impact on futures pricing.

Chapter 7 by Andrea Bastianin, Matteo Manera, Anil Markandya and Elisa Scarpa, "Evaluating the Empirical Performance of Alternative Econometric Models for Oil Price Forecasting", reviews the empirical literature on the alternative econometric specifications to capture the dynamics of oil prices. The empirical literature is far from reaching consensus about the appropriate model to be used in forecasting since conclusions change in line with the specifications, time periods, and frequencies. This is a very useful chapter for researchers or postdoc students since it provides an overview of the topic by exhaustively testing and evaluating the forecasting performance of several econometric specifications found in the literature: some related with the relationship between spot and futures prices that the authors call "financial" models and others involving economic fundamentals known as "structural" models, and a new class of models that combines "financial" and "structural" models. The forecasting performance of the various models is done for different frequencies of the data and the main finding is that forecasting in "financial" models and time series models performs better than "structural" and "mixed" models.

The last part of the book addresses the financial issues in Electricity Markets such as market design and price dynamics.

Determining the social cost of carbon emissions and incorporating the cost of polluting activities is an important economic issue. Over the last decade, a European carbon market has been established where the price of carbon emission allowances is determined by balancing supply and demand and these are traded in exchanges across Europe. Chapter 8 "Commodity Price Interaction: Co_2 Allowances, Fuel Sources and Electricity", by Mara Madaleno, Carlos Pinho and Cláudia Ribeiro, examines interactions between carbon, electricity, and fossil fuel (coal, oil and natural gas) returns, analyzing the impacts of emissions trading using data from the German and French markets. Results reveal that the price does not depend entirely on the energy mix of the country under analysis. Market power influences the correct transfer of prices, i.e., limiting cost increases but less carbon coercion in the European Energy Exchange (EEX) and innovations in carbon are not strongly reflected in electricity prices.

Although the electricity sector was state-regulated in many countries, there has been a wave of deregulation since the 1980s notably in developed countries. Chapter 9 "An Overview of Electricity Price Regimes in the U.S. Wholesale Markets", by José G. Dias and Sofia Ramos—describes the case of the deregulation of the United States wholesale electricity markets and studies the price dynamics in several regional markets. The case for deregulation was supported by the arguments that users would benefit from more competition between market participants, i.e., the classical economic argument that the larger the competition, the lower the prices would apply to electricity markets. However, to establish a competitive wholesale electricity market is a challenge for many reasons: the power of incumbent firms and the large upfront investments necessary to enter the market. Moreover, the inherent features of the electricity business hinder competition; first, although electricity can be generated by multiple operators, transmission tends to be a natural monopoly because duplicating the grid is not efficient. The dynamics of electricity prices are therefore a consequence of the regulation of the market. As prices are no longer curbed by regulators, and fluctuate according to supply and demand, prices spike with demand shocks because the supply is rigid. The chapter shows that the use of multi-regime switching models is suitable for representing the price dynamics of electricity prices because it can capture the various volatility regimes that characterize electricity in an endogenous way. Moreover, it compares the price dynamics in the different geographical markets of the US. The study shows that despite geographic distance, there is slight synchronization of regimes in a high and a low volatility regime.

The liberalization of the power markets in Europe and worldwide led to markets for the purchase of electrical energy that functioned well in many European countries. Markets in Europe work with a day-ahead spot market and financial contracts for future delivery of electricity. Chapter 10 by Fred Espen Benth and Maren Diane Schmeck, "Pricing Futures and Options in Electricity Markets", studies the relation between spot, futures, and options in electricity markets by pricing them simultaneously. The authors find evidence that supports the existence of different risk-neutral pricing measures: one for options and another for deriving futures prices from the spot dynamics.

Chapter 11 "Switching from Feed-in Tariffs to a Tradable Green Certificate Market" by Aitor Ciarreta, Maria Paz Espinosa, and Cristina Pizarro-Irizar discusses the creation of tradable green certificate markets focusing on the role of risk-sharing in markets. Electricity generated from renewable sources leads to high production costs, which often makes this energy unprofitable in a free market framework. Nevertheless, green energy brings benefits to the economy and society since it not only seeks to improve market efficiency and internalize external costs but also brings new research and technologies for future commercialization. Clean energy has been promoted in some countries through a feed-on tariff or feed-in premium. These systems allow the producers to sell their entire production at a fixed guaranteed price that is settled above the market price for electricity. The consumers assume all the risk. A different way of promoting clean energy is to create tradable green certificate markets. This system is distinct from the previous one because it separates common electricity from clean electricity, which is treated as a new product in this new market.

The electricity distribution sector has recently undergone several regulatory reforms aimed to improve its efficiency; moreover, concepts and estimated measures of scale, scope, and cost efficiency have become very important to compare firms within the sector. Chapter 12 by Per J. Agrell, Mehdi Farsi, Massimo Filippini and Martin Koller, "Unobserved Heterogeneous Effects in the Cost Efficiency Analysis of Electricity Distribution Systems", focuses on the cost efficiency of electricity distribution systems and considers the impact of unobserved heterogeneity on these estimates. The authors use heterogeneity to mean the characteristics inherent to each firm arising from the fact the firms operate in different regions with different environmental and network characteristics. This chapter uses an alternative strategy to distinguish cost efficiency from unobserved heterogeneity by decomposing the benchmark process in two steps. They start by identifying similar companies and including them in the same class in order to reduce the unobserved heterogeneity, and then apply the best methodology in each class. This two-step procedure is shown to provide more realistic efficiency estimates than the conventional one-step analysis because it reduces the unobserved heterogeneity within classes and, consequently, the unexplained variance identified previously as inefficiency. These results suggest that cross-sectional or pooled models might underestimate the real cost efficiency if they do not account for unobserved heterogeneity.

Part I
Oil Industry

Risk Factors in the Oil Industry: An Upstream and Downstream Analysis

Sofia B. Ramos, Helena Veiga and Chih-Wei Wang

Abstract In this paper we examine the drivers of stock market value in the upstream (producers) and downstream segments (petroleum refiners) of the oil industry. Using a sample of U.S. firms we find that stock returns of upstream and downstream firms follow stock market and oil price returns. Moreover, the upstream firm stock returns are sensitive to changes in the Canadian dollar, an important oil trade partner of the U.S., to natural gas returns and its volatility, but not to oil return volatility. Both the upstream and downstream segments present asymmetric changes regarding oil return changes. Stock returns of the oil industry respond asymmetrically to oil returns, i.e., positive oil returns had a greater impact than oil price drops in the period 1998–2004. Before 1997 we do not find any asymmetric effects, and after 2004, they are only statistically significant in the upstream segment. Overall, the evidence for asymmetric effects is more consistent across measures and time in the upstream than in the downstream segment.

Keywords Asymmetric effects · Oil and natural gas companies · Oil prices · Oil volatility

---

The authors acknowledge financial support from Fundação para a Ciência e Tecnologia PEst-OE/EGE/UI0315/2011, from the Spanish Ministry of Education and Science, research projects ECO2012-32401 and MTM2010-17323.

S.B. Ramos (✉)
Business Research Center/UNIDE, Instituto Universitario de Lisboa (ISCTE-IUL), Avenida Das Forças Armadas, 1600-083 Lisbon, Portugal
e-mail: sofia.ramos@iscte.pt

H. Veiga
Department of Statistics and Instituto Flores de Lemus, Universidad Carlos III de Madrid, C/Madrid 126, 28903 Getafe, Spain. BRU/UNIDE, Avenida Das Forças Armadas, 1600-083 Lisbon, Portugal

C.-W. Wang
Chinese Academy of Finance and Development, Central University of Finance and Economics, 39 South College Road, Haidian District, Beijing 100081, China

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_1

1 Introduction

Several studies have documented a relation between stock returns of the oil and gas industry and risk factors other than stock market returns (see Boyer and Filion 2007; El-Sharif et al. 2005; Nandha and Faff 2008; Ramos and Veiga 2011; Sadorsky 1999, 2001). A common finding emerging from these studies is that industry stock returns follow oil returns, the pivot commodity of this industry. Moreover, Ramos and Veiga (2011) find evidence of the perception of a pass-through effect of oil price hikes, i.e., industry returns increase more with oil price hikes than proportionally decrease with oil price drops, which is not found in other commodity dependent industries. In the discounted cash flow framework, this can be due to the fact that the demand is not sufficiently depressed by price hikes or the market power of firms in this industry is large.

Although previous studies have investigated risk factors and asymmetric effects of oil returns for oil and gas industry, they have not determined whether impacts are different for upstream (producers) and for downstream (petroleum refiners) segments. Oil is simultaneously the main output of the upstream industry and an input for the downstream segment. Thus, depending on the price elasticity of oil, oil price hikes may generate higher revenues for the upstream segment and an increase of input cost followed by a reduction of supply in the petroleum refinery (Lee and Ni 2002). Therefore, the impact of oil price changes might not be the same along the oil activity value chain.

This paper investigates whether risk factors are different along the oil and gas industry chain. As far as we know, this is the first work analyzing a panel of 260 U.S. companies engaged in operations in the oil and natural gas industry in the period 1988–2010, covering all the value chain from upstream to downstream activities: production, drilling and other related services, refining and even field machinery and equipment.

We find differences in the sensitivity to factors of the industry segments. Oil producers and petroleum refiners are not very sensitive to stock market variations and their stock market betas range between 0.5 and 0.6. On the other hand, the other industry segments, such as complementary segments of services, equipment and machinery are more market sensitive and have market betas close to one.

Oil producers show sensitivity to exchange rate changes, suggesting that cash flows are likely to be affected by revenues from foreign sales.

Firm returns are sensitive to oil price changes. Interestingly, oil shocks affect more complementary industries than upstream and downstream segments. Oil price hikes have an impact ranging from 1.45 (downstream) to 4.19 % (machinery and equipment) depending on the industry. This might suggest that when the oil price increases, oil producers might increase their production of oil, fostering business activity in complementary industry segments.

Previous work has not analyzed the effects of natural gas price changes and its volatility on industry returns. We find that in addition to oil sensitivity, firm value is also sensitive to natural gas price changes and its volatility. For instance, one standard deviation shock has an impact on stock returns of oil and gas producers of 2.34 %.

To gather evidence on the perception of a pass-through effect, we analyze whether industry returns change asymmetrically with oil and natural gas price changes. We find that industry returns for oil and gas producers (upstream) and petroleum refiners (downstream) change asymmetrically with oil price changes, but not for services, equipment and machinery. Industry returns also follow asymmetrically natural gas returns but only in the upstream segment.

An analysis through time shows that there were no asymmetric effects before 1998, and that they prevail in the whole industry in the period 1998–2004; after that, we only find a weak presence for the upstream segment.

Our work provides new contributions for the literature. We have documented differences on the risk factors along the value chain of the oil and gas industry. The analysis confirms that at the upstream, firm value increases proportionally more with oil price hikes than with oil price drops. Our results are of interest for corporate managers and investors that care about the oil industry's exposure to interest rates, exchange rates and oil and natural gas prices.

The structure of the paper is as follows. Section 2 reviews the literature. Section 3 presents the data. Section 4 describes the methodology and details about the estimation. Section 5 presents the estimation results for portfolios. Section 6 checks if the asymmetric effects are time-varying. Section 7 provides a series of robustness tests for the analysis, and Sect. 8 concludes.

2 Review of Literature

Research work has analyzed the impact of oil price changes on stock markets but with conflicting results. Early works were not supportive of oil as a significant factor in financial markets. Huang et al. (1996), Chen et al. (1986) and Ferson and Harvey (1994) find that oil futures returns do not have much impact on stock market indices and that there is no reward for oil price risk in stock markets. Jones and Kaul (1996), however, provide evidence that aggregate stock market returns in the U.S., Canada, Japan and the U.K. are negatively sensitive to the adverse impact of oil price shocks on their economies. More recently, Driesprong et al. (2008) find some predictive power in oil returns.

A strand of literature has also examined whether oil asymmetric effects found in the macro literature pertain to stock market returns. Cong et al. (2008), Nandha and Faff (2008) and Park and Ratti (2008) do not find evidence of such effects, while Basher and Sadorsky (2006) find evidence of oil asymmetric effects, and Ramos and Veiga (2013) find that they exist both for oil-importing and oil-exporting countries, but run in opposite directions. Oil price rises have a negative effect on stock markets of oil-importing countries, while the impact is positive for stock markets of oil-exporting countries.

The literature has also investigated whether oil return volatility impacts stock returns. Bernanke (1983) argues that companies should postpone irreversible investment expenditures when they experience increased uncertainty about the future oil price. Thus, the falling energy prices' tendency to stimulate output might be dampened if firms are uncertain whether the fall in energy prices is permanent or transitory.

With the growing profitability of oil companies, a number of works have looked for drivers of oil and natural gas industry returns, such as the market index, interest and currency rates and naturally oil prices. The evidence has unanimously showed that stock returns of oil and natural gas companies follow market and oil returns, but other variables such as interest and currency rates have different results depending on the sample.[1]

Scholtens and Wang (2008) do not find differences between oil and gas producers (upstream) and the equipment, services and distribution companies (downstream). Elyasiani et al. (2011) study several sector returns in the U.S. that are oil related. They find that the oil-user industries (building, chemical, plastic and rubber, metal, industrial machinery, transport equipment, and air transportation) are more likely to be affected by changes in the volatility of oil returns, while the level of oil-futures return exerts a greater impact on the oil-substitute (coal and electric and gas services) and oil-related (oil and gas extraction and petroleum refinery) industries. The above studies have not, however, analyzed the presence of asymmetric effects.

3 Data

Our work aims to shed light on the understanding of the impact of oil price changes in the upstream and downstream segments of the oil industry.[2]

3.1 Oil and Gas Companies

Our data sample is an unbalanced panel of 260 oil and gas companies drawn from CRSP and Compustat. The sample period runs from January 1988 to December 2010 and we use adjusted returns at end of the month.

---

[1] See Faff and Brailsford (1999) for evidence on Australian oil and gas industry equity returns, Sadorsky (2001) and Boyer and Filion (2007) for Canada, El-Sharif et al. (2005) for U.K., Al-Mudaf and Goodwin (1993) and Hammoudeh et al. (2004) for the U.S., Park and Ratti (2008) and Oberndorfer (2009) for Europe and Ramos and Veiga (2011) for evidence on a sample of 34 countries. Ramos and Veiga (2011) also find that the oil and gas sector in developed countries responds more strongly to oil price changes than in emerging markets.

[2] The petroleum industry consists of three main segments commonly known as the upstream, midstream and downstream, though the midstream is usually grouped with the upstream.

As customary in the financial literature, returns are computed as $r_t^i = \left[\ln(P_t^i) - \ln(P_{t-1}^i)\right]$, where P_t^i is the price of firm i at time t. Returns are expressed in U.S. dollars.

The upstream oil sector refers to the initial part of the oil value chain, namely the exploration and production activity. The goal of these activities is to find new oil resources and bring them to the surface. Oil is produced through the mining and extraction of oil from oil shale and oil sands, and gas and hydrocarbon liquids are produced through gasification, liquid faction, and pyrolysis of coal at the mine site. Oil and natural gas are the main outputs of this industry.

The downstream oil sector refers to the refining of oil crude and the selling and distribution of natural gas and products derived from crude oil. The refining process is very complex and involves both chemical reactions and physical separation. Oil is transformed into an array of products, including gasoline, distillate fuels, and jet fuel, which are used in other industry businesses. As research work documents a relation between the price of crude oil and refined products (Asche et al. 2003; Girma and Paulson 1999; Serletis 1994), we hypothesized that the refining industry shows some sensitivity to oil prices changes.

We use the following SIC codes to select firms from the oil sector: crude petroleum & natural gas companies (SIC code 1311) and oil and gas field services (SIC code 138). These sectors can be classified as being in the upstream segment. The main difference is that SIC code 138 corresponds to oil and gas field services which are engaged in performing oil field services for operators on a contract or fee basis.[3] For the downstream sector, we take petroleum refining firms using SIC code 2911. In addition, we drew data from oil and gas field machinery & equipment companies (SIC code 3533). The sample comprises 260 firms.[4,5]

Table 1 reports the descriptive statistics by industry. The table provides a snapshot of the market capitalization of firms and the industry returns.

Gross returns are higher for oil machinery and equipment and oil related services, 1.61 and 1.31 % respectively. The standard deviation of returns is lower for petroleum refiners. Second, the market capitalization of firms is very disperse; petroleum refiners stand out because the average market capitalization is almost twenty million dollars while other industries are all less than three million dollars.

---

[3] Because there were few companies in each subsector, we group SIC code 1381 that corresponds to drilling oil & gas wells; SIC code 1382 to oil & gas field exploration services and SIC code 1389 that corresponds to oil & gas field services companies, not elsewhere classified (nec). Some sectors were excluded due to few observations (SIC codes 299 and 517).

[4] The petroleum industry has been studied in the literature of hedging because it is a good illustration of the usage of derivatives. For instance, Mackay and Moeller (2007) use the SIC code 2911, and Haushalter et al. (2002) use the SIC code 1311.

[5] It is noted that some firms changed SIC code during the period of analysis.

Table 1 Summary statistics of oil and natural gas firms by SIC code

SIC codes		Variable	Mean	Median	St. Deviation	N
1311	Crude petroleum and natural gas	Returns	1.30 %	0.00 %	16.36 %	29,425
		Market capitalization	22,86,681	283,311	6,378,120	29,486
138	Oil and gas drilling and other services	Returns	1.38 %	0.22 %	16.08 %	11,288
		Market capitalization	2,646,157	447,020	7,833,220	11,312
2911	Petroleum refining	Returns	1.29 %	0.99 %	10.92 %	6,096
		Market capitalization	18,900,000	1,893,735	54,700,000	6,104
3533	Oil and gas field machinery and equipment	Returns	1.61 %	1.30 %	14.92 %	2,069
		Market capitalization	2,720,160	732,485	4,872,311	2,072

This table presents the mean, median and standard deviation of returns and market capitalization by SIC code. N is the number of firms monthly observations. *Source* CRSP/Compustat

3.2 Risk Factors

We next describe the risk factors. Table 2 presents the variables' summary statistics.

Stock Market Excess returns Asset pricing models like the Capital Asset Pricing Model posit that stock returns are explained by the market portfolio variations. We proxy the U.S. stock market using the returns of the S&P 500, one of the main stock market indexes. *market* is the local market excess return, computed as the logarithmic changes in the local market index. Returns are in excess of the same short-term interest rate, namely the one-month Eurodollar interest rate. Both are drawn from Datastream. We expect that the coefficient *market*, also known as the market beta, is positive and statistically significant, similar to previous studies.

3.2.1 Oil and Natural Gas Prices

Oil Futures Prices Oil prices are from the settlement price of the NYMEX oil futures contract, the most widely traded futures contract on oil. The underlying is the West Texas Intermediate oil a light crude oil widely used as a current

Table 2 Summary statistics of independent variables

Variable	Description	Mean	SD	Kurtosis	Skewness	Jarque-Bera
market	S&P 500 log returns	0.002	0.043	4.569	−0.777	54.360***
exchrate	Canadian dollar log variations against the U.S. dollar	0.000	0.022	8.937	−0.468	413.876***
int_rate	(monthly) one-month Eurodollar interest rate	0.004	0.002	2.372	0.025	4.797*
oil	Settlement price of the NYMEX oil futures contract (log returns)	0.006	0.094	4.912	−0.176	26.995***
gas	Settlement price of the NYMEX natural gas futures contract (log returns)	0.003	0.162	3.592	−0.131	3.988
vol_oil	Estimated volatility from a GARCH(1,1) model on *oil*	0.092	0.029	8.759	2.171	585.274***
vol_gas	Estimated volatility from a GARCH(1,1) model on *gas*	0.161	0.024	9.530	2.385	662.592***

This table reports the summary statistics of the variables S&P 500 returns (*market*), oil futures returns (*oil*), Canadian dollar variations against the U.S. dollar (*exchrate*), U.S. interest rate (*int_rate*), oil return volatility (*vol_oil*), natural gas futures returns (*gas*), natural gas volatility (*vol_gas*). The sample period ranges from 1988:02 through 2010:12. By column, we report the mean, the standard deviation (SD), the kurtosis, the skewness and the Jarque-Bera test statistics. The returns are the first differences of the logarithm of prices in percentage. Superscripts *, ** and *** denote statistical significance at 10, 5 and 1 % levels respectively. *Source* Datastream

benchmark for U.S. crude production. Prices are in U.S. dollars per barrel (U$/BBL). The variable *oil* is the logarithmic difference of oil prices.

Summary statistics are displayed in Table 2. Oil returns show a positive mean during the period, around 0.6 % a month, with a standard deviation of almost 9.4 % monthly. We note that the stock returns of oil firms for the same period had a higher average return, and a higher standard deviation than oil returns.

Figure 1 depicts the oil futures prices. The price of oil fluctuates little until around 1998. There is some turbulence in the summer of 1990, which coincides with the beginning of the invasion of Kuwait and the Gulf War, but prices drop to normal levels after the end of the war in February 1992. In 1999, prices rise again, but then fall after 2000 and 2001. They increase again to over $50/BBL in 2005, $100/BBL in 2007 and almost $150/BBL in July 2008. In the second half of 2008 when many countries worldwide experienced economic recession, prices continue to slide until the end of the year, to peak again during 2009. The value in June 2009 is again close to $70/BBL but then rebounds to reach $91/BBL at the end of 2010.

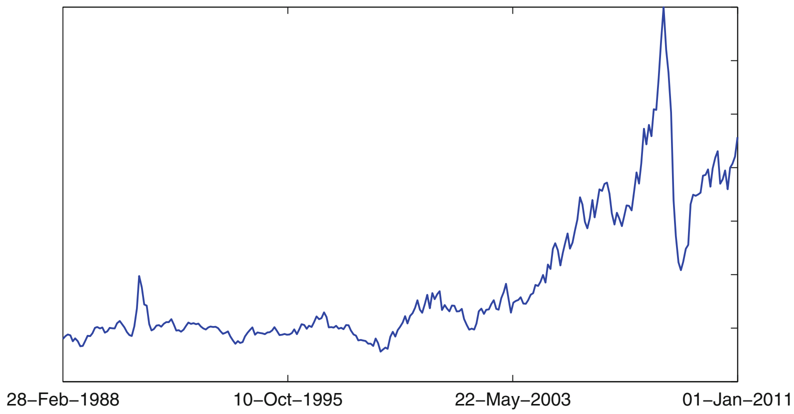


Fig. 1 Oil futures price (settlement price of the NYMEX oil futures contract- U$/BBL)

Natural Gas Prices Natural gas prices are from the NYMEX gas futures contracts (*gas*).[6] The underlying asset of one contract is 10,000 million British thermal units (MMBtu) of natural gas delivered at Henry Hub, Louisiana. The Henry Hub is the largest centralized natural gas trading hub in the U.S. It interconnects nine interstate and four intrastate pipelines. Collectively, these pipelines provide access to markets throughout the U.S. East Coast, the Gulf Coast, the Midwest, and up to the Canadian border. Natural gas production from areas around the Henry Hub accounts for about 50 % of total U.S. production.

Summary statistics about natural gas returns are displayed in Table 2. Natural gas returns have a positive mean during the period, around 0.3 % a month, with a standard deviation of almost 16.2 % monthly. The average return is lower than that of oil but the standard deviation is higher. Compared to other explanatory variables, the distribution of natural gas returns seems to be closer to normal since the hypothesis of normality is not rejected.

Figure 2 depicts natural gas futures prices. The price trend is similar to that of oil. Prices are stable until the end of the 1990's and then increase. Natural gas prices present several price spikes mainly around 2000–2001. Then there is a strong upward trend after 2003. Prices fall substantially after 2008 and then surge again.

Oil and Natural Gas Return Volatility We also analyze the exposure to oil and gas volatility. Oil price volatility is a source of uncertainty that affects the cost of an important input or output of firms. This creates uncertainty regarding firm profitability, firm value and investment decisions. Sadorsky (1999) finds that either an oil price change or its volatility has an impact on real stock returns. Haushalter et al. (2002) finds that although stock returns of U.S. oil producers have a negative

[6] Natural gas is one of the cleanest burning fuels, producing primarily carbon dioxide, water vapor, and small amounts of nitrogen oxides. Natural gas is a source of energy used for both heating and also producing electricity.

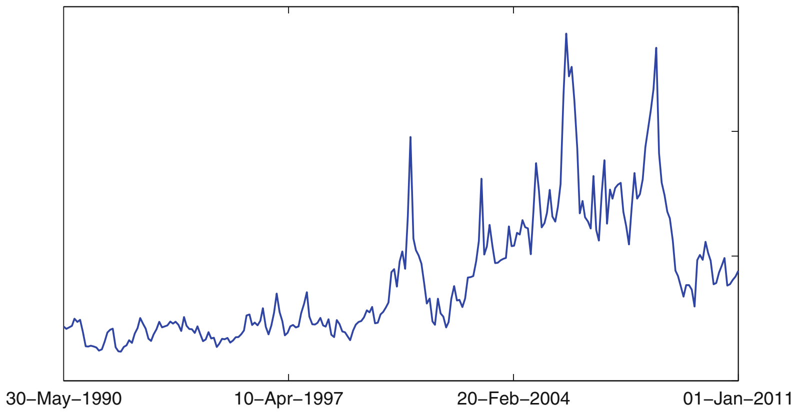


Fig. 2 Gas futures price (settlement price NYMEX gas futures contracts—MMBtu)

sensitivity to oil implied volatility, it is not statistically significant. Oberndorfer (2009) also finds that oil volatility has a negative impact on European oil and natural gas stock returns.

To compute oil volatility, we use the estimated volatility (*vol_oil*) obtained from a GARCH(1,1)

$$oil_t = \mu + \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $\varepsilon_t = \sigma_t \epsilon_t$ is the prediction error, $\sigma_t > 0$ is the conditional standard deviation of the underlying oil return (denoted volatility) and the innovation $\epsilon_t \sim NID(0, 1)$. We impose the conditions $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta_1 \geq 0$ to guarantee that the conditional variance is positive and $\alpha_1 + \beta_1 < 1$ to assure its stationarity. The volatility of natural gas (*vol_gas*) is estimated in the same way.

Oil and natural gas volatility are depicted in Fig. 3, respectively, and summary statistics are in Table 2. Table 2 reveals that the kurtosis of oil and gas volatility are both larger than 8.7, and both show positive skewness.

Interest Rate In the framework of the discounted cash flow method, shocks in interest rates affect firm value, because the present value of cash flows is smaller. Moreover, the interest rate is a factor that also proxies for macroeconomic conditions. Empirical evidence on the importance of interest rates for oil and gas industry returns is mixed. Boyer and Filion (2007) finds that interest rates have a negative significant impact on stock returns in the Canadian oil and gas industry. The term premium, the difference between the three-month and one-month interest rate, has a positive impact on industry returns. El-Sharif et al. (2005) and Oberndorfer (2009), however, do not find statistical significance for interest rates. To proxy the interest rate factor, we use the one-month Eurodollar interest rate (int_rate).

Exchange Rate The theoretical model of Adler and Dumas (1983) and Solnik (1974) supports the pricing of exchange rate fluctuations in a global setting.

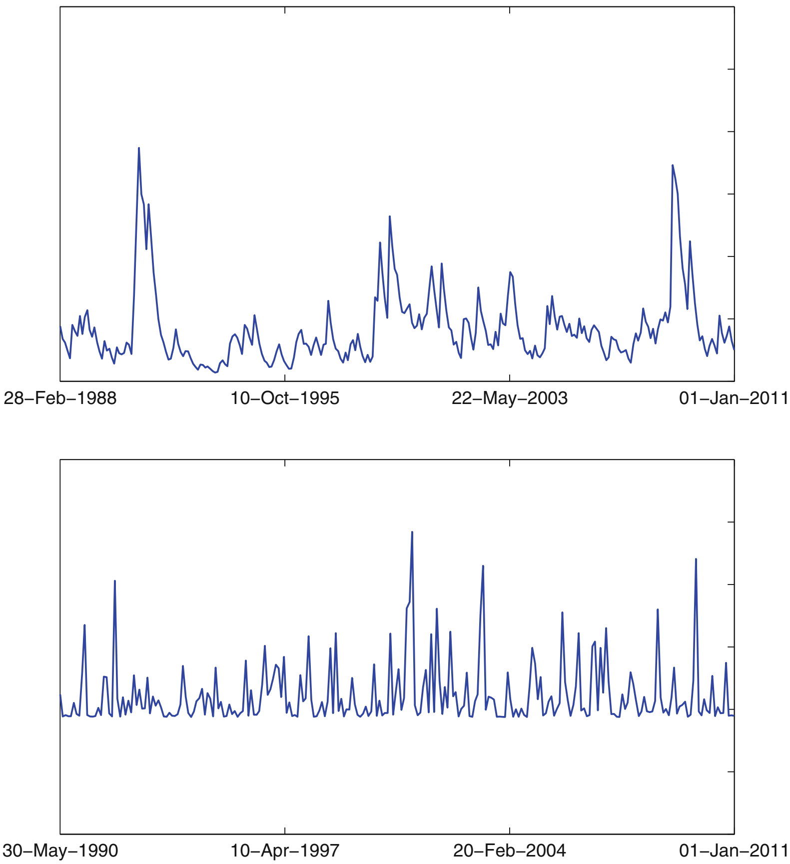


Fig. 3 Oil price volatility (*first panel*); gas price volatility (*second panel*)

Therefore, we test whether excess returns of the oil industry show some sensitivity to changes in the currency rates against the U.S. dollar.

As Canada is one of the main oil trade partners of the U.S., we computed the logarithmic changes in Canadian dollar rates against the U.S. dollar (*exchrate*). *exchrate* is expressed in foreign currency per unit of U.S. dollars; thus a positive change in the rate means that the U.S. dollar depreciated with respect to the foreign currency, which might increase revenues if the firm is foreign sales oriented. Contrary to expectations, Sadorsky (2001) and Boyer and Filion (2007) studying the Canadian oil and gas industry find that a weakening of the Canadian dollar against the U.S. dollar has a negative impact on stock returns.

Table 2 presents summary statistics for the currency rate. The exchange rate volatility is on average lower than the volatility of oil and gas companies and market returns. Since kurtosis is higher than three and there is negative skewness; the Jarque-Bera test rejects the null of Gaussian currency returns.

4 Methodology and Estimation

4.1 Methodology

Following previous studies on oil industry risk factors, we construct industry portfolios for each SIC code, weighting firm returns in each SIC code by their market capitalization.

The analysis of the risk factors on the oil industry has been undertaken using both multifactor models (see Basher and Sadorsky 2006; Nandha and Faff 2008; Ramos and Veiga 2011; Sadorsky 2008) and vector autoregressive models (see Cong et al. 2008; Park and Ratti 2008; Sadorsky 1999). We follow the literature that uses factor models to examine the impact of systematic factors on stock returns (see Ferson and Harvey 1994; Haushalter et al. 2002; Jin and Jorion 2006; Karolyi and Stulz 2003; Tufano 1998). According to Ferson and Harvey (1994), factor regressions provide information about the usefulness of factors in controlling for the risks of investments. The models that we estimate for the oil and gas industries are:

$$r_t^i = \alpha_i + \beta_{1i} market_t + \beta_{2i} exchrate_t + \beta_{3i} int\_rate_t + \beta_{4i} oil_t + \beta_{5i} vol\_oil_t + u_t^i \quad (1)$$

and

$$\begin{aligned} r_t^i = \alpha_i + \beta_{1i} market_t + \beta_{2i} exchrate_t + \beta_{3i} int\_rate_t + \beta_{4i} oil_t + \beta_{5i} gas_t \\ + \beta_{6i} vol\_gas_t + u_t^i, \end{aligned} \quad (2)$$

where r_t^i is the excess return of the SIC industry index i at time t. These are time series regressions for each i with $i = 1, ..4$. β_{ji} are the coefficients of r_t^i on several risk factors, such as *market*, *exchrate*, *int_rate*, *oil*, *vol_oil*, *gas* and *vol_oil*. The α_i's are the intercepts and u_t^i is the error term and represents the excess return not explained by the factors in each model. The models are estimated by Ordinary Least Squares. The correlation among explanatory variables is low with the exception of stock market returns (*market*) and the exchange rate (*exchrate*). Results are not shown for brevity's sake but are available from the authors.

4.2 Asymmetry Measures and Tests

To test for asymmetric effects of oil price changes, we define nonlinear measures of oil price changes. The traditional approach is based on a dummy variable that differentiates positive from negative oil price changes and multiplies the variable oil price changes, which is equivalent to the following variables:

$$oilp_t = \max(0, oil_t)$$
$$oiln_t = \min(0, oil_t),$$

at time t, the *oilp* (*oiln*) variable assumes positive (negative) values each time changes are positive (negative) and zero otherwise.

Similarly we also define

$$gasp_t = \max(0, gas_t)$$
$$gasn_t = \min(0, gas_t),$$

at time t, for positive and negative changes of natural gas.

Figure 4 depicts *oilp* and *oiln*. Many large monthly changes are as great as $+/-20\,\%$. There are four large declines in prices that correspond to December 1990 and January 1991 with the end of the Gulf War; December 2000; March 2003 and more recently October and December 2008. Price spikes can be seen in July, August and September of 1990 (the beginning of the Gulf War), March 1999, May 2000, March 2002, January 2005 and May 2009.

We use tests of asymmetry that favor asymmetry if we reject the null hypothesis of symmetry $H_{01} : \beta_{oilp} = \beta_{oiln}$.[7] Thus, the null is rejected if coefficients are statistically different. A possibility is that the null is not rejected when the two coefficients are not significant, i.e., although oil has a non significant effect on stock markets, the null leads us to conclude for symmetry. To account for this issue, a second hypothesis is formulated $H_{02} : \beta_{oilp} = 0$ and $\beta_{oiln} = 0$. Asymmetry is therefore assured by the joint rejection of the two null hypotheses: H_{01} and H_{02}.

Another two measures are introduced with the aim of capturing oil price shocks or innovations. Hamilton (1996) advocates that it is more appropriate to compare the current price of oil with its value over the last year, rather than during the previous month alone, to measure how unsettling an increase in the price of oil is likely to be for the spending decisions of consumers and firms. Net oil price increase (*nopi*) at time t is defined as:

$$nopi_t = max(0, ln(p\_oil_t) - ln(max(p\_oil_{t-1}, \ldots, p\_oil_{t-12}))).$$

nopi can be interpreted as the amount by which the log oil futures price exceeds its maximum value over the last year (here, p_oil_t is used for oil futures price). Note that *nopi* would be small in a period of consistently escalating oil prices, but if prices soar sharply then *nopi* is high. Kilian (2008) refers that *nopi* has the advantage of being a better measure to extract the exogenous component of oil price fluctuations. Likewise, we define net oil price decline (*nopd*) at time t as $nopd_t = \min(0, ln(p\_oil_t) - ln(max(p\_oil_{t-1}, \ldots, p\_oil_{t-12})))$. *nopd* is negative when oil prices are below its peak value over the last year.

[7] See (Nandha and Faff 2008; Park and Ratti 2008; Ramos and Veiga 2011, for instance).

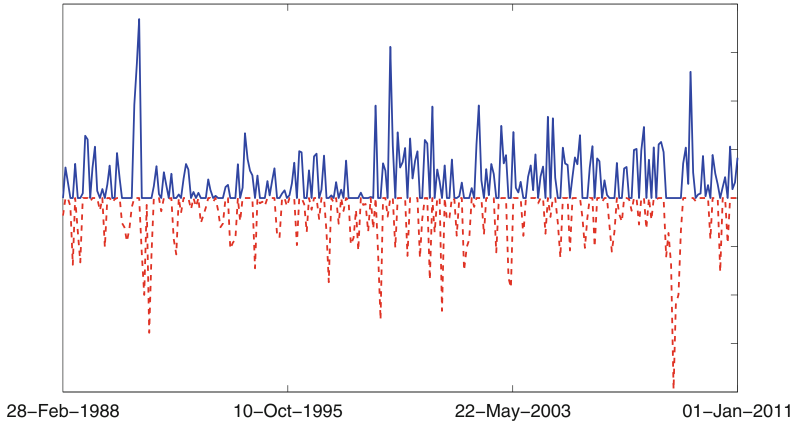


Fig. 4 Oil futures price increases (*continuous line*) and oil futures price decreases (*dotted line*)

Similarly for natural gas, we define the natural gas price increase (*ngpi*) and natural gas price decrease (*ngpd*) measures where *p_gas* is the price of gas.

$$ngpi_t = max(0, ln(p\_gas_t) - ln(max(p\_gas_{t-1}, \ldots, p\_gas_{t-12})))$$
$$ngpd_t = min(0, ln(p\_gas_t) - ln(max(p\_gas_{t-1}, \ldots, p\_gas_{t-12}))),$$

Asymmetry is also checked using a measure developed by Lee et al. (1995) called scaled oil price increases (*sopi*). According to these authors, what matters is how unexpected an oil price increase is based on the observed changes. An unexpected oil price change, it will have less impact when conditional variances are large because much of the change in oil price will be regarded as transitory.

In order to calculate this measure, we estimate a GARCH(1,1) model similar to the one presented in the previous subsection. Therefore, a measure that reflects the size and variability of the unexpected oil shock might be defined as $\hat{\varepsilon}_t^* = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t}$ and consequently *sopi* at time *t* is given by:

$$sopi_t = max(0, \hat{\varepsilon}_t^*).$$

In a similar manner scaled oil price declines (*sopd*) at time *t* are defined as $sopd_t = min(0, \hat{\varepsilon}_t^*)$. Therefore, oil price increases and decreases are scaled by the oil return conditional standard deviation. *sopi* and *sopd* will be large (in absolute value) when the oil innovation is large (in absolute value). Similarly for natural gas, we define the scaled natural gas price increases (*sgpi*) and scaled natural gas price decrease (*sgpd*) measures.[8,9]

$$sgpi_t = max(0, \hat{\varepsilon}_t^*)$$
$$sgpd_t = min(0, \hat{\varepsilon}_t^*).$$

Figure 5 plots *nopi* and *nopd* for the sample period December–1988 to December–2010. We see episodes of peaking prices that seem to cluster in some periods of time. *nopd* also has some peaks and slumps, and we can see the dramatic fall of oil futures prices during 2009. Figure 6 graphs *sopi* and *sopd*. Although the figure has some similarities with returns (see Fig. 4), we also see some differences, namely the variables *sopi* and *sopd* have a more shrinking scale due to the standardization. Figures 7, 8 and 9 depict the asymmetry measures for natural gas. In Fig. 8, we observe three strong downward trends in 1997, 2001, and 2009. Also, *ngpi* seems very smooth compared to *ngpd*; this may be because there is some correction in the price after prices spikes, common in natural gas.

The equations for testing asymmetric effects are:

$$r_t^i = \alpha_i + \beta_{ji} f_t + \beta_{4i} x_t^+ + \beta_{5i} x_t^- + u_t^i, \tag{3}$$

where $j = 1, \ldots, 3,\ i = 1, \ldots, 4$ SIC codes, f are the factors defined in Eqs. (1) and (2), excluding oil and gas, and the following asymmetry measures $x^+ \in \{oilp, nopi, sopi, gasp, ngpi, sgpi\}$ and $x^- \in \{oiln, nopd, sopd, gasn, ngpd, sgpd\}$.

5 Empirical Results

Table 3 displays the results of the analysis of the risk factors for each of the industry portfolios. Each column shows the result for each segment defined by SIC code.

The first row shows the impact of stock market changes. Oil and gas producers (SIC code 1311) and the refining industry (SIC code 2911) have the lowest market betas, 0.619 and 0.535. Results are consistent with previous work. As an example, Elyasiani et al. (2011) find a beta lower than one for oil-gas extraction and petroleum refinery in the U.S. economy and Sadorsky (2001) a beta of 0.7 for Canadian firms. This means that industry returns change less than proportionally

---

[8] Both *nopi* and *sopi* (*ngpi* and *sgpi*) are nonlinear and time-dependent measures. The time dependence of *nopi* comes from the fact that if a shock is not large enough to increase prices above their value of the previous year, then the shock is scaled down to zero. *sopi* scales the shocks that occur in a volatile period down and scales those that occur in a less volatile period up.

[9] Given that *sopi* also accounts for oil volatility by scaling the shocks down and up according to the volatility of the period, we do not include oil volatility in the models.

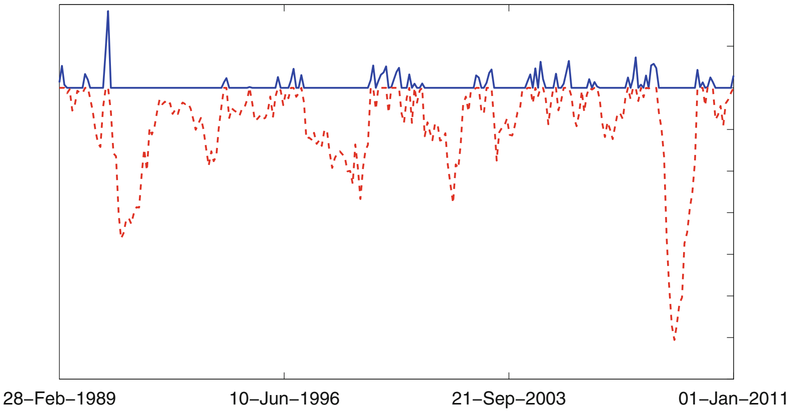


Fig. 5 Net oil futures price increases (*continuous line*) and net oil futures price decreases (*dotted line*)

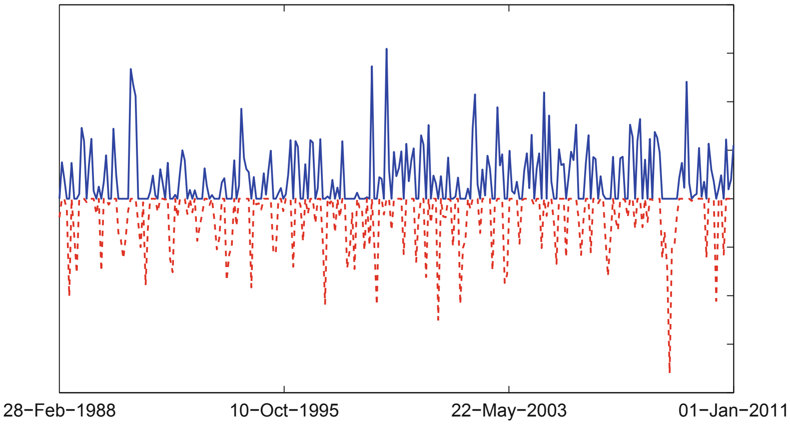


Fig. 6 Scaled oil futures price increases (*continuous line*) and scaled oil futures price decreases (*dotted line*)

with stock market returns. The market beta is around one for SIC codes 138 and 3533, which suggests that market shocks have a similar impact on the segments.

To examine the economic significance of these effects, we calculate the effect of one standard deviation change in the variables on each of the industry returns considered. These effects can be calculated by multiplying the coefficient estimated for the variable by its respective standard deviation reported in Table 2. Regarding economic significance, one standard deviation shock to the market returns increases stock returns of crude petroleum and natural gas (SIC code 1311) in 2.663 % and of oil and gas field machinery and equipment (SIC code 3533) 4.693 %.

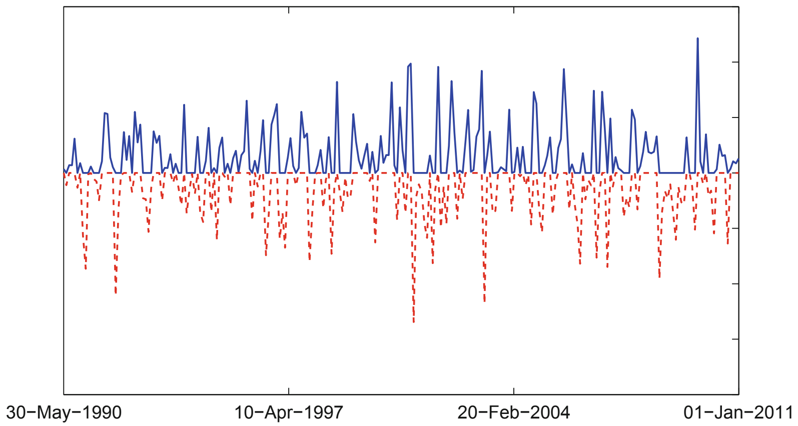


Fig. 7 Gas futures price increases (*continuous line*) and gas futures price decreases (*dotted line*)

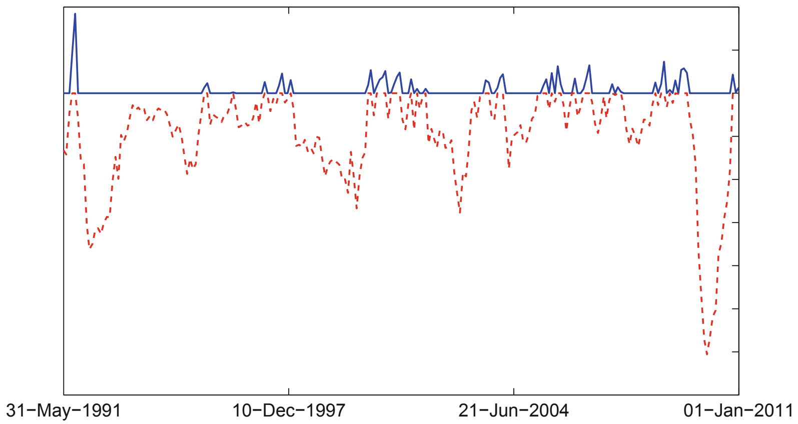


Fig. 8 Net gas futures price increases (*continuous line*) and net gas futures price decreases (*dotted line*)

Exchange rate coefficients are positive, but are statistically significant only for crude petroleum and natural gas (SIC code 1311). This seems to indicate that the depreciation of the U.S. dollar has a positive effect on stock returns; this may be because it increases foreign sales as crude is traded internationally in U.S. dollars.

Interest rates have a negative coefficient in line with previous studies that can result from the contractionary effect that high interest rates have in the economy or in the present value of the firm, but it is not statistically significant.

Oil price changes are statistically significant for all industry segments. Coefficients range from 0.153 in petroleum refiners (downstream) to 0.444 in oil and gas

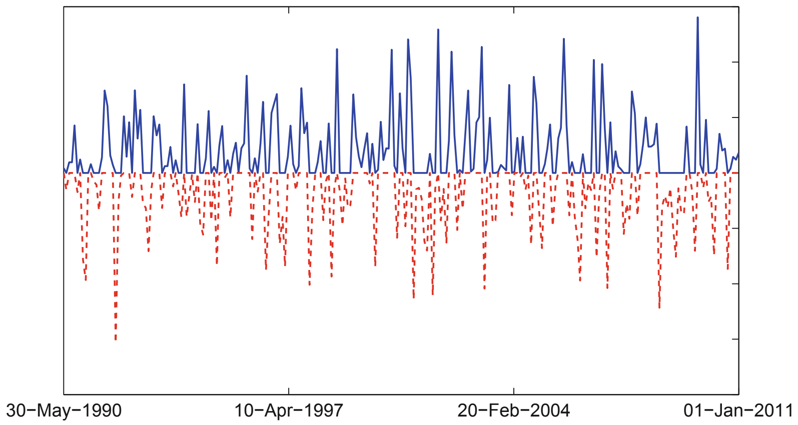


Fig. 9 Scaled gas futures price increases (*continuous line*) and scaled gas futures price decreases (*dotted line*)

Table 3 Risk factors of the oil industry

	Crude and nat. gas producers		Drilling and other services		Petroleum refineries		Machinery and equipment	
SIC codes	1311		138		2911		3533	
market	0.619	0.645	0.962	1.006	0.535	0.548	1.091	1.123
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
exchrate	0.659	0.570	0.386	0.268	0.073	0.047	0.314	0.229
	(0.000)	(0.000)	(0.130)	(0.252)	(0.524)	(0.677)	(0.315)	(0.431)
int_rate	−1.382	−1.514	0.342	−0.008	1.475	2.060	1.529	0.437
	(0.235)	(0.301)	(0.840)	(0.997)	(0.156)	(0.147)	(0.514)	(0.886)
oil	0.273	0.206	0.345	0.294	0.153	0.118	0.444	0.375
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
vol_oil	-0.059		-0.029		−0.004		0.020	
	(0.586)		(0.839)		(0.967)		(0.927)	
gas		0.144		0.141		0.059		0.151
		(0.000)		(0.000)		(0.000)		(0.001)
vol_gas		0.110		0.393		0.011		0.317
		(0.415)		(0.024)		(0.919)		(0.211)
Constant	0.023	0.000	0.013	−0.053	0.005	0.001	0.009	−0.038
	(0.035)	(0.985)	(0.374)	(0.084)	(0.635)	(0.959)	(0.675)	(0.379)
Observations	269	242	275	248	266	239	275	248
R^2	0.534	0.642	0.477	0.546	0.381	0.415	0.387	0.443

This table reports estimation results of Eqs. ((1) and (2)) where the dependent variable is the monthly excess returns of value weighted portfolios formed by SIC code. Sample period: 1988:02 through 2010:12. Explanatory variables are described on Table 11. Robust p-values are reported below coefficients field machinery and equipment (SIC code 3533). For instance, one standard deviation positive shock results in an approximately 1.45–4.19 % significant increase in stock returns, respectively. Note that the impact is greater for complementary activities and lower for oil producers and refiners. Oil volatility has a negative coefficient but is not statistically significant, a result similar to that of Haushalter et al. (2002).

We next analyze the impact of natural gas price changes (second column of Table 3 for each SIC code). Given the close relation between natural gas and oil, we also include *oil* in the regression to control for its effects. Coefficients of other variables are similar, but that of *oil* is slightly smaller. The coefficient of natural gas is positive and ranges between 0.059 and 0.151; the small interval means that sensitivity does not differ much among industries. The impact is also economically significant; one standard deviation shock has an impact on stock returns of producers of 2.34 % (upstream). The volatility of natural gas does not have an impact on the returns of the industry except for oil and gas field services (SIC code 138). The R^2 of the estimation increases when natural gas is included as an explanatory variable.

Table 4 shows the results of the asymmetry tests for oil in order to see whether markets perceive a pass-through effect of the industry. We repeat the estimation of Eq. (3). The different subpanels show the results for the various measures. For the sake of brevity, we only present the coefficients of asymmetric measures. We recall that the null must be rejected in both hypotheses H_{01} and H_{02} to substantiate the statistical significance of asymmetry at standard levels of significance. The results confirm that crude petroleum & natural gas (SIC code 1311) present asymmetric effects. While *oilp* and *nopi* coefficients are larger than *oiln* and *nopd*, *sopi* is smaller than *sopd*, i.e., the asymmetry is scaled effects goes in the opposite direction.

Petroleum refining (SIC code 2911) shows asymmetric effects using *sopi* and *sopd* and using *oilp* and *oiln* (only weaker evidence). This is consistent with Lee and Ni (2002) results which conclude that although the output of the petroleum refinery is not significantly affected by oil price shocks, its price increases significantly. Machinery and equipment (SIC code 3533) also shows asymmetric effects for *nopi* and *nopd*.

Table 5 presents the results of the asymmetry tests for natural gas prices. The tests confirm the existence of asymmetry effects in the upstream segment using *ngpi* and *ngpd* and *sgpi* and *sgpd*, and in the downstream segment using *ngpi* and *ngpd*.

Overall, the results show that oil and natural gas price changes are risk factors for all industry segments. While the sensitivity to natural gas returns is very similar across industries, the sensitivity to oil differs for industry segments. Crude petroleum & natural gas industry returns (upstream) change asymmetrically with oil and natural gas price changes but petroleum refining industry returns (downstream) also change asymmetrically with oil price changes; this suggests that both upstream and downstream segments might have a pass-through effect.

Table 4 Asymmetric effects of oil futures returns

	Crude and nat. gas producers	Drilling and other services	Petroleum refineries	Machinery and equipment
SIC codes	1311	138	2911	3533
Panel A oilp and oiln				
oilp	0.399	0.426	0.230	0.587
	(0.000)	(0.000)	(0.000)	(0.000)
oiln	0.149	0.266	0.078	0.303
	(0.047)	(0.034)	(0.102)	(0.062)
Observations	269	275	266	275
R^2	0.547	0.48	0.39	0.393
Test 1	4.38	0.68	3.61	1.17
p-value	(0.037)	(0.410)	(0.059)	(0.280)
Test 2	23.88	15.02	15.47	14.97
p-value	(0.000)	(0.000)	(0.000)	(0.000)
Panel B nopi and nopd				
nopi	0.231	0.179	0.063	0.260
	(0.012)	(0.180)	(0.413)	(0.054)
nopd	0.033	0.053	0.027	0.090
	(0.076)	(0.096)	(0.069)	(0.024)
Observations	257	263	254	263
R^2	0.431	0.373	0.316	0.306
Test 1	3.9	0.74	0.22	3.9
p-value	(0.049)	(0.392)	(0.637)	(0.049)
Test 2	7.55	3.55	3.9	7.55
p-value	(0.001)	(0.030)	(0.022)	(0.001)
Panel C sopi and sopd				
sopi	0.039	0.042	0.024	0.059
	(0.000)	(0.000)	(0.000)	(0.000)
sopd	0.016	0.026	0.010	0.031
	(0.019)	(0.024)	(0.012)	(0.036)
Observations	269	275	266	275
R^2	0.566	0.494	0.424	0.417
Test 1	4.95	0.91	5.02	1.47
p-value	(0.027)	(0.340)	(0.026)	(0.227)
Test 2	37.06	23.18	39.37	21.1
p-value	(0.000)	(0.000)	(0.000)	(0.000)

This table reports a summary of estimation results of Eq. (3) and analogously for the other measures of asymmetry from 1988:02 through 2010:12. The dependent variable is the monthly excess returns of value weighted portfolios in U.S. dollars. Explanatory variables are described on Table 2. Measures of asymmetry are oil futures price increases (*oilp*), oil futures price decreases (*oiln*), net oil futures price increases (*nopi*), net oil futures price decreases (*nopd*), scaled oil futures price increases (*sopi*), scaled oil futures price decreases (*sopd*). Robust p-values are reported below coefficients. Test 1 corresponds to the null hypothesis $H_{01}: \beta_{oilp} = \beta_{oiln}$ and test 2 to $H_{02}: \beta_{oilp} = 0$ and $\beta_{oiln} = 0$, and analogously for the other measures of asymmetry

Table 5 Asymmetric effects of natural gas returns

	Crude and nat. gas producers	Drilling and other services	Petroleum refineries	Machinery and equip.
SIC codes	1311	138	2911	3533
Panel A gasp and gasn				
gasp	0.167	0.136	0.055	0.172
	(0.000)	(0.037)	(0.028)	(0.043)
gasn	0.121	0.145	0.062	0.132
	(0.001)	(0.000)	(0.023)	(0.029)
Observations	242	248	239	248
R^2	0.643	0.546	0.415	0.444
Test 1	0.69	0.01	0.03	0.12
p-value	(0.409)	(0.920)	(0.873)	(0.733)
Test 2	26.44	11.2	8.82	6.23
p-value	(0.000)	(0.000)	(0.002)	(0.002)
Panel B ngpi and ngpd				
ngpi	0.213	0.184	0.073	0.217
	(0.009)	(0.198)	(0.036)	(0.215)
ngpd	0.010	0.011	0.001	0.020
	(0.277)	(0.507)	(0.937)	(0.302)
Observations	231	236	227	236
R^2	0.587	0.515	0.387	0.435
Test 1	5.85	1.33	3.49	1.18
p-value	(0.016)	(0.246)	(0.063)	(0.278)
Test 2	5.33	1.4	2.81	1.82
p-value	(0.006)	(0.249)	(0.062)	(0.164)
Panel C sgpi and sgpd				
sgpi	0.026	0.020	0.009	0.026
	(0.000)	(0.038)	(0.022)	(0.045)
sgpd	0.021	0.024	0.010	0.022
	(0.000)	(0.000)	(0.027)	(0.023)
Observations	242	248	239	248
R^2	0.641	0.532	0.415	0.438
Test 1	0.25	0.13	0.01	0.03
p-value	(0.016)	(0.723)	(0.912)	(0.856)
Test 2	27.38	10.44	8.68	6.39
p-value	(0.000)	(0.000)	(0.000)	(0.002)

This table reports estimation results of Eq. (3) and analogously for the other measures of asymmetry except from 1988:02 through 2010:12. The dependent variable is the monthly excess returns of value weighted portfolios in U.S. dollars. Explanatory variables are described on Table 2. Measures of asymmetry are gas futures price increases (*gasp*), gas futures price decreases (*gasn*), net gas futures price increases (*ngpi*), net gas futures price decreases (*ngpd*), scaled gas futures price increases (*sgpi*), scaled gas futures price decreases (*sgpd*). Robust p-values are reported below coefficients. Test 1 corresponds to the null hypothesis $H_{01}: \beta_{gasp} = \beta_{gasn}$ and test2 to $H_{02}: \beta_{gasp} = 0$ and $\beta_{gasn} = 0$, and analogously for the other measures of asymmetry

6 Time Variation of Asymmetric Effects

To our knowledge, the literature of asymmetric effects has not addressed whether and how, asymmetric effects change over time. To analyze this time issue, we estimate Eq. (3) using a rolling window of fixed width, with 30 observations.

Figures 10 and 11 depict the estimations of the coefficients of *oilp* and *oiln* and *gasp* and *gasn*, respectively, using rolling windows. The figures show that the coefficients are not stable over time. A visual inspection allows us to see that *oilp* is larger than *oiln* mainly in the period 1998–2004. To substantiate this hypothesis, we divide the sample into three subperiods: 1989–1997, 1998–2004 and 2005–2010 and we redo the estimations of Eq. (3).

Table 6 shows the results for the period ranging from 1989 to 1997. The coefficient of *oiln* tends to be larger than *oilp*, and *sopd* is also larger than *sopi*, but there are no asymmetric effects. *nopi* is larger than *nopd* for crude petroleum and natural gas (SIC Code 1311) but the difference is not statistically significant.

Table 7 shows the results for the period 1998–2004. Positive changes of oil prices (*oilp*) always have a larger coefficient than negative changes (*oiln*) for industry segments. Tests confirm statistically that crude petroleum and natural gas (SIC Code 1311), oil and gas services wells (SIC Code 138) and petroleum refiners (SIC code 2911) present asymmetric effects using *oilp* and *oiln* and *sopi* and *sopd* at standard levels of significance. The results for *nopi* are surprising. *nopd* tends to be statistically significant with a positive coefficient and there is no asymmetry.

Table 8 analyzes the period that ranges from 2005 to 2010. *oilp* is larger than *oiln* for the upstream and downstream industry segments and the difference is statistically significant. Although *nopi* is larger than *nopd*, the difference is only statistically significant for crude petroleum and natural gas (SIC Code 1311). Using *sopi* and *sopd*, there is no evidence of asymmetry at standard levels of significance. Concerning natural gas and looking at Fig. 11, it seems that there may only be asymmetric effects in the period 1997–2003.

7 Robustness Analysis

In this section, we check whether the results are robust to the use of spot prices instead of futures prices. We use the price of West Texas Intermediate (WTI) crude traded in the spot market at the Cushing, Oklahoma Center (*oil)*. Prices are in U.S. dollars per barrel ($/BBL).[10] We recompute all oil related asymmetric variables with oil spot prices. Table 9 presents the results and all coefficients are similar to those obtained previously. Table 10 presents the results of asymmetric tests. Using

[10] WTI is a type of crude oil used as a benchmark in oil pricing and is the underlying commodity of New York Mercantile Exchange's (NYMEX) oil futures contracts. WTI is a light crude and is refined in Gulf Coast regions in the United States.

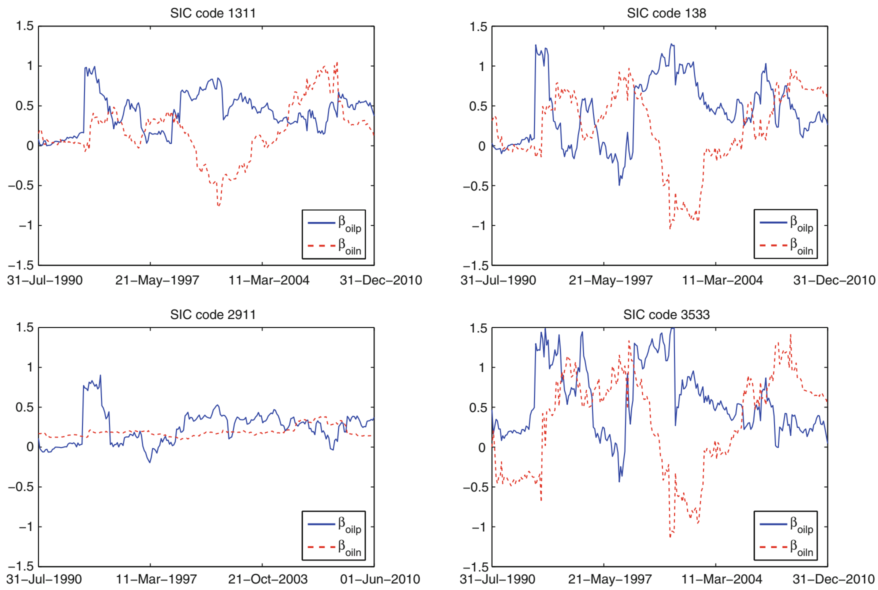


Fig. 10 Time-varying asymmetry. Coefficient estimates of oil futures price increases (*continuous line*) and oil futures price decreases (*dotted line*)

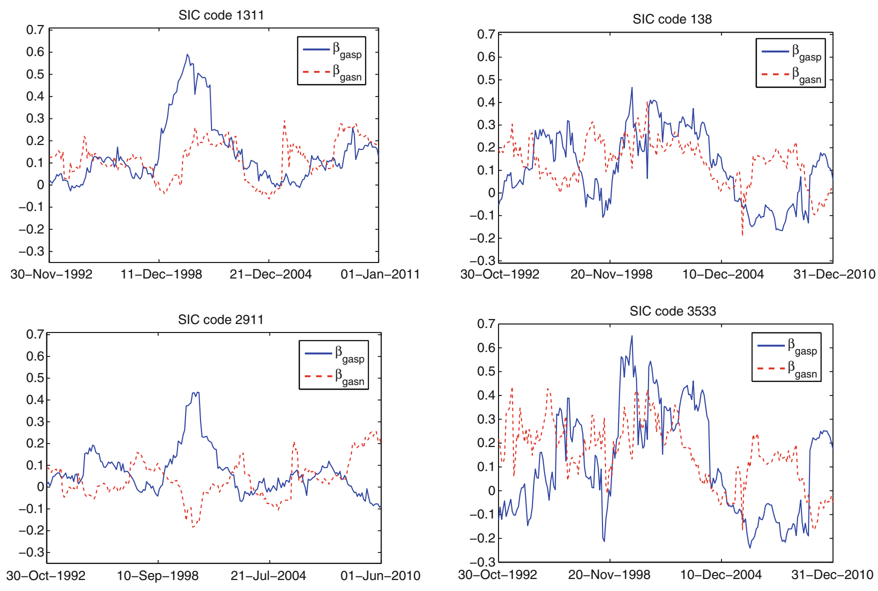


Fig. 11 Time-varying asymmetry. Coefficient estimates of gas futures price increases (*continuous line*) and gas futures price decreases (*dotted line*)

Table 6 Asymmetry effects of oil returns: period 1989–1997

	Crude and nat. gas producers	Drilling and other services	Petroleum refineries	Machinery and equipment
SIC codes	1311	138	2911	2533
Panel A oilp and oiln				
oilp	0.211	0.134	0.110	0.416
	(0.001)	(0.205)	(0.118)	(0.005)
oiln	0.294	0.418	0.221	0.410
	(0.000)	(0.000)	(0.000)	(0.026)
Observations	104	108	108	108
R^2	0.539	0.418	0.535	0.285
Test 1	0.58	2.43	1.08	0
p-value	(0.448)	(0.123)	(0.300)	(0.982)
Test 2	27.8	11.67	19.34	11.56
p-value	(0.000)	(0.000)	(0.000)	(0.000)
Panel B nopi and nopd				
nopi	0.138	0.012	0.049	0.173
	(0.081)	(0.907)	(0.422)	(0.290)
nopd	0.039	0.067	0.023	0.139
	(0.095)	(0.075)	(0.163)	(0.023)
Observations	103	107	107	107
R^2	0.341	0.308	0.398	0.226
Test 1	1.14	0.2	0.15	0.03
p-value	(0.289)	(0.657)	(0.703)	(0.867)
Test 2	6.25	2.37	2.08	6.76
p-value	(0.003)	(0.099)	(0.130)	(0.002)
Panel C sopi and sopd				
sopi	0.021	0.016	0.013	0.044
	(0.000)	(0.106)	(0.025)	(0.001)
sopd	0.028	0.043	0.021	0.042
	(0.000)	(0.000)	(0.000)	(0.003)
Observations	104	108	108	108
R^2	0.575	0.449	0.576	0.318
Test 1	0.75	3.89	0.75	0.01
p-value	(0.388)	(0.082)	(0.388)	(0.933)
Test 2	48.56	25.78	38.69	16.93
p-value	(0.000)	(0.000)	(0.000)	(0.000)

This table reports estimation results of Eq. (3) from 1988:02 through 1997:12. The dependent variable is the monthly excess returns of value weighted portfolios in U.S. dollars. Explanatory variables are described on Table 2 and oil futures price increases (*oilp*), oil futures price decreases (*oiln*), net oil futures price increases (*nopi*), net oil futures price decreases (*nopd*), scaled oil futures price increases (*sopi*), scaled oil futures price decreases (*sopd*). Robust p-values are reported below coefficients. Test 1 corresponds to the null hypothesis $H_{01} : \beta_{oilp} = \beta_{oiln}$ and test 2 to $H_{02} : \beta_{oilp} = 0$ and $\beta_{oiln} = 0$, and analogously for the other measures of asymmetry

Table 7 Asymmetry effects of oil returns: period 1998–2004

	Crude and nat. gas producers	Drilling and other services	Petroleum refineries	Machinery and equipment
SIC codes	1311	138	2911	2533
Panel A oilp and oiln				
oilp	0.558	0.772	0.328	0.973
	(0.000)	(0.000)	(0.000)	(0.001)
oiln	−0.114	−0.131	−0.079	−0.061
	(0.492)	(0.690)	(0.206)	(0.889)
Observations	82	84	83	84
R^2	0.475	0.478	0.377	0.406
Test 1	8.5	4.2	20.74	2.7
p-value	(0.005)	(0.044)	(0.000)	(0.105)
Test 2	14.42	15.22	19.36	8.29
p-value	(0.000)	(0.000)	(0.000)	(0.001)
Panel B nopi and nopd				
nopi	−0.130	0.271	−0.078	0.204
	(0.615)	(0.437)	(0.760)	(0.619)
nopd	0.166	0.170	0.064	0.264
	(0.008)	(0.064)	(0.113)	(0.055)
Observations	82	84	83	84
R^2	0.385	0.389	0.278	0.326
Test 1	1.01	0.06	0.26	0.01
p-value	(0.317)	(0.800)	(0.610)	(0.904)
Test 2	4.24	3.48	1.47	3.63
p-value	(0.018)	(0.036)	(0.237)	(0.031)
Panel C sopi and sopd				
sopi	0.051	0.066	0.031	0.089
	(0.000)	(0.000)	(0.000)	(0.000)
sopd	−0.008	−0.006	−0.008	0.000
	(0.585)	(0.819)	(0.211)	(0.991)
Observations	82	84	83	84
R^2	0.488	0.472	0.39	0.42
Test 1	8.13	3.76	18.88	3.08
p-value	(0.006)	(0.056)	(0.000)	(0.083)
Test 2	15.43	12.72	21.61	9.66
p-value	(0.000)	(0.000)	(0.013)	(0.000)

This table reports estimation results of Eq. (3) from 1998 until 2004. The dependent variable is the monthly excess returns of value weighted portfolios in U.S. dollars. Explanatory variables are described on Table 2. oil futures price increases (*oilp*), oil futures price decreases (*oiln*), net oil futures price increases (*nopi*), net oil futures price decreases (*nopd*), scaled oil futures price increases (*sopi*), scaled oil futures price decreases (*sopd*). Robust p-values are reported below coefficients. Test 1 corresponds to the null hypothesis $H_{01} : \beta_{oilp} = \beta_{oiln}$ and test 2 to $H_{02} : \beta_{oilp} = 0$ and $\beta_{oiln} = 0$, and analogously for the other measures of asymmetry

Table 8 Asymmetric effects of oil returns: period 2005–2010

	Crude and natural gas producers	Drilling and other services	Petroleum refineries	Machinery and equipment
SIC codes	1311	1381	1382	1389
Panel A oilp and oiln				
oilp	0.637	0.507	0.456	0.353
	(0.000)	(0.003)	(0.000)	(0.046)
oiln	0.258	0.591	0.045	0.703
	(0.084)	(0.000)	(0.706)	(0.000)
Observations	72	72	64	72
R^2	0.709	0.68	0.476	0.679
Test 1	3.4	0.12	4.94	1.7
p-value	(0.070)	(0.731)	(0.030)	(0.197)
Test 2	19.82	21.15	10.36	18.86
p-value	(0.000)	(0.000)	(0.000)	(0.000)
Panel B nopi and nopd				
nopi	0.572	0.327	0.245	0.189
	(0.002)	(0.269)	(0.263)	(0.573)
nopd	0.011	0.008	0.060	0.063
	(0.783)	(0.892)	(0.198)	(0.157)
Observations	72	72	64	72
R^2	0.577	0.474	0.399	0.504
Test 1	8.04	1.01	0.57	0.13
p-value	(0.006)	(0.318)	(0.455)	(0.720)
Test 2	6.68	0.75	3.48	1.62
p-value	(0.002)	(0.475)	(0.038)	(0.205)
Panel C sopi and sopd				
SIC codes	1311	1381	1382	1389
sopi	0.054	0.045	0.037	0.032
	(0.000)	(0.003)	(0.001)	(0.043)
sopd	0.033	0.060	0.016	0.075
	(0.023)	(0.000)	(0.157)	(0.000)
Observations	72	72	64	72
R^2	0.719	0.683	0.489	0.696
Test 1	1.17	0.48	1.25	3.07
p-value	(0.283)	(0.493)	(0.269)	(0.085)
Test 2	19.97	24.19	12.77	28.42
p-value	(0.000)	(0.000)	(0.000)	(0.000)

This table reports estimation results of Eq. (3) from 2005 through 2010. The dependent variable is the monthly excess returns of value weighted portfolios in U.S. dollars. Explanatory variables are described on Table 2. oil futures price increases (*oilp*), oil futures price decreases (*oiln*), net oil futures price increases (*nopi*), net oil futures price decreases (*nopd*), scaled oil futures price increases (*sopi*), scaled oil futures price decreases (*sopd*). Robust p-values are reported below coefficients. Test 1 corresponds to the null hypothesis $H_{01}: \beta_{oilp} = \beta_{oiln}$ and test 2 to $H_{02}: \beta_{oilp} = 0$ and $\beta_{oiln} = 0$, and analogously for the other measures of asymmetry

Table 9 Risk factors of the oil industry using oil spot returns

	Crude and natural gas producers		Drilling and other services		Petroleum refineries		Machinery and equipment	
SIC codes	1311		138		2911		3533	
market	0.614	0.564	0.953	0.913	0.531	0.500	1.081	0.946
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
exchrate	0.659	0.693	0.396	0.347	0.078	0.073	0.319	0.406
	(0.000)	(0.000)	(0.120)	(0.184)	(0.492)	(0.554)	(0.306)	(0.203)
int_rate	−1.380	−0.399	0.343	1.097	1.484	2.615	1.552	2.360
	(0.234)	(0.819)	(0.840)	(0.684)	(0.153)	(0.111)	(0.506)	(0.492)
oil	0.272	0.227	0.337	0.317	0.148	0.126	0.437	0.417
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
vol_oil	−0.054		−0.043		−0.007		0.017	
	(0.618)		(0.777)		(0.952)		(0.940)	
gas		0.114		0.118		0.051		0.131
		(0.000)		(0.000)		(0.000)		(0.001)
vol_gas		0.013		0.111		0.035		0.100
		(0.828)		(0.052)		(0.455)		(0.236)
Constant	0.023	0.013	0.014	−0.013	0.005	−0.005	0.009	−0.009
	(0.037)	(0.317)	(0.345)	(0.417)	(0.623)	(0.660)	(0.672)	(0.642)
Observations	269	201	275	205	266	196	275	205
R^2	0.533	0.633	0.471	0.551	0.375	0.411	0.384	0.473

This table reports estimation results of equations ((1) and (2)) from 1988:02 through 2010:12. The dependent variable is the monthly excess returns of value weighted portfolios in U.S. dollars. Explanatory variables are described on Table 2. Oil (natural gas) price variations are computed using oil (natural gas) spot prices. Robust p-values are reported below coefficients

oilp and *oiln*, we find asymmetric effects for all industry segments. Thus, the market seems to react to contemporaneous information on spot prices. However, exogenous measures like net price variations and scaled price variations confirm asymmetric effects only for oil producers and petroleum refiners.

The procedure is repeated for natural gas. Spot prices are from Henry Hub, Louisiana (*gas*) and the settlement price for the NYMEX natural gas futures contract. Table 10 presents the results and all coefficients are similar to those obtained previously. Table 11 presents the results of asymmetric tests, but we do not find statistical support for the asymmetry hypothesis.

A second test of robustness consists of using the market portfolio computed by Datastream, which weights all stocks of the U.S. market. Results are kept unchanged and we do not present them for the sake of brevity.

Table 10 Results of asymmetric tests of oil using spot prices

	Crude and natural gas producers	Drilling and other services	Petroleum refineries	Machinery and equiment
SIC codes	1311	138	2911	2533
Panel A oilp and oiln				
oilp	0.425	0.490	0.244	0.668
	(0.000)	(0.000)	(0.000)	(0.000)
oiln	0.052	0.068	0.031	0.066
	(0.177)	(0.253)	(0.368)	(0.438)
Observations	269	275	266	275
R^2	0.535	0.459	0.386	0.376
Test 1	16.2	9.21	9.83	8.77
p-value	(0.000)	(0.003)	(0.002)	(0.003)
Test 2	23.16	13.9	14.39	14.87
p-value	(0.000)	(0.000)	(0.000)	(0.000)
Panel B nopi and nopd				
nopi	0.229	0.182	0.068	0.266
	(0.012)	(0.168)	(0.376)	(0.048)
nopd	0.033	0.051	0.025	0.087
	(0.072)	(0.100)	(0.086)	(0.025)
Observations	257	263	254	263
R^2	0.43	0.372	0.315	0.306
Test 1	3.91	0.82	0.26	1.38
p-value	(0.049)	(0.358)	(0.609)	(0.240)
Test 2	7.48	3.5	2.8	6.99
p-value	(0.001)	(0.032)	(0.063)	(0.011)
Panel C sopi and sopd				
sopi	0.038	0.041	0.024	0.058
	(0.000)	(0.000)	(0.000)	(0.000)
sopd	0.017	0.026	0.009	0.031
	(0.016)	(0.028)	(0.023)	(0.039)
Observations	269	275	266	275
R^2	0.562	0.489	0.421	0.414
Test 1	4.11	0.86	5.41	1.4
p-value	(0.044)	(0.354)	(0.021)	(0.239)
Test 2	35.66	22.06	37.55	20.76
p-value	(0.000)	(0.000)	(0.000)	(0.000)

This table reports estimation results of Eq. (3) and analogously for the other measures of asymmetry from 1988:02 through 2010:12. The dependent variable is the monthly excess returns of value weighted portfolios in U.S. dollars. Explanatory variables are on Table 2. Asymmetry measures are oil price increases (*oilp*), oil price decreases (*oiln*), net oil price increases (*nopi*), net oil price decreases (*nopd*), scaled oil price increases (*sopi*), scaled oil price decreases (*sopd*). Robust p-values are reported below coefficients. Test 1 corresponds to the null hypothesis H_{01} : $\beta_{oilp} = \beta_{oiln}$ and test 2 to $H_{02} : \beta_{oilp} = 0$ and $\beta_{oiln} = 0$, and analogously for the other measures of asymmetry

Table 11 Results of asymmetric tests of natural gas using spot prices

	Crude and natural gas producers	Drilling and other services	Petroleum refineries	Machinery and equipment
SIC codes	1311	138	2911	2533
Panel A gasp and gasn				
gasp	0.117	0.104	0.030	0.130
	(0.001)	(0.072)	(0.182)	(0.076)
gasn	0.110	0.133	0.073	0.133
	(0.004)	(0.000)	(0.014)	(0.007)
Observations	201	205	196	205
R^2	0.633	0.551	0.414	0.473
Test 1	0.01	0.16	0.90	0.00
p_value	(0.918)	(0.690)	(0.344)	(0.979)
Test 2	15.88	11.82	7.13	7.23
p_value	(0.000)	(0.000)	(0.001)	(0.009)
Panel B ngpi and ngpd				
ngpi	0.142	0.162	0.033	0.140
	(0.063)	(0.137)	(0.249)	(0.274)
ngpd	0.014	0.013	0.004	0.024
	(0.135)	(0.377)	(0.605)	(0.153)
Observations	190	193	184	193
R^2	0.573	0.519	0.37	0.438
Test 1	2.61	1.75	0.76	0.78
p_value	(0.108)	(0.187)	(0.385)	(0.379)
Test 2	4.31	1.9	1.4	2.04
p_value	(0.015)	(0.152)	(0.250)	(0.133)
Panel C sgpi and sgpd				
sgpi	0.020	0.016	0.005	0.019
	(0.002)	(0.116)	(0.306)	(0.131)
sgpd	0.027	0.030	0.017	0.032
	(0.000)	(0.000)	(0.003)	(0.002)
Observations	201	205	196	205
R^2	0.637	0.54	0.414	0.466
Test 1	0.47	1.17	1.99	0.54
p_value	(0.495)	(0.281)	(0.160)	(0.464)
Test 2	21.79	10.2	8.12	7.56
p_value	(0.000)	(0.000)	(0.000)	(0.001)

This table reports estimation results of Eq. (3) and analogously for the other measures of asymmetry from 1988:02 through 2010:12. The dependent variable is the monthly excess returns of value weighted portfolios in U.S. dollars. Explanatory variables are on Table 2. Asymmetry measures are natural gas price increases (gasp), natural gas price decreases (*gasn*), net natural gas price increases (*ngpi*), net natural gas price decreases (*ngpi*), scaled natural gas price increases (*sgpi*), scaled natural gas price decreases (*sgpd*). Robust p-values are reported below coefficients. Test 1 corresponds to the null hypothesis $H_{01} : \beta_{oilp} = \beta_{oiln}$ and test 2 to $H_{02} : \beta_{oilp} = 0$ and $\beta_{oiln} = 0$, and analogously for the other measures of asymmetry

8 Conclusion

This work analyzes the drivers of firm value along the value chain of the oil industry. We find some similarities in the upstream and downstream sectors, such as, similar levels of sensitivity to market returns and oil price shocks. The upstream sector is also sensitive to exchange rate shocks, while oil related services and machinery have greater sensitivity to the state of the market and the price of oil as revenues are driven by the production of oil and natural gas.

With regard to asymmetric effects, we find that the upstream and downstream sectors are affected differently by positive and negative oil price variations, which suggests that firms are able to capture the value generated by oil price hikes. Notwithstanding, in the upstream segment the evidence supporting asymmetric effects of oil is more consistent across different measures and time subperiods. It might be that the petroleum refining industry also shows asymmetric effects regarding oil subproducts such as gasoline, but this remains for future research.

References

Adler M, Dumas B (1983) International portfolio selection and corporation finance: a synthesis. J Finan 46:925–984

Al-Mudaf A, Goodwin TH (1993) Oil shocks and oil stocks: an evidence from 1970s. Appl Econ 25:181–190

Asche F, Gjolber O, Volker T (2003) Price relationships in petroleum markets: an analysis of crude oil and refined product prices. Energy Econ 25:289–301

Basher SA, Sadorsky P (2006) Oil price risk and emerging stock markets. Glob Finan J 17:224–251

Bernanke BS (1983) Irreversibility, uncertainty, and cyclical investment. Q J Econ 98:85–106

Boyer MM, Filion D (2007) Common and fundamental factors in stock returns of Canadian oil and gas companies. Energy Econ 29(3):428–453

Chen N-F, Roll R, Ross S (1986) Economic forces and the stock market. J Bus 59:327–383

Cong R-G, Wei Y-M, Jiao J-L, Fan Y (2008) Relationships between oil price shocks and stock market: An empirical analysis from China. Energy Policy 36:3544–3553

Driesprong G, Jacobsen B, Maat B (2008) Striking Oil: Another Puzzle. J Finan Econ 89:307–327

El-Sharif I, Brown D, Burton B (2005) Evidence on the nature and extent of the relationship between oil prices and equity values in the U.K. Energy Econ 27:810–830

Elyasiani E, Mansur I, Odusami B (2011) Oil price shocks and industry stock returns. Energy Econ 33:966–974

Faff R, Brailsford T (1999) Oil price risk and the Australian stock market. J Energy Finan Dev 4:69–87

Ferson W, Harvey C (1994) An exploratory investigation of the fundamental determinants of national equity market returns. In: Frankel J (ed) The internationalization of equity markets. University of Chicago Press, Chicago

Girma PB, Paulson A (1999) Risk arbitrage opportunities in petroleum futures spreads. J Futur Markets 19:931–955

Hamilton J (1996) This is what happened to the oil price-macroeconomy relationship. J Monet Econ 38:215–220

Hammoudeh S, Dibooglu S, Aleisa E (2004) Relationship among US oil prices and oil industry equity indices. Int Rev Econ Finan 13:427–453

Haushalter D, Heron R, Lie E (2002) Price uncertainty and corporate value. J Corp Finan 8:271–286
Huang R, Masulis R, Stoll H (1996) Energy shocks and financial markets. J Futur Markets 16:1–27
Jin Y, Jorion P (2006) Firm value and hedging: evidence from U.S. oil and gas producers. J Finan 61:893–919
Jones C, Kaul G (1996) Oil and stock markets. J Finan 51:463–491
Karolyi A, Stulz R (2003) Are financial assets priced locally or globally? In: Constantinides MHD, Stulz R (eds) Handbook of the economics of finance. North-Holland, Amsterdam
Kilian L (2008) The economic effects of energy price shocks. J Econ Lit 46:871–909
Lee K, Ni S (2002) On the dynamic effects of oil price shocks: a study using industry level data. J Monet Econ 49:823–852
Lee K, Ni S, Ratti R (1995) Oil shocks and the macroeconomy: the role of price variability. Energy J 16:39–56
Mackay P, Moeller S (2007) The value of corporate risk management. J Finan 62:1379–1419
Nandha M, Faff R (2008) Does oil move equity prices? A global view. Energy Econ 30:986–997
Oberndorfer U (2009) Energy prices, volatility, and the stock market: evidence from the Eurozone. Energy Policy 37:5787–5795
Park J, Ratti RA (2008) Oil price shocks and stock markets in the U.S. and 13 European countries. Energy Econ 30:2587–2608
Ramos S, Veiga H (2011) Risk factors in oil and gas industry returns: international evidence. Energy Econ 33:525–542
Ramos S, Veiga H (2013) Oil price asymmetric effects: Answering the puzzle in international stock markets. Energy Econ 38:136–145
Sadorsky P (1999) Oil price shocks and stock market activity. Energy Econ 21:449–469
Sadorsky P (2001) Risk factors in stock returns of Canadian oil and gas companies. Energy Econ 23:17–21
Sadorsky P (2008) Assessing the impact of oil prices on firms of different sizes: It's tough being in the middle. Energy Policy 36:3854–3861
Scholtens B, Wang L (2008) Oil risk in oil stocks. Energy J 29:89–111
Serletis A (1994) A cointegration analysis of petroleum futures prices. Energy Econ 16:93–97
Solnik B (1974) The international pricing of risk: an empirical investigation of the world capital market structure. J Finan 29:365–378
Tufano P (1998) The determinants of stock price exposure: financial engineering and the gold mining industry. J Finan 29:1015–1052

Cost, Risk-Taking, and Value in the Airline Industry

Paul A. Laux, He Yan and Chi Zhang

Abstract This chapter develops empirical measurements of the shape of airline firms' cost functions as they relate to price variation of oil-based inputs and outputs during the 1998–2009 periods. Using the estimates, we assess the value-added potential for hedging and risk taking with respect to oil prices. We find reasons to believe that the potential value-added of hedging fuel costs with oil derivatives is somewhat limited on average, but that it varies across the business cycle. Our evidence helps explain why, although many airlines hedge, also many do not hedge, why hedging is incomplete, and why hedging intensity varies over time within many airlines.

Keywords Airlines · Risk management · Hedging · Industry studies

1 Introduction

This chapter develops empirical measurements of airline firms' cost functions as they relate to price variation of oil-based inputs and outputs during the 1998–2009 periods. Using the estimates, we assess the value-added potential for hedging and risk taking with respect to input prices.

P.A. Laux (✉)
Department of Finance, University of Delaware, Newark, DE 19707, USA
e-mail: laux@udel.edu

H. Yan
Department of Applied Economics and Statistics, University of Delaware, Newark, DE 19707, USA
e-mail: yhoo@udel.edu

C. Zhang
Department of Finance, Temple University, Philadelphia, PA 19122, USA
e-mail: tue82892@temple.edu

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_2

The potential for corporate risk management to add value is a classic issue in finance. Under the perfect markets reasoning of Modigliani and Miller (1958), that potential is nonexistent because investors can replicate the firm's risk management choices and arbitrage away any benefits. In more realistic settings, Smith and Stulz (1985) show that concavity in the relationship of a firm's value with respect to a particular source of uncertainty opens the door to valuable risk management. Intuitively, if the probability-weighted downside effect on value when the uncertainty is resolved unfavorably more than offsets the upside effect of favorable resolution, then it can make sense to lay off the risk and take the value outcome associated with the expected value outcome of the risk driver. Mathematically, the point follows from Jensen's inequality.

Smith and Stulz (1985) focus on specific examples such as progressive corporate tax codes (the tax bite is disproportionately greater on the pre-tax profits upside) and financial distress costs (the distress costs are disproportionately greater on the profits downside). Investors cannot replicate the firm's managed tax or financial distress risk positions.[1] The converse reasoning also applies. If the probability-weighted upside effect on firm value when a risk is resolved favorably is greater than the probability-weighted downside effect if the risk is resolved badly, then expected value would not be enhanced by hedging. In that instance, it is better to let the risk take its course and accept the average results over time.

Froot et al. (1993) powerfully extend this intuition. Assuming that the availability of internal financing best enables firms' optimal real investment plans, then it is valuable to hedge risks that tend to restrict internal funds available at times when investment opportunities are apt to arise. The key to valuable risk management is the correlation between the risk source and the firm's investment opportunity set. A positive correlation between the value-effect of the risk outcome and the investment opportunity set implies that hedging can be beneficial. In other words, the value of hedging arises from an increased cost of capital (i.e., due to the need to finance externally) if risks turn out badly. The managerial implication is this: the cost drivers that ought to be hedged are the ones that bite harder at times when important investment projects ought to be undertaken. Similar to the Smith and Stulz (1985) reasoning, non-linearities are seen to be at the heart of the potential for value-added risk management. Under the Froot et al. (1993) logic, the non-linearities operate in the sense of shifts in the intensity of risk exposures across economic states. Firms with cost functions that are effectively convex because of greater sensitivity to a detrimental risk source during high-cost times would then benefit from hedging.

This chapter works directly from these seminal arguments to characterize the possibilities for and achievement of value-added risk management in an important industry facing a specific risk. Our focal industry is commercial airlines. Airlines

---

[1] Even so, if the firm does enact hedges, the value added would be reduced to the extent that the risk being hedged corresponds to a priced factor in the security markets. Effectively, the firm would need to pay an insurance premium for the risk it lays off; this effect would be offset if the firm's reduced-risk profile results in a lower discount rate being applied to its cash flows by the market.

are an interesting case because the direct effect of a clearly identifiable and economically predominant source of risk resides squarely within the cost function. There is no offset in revenue functions (unlike for oil producers, for example) so value effects from costs feed directly into equity value.

Most directly, the risk source is fuel costs. Jet fuel is, of course, a derivative product of crude oil, so airlines indirectly face oil price risk. There are reasons to expect that airlines' fuel costs might be convex in oil price (i.e., absent any hedging). For example, oil prices, being generally pro-cyclical in recent times, tend to be highest when airline demand is strong. Airlines are therefore apt to use more high-priced fuel than low-priced fuel over time. Airlines can and do raise their prices when fuel cost is high, of course, but this offsetting benefit is limited by the elasticity of demand. Second, because airlines cannot shift away from jet fuel, they will bear higher costs when refining margins widen, which tends to happen at times of high oil prices. Finally, cost functions could be convex to the extent that fuel cost spikes correspond to upturns in economic activity overall (due to demand pressures on oil-related prices), straining airlines' capacity to deliver their services given their level of fixed capital.

On the other hand, it is also plausible that cost functions might be concave. For example, cost functions would be concave to the extent that airlines can use oil price spikes as leverage to negotiate reductions in other cost factors (including by invoking bankruptcy). Moreover, if oil price shocks are an underlying cause of recessions, as described by, for example, Hamilton (2008) and Kilian (2008), then times of high oil prices might correspond to a weak investment opportunity set. Finally, cost functions could even be be roughly linear to the extent that airlines can operationally offset these various economic effects of variable fuel costs. A priori, then, it seems possible that either risk-taking or hedging with respect to oil prices might be the value-added recommendation for airlines.

Airlines do make fairly extensive use of energy-linked futures, forwards, options and swaps, suggesting that they are actively managing their operational risks with financial instrument hedges.[2] Importantly, however, it is not typically practical to hedge jet-fuel exposure directly, except perhaps over short horizons. In practice, airlines resort to hedging via derivatives linked to other oil markets, such as crude oil or refined distillate such as heating oil or even gasoline, all of which exhibit greater liquidity over a larger range of expiry dates. The implication is that airlines necessarily remain exposed to the basis risk. The essence of their basis risk in the case of jet fuel is essentially the time-profile of the refining margin between crude and jet fuel, or the time-profile of the price differential between other refined distillates and jet fuel. Thus, it is far from clear that risk management with oil derivatives is sure to add value.[3]

---

[2] Airlines typically state in annual reports that their derivatives use is for hedging purposes.

[3] News reports sometimes emphasize this point of view. For example, Freed (2012) reports that hedging losses turned a strong second quarter of 2012 operating profit into a loss overall for Delta Airlines.

Perhaps consistent with limits to value-added hedging, there does not seem to be any standard for hedging within the industry. Instead, hedge ratios vary across firms and time. News reports often put airlines hedging activities in the 20–40 % range [for example, see Peterson and Reiter (2008)]. Morrell and Swan (2006) report on hedging at a selection of airlines and over time. In 2004, hedge ratios varied cross sectionally from 0 to 82 % of fuel purchases. Over 1989–2003, American Airlines hedge ratio varied in the time series from 12 to 48 % of fuel purchases. The International Air Transport Association (IATA), which represents airlines around the globe, estimated that carriers would hedge 30 % of their fuel purchases in 2011, up from 10–20 % the previous year. Additionally, airlines are said to sometimes be discouraged from hedging on account of poor outcomes due to basis divergence between jet fuel and oil-related prices in the specific hedging markets [see, for example, Blas and Clark (2011) and Credeur et al. (2011)]. Morrell and Swan (2006) quote the CEO of British Airways as saying in a news report that no "sensible airline" believes hedging saves on fuel bills. For a recent industry-oriented discussion of airline hedging practice, considering both pros and cons, see Rivers (2012).

Building on the core ideas from risk management theory, our goal is to assess the potential for value-added oil price hedging/risk-taking in the airline industry, and develop evidence on whether the potential is being realized. To accomplish this, we use data on airlines' physical fuel consumption, market prices, reported fuel costs, and reported non-fuel costs. We estimate three industry-level cost functions for airlines. Specifically, these are unhedged fuel cost functions, fuel cost functions that include important hedging effects, and total cost functions.[4]

Estimated unhedged fuel cost functions are concave in oil prices. To the extent that our estimated functions fully capture the shape of actual cost functions, the unhedged cost function estimates suggest that airlines' values would benefit from *not* hedging the risk.

A more conservative interpretation must acknowledge that our estimated functions can only map the shape of estimated functions within the range of the data, and hedging could be valuable because of the possibility of outcomes that are scarce in our sample. Additionally, our simple cost function specifications might be incorrect. To the extent that excluded influences correlate to the oil price factors we do include, actual cost functions could be less concave than we estimate.

Even acknowledging all this, a comparison across unhedged and hedged cost functions is highly informative. Our estimates imply that airlines do change the sensitivity and shape of their fuel cost functions by hedging. Unhedged fuel cost functions are flatter and less concave than hedged fuel cost functions. Moreover, oil price variation explains only about half as much of the variation in hedged fuel

---

[4] We measure costs relative to the asset size of the firm. Not only does this facilitate comparisons across airlines, which vary dramatically in size, but it also makes our estimates more economically meaningful. To the extent that a firm can adjust its asset base over time to match the rise and fall of its dollar costs, then it essentially already hedged against changes in the investment opportunities that those assets represent.

costs as compared to unhedged costs. Total costs are even less closely linked to oil prices. While it is natural that oil would explain less variation in total costs, given that non-fuel costs are also important, the distinction goes even further in that it is difficult to reject the null hypothesis of no oil price linkage. The clear implication is that both hedging activities and outcomes on non-fuel cost factors tend to offset the natural effect of oil prices on airlines unhedged fuel costs.

Further evidence consistent with this interpretation of strong hedge effects on costs comes from estimates of quarterly seasonals in airlines' oil costs. Controlling for other factors, unhedged fuel costs are significantly higher in the third quarter (summertime) versus any other quarter of the year. This makes sense, as the press of business is traditionally greatest in the summertime, and it seems natural that resources usages would be stretched beyond their most efficient levels. In contrast, our evidence reveals no such pattern in hedged fuel cost or total cost. The implication is again that hedging tends to offset the natural effects of oil price on unhedged fuel cost.

We also find that airline costs tend to increase at the same time they invest the most in fixed capital. This suggests that management perceptions of investment opportunities are more optimistic during high-cost periods. This is about equally so whether we consider unhedged fuel costs, hedged fuel costs, or total costs. To the extent that growth in fixed capital reflects perceived investment opportunities, this comparison means airline managements do not hedge, on average, in a pattern that corresponds to changes in investment opportunities. Since airline investment opportunities may vary business conditions over time and across different classes of firms at a point in time, we double-check that our findings are robust to including controls for GDP and market share. We find the results robust.

We develop additional evidence to understand the potential for value-added by risk management, and indications as to whether the potential is realized, by considering airlines cost functions across regimes that likely correspond to different investment opportunities. Specifically, we consider high versus low oil price periods, and expansion versus contraction macroeconomic phases. If recessions tend to follow high oil price periods [see, for example, Hamilton (2008); Kilian (2008)], then high oil price periods signal weak investment opportunities. In principal recession periods themselves might correspond to weak opportunities (if the bad conditions are secular) or strong ones (if the bad conditions will soon lead to strong business conditions) or even a mix (if different recessions are, in fact, qualitatively different). Campbell et al. (2012) study recent decades from this point of view. Taking stock market booms-busts as a rough economic indicator, their evidence shows that downturns in the early 1990s were most connected with weak market-wide cash flow news (i.e., secular bad news). For a short time in early 2000s, downturns were most connected with weak market wide sentiment (i.e., temporary bad news). In the late 2000s, the cause was again bad cash-flow news. Thus, during our sample period most stock market downturns were linked to weak cash flow news, a secular problem that does not sow the seeds of its own reversal. Airlines are a very pro-cyclical industry. Viewed through this lens, recessions during our sample likely correspond mostly to weak investment opportunities for airlines.

We find that unhedged airline cost functions are less concave and less sensitive to oil price during sustained periods of high oil prices, and similarly during recessions. From the reasoning above, airlines investment opportunities were probably weak during both high oil price times and recession times. The implication for risk management from this second point of view [i.e., motivated by Froot et al. (1993)-type reasoning] contrasts somewhat with that obtained from simply examining the curvature of the cost function [i.e., motivated by Smith and Stulz (1985)-type reasoning]. Unhedged airline cost functions depend most on oil during strong investment opportunity periods–i.e., when low cash flows are most likely to constrain important future-oriented choices. The strength of this implication is offset somewhat in that unhedged cost functions are also the most concave at such times. We can see why different airlines might choose differently on risk management, and why the intensity of industry hedging would vary over time. In fact, hedge fuel cost function are different during recessions (versus expansions), but are not different during high oil price periods, consistent with such heterogeneity in managers' choices.

Overall, the statistical evidence suggests that airlines can and do enact operational and hedging mechanisms for dealing with price variability that blunts the cost effect of fuel price spikes, on average. It is less clear that these activities add to value. On average, the shape of estimated cost functions suggests that it would be better not to hedge, though our estimates also suggest that hedging at some times and under some conditions can be more beneficial. Our results help provide an explanation for the fact that airlines tend not to fully hedge their fuel costs, and why their hedging behavior shifts around over time, even though a variety of fairly appropriate financial contracts are available.

A related strain of the empirical corporate risk management literature focuses on firms' use of derivatives. Such studies assess the extent to which firms use derivatives or whether that use succeeds in reducing risk. Overall, these studies establish extensive derivatives use, but findings of value added are, in general, less conclusive. Some of these studies have focused on airlines in particular. For example, Carter et al. (2006) have investigated the relationship of stock price to reported derivatives usage and firm's 10-K statements about the extent of hedging, finding the relationship to be positive. Another strain focuses on risk exposures, assessing whether firms show stock price exposure to specific and intuitive sources of risk (gold for gold miners, exchange rates for international firms, and so on). These studies are informative about firms actions and their outcomes. They have limited potential, however, to inform about the potential for risk management to add value.

We are not the first to apply the risk management theory to understand cost functions as we do. Our paper adds to a strain of the risk-management literature, exemplified by Mackay and Moeller (2007), that takes a very different and more direct approach from the papers mentioned above. Mackay and Moeller use the core risk management theory to motivate an extensive and insightful set of measurements along the same line as in our study, but for oil refiners. Our contribution for airlines is useful for the reasons outlined above, and additionally because it makes use of a special industry dataset for the first time we are aware in the literature. It is this dataset that allows us to compare unhedged costs versus hedging-influenced costs. At the same time, the nature of the dataset limits us from using econometric methods as sophisticated as Mackay and Moeller's, we are more possibly subject to statistical biases, and thus our conclusions must be treated with some caution.

The remaining sections of this paper provide an informal discussion on the economics of hedging in the airline industry, describe our data set, report and discuss our empirical results, and conclude, respectively.

2 Airline Costs and Airline Risk Management

Airlines face substantial risk from many external sources, including jet fuel price volatility, interest rate and foreign currency changes, and macroeconomic revenue drivers. Among these, fuel price risk may be the most severe, at least over short periods, for two reasons. First, fuel prices are highly volatile. Second, fuel is the largest or second largest cost for most airlines. Due to the competitiveness of the industry, it is not always possible to pass higher fuel prices on to passengers by raising ticket prices over the short run. This suggests that fuel risk management might be a central issue for airlines. Airlines that want to stabilize operating expenses and assure bottom line profitability might seek to hedge fuel price exposure. Airlines that hedge could do so using either operational hedging mechanisms or financial derivatives mechanisms.

Operational hedging mechanisms include engaging in long-term contracts for fuel purchases, attempting to raise ticket prices in response to high fuel prices, and flying slower or less into-the-wind to preserve fuel when fuel is expensive. Airlines may also engage in some operational practices that have the same effect as forward contracts. For example, some airlines negotiate fuel pass-through arrangements with other airlines, whereby a larger airline assumes the risk of fluctuating fuel prices and shields a smaller airline. One major airline has even acquired a refinery as an operational hedge (Staff 2012).

Financial derivatives hedging mechanisms include futures, options, swaps and collars on jet fuel or other petroleum products such as crude oil, heating oil, or even gasoline. Hedging with jet fuel derivatives tends to be limited in quantity and time-horizon. Among the reasons are that over-the-counter derivatives do not trade in sufficient quantities to hedge all of the airlines jet fuel consumption, and jet fuel derivatives markets are rather illiquid in general. Further, no exchange-traded derivatives for jet fuels exist in the United States.

When refiners process crude oil, the main products are gasoline, middle distillates (heating oil, diesel fuel, and jet kerosene) and residual fuel oil. Since jet fuel is refined from crude oil, and heating oil is from the same part of the barrel during refining process, both of them are among the top choices to hedge jet fuel prices. Historically, most U.S. airlines have tended to hedge their exposure to energy costs mostly through the Chicago Mercantile Exchange futures contract on West Texas Intermediate crude and the NYMEX futures contract on heating oil. Options have been somewhat less popular, perhaps because of the cash outlay required to cover option premia. Most hedging is in plain-vanilla contracts, though exotics do exist. Such hedges are not perfect, and significant basis risk remains. For example, in the late 2000s, jet fuel tended to track more closely to Brent crude prices, even though U.S. airlines' hedging was concentrated in derivatives settled to West Texas Intermediate crude prices. Getting the right mix of hedging strategies is claimed to be especially difficult.

As discussed in the introduction, finance theory indicates that risk-management with financial contracts and instruments can have value-added under at least two circumstances. First, risk management theory implies that the potential for that risk taking (hedging) to add value when cost functions are concave (convex) in the underlying source of risk, all else equal. Second, theory implies that valuable risk management adjusts the correlation of internal cash flows to investment opportunities to free the firm from dependence on more-costly external capital. Mackay and Moeller (2007) provide a rigorous example of the application of this theoretical reasoning to assess the potential for hedging value-added in an industry. They study oil producers over the period 1985–2000 to assess the extent to which cost, revenue, and profit functions are concave or convex in the price of oil. They report a two-edged potential for value added hedging in that both revenue and cost functions are concave. The recommendation for hedging policy, based on theory, would therefore be: hedge the revenues and leave the costs unhedged. Our study is an application of their basic idea to the airline industry. Because the oil risk exposure of the airline industry is essentially on the cost side, we focus there.

Some existing evidence suggests that such risk management adds value for airlines. For example, using a panel-data design, Carter et al. (2006) assess the relation between airlines hedging intensity, as reported in financial statements, and Tobin's q (an index of firm value in excess of replacement cost). They conclude that airline firm value is positively related to hedging of future jet fuel needs. Their study includes controls for investment opportunities and derivatives usage overall. Studies like this would seem to establish that hedging is valuable in airlines. Yet airlines do not uniformly or completely hedge their fuel risk, according to news sources such as those referenced in the introduction. And there seems to be some pattern to the time series variation, where hedging increases as oil prices rise. For example, hedging was said to be more widespread during the pre-financial-crisis global run-up in oil prices, and many airlines ceased to hedge after prices fell in the crisis.

Other research (not focused on airlines) finds that it can be difficult to conclude whether firms are hedging or speculating using data that selectively characterizes their actions (e.g., derivatives use). As argued in, for example, Faulkender (2005), the problem is that a firm's risk position is the amalgam of its real and financial market activities and choices. Focusing on, say, derivative market activities might mask other offsetting choices. In airlines, there is some reason to think that success in real activities correlates negatively to oil prices as far as costs are concerned, which alleviates some concerns from this perspective.[5] However, the point only pertains to the cost side. If high oil prices and strong economic activity tend to coincide, then airlines total value may be more positively correlated with oil prices. This reasoning is another motivation for our focus on the potential for risk management value added from the cost side.

Finally, it is worth noting that although classic empirical results in finance such as Chen et al. (1986) suggest that oil price risk is not a priced factor, some recent results such as Chiang et al. (2012) provide contrasting evidence. If systematic oil risk is priced in the financial markets, then widely-held firms that hedge it may be forced to pay as much for the insurance as it is worth to their investors. The best chance to understand the value of hedging might then be in private firms, whose owners may be less fully diversified. There the benefit might be perceived to outweigh the cost. One attractive feature of our data set is that it includes private firms. With this in mind, we turn to a discussion of our data.

3 Data

Our central data source is from the Research and Innovative Technology Administration (RITA), which provides a database suite organized by U.S. Department of Transportation. As regulated carriers, airlines are obliged to report a wide variety of operating, safety, ownership and financial data to the Department of Transportation on a quarterly basis. Among various databases from RITA, the central one for our purpose is the Air Carrier Financial Reports (Form 41 Financial Data) database, which provides detailed financial information on public and private airline companies. We have downloaded quarterly balance sheets (Schedules B-1 and B-1.1) and quarterly income statement (Schedules P-1.1 and P-1.2). In addition, we have downloaded quantity data on airlines' jet fuel usage. After consolidating the various tables from RITA and eliminating those with extensive missing data, our sample includes 141 airline companies. The sample, based on the availability of RITA data, covers the first quarter of 1990 through the fourth quarter of 2010.

The RITA data provides two special advantages for a study like ours. The first is universality of coverage. All commercial airlines in the U.S. are regulated by the Department of Transportation, and must report their data. This means that we are able to include unlisted firms and smaller airlines that would be missed using other data sources. Second, the RITA data includes information on physical fuel usage as well as dollar fuel expense from financial statements. With this extra information, we can compute the fuel cost airlines would face if fully unhedged. Thus, we can assess the fuel cost risk that is inherent in their production process by calculating

[5] That is, if oil price is high at the time of high demand by Western macroeconomies, and therefore high demand for air travel also.

their "unhedged fuel cost" as the product of physical fuel usage and the market price of jet fuel.

From RITA financial statements, we also obtain another measure of fuel cost– one that is affected by hedging. For convenience, we will refer to this as the "hedged fuel cost". Our terminology should be understood in light of GAAP hedge accounting rules. Generally Accepted Accounting Principles dictate that derivative assets, such as futures, options, and swaps on oil and oil products, should be marked to market on a regular basis. Gains and losses on every position in such assets must be reflected directly on the income statement, *unless* the specific position is pre-qualified for special hedge accounting treatment, and *unless* the position is periodically tested to assure that it continues to qualify. If the formal pre-qualification and the continuing qualification requirements are satisfied, then the gains or losses due to the derivative can be held away from the core of the income statement, and do not immediately affect net income (for example, being reflected in the broader "other comprehensive income" category). Under hedge accounting treatment, the qualified derivative asset gains or losses flow to the net income only at the same time as the realization of the cash flows that motivated the hedge. Thus, under hedge accounting treatment, the income (loss) due to the fundamental business activity associated with a hedge would be realized in income at the same time as the loss (income) associated with the offsetting hedge (except to the extent that the hedges are judged "ineffective"). The result of this offset is the source of our term "hedged fuel cost".

Obtaining hedge accounting treatment for derivative positions is a rigorous undertaking for an airline. It requires a fairly sophisticated accounting function within the firm, for hedge accounting is a matter of substantial focus for auditors. Moreover, the accounting rules are complex, require frequent judgment calls and testing, and were under more or less continuous development during the years of our sample. Hedge accounting treatment tends to be used more by larger firms in general, and this seems likely to be the case with airlines as well. Also, hedge accounting treatment requires that the specific derivatives position can be linked in advance to a specific risk, as qualified by specific rules. In practice, the asymmetric nature of options gains and losses on some hedges (options, for example) often disqualifies them from hedge accounting treatment because they do not track symmetric cost effects sufficiently closely.

Importantly for our purposes, even without hedge accounting treatment, a firm that is, in fact, hedging its fuel costs will have somewhat similar gain-loss offsets on the income statement to the extent that it is partially hedged and that its current derivatives realizations (i.e., positions unwound in the quarter) are similar to the mark-to-market effects of its forward hedges. To the extent that firm's hedging intensity is fairly stable over time (quantities and directions) and that oil price changes have a similar effect across the hedging term structure, a firm that follows a consistent policy of hedging will experience cash flows and derivatives gains/losses in offsetting directions. Thus, even for a firm that does not obtain hedge accounting treatment for its positions, the reported fuel cost on the income statement will, to an extent, have the nature of a "hedged fuel cost".[6] For simplicity, we use this term, rather than the more-fully-descriptive but more-cumbersome "hedging-affected fuel cost". The RITA data also provides each airline's total cost, about which similar points can be made.

Table 1 provides summary statistics on some of the central variables for our cost function estimates. For Table 1 only, we report raw measures, i.e., not normalized by asset value. From the table, it is apparent that airlines' sizes differ substantially. For the average firm/quarter over our sample period, sales revenue is about $500 million, with a standard deviation equal to about one-fifth of that average. The smallest firm/quarter observation on revenue in the data is only $560 thousand, and the largest is over $9 billion. Fuel usage also varies greatly, with a mean of almost 77 thousand gallons of jet fuel in a firm/quarter, a standard deviation of about 16 thousand gallons, but a maximum of 737 thousand gallons. Airlines' market shares (based on costs) vary from almost 0 to 4 %. The table also reports some other relevant aspects of airlines' financial statements, and similar wide variation is apparent.

Because we are interested in characterizing costs and potential value effects for the industry overall, we need some way of comparing and summarizing across these disparate-sized firms. Specifically, we need a normalization factor to use in regression analysis. Otherwise, regression error variances would vary according to firm size, violating standard assumptions. We choose to normalize by total asset value, so that the various types of costs are all expressed per thousand dollars of asset value in the regression data. This choice is driven by our purpose. We want to understand how the nature of cost functions and their non-linear sensitivity to oil prices might impact firm value. The most appealing concept of value in this setting is Tobin's q, i.e., firm value per unit of replacement value. With a total assets normalizing factor, we are measuring costs on a similar basis.[7]

4 Cost Function Curvature and Sensitivity to Oil Prices

In this section, we present and discuss the implications of estimates of cost determinants for our three categories of airline cost: unhedged fuel cost, hedged fuel cost, and total cost. We estimate a simple specification of costs, as normalized by total assets. We chose this specification on economic grounds, but also considering

---

[6] There is the logical possibility that it is an over-hedged fuel cost, i.e., that the total effect of the realizations and the marking-to-market takes cost exposure in the opposite direction from its natural one.

[7] Seat-miles, or the number of miles flown times the number of passengers carried on a flight, summed across all flights, is a commonly-used normalizing factor in the industry. However, a seat-mile normalization would leave a systematic firm-size effect, for smaller airlines inherently have fewer seat-miles across which to spread costs that do not vary directly with business activity (such as headquarters costs and even lumpy aviation equipment costs).

Table 1 Summary statistics for the panel of airline firms over 1998–2010

Central variables	Mean	Median	Standard Deviation	Minimum	Maximum
Sales	596,763	103,350	1,262,664	560	8,987,339
Fuel usage	76,700,000	15,900,000	152,000,000	44,784	737,000,000
Hedged fuel cost	118,581	20,180	268,800	0	2,403,209
Total cost	563,169	97,315	1,198,521	413	7,429,244
Fixed capital	1,623,509	83,182	3,573,726	13	19,700,000
Working capital	(143,808)	(2,942)	630,824	(5,334,403)	1,224,158
Capital expenditures	14,540	907	390,976	(10,800,000)	8,785,411
Total debt	869,452	50,885	2,044,791	0	15,900,000
Market share	0.406 %	0.072 %	0.853 %	0.002 %	3.921 %

what can be implemented using the RITA data. As noted above, the data presents a special opportunity to include all US airlines, including privately owned ones, but, at the same time, does not provide for such a broad variable coverage as if we were to restrict the sample to large public companies.

The key cost factors of interest for our purposes are oil price and the square of oil price, including squared oil price is a simple way to allow for non-linearly oil sensitive costs. We also include some additional cost determinants in the specification based on several economic considerations. Bolton et al. (2011), in recent research, reason that firms' cash holding and liquidity management policies interact with risk management to determine firm value. Therefore, we also include a measure of the growth in working capital as a control variable in our specification. Chen et al. (2011) argue that executive compensation is another important element of the linkage to value. Given the nature of our sample, we do not have compensation data available. Relying on the evidence in Emans et al. (2009) that compensation scales up with firm size and at a rate different than costs, we include a measure of the growth in fixed capital in our specification. Working capital and fixed assets also make sense in the cost specification on microeconomic grounds. Working capital relates to the firm's efficiency, and fixed capital relates to economies of scale. Fixed costs are apt to increase when management is optimistic about business opportunities, and so this regressor also ties to hedging considerations. Finally, because the airline industry is subject to large seasonal swings in activity, we include in our cost specification dummy variables for the first, second, and third calendar quarter of the year, leaving the fourth quarter effect to be subsumed in the constant term. Because we use asset-scaled versions of all our measures in regressions, henceforth, we will use italics to indicate when we are referring to a variable that is normalized by assets.

4.1 Base-Case Model Estimates

Table 2 reports regression estimates for our base cost model as applied to the three cost measures, *Unhedged fuel cost*, *Hedged fuel cost*, and *Total cost*, respectively, in columns (1)–(3). Panel A reports OLS estimates with heteroskedasticity and firm-cluster robust standard errors. In Panel A, we report estimates for all regressors, to fully catalog the base case findings.

The estimated coefficients on *Oil price* are positive for all three cost functions, and estimated coefficients on *Squared oil price* are negative for all three cost functions. Thus, our estimated cost functions all tend toward concavity in oil price. Economically, the suggestion is that hedging the cost effects of oil price variation is apt to be counterproductive for value: the extra cost incurred for high oil price outcomes is more than offset with the cost savings for low oil price outcomes.

Several caveats are appropriate. We do have fully-specified cost functions, so we cannot say for sure that costs are literally concave in oil. There might be omitted variables, under the control of the firm, which happen to correlate with the price of

Table 2 Regressions of airlines costs on oil prices

Central variables	(1)	(2)	(3)
Panel A. OLS estimates			
Central variables	*Unhedged fuel cost*	*Hedged fuel cost*	*Total cost*
Oil price	2.825	1.815	0.538
	(3.584)	(1.680)	(0.839)
Squared oil price	−0.266	−0.156	−0.068
	(−2.764)	(−1.157)	(−0.874)
Q(1)	0.050	0.022	0.020
	(1.809)	(0.639)	(1.159)
Q(2)	0.040	−0.013	−0.005
	(1.448)	(−0.377)	(−0.352)
Q(3)	0.063	−0.006	0.003
	(2.594)	(−0.203)	(0.224)
Δ *Fixed capital*	1.884	1.810	2.546
	(3.081)	(2.189)	(7.838)
Δ *Working capital*	0.520	−0.145	−0.391
	(1.024)	(−0.277)	(−0.933)
Constant	−2.673	−7.393	−2.041
	(−1.675)	(−3.403)	(−1.554)
R-squared	0.101	0.048	0.011
F-test (oil effects)	(29.08)	(11.05)	(0.41)
	[0.00]	[0.00]	[0.66]
Observations	2,607	2,607	2,607
Panel B. Selected panel regression estimates			
Oil price	2.861	2.193	0.622
	(5.147)	(2.312)	(1.222)
Squared oil price	−0.258	−0.188	−0.064
	(−3.779)	(−1.595)	(−1.040)
Q(1)	0.034	0.013	0.008
	(1.352)	(0.464)	(0.604)
Q(2)	0.034	−0.002	−0.008
	(1.384)	(−0.056)	(−0.639)
Q(3)	0.044	−0.021	−0.010
	(1.835)	(−0.799)	(−0.903)
Δ *Fixed capital*	1.848	2.305	2.178
	(2.986)	(4.527)	(5.616)
Number of airlines	95	95	95
R-squared	0.294	0.157	0.057
F-test (oil effects)	(62.21)	(33.08)	(2.02)
	[0.00]	[0.00]	[0.14]
Observations	2,607	2,607	2,607

oil within our sample. Therefore, we will focus more on comparisons of the extent of the concavity accords the various types of costs. The assumption is, then, that any such biases are stable across the cost categories, which seems reasonable.

Comparing the coefficients on *Oil price* and *Squared oil price* across the cost function types in Panel A, both the size and statistical significance of the estimated coefficients are attenuated. Additionally, the regression R-squared statistics decline across the cost function types. Finally, the Panel reports F-tests as to the combined influence of the two oil price regressors across each cost type, with the finding that the combined effect is statistically significant for *Unhedged fuel cost* and *Hedged fuel cost* but not for *Total cost*. Overall, the tendency of the cost function toward positive slope and concavity is smaller and weaker as we move from considering *Unhedged fuel cost* to *Hedged fuel cost* to *Total cost*. The implication is that firms in the airline industry, on average, use hedging to offset the sensitivity to oil price that is clearly apparent in their unhedged fuel costs. This is so to some extent as it impacts their hedged fuel costs, and to a more complete extent as it impacts their total cost. Given that the exposure offset involves a concave cost function, this may not be value enhancing.

Coefficients on the seasonal dummies are informative about hedging also. The summer (*Q3*) coefficient is significantly positive in the *Unhedged fuel cost* regression in Panel A, but not for the other cost types. Summer is the time of year when airlines' passenger flow, and therefore fuel demand, is the strongest. It is not surprising that the demand pressure from the prime jet-fuel consuming industry would then lead to higher unhedged fuel costs in the summer: the industry would not be expected to operate its jets most efficiently at its time of greatest strain. These higher costs are apparently offset by financial or operating hedges before impacting *Hedged fuel cost* or *Total cost*, again evidence of industry-wide hedging on average.

The coefficients on one of the control variable provides interesting additional information about airline's hedging choices. The coefficients on Δ *Fixed Capital* in Panel A are not statistically significant for either fuel cost measure, but are positive and significant in the *Total cost* regression. This suggests that airlines' non-fuel costs, but not their fuel costs, are highest at the time they invest in their fixed capital. In the view of airlines' managements, and based on these OLS results, it does not appear that capital investment opportunities are strongest at the time of high fuel costs.

To check the robustness of these findings to reasonable variations in the estimation method, Panel B reports panel regression estimates of the same specification, now including firm fixed effects, with heteroskedasticty-robust standard errors. Findings in Panel B thus rely on time-series effects within each airline, preventing inference from being driven by differences in the economics of different airlines. For example, results on *Squared oil price* effects like those in Panel A could result if airlines that hedge happen to be ones that are also more flexible in dealing with extreme oil prices. In Panel B and subsequent regression tables, we suppress reporting on some control variables coefficients to save space and focus on more central coefficients.

Panel B reports that findings as to oil price effects and seasonal effects are robust to this alternative method. The positive slope of costs with respect to *Oil price* is somewhat more strongly statistically significant. However, findings on the relationship of costs to fixed capital changes are altered. The new finding is that all categories of cost are positively correlated with fixed capital growth. This suggests that management perceptions of investment opportunities are more optimistic during high-cost periods. Because the link between oil prices and investment opportunities is an important consideration under risk management theory, we are motivated to investigate the robustness of our findings extensively.

4.2 Costs and the Firm's Industry and Economic Situation

Strictly speaking, airlines' cost functions are derived from the firm's production function, as conditioned on input prices. In the previous section, we have estimated a version of such a cost function. In this section, we amend the specification to include some useful additional conditioning variables regarding the firm's situation and environment.

We are motivated by the finding in the previous section that airlines' costs vary with their fixed capital growth. Capital investment choices are in turn motivated by firms' situation in the product market. This endogeneity or omitted variable issue could have an effect on our estimates of oil price coefficients. To take an example rooted in the time series of the data, during business cycle upswings, oil prices might tend to be larger at the same time as airlines' capacity is strained with business. The strain might contribute to costs being higher at the time of high oil prices. Alternatively, it might be the case that when firm's product market positions are stronger they are better able to enact flexibilities to deal with high oil prices, leading to reduced costs. Overall, we want to be sure that our estimated oil price cost sensitivity parameters are more than an reflection of these or other similar non-oil effects. Therefore, in this section, we extend our estimates of cost determinants to include recent GDP growth (as a business cycle indicator) and market share (as an indication of firm's product market success).

Table 3 reports the estimates of these extended cost functions. In Panel A, we add *GDP growth* as a regressor, and in Panel B we additionally include *Market share* as a regressor. Our findings above are not strongly driven by endogeneity/omitted variable issues of the type just discussed. The key specific finding, common to both panels, is that conclusions regarding the positive slope and concavity of the various cost functions are unchanged. *Unhedged fuel cost* is increasing in oil price and concave. *Hedged fuel cost* is increasing in oil price, but less reliably concave, suggesting that hedging has reduced some extreme cost outcomes. *Total cost* is not closely related to oil price, suggesting both that oil price sensitivity is offset in other operational ways, and that there are many other costs drivers besides oil prices.

The table also reports that the specific coefficients on *GDP growth* and *Market share* are at most weakly statistically significant, but that the statistical significance

Table 3 Cost functions as extended to include the effects of business cycle and product market situation

Central variables	(1)	(2)	(3)
Panel A. Selected panel regression estimates of cost functions with business cycle effects			
Central variables	*Unhedged fuel cost*	*Hedged fuel cost*	*Total cost*
Oil price	2.540	2.018	0.406
	(4.933)	(2.382)	(0.902)
Squared oil price	−0.213	−0.163	−0.034
	(−3.306)	(−1.561)	(−0.638)
Δ *Fixed capital*	1.863	2.313	2.188
	(3.021)	(4.561)	(5.633)
GDP growth	2.974	1.625	2.002
	(1.490)	(0.591)	(1.596)
Constant	−2.364	−8.050	−2.024
	(−2.284)	(−4.723)	(−2.173)
Observations	2,607	2,607	2,607
R-squared	0.295	0.158	0.058
Number of airlines	95	95	95
F-test (oil effects)	(57.03)	(28.22)	(2.38)
	[0.00]	[0.00]	[0.10]
Panel B. Selected panel regression estimates of cost functions with business cycle and product market effects			
Oil price	2.584	2.079	0.434
	(5.020)	(2.463)	(0.962)
Squared oil price	−0.220	−0.172	−0.039
	(−3.407)	(−1.654)	(−0.713)
Δ *Fixed capital*	1.858	2.306	2.184
	(3.031)	(4.618)	(5.655)
GDP growth	2.970	1.618	1.999
	(1.494)	(0.591)	(1.603)
Market share	27.073	37.578	17.283
	(1.305)	(1.597)	(1.736)
Constant	−2.552	−8.310	−2.144
	(−2.485)	(−4.926)	(−2.289)
Observations	2,607	2,607	2,607
R-squared	0.297	0.161	0.061
Number of airlines	95	95	95
F-test (oil effects)	(54.89)	(26.66)	(2.20)
	[0.00]	[0.00]	[0.116]

of the *GDP growth* effect is stronger for total costs than for fuel costs. This fits with evidence and reasoning in Chung et al. (2012), who find that cost of equity in unionized industries (a group that includes most airlines) has a countercyclical pattern, suggesting that risk is also countercyclical. Their interpretation is that unions are a serious impediment to operating flexibility, preventing firms from making appropriate adjustments in downturns, and that investors rationally take into account in discounting the stock. Pulvino (1998) shows that airlines bear the cost of fire-sales of assets in financial distress, which also suggests that risk is counter-cyclical to the extent that airlines incorporate systematic distress risk. Financial distress might aggravate any negative effects of high oil prices: Morrell and Swan (2006) cite their personal experience with airlines near or in bankruptcy, saying that bad credit make hedging impossible. Our *Total cost* results are weakly consistent with this reasoning, and the fact that we do not find such effects for fuel costs reinforces the notion that labor costs are the underlying source of the effect.

4.3 Airline Costs and the Investment Opportunity Set

4.3.1 Oil Price Regimes

To this point, our results suggest that risk-taking regarding the sensitivity of fuel costs with respect to oil prices would be more valuable on average than would hedging. Nonetheless, it also seems that some of the overall sensitivity and con-cavity of fuel costs with respect to oil prices is in fact offset by hedging. The overall picture so far is not clearly value-maximizing. Our estimates to this point are on average over time. In this section, we develop more time-and-condition specific estimates of the industry cost function to establish if the time-pattern of hedging improves the value-maximization picture. Anecdotally, the airline industry is known to incompletely and sometimes sharply scales back the overall level of hedging. Thus, in this section we are interested to learn if the mix of hedging and risk taking time periods is appropriate for value maximization.

Industry-wide investment opportunities might be stronger or weaker in high oil price regimes. Carter et al. (2006) provide evidence to suggest that airlines' investment opportunities may be stronger during high oil price regimes. If so, then such periods would be the times when hedging has the most potential to add value, according to risk management theory (i.e., because bad cost outcomes during periods of strong investment opportunities might mean failing to exploit them due to lack of funds). On the other hand, Hamilton (2008) and Kilian (2008) have posited that oil price shocks may be an underlying cause of recessions to come. In that case, times of high oil prices might correspond to a weak investment opportunity set.

We estimate our full cost function (including the GDP and market share effects) as augmented to allow for different oil price effects during high oil price regimes. We accomplish this by adding two interaction-term regressors, both of which involve an indicator variable for quarters in which oil price is above the sample median. *High oil interaction* is defined as the product of the indicator variable and *Oil price*, and *High oil square interaction* is analogously defined as the product of the indicator variable and *Square oil price*.

Table 4 presents the results of estimating this augmented cost function as a firm fixed effects panel regression. In column (1) of the table, containing estimates for the *Unhedged fuel cost* function, our earlier finding remains intact overall–costs are increasing and concave in oil price, with a large and statistically significant positive estimated coefficient on *Oil price* and a large and statistically significant negative coefficient on *Square oil price*. The F-test statistics labelled "F-test (oil effects)" tests the joint hypothesis that the *Oil price* coefficient and the *Square oil price* coefficient are both equal to zero. That F-statistic is very large and soundly rejects that null hypothesis.

At the same time, the *High oil interaction* is strongly statistically significantly negative, and the *High oil square interaction* is strongly statistically significantly positive. These coefficients indicate that the cost function leans more toward convexity during high oil price regimes. The economic implication is that hedging might be more beneficial during such times. Anecdotal stories from the airline industry often suggest that hedging is more prominent during the high-price times, suggesting that the time pattern of hedging may make sense. The statistical conclusion is confirmed by a joint F-test on the interaction coefficients, labelled as "F-test (oil shift effects)" in the table, which strongly reject the null of no oil coefficient shift.

Table 4 Cost regressions in different oil price regimes

Central variables	(1)	(2)	(3)
	Unhedged fuel cost	*Hedged fuel cost*	*Total cost*
Oil price	4.943	3.735	1.203
	(4.236)	(2.324)	(1.696)
Square oil price	−0.543	−0.399	−0.144
	(−3.414)	(−1.829)	(−1.496)
High oil interaction	−1.147	−1.034	−0.431
	(−2.493)	(−2.458)	(−1.998)
High oil square interaction	0.267	0.237	0.100
	(2.473)	(2.399)	(1.970)
Δ *Fixed capital*	1.835	2.294	2.178
	(2.933)	(4.550)	(5.538)
Observations	2,607	2,607	2,607
R-squared	0.301	0.163	0.063
Number of airlines	95	95	95
F-test (oil effects)	(58.18)	(21.42)	(2.83)
	[0.00]	[0.00]	[0.64]
F-test (oil effect shifts)	(3.39)	(3.92)	(2.20)
	[0.04]	[0.02]	[0.17]

The same effects follow through to the *Hedged fuel cost* column, where the results are qualitatively similar but economically a bit smaller. This is consistent with the idea that hedging has offset some of the oil sensitivity. No strong oil shift effects are apparent for *Total cost*. Neither are strong oil effects present, similar to our earlier findings.

4.3.2 Business Cycle Regimes

If airlines' investment opportunities and the nature of their cost functions vary across oil price regimes, it may also be the case that they vary across business cycle stages (expansions and contractions in macroeconomic activity). In Table 5 we show that this is exactly the case. Oil effect shifts in recessions are of a similar nature to oil effect shifts in high oil price regimes. That is, the industry *Unhedged fuel cost* function is less concave at such time, suggesting that hedging could be more valuable at such times. If recessions are a time when financial distress risk is enhanced, and/or when investment opportunities need to be taken for post-recession gain, then this makes economic sense. We note that the industry *Hedged fuel cost* function does not show such a shift, which suggests that the hedging policies are adjusted across expansion and contraction regimes to offset their effects on the oil sensitivity.

Table 5 Cost regressions in different macroeconomic conditions

Central variables	(1)	(2)	(3)
	Unhedged fuel cost	*Hedged fuel cost*	*Total cost*
Oil price	2.391	2.277	0.332
	(3.449)	(2.232)	(0.636)
Square oil price	−0.199	−0.204	−0.028
	(−2.214)	(−1.566)	(−0.435)
Recession interaction	−0.123	−0.127	−0.085
	(−1.417)	(−1.437)	(−1.861)
Recession square interaction	0.025	0.030	0.017
	(1.096)	(1.371)	(1.505)
Δ *Fixed capital*	1.843	2.290	2.174
	(2.951)	(4.485)	(5.547)
Observations	2,607	2,607	2,607
R-squared	0.300	0.162	0.065
Number of airlines	95	95	95
F-test (oil effects)	(49.78)	(23.45)	(1.33)
	[0.00]	[0.00]	[0.27]
F-test (oil effect shifts)	(4.24)	(1.12)	(4.49)
	[0.02]	[0.33]	[0.01]

5 Conclusion

Airlines are an important industry, and historically a somewhat unstable one. Airline bankruptcies and recombinations are commonplace in the record. Additionally, airlines face serious commodity price risks in that jet fuel is one of their largest cost factors. For an already-stressed airline, it seems that a fuel price spike might take costs past a breaking point. Even for less-stressed airlines, there are strong possibilities that a fuel price spike could occasion a more-than-proportional cost increase, as the strain on the company's operating capabilities increases, or could lead to cash shortfalls that, in turn, cause an airline to pass by on beneficial investments. The *prima facie* case that hedging might add value is easy to make.

When airlines hedge, it is typically using derivatives on oil products other than jet fuel. Thus, they face significant basis risks. Additionally, airlines have available a variety of operational risk-offsetting mechanisms that also might limit the marginal the value-added of financial hedges. Finally, airlines' investment opportunities vary over time and with the business cycle, as do oil prices. It seems likely that the correlations among oil prices, fuel prices, investment opportunities, and business conditions may not be stable over time. All these factor complicate airlines hedging, and may limit its potential for value added.

Airlines do hedge significantly, but hedging is not universal within the industry nor do any firms hedge fully. Further, hedging intensity varies substantially over time for many airlines. We have developed empirical evidence for an explanation: airlines' cost structures are such that the value-added to hedging is limited. Specifically, fuel costs on average tend toward concavity, suggesting that cost savings when oil prices drop exceed cost increases when oil prices spike. Furthermore, airlines total costs apparently include significant operational hedges to oil prices and significant basis differential effects between fuel costs and oil prices. We also develop evidence that the value-potential for hedging varies across the business cycle, helping to explain why airlines' hedging intensity is dynamic.

Acknowledgments We acknowledge and greatly appreciate the contributions of Jiao Tu, who worked with us at the early stages of this project. We are extremely grateful for the advice and comments of the editors.

References

Blas J, Clark P (2011) US airlines rethink fuel hedging policy, Financial Times (2 March 2011)
Bolton P, Chen H, Wang N (2011) A unified theory of Tobin's q, corporate investment, financing, and risk management. J Finan 66:1545–1578
Campbell JY, Gigli S, Polk C (2012) Hard times, University of Chicago Working Paper
Carter DA, Rogers DA, Simkins BJ (2006) Does hedging affect firm value? Evidence from the US airline industry. Finan Manage 35:53–86
Chen C, Jin Y, Wen M-M (2011) Executive compensation, hedging, and firm value. California State University Northridge working paper

Chen N-F, Roll R, Ross S (1986) Economic forces and the stock market. J Bus 59:383–403
Chiang I-HE, Hughen WK, Sagi JS (2012) Estimating oil risk factors using information from equity and futures markets. Vanderbilt University working paper
Chung R, Lee W-J, Sohn BC (2012) Labor unions and investment efficiency. Griffith Business School working paper
Credeur MJ, Schlangenstein M, Burkhardt P (2011) United, Delta profit at risk on 'silent killer' hedges, Bloomberg News Service (31 January 2011)
Emans A, Gabaix X, Landier A (2009) A multiplicative model of optimal CEO incentives in market equilibrium. Rev Finan Stud 22:4481–4917
Faulkender M (2005) Hedging or market timing? Selecting the interest rate exposure of corporate debt. J Finan 55:931–962
Freed J (2012) Hedging drives delta air lines to 2Q loss. SeattlePI.com (26 July)
Froot KA, Scharfstein DS, Stein JC (1993) Risk management: coordinating corporate investment and financing policies. J Finan 48:1629–1658
Hamilton JD (2008) Oil and the macroeconomy. In: Durlauf SN, Blume LE (eds) The new Palgrave dictionary of economics, 2nd edn. Palgrave MacMillan, Basingstoke, UK
Kilian L (2008) The economic effects of energy price shocks. J Econ Lit 46:871–909
Mackay P, Moeller SB (2007) The value of corporate risk management. J Finan 62:1379–1419
Modigliani F, Miller M (1958) The cost of capital, corporate finance and the theory of corporation finance. Am Econ Rev 48:261–297
Morrell P, Swan W (2006) Airline fuel hedging: theory and practice. Transp Rev 26:713–730
Peterson K, Reiter C (2008) U.S. airlines hedge less, let fuel cost exposure grow. Reuters News Service (March 13)
Pulvino TC (1998) Do asset fire sales exist? An empirical investigation of commercial aircraft transactions. J Finan 53:939–978
Rivers M (2012) Should airlines hedge their bets on fuel? Airline Business (July 25)
Smith C, Stulz R (1985) The determinants of fims' hedging policies. J Finan Quant Anal 20:391–405
Staff (2012) (Investor Guide) Delta airlines (DAL) buys an oil refinery as the "ultimate fuel hedge". InvestorGuide.com (15 May)

Part II
The Impact of Oil Shocks

Oil Prices, Volatility, and Shocks: A Survey

Ulrich Oberndorfer

Abstract This paper surveys the literature on the economic effects of oil market developments. It assesses the economic theory behind oil price impacts and presents how the existing literature has analysed the link between oil markets—oil prices, oil price shocks, and oil price volatility—and economic outcomes. This review documents the general consensus amongst economists that the significance of moderate oil price movements is low if not inexistent, with clear impacts only present on financial markets. However, the evidence for significant macroeconomic effects of energy price shocks is strong, although methodological challenges such as causality and endogeneity remain an issue.

Keywords Oil price · Oil price shocks · Oil price volatility

1 Introduction

In the last decade, the oil price has returned to the political agenda. Against the background of price hikes and strong price volatility, international fora such as the energy consumer organization International Energy Agency (IEA), the producer-consumer-dialogue organization International Energy Forum (IEF) as well as G8 and G20 have been dealing intensively with oil market issues. Due to growing evidence that trading activities may foster significant oil price volatility, energy policy makers around the globe have been discussing whether and how stricter oil market regulation can limit excessive oil price fluctuations.

The implicit assumption underlying the policy makers' strong efforts is the economic significance of oil prices and their variations. To put it in simple terms: Oil prices are the focus of global policy because of their economic importance. In this context and in the public debate in general, the emotional scientific dispute on

U. Oberndorfer (✉)
Permanent Mission of the Federal Republic of Germany to the OECD,
9 rue Maspéro, 75116 Paris, France
e-mail: ulrich.oberndorfer@diplo.de

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_3

the economic significance of oil prices has been ignored, however. Few economic questions have been tackled empirically with such a variety of approaches regarding measurement as well as period, region or market of analysis, leading to numerous research articles, comments and replies—and to very different results with regard to the estimated economic relevance of the oil price.

This paper surveys the broad literature available in the field. It assesses the economic reasoning behind oil price impacts and presents how the existing literature has analyzed the relationship between oil market developments and economic outcomes. *Oil price shocks* and *oil price volatility* are the buzzwords with regard to such energy market developments. Definitions and measurement approaches of those phenomena are presented in the paper. The focus on the economic side of the *oil to macroeconomy-relationship* has been given to standard macroeconomic indicators such as production, GDP and (un-) employment, as well as to financial market developments.

The remainder of the paper is structured as follows: Sect. 2 presents popular definitions of oil market—price, price shock and volatility—measures. Section 3 summarizes the theoretical background as well as empirical findings regarding the oil-to-macroeconomy relationship, while Sect. 4 tackles theoretical and empirical findings regarding the role of the oil price in financial markets. Section 5 concludes.

2 Oil Market Measures

2.1 Oil Prices and Oil Price Shocks

Previous literature has noted that the nature of the movements of oil prices must be adequately addressed in order to accurately measure the economic effects of these prices (e.g., Löschel and Oberndorfer 2009). In his pioneering work on the oil-to-macroeconomy relationship, Hamilton (1983) makes use of an *oil price series in 1st differences*.[1] This approach is still common today, often in a log-differenced form in order to avoid non-stationarity problems by differencing and to receive easily interpretable estimation results (elasticities) by using logs. This series is constructed as

$$dloil_t = \log(oil_t) - \log(oil_{t-1}). \quad (1)$$

Here, the price of oil at time t is denoted oil_t.[2] In the search for more adequate and economically relevant oil price measures, Mork (1989) introduces the use of an *asymmetric oil price variable* that is defined as

[1] For the analysis of certain periods Hamilton (1983) makes use of detrended oil price change variables.

[2] Most authors use real oil prices, i.e., deflated nominal oil prices, in their analyses. In this sense, I refer to oil_t as a real oil price series throughout this paper.

$$dloilpos_t = max(0, \log(oil_t) - \log(oil_{t-1})). \quad (2)$$

$dloilpos_t$ gives the value of 0 if the oil price has decreased at time t compared to $t-1$. In contrast, if the oil price increased within this period, $dloilpos_t$ gives the first differenced logged oil price series ($dloil_t$).

Hamilton (1996) proposes the *net oil price increase* (*nopi*) as a further definition of an oil price variable. It compares the current price of oil with the maximum value of the previous year rather than its value at $t - 1$ (i.e., at the previous quarter, month, etc.) alone. If the current value of the oil price exceeds the previous year's maximum, the value of $nopi_t$ is assigned to the change of the current value over the previous year's maximum. If the price of oil at the current point in time is lower than it had been at any point during the previous year, the series is assigned the value of zero.

$$nopi_t = max(0,\ \log(oil_t) - max(\log(oil_{t-1}), \log(oil_{t-2}), \ldots, \log(oil_{t-m}))), \quad (3)$$

m gives the number of observed periods per year (i.e., in case of quarterly data, $m = 4$, in case of monthly data, $m = 12$).

In the following, these three oil price variables are presented in real terms on a monthly basis for the period 10/1973–01/2008. They give the oil price variables from a German perspective, with the oil import cost data published by the EIA deflated using the German consumer price index, and converted to domestic currency using exchange rates from the time series database of Deutsche Bundesbank (German Central Bank; based on data from the German Federal Statistical Office). The graphs (Figs. 1–3) are based on the dataset used by Löschel and Oberndorfer (2009).

The graphs illustrate the measurement differences between the three oil price variables. Both the oil price increase and the net oil price increase exclude oil price decreases. In particular, the net oil price increase series, takes the value of 0 for many points in time.

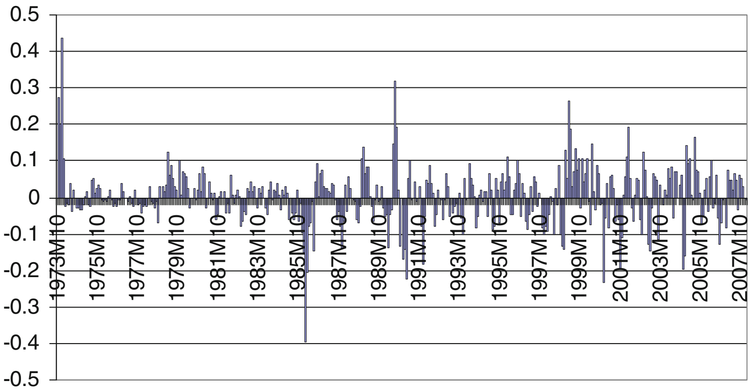


Fig. 1 Oil price change (dloil)

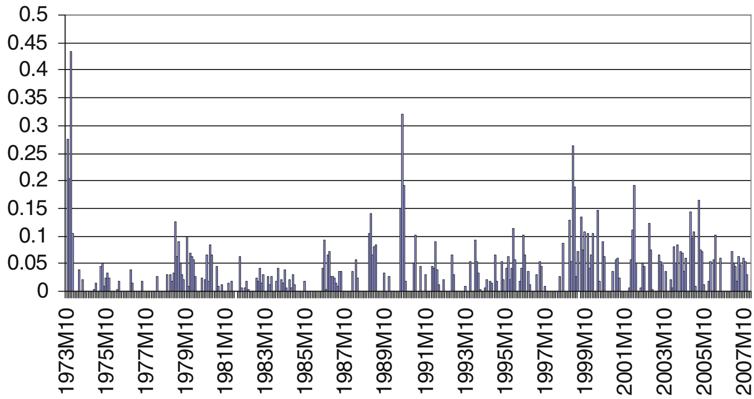


Fig. 2 Oil price increase (dloilpos)

The notion of oil price shocks is not very well defined. It is obvious though that a simple oil price change series includes all price changes in the sample period, whether or not they are shocks. If—as often argued—the economic effect of the oil price varies, depending on whether the oil price is rising or falling, or whether it is changing moderately or substantially in a shock-like manner, the inclusion of a simple oil price change series in an empirical analysis may not accurately reflect the effects of an oil shock.

As the term of a price shock is associated with the idea of rising prices, the definition of the asymmetric oil price increase variable comes closer to the phenomenon of an oil price shock. Finally, the net oil price increase may be the natural empirical implementation of the oil price shock idea, although seen rather arbitrary by some scholars. Only shock-like oil price increases, defined as one-year highs, are considered. Accordingly, Hooker (1996) criticises *nopi* definitions as being ad hoc,

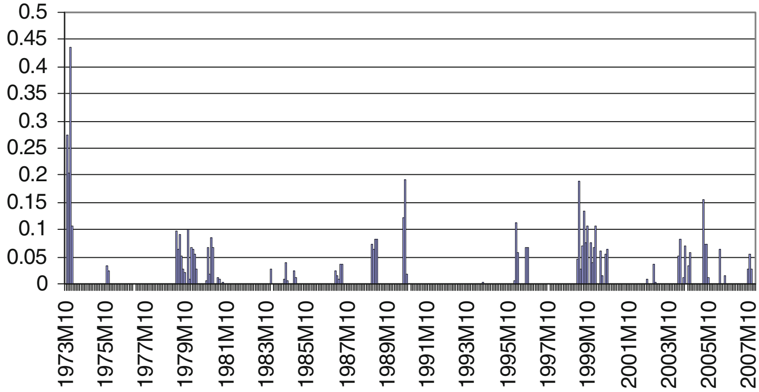


Fig. 3 Net oil index increase (nopi)

although he admits being "sympathetic to the argument that oil price increases which cancel out recent decreases have different effects than those which occur in a relatively stable environment".

2.2 Oil Price Volatility

The standard measure of oil price volatility at time (period) t, denoted as $oilvol_t$, is the estimation of the *standard deviation of the oil price* in a given period (cp., e.g., Ferderer 1996).

$$oilvol_t = \left[(1/(n-1)) \sum_{p=1}^{q} \left(oil_{t,p} - \mu_t \right)^2 \right]^{0.5}, \tag{4}$$

Here, μ_t is the mean of the oil price $oil_{t,p}$ in time (period) t. n gives the number of observations. t can be divided into subperiods from $p = 1$ to q, i.e., $\mu_t = (1/q) \sum_{p=1}^{q} oil_{t,p}$. E.g., in case $oilvol_t$ refers to monthly oil price volatility as in Ferderer's (1996) analysis, $oil_{t,p}$ could give daily oil prices. In this case, $oil_{t,q}$ would refer to the oil price on last day q of the month t.

However, this standard estimation approach not only requires values for the oil price to be available for period t (i.e., oil_t), but also for shorter subperiods p (i.e., $oil_{t,p}$). Depending on the choice of period and subperiod, this may prove to be difficult, i.e., if period t represents days/daily observations. In such a case, it is not possible to estimate the standard deviation unless intraday prices are available.

In order to cope with this data challenge, Oberndorfer (2009a) uses *squared oil price changes* as an oil price volatility proxy variable.

$$oilsp_t = (oil_t - oil_{t-1})^2. \tag{5}$$

Squared price changes can be seen as good indicators of market volatility as they give the deviation of the changes of the respective price from its mean (which is often 0). However, volatility terms defined as squared changes, such as $oilsp_t$, are positive by definition and therefore often exhibit highly significant positive means. This means that these volatility variables do not indicate volatility surprises (or unexpected volatility), i.e., volatility innovation, and they can be predicted to a certain extent.[3]

This may be problematic in financial market analyses: If capital markets work efficiently, only innovations, i.e., unexpected movements of selected systematic variables, can affect them. The use of these volatility variables that at least partly

[3] This is illustrated by the success of estimators of the ARCH-class (cp., e.g., Engle 2001) that model volatility by dynamic processes.

represent expected volatility could therefore induce an errors-in-variables problem in financial market analyses (Chen et al. 1986).

In order to cope with this problem, Oberndorfer (2009a) additionally proposes using errors of AR(K) processes of squared oil price changes as *oil volatility innovations* ($oilvi_t$). This is done by estimating an AR(K) model for the squared oil price change series.

$$oilsp_t = a + \sum_{k=1}^{K} b_k oilsp_{t-k} + e_t. \tag{6}$$

e_t is the noise disturbance with zero mean and variance v_t^2. a and the b_k, besides v_t^2 are the unknown parameters that have to be estimated by OLS. The lag lengths (K) of the respective regressions can be determined according to information criteria such as the Bayesian Schwarz Information Criterion. e_t is the estimated error term of the model and therefore the oil volatility innovation that can be used as an explanatory variable within a regression analysis.

$$oilvi_t = e_t.$$

According to Pagan (1984), the use of current levels of "generated regressors" such as $oilvi_t$ within a two-step analysis should yield consistent and efficient estimates in an empirical analysis.

3 Oil Prices and the Macroeconomy

3.1 Theoretical Background

The theory on the macroeconomic role of oil prices is complex. Numerous channels have been proposed and several survey articles exist. This section can only briefly summarize the main channels of the oil-to-macroeconomy-relationship. An excellent and extensive review is provided by Brown and Yücel (2002); I would like to refer the interested reader to that manuscript for further information and references. If not designated otherwise, the argumentation in this subsection is based on Brown and Yücel (2002).

The traditional and most common explanation for macroeconomic oil price impacts is the *supply-side effect*. It describes rising oil prices as an indicator of the reduced availability of a basic input—oil—for production (e.g., Brown and Yücel 2000). As a consequence of this increased scarcity, prices rise in general, creating inflationary pressure. Moreover, the growth of output and productivity are slowed down. The decline in productivity growth lessens real wage growth, and increases unemployment. Negative effects can be expected to be stronger if wages are nominally sticky downward and therefore cannot fully adjust. Consumption smoothing on behalf of the consumers—decreased savings, increased borrowings—together with lower output growth, may increase the real interest rate and, consequently, the inflation rate. This effect can add to the direct inflationary impact of rising oil prices. Thus, the supply-side effect is the best explanation for a double negative macroeconomic effect of rising oil prices: slowed economic growth and increasing inflation. *The real balance effec*t is in some way related to the aspect of consumption smoothing, suggesting that an oil price rise increases the demand for money, leading to higher interest rates and, consequently, to lower GDP growth. Additionally, the *income transfer channel* describes the shift in purchasing power from oil importing countries to oil exporting countries when oil prices rise. A reduction in demand for goods produced in oil-importing countries can be the net effects.

Going beyond these simple mechanisms, at least four channels suggest oil prices have an *asymmetric effect* on the economy. Accordingly, whereas oil price rises would harm the economy, comparable oil price decreases would not (fully) compensate for those effects, respectively. Firstly, it is argued that monetary policy that fails to hold GDP constant can constitute such a channel if wages are nominally sticky downward. Secondly, adjustment costs that occur within a sectoral shift in the economic production or structure in general from energy-intensive to energy-extensive sectors can play a role in this respect. Thirdly, asymmetric cost pass-through in the oil-intensive production chain can be responsible or such asymmetric effects. Finally, authors like Hamilton (1996) argue that historical oil crises have been characterized by widespread *concern about the price and availability of energy*, potentially causing irreversible investment decisions to be postponed in case of oil price appreciations. This argument is referred to and is particularly relevant in cases where the oil prices rise significantly, i.e., in so-called oil shock situations (see also previous section). A similar channel is put forward by Bernanke (1983), who argues that investment will be postponed in a situation of oil price increases as firms attempt to find out whether or not the observed price rise is permanent.

Similarly, Sauter and Awerbuch (2003) argue that since the 1980s *oil price volatility* has had a more significant effect on economic activity than the oil price level. In their assessment, however, no clear definitory distinction is made between notions such as oil price volatility, oil price increases and oil price shocks. However, Sauter and Awerbuch's (2003) claim seems to be motivated by the reasoning of negative economic implications of oil shocks or asymmetric oil price effects. They identify two different negative implications of oil price volatility, uncertainty in investment and sectoral shift, which other authors also associate to oil price increases or shocks (see above).

Generally, it is obvious that the above mentioned effects strongly depend on *fiscal and monetary policy* reactions to oil price increases. Both can contribute to a demand stimulation, e.g., in oil shock situations. However, there are good reasons to believe in Brown and Yücel's (2002) claim that oil price shocks "increase the potential for errors in monetary policy", as well as in fiscal policy. From a theoretical standpoint, it is crucial in this respect whether money illusion is present or not.

If so, an accommodative monetary policy has the potential to offset, at least partly, losses in GDP growth that are due to oil price rises. A restrictive monetary policy can aggravate negative macroeconomic oil price effects in such a setting. In the absence of a money illusion, on the other hand, monetary policy is simply mirrored by inflation without having real effects, apart from—potentially negative—impacts of the inflation caused by monetary policy. Wage policy is a further aspect to be observed in this regard; in contrast to the mechanism of the supply-side effect described above, a so-called *wage-price-spiral* implying inflationary pressure can evolve if nominal wages are set in line with observed (oil) price increases and if prices in general reflect past wage increases (Barsky and Kilian 2004).

Several authors argue that macroeconomic oil price effects have diminished in recent years, an occurrence that is difficult to describe with any of the above mentioned channels at hand. It is perceivable, though, that (monetary and/or fiscal) policy makers have drawn lessons from past oil crises and consequently improved their responses to oil shocks. Moreover, as recent world oil consumption is particularly boosted by the dramatic gains in oil consumption outside the advanced economies of the OECD, with the strongest gains in emerging (mostly Asian) economies, the predominance of *oil demand rather than supply shocks* in recent years can offer an explanation in this regard. The boost in oil consumption seems to be driven, as well as accompanied by, an economically beneficial general rise in demand for goods and services on the world markets.[4] Such a stimulating effect can at least partly offset the negative impacts of oil price rises.

A further reason why the oil price effect on employment could have diminished over the past decades is the decline in *energy intensity* observed almost all over the world (e.g., IEA 2011a),[5] which could go hand in hand with a reduced oil price impact on the economy (e.g., Barsky and Kilian 2004, or Schmidt and Zimmermann 2007). (Differences in) Energy efficiency could indeed constitute a central factor for the specific macroeconomic effects in respective countries or regions and over time, with efficiency improvements being a potential tool to diminish the economic vulnerability to oil prices.

3.2 Insights from Empirical Analyses

Different empirical analyses tell different stories about how (much) the oil price matters to economic development. Interestingly, this is also true for the available literature reviews. In their prominent survey paper, Barsky and Kilian (2004) argue that there was little evidence that the oil price significantly affected the

[4] In this respect, e.g., Lin (2008) emphasizes the role of recently rising Chinese demand for increases in the oil price.

[5] However, recent energy efficiency data provided by the IEA suggest that this global trend towards energy efficiency halted or at least paused in 2008 and 2009.

macroeconomic performance in the United States. Their findings suggest that oil price shocks are neither necessary nor sufficient to explain the weak macroeconomic performance in the US and generally conclude that the economic influence of oil prices changes was rather small or even nonexistent. In contrast, the review provided by Sauter and Awerbuch (2003) states that the "idea that rising oil prices and price volatility serve to stifle economic activity … has by now become widely accepted in the literature and seems virtually axiomatic". Making reference to influential studies using US data for the post-WW II-period, Sauter and Awerbuch (2003) argue that oil price increases of 10 % could be followed by GDP decreases of around 1.5 %.

The academic discussion about the robustness of an oil-to-macroeconomy-relationship has led to the development of different measures representing oil shocks. This term is rather vague, but implies oil price increases that are greater than usual variations. Accordingly, apart from the discussion about channels or transmission mechanisms of oil price shocks on the economy, the available literature suggests that the actual nature of oil price movements has to be adequately addressed in empirical analyses in order to correctly measure the effects of oil price shocks (Löschel and Oberndorfer 2009; see above).

While Hamilton (1983) establishes that oil prices have a linear negative effect on GDP, subsequent research has called this result into question on the grounds that the 1973 oil crisis included in Hamilton's (1983) data set would be an outlier and impact his results. Going beyond the linear oil-to-macroeconomy relationship, Mork (1989) makes use of asymmetric oil price variables. This research shows that the negative relationship between GDP growth and oil prices found by Hamilton (1983) was robust in the case of *oil price increases*, but that the correlation between the change in GDP and *oil price decreases* was significantly different or even zero, indicating an asymmetric relationship between oil prices and economic activity. In response to further scepticism with regard to the oil-to-macroeconomy relationship, expressed particularly by Hooker (1996), Hamilton (1996) goes beyond the simple asymmetric oil price effect as applied by Mork (1989). The net oil price increase proposed by Hamilton (1996) gives values different from zero only if current oil prices exceed the previous year's maximum (for definitions, see Sect. 2) and outperforms other oil price measures in causality tests provided by Hamilton (1996).

Subsequently to Hamilton's (1996) work, the net oil price increase has been widely used as an oil shock variable. Net oil price increases have been shown to have a significant impact on economic performance for markets outside the US (e.g., for Germany, see Löschel and Oberndorfer 2009, for other European countries, see Cuñado and Pérez de Gracia 2003, and for Asian economies, see Cuñado and Pérez de Gracia 2005). According to Du et al. (2010), a complex asymmetric relationship between oil price and the macroeconomy is also present in China; moreover, reforms of the national oil pricing mechanism need to be taken into account here. All in all, evidence for economic effects of oil shocks seems therefore convincing.

As mentioned above, critics of the empirical finding of a significant oil-to-macroeconomy-relationship are nevertheless widespread (see, e.g., Barsky and Kilian 2004, for the US and Schmidt and Zimmermann 2007, for Germany). Apart from the theoretical reasoning elaborated above, empirical arguments against a strong economic role of the oil price include the empirically challenging questions about the exogeneity of oil prices: Do oil prices move the economy or does the economy move the oil price? They also address causality. In the past, oil price shocks have often occurred in times of geopolitical crises in the Middle East. This raises the question whether oil price hikes or respective geopolitical crises themselves affect the economy (Barsky and Kilian 2004).

4 Oil Prices and Financial Markets

4.1 Theoretical Background

Based on the common representation of stock prices of corporation *i* (p_i) as expected future cash flows of the corporation ($E(cf_i)$) that are discounted by the discount rate δ, the argument that oil prices may affect the respective corporation's stock returns is straightforward. Such representation is proposed in a general context regarding the *systematic effect of macroeconomic variables on stock returns* by Chen et al. (1986). They accordingly define stock prices as

$$p_i = E(cf_i)/\delta, \tag{7}$$

implying stock returns of corporation *i* of

$$dp_i/p_i = d[E(cf_i)]/E(cf_i) - d\delta/\delta.$$

Within this framework, Chen et al. (1986) argue that both changes in the discount rate δ and in the expected future cash flows $E(cf_i)$ determine the stock returns of corporation *i*. Following Oberndorfer (2009a), this suggests in particular that *stock returns of companies directly involved in the oil business* or dealing with oil products and services are affected by oil price changes. Rising oil prices would upvalue the resource stocks of companies related to the oil business or their products and services. Consequently, their expected future cash flows should rise. Stock returns of utilities and other companies that use fossils fuels as an input, e.g., for electricity generation, would be negatively affected, as the price of their most important input for production rises, with an ex-ante unclear ability to pass through those cost increases to consumers.

However, based on the framework presented above, the *impact of oil price changes on stock returns can be generalized* to corporations from other sectors if it is assumed that the oil price has a direct or indirect effect on their cash flows. Given the role oil prices play for the macroeconomy as such—as set out above from a theoretical perspective—, the channels include induced changes of prices of oil intensive goods, interest rates, production and wages. Oil price effects on stock returns are therefore perceivable for corporations stemming from practically any sector. In particular, a negative relationship between oil prices and stock prices is expected for corporations outside of the energy sector, given the general negative oil-to-macroeconomy-relationship.

As described above, Sauter and Awerbuch (2003) allege that, *oil price volatility* has had a more significant effect on economic activity than the oil price level since the 1980s. Based on this claim, and against the background that the energy industry is strongly exposed to energy price risks even though option trading is available (Hampton 1995), it may not only be appreciations and depreciations in oil price levels that, to the market developments of energy stocks but also oil price volatility. Oberndorfer (2009a) argues that oil market volatility can lead to augmented expenditures for affected corporations, and may for example induce hedging costs. Moreover, following Pindyck (2004), an increase in price volatility may decrease the production of the respective commodity. Therefore, Oberndorfer (2009a) concludes that oil market volatility may equally impact the discounted expected future cash flows of corporations and therefore affect stock prices as shown above in the theoretical framework based on Chen et al. (1986).

4.2 Insights from Empirical Analyses

A number of authors have made an in-depth analysis of the role of oil prices for financial markets, but the available literature is not as broad as in the field of the oil-to-macroeconomy relationship. The—more intuitive—relationship between *oil prices and energy corporations' stock prices* has received more scientific attention than possible oil price effects on corporations from other sectors (or on stock prices in general, as measured by stock indexes).

The main result with regard to oil prices and oil corporations' stocks is that they are—as expected—positively related. This result has for example been produced for the UK oil industry by Manning (1991), who also shows that the effect is largest for corporations purely engaged in oil exploration and production. Faff and Brailsford (1999) reproduce the general positive relationship for the Australian oil and gas sector, Sadorsky (2001) for Canadian, and Oberndorfer (2009a) for Eurozone oil and gas firms.

Amongst that studies that have analyzed the energy sector from a broader perspective, interesting contributions include that of Henriques and Sadorsky (2008) who found that the stock prices of alternative energy companies are positively affected by oil prices (although this result shows only little significance). Their interpretation of this finding is that oil price movements are not as important as once thought with regard to alternative energy companies because investors may view the sector as similar to other high technology branches.[6] Oberndorfer (2009a) finds

[6] This result may also be explained by the fact that most renewable energy sources are not competitive in many energy markets and therefore profit from different kinds of public support.

that oil prices negatively impact stock returns of European utilities that use fossil fuels as a main input for electricity production. In their international analysis of the risk factors of the oil and gas industry, Ramos and Veiga (and Veiga 2011) conclude that the oil and gas sector in advanced countries responds more strongly to oil price changes than in emerging markets, and that oil and gas industry returns respond asymmetrically to oil price changes: Oil price rises have a greater impact than oil price drops.

Sadorsky's (1999) findings suggest that oil price movements are an important determinant of *stock returns in general*. Based on an analysis of stock return data for the S&P 500—i.e., the biggest US corporations—Sadorsky (1999) draws the conclusion that positive shocks to oil prices depress real stock returns. Similar results are produced by Nandha and Faff (2008) who analyze different industry indices. Their findings indicate that oil price rises have a negative impact on equity returns for all sectors except energy industries.

Very few authors have integrated *oil price volatility in empirical analyses of stock markets*. This is even more striking given the fact that at least two analyses have produced statistically significant results on this. Oberndorfer (2009a) finds that oil market volatility negatively affects European oil and gas stocks. He specifically shows that also for castable oil market volatility impacts the stock market, implying profit opportunities for strategic investors. According to Sadorsky (2003), technology stock return volatility is positively affected by oil price volatility. The analysis of Arouni et al. (2011) for the stock markets of the Gulf Cooperation Council countries suggests that there are both return and volatility spillovers between oil and stock markets.

5 Conclusion

The oil price is back on the political agenda. This makes it even more important to analyze the relevance of oil prices for the economy and financial markets. This paper intends to contribute to that debate—not by estimating the magnitude of oil price effects, but rather by shedding some light on the arguments, notions and definitions underlying the existing analyses.

Apart from these theoretical and technical aspects, this review documents the general consensus amongst economists that the economic significance of moderate oil price movements is low or even nonexistent. Even minor oil price changes affect financial markets, however: Stock prices of energy companies as well as those operating in other sectors are shown to be very sensitive to price movements and volatility of the oil market. As long as oil prices remain within a price floor,

(Footnote 6 continued)
Several renewable energy support schemes such as feed-in-tariffs applied in many countries eliminate price risks for renewable energy generation so that the prices of fossil fuels such as oil should have a minor impact—or no impact at all—on renewable companies' businesses.

macroeconomic indicators seem to be barely affected. However, there is strong evidence that energy price shocks, i.e., massive price movements, have significant economic effects. Macroeconomic impacts may depend on their own "nature"—the differentiation between supply and demand shocks provides an example in this regard—, on the national or regional particularities—such as the level of energy efficiency—, as well as on policy responses.

The debate about the oil-to-macroeconomy-relationship is not over. Methodological challenges such as causality and endogeneity remain an issue and the question about how strongly the oil price affects economic outcomes is far from being resolved.[7] These questions call for further empirical analysis, based on approaches presented in this paper as well as on modern econometric techniques. As oil prices are expected to continue rising in the mid- and long-term (e.g., IEA 2011b), the question about their economic implications will remain highly relevant.

The age of cheap oil may be over, as stated by both the IEA and the *peak oil hypothesis*. As a result, and in combination with the rise of shale gas and progress on renewable energy technologies, the world might be entering a *golden age of gas* as well as a period of *electrification*. This suggests the increasing importance of analyses that deal with non-oil energy market segments such as electricity and gas markets. The approaches presented in this paper should be well suited also for these kinds of assessments. For this analytical purpose, the integration of macroeconomic energy cost indicators (Oberndorfer 2012) could also be considered. Finally, carbon markets such as the European Union Emission Trading Scheme (EU ETS) as further evolving energy markets could be assessed (cp. Oberndorfer 2009b or Chevallier 2011).

Conflict of Interest This article represents the personal opinion of the author and does not reflect the official position of the institution he is affiliated with.

References

Arouni MEH, Lahani A, Nguyen DK (2011) Return and volatility transmission between world oil prices and stock markets of the GCC countries. Econ Model 28(4):1815–1825

Barsky RB, Kilian L (2004) Oil and the macroeconomy since the 1970s. J Econ Perspect 18 (4):115–134

Bernanke B (1983) Irreversibility, uncertainty, and cyclical investment. Quart J Econ 98(1):85–106

Brown SPA, Yücel MK (2002) Energy prices and aggregate economic activity: an interpretative study. Q Rev Econ Finan 42(2):193–208

Brown SPA, Yücel MK (2000) Oil prices and the economy. Southwest Econ 4:1–6

Chen N-F, Roll R, Ross SA (1986) Economic forces and the stock market. J Bus 59(3):383–403

Chevallier J (2011) Evaluating the carbon-macroeconomy relationship: evidence from threshold vector error-correction and Markov-switching VAR models. Econ Model 28(6):2634–2656

[7] Further methodological issues that can drive the results of oil-to-macroeconomy and oil-to-financial markets analyses are the choice of empirical approaches and specifications.

Cuñado J, Pérez de Gracia F (2005) Oil prices, economic activity and inflation: evidence for some Asian countries. Q Rev Econ Finan 45(1):65–83
Cuñado J, Pérez de Gracia F (2003) Do oil price shocks matter? evidence for some European countries. Energy Econ 25(2):137–154
Du L, Yanan H, Wei C (2010) The relationship between oil price shocks and China's macro-economy: an empirical analysis. Energy Policy 38(8):4142–4151
Engle RF (2001) GARCH 101: the use of ARCH/GARCH models in applied econometrics. J Econ Perspect 15(4):157–168
Faff RW, Brailsford TJ (1999) Oil price risk and the Australian stock market. J Energy Finan Dev 4(1):69–87
Ferderer JP (1996) Oil price volatility and the macroeconomy. J Macroecon 18(1):1–26
Hamilton JD (1996) This is what happened to the oil price-macroeconomy relationship. J Monetary Econ 38(2):215–220
Hamilton JD (1983) Oil and the macroeconomy since World War II. J Polit Econ 91(2):228–248
Hampton M (1995) Options. In: Jameson R (ed) Managing energy price risk. Risk Magazine Publications, London, pp 97–116
Henriques I, Sadorsky P (2008) Oil prices and the stock prices of alternative energy companies. Energy Econ 30(3):998–1010
Hooker MA (1996) What happened to the oil price-macroeconomy relationship? J Monetary Econ 38(2):195–213
IEA (2011a) IEA Scoreboard 2011: implementing energy efficiency policy: progress and challenges in IEA member countries, Paris
IEA (2011b) World energy outlook 2011, Paris
Lin SX (2008) Effect of Chinese oil consumption on world oil prices. Cass Business School, City University London
Löschel A, Oberndorfer U (2009) Oil and unemployment in Germany. J Econ Stat 229(2–3):146–162
Manning N (1991) The UK oil industry: Some inferences from the efficient market hypothesis. Scott J Polit Econ 38(4):324–334
Mork KA (1989) Oil and the macroeconomy when prices go up and down: an extension of Hamilton's results. J Polit Econ 97(3):740–744
Nandha M, Faff R (2008) Does oil move equity prices? A global view. Energy Econ 30(3):986–997
Oberndorfer U (2012) Developing macroeconomic energy cost indicators. Energy Policy 42:1–3
Oberndorfer U (2009a) Energy prices, volatility, and the stock market: evidence from the Eurozone. Energy Policy 37(12):5787–5795
Oberndorfer U (2009b) EU emissions allowances and the stock market: evidence from the electricity industry. Ecol Econ 68(4):1116–1129
Pagan A (1984) Econometric issues in the analysis of regressions with generated regressors. Int Econ Rev 25(1):221–247
Pindyck RS (2004) Volatility and commodity price dynamics. J Futures Markets 24(11):1029–1047
Ramos SB, Veiga H (2011) Risk factors in oil and gas industry returns: international evidence. Energy Econ 33(3):525–542
Sadorsky P (2003) The macroeconomic determinants of technology stock price volatility. Rev Financ Econ 12(2):191–205
Sadorsky P (2001) Risk factors in stock returns of Canadian oil and gas companies. Energy Econ 23(1):17–28
Sadorsky P (1999) Oil price shocks and stock market activity. Energy Econ 21(5):449–469
Sauter R, Awerbuch S (2003) Oil price volatility and economic activity: a survey and literature review. IEA Research Paper, Paris
Schmidt T, Zimmermann T (2007) Why are the effects of recent oil price shocks so small? Ruhr Economic Paper No. 21, Essen

Oil Shock Transmission to Stock Market Returns: Wavelet-Multivariate Markov Switching GARCH Approach

Rania Jammazi

Abstract Our understanding of the nature of crude oil price shocks and their effects on the stock market returns has evolved noticeably in recent years. Evidence of spillover effects between several kinds of markets has been widely discussed around the globe, and yet the transmission of shocks between crude oil market and stock market returns has received little attention. Extending earlier work in the literature, we use data on monthly crude oil returns and stock market returns of five developed countries (USA, UK, Japan, Germany and Canada) to investigate two issues that have been at the centre of recent debates on the effect of crude oil shocks on the stock market returns. First, we analyse whether shocks and or volatility emanating from two major crude oil markets are transmitted to the equity markets. We do this by decomposing monthly real crude oil prices and analysing the effect of the smooth part on the degree of the stock market instability. The motivation behind the use of this method is that noises can affect the quality of the shock and thus increase erroneous results of the shock transmission to the stock market. Second, under the hypothesis of common increased volatility, we investigate whether these states happen around the identified international crises. In doing so, flexible model is implemented involving the dynamic properties of the Trivariate Markov switching GARCH model and the recent Harr A trous wavelet decomposition, in order to achieve a strong prediction of the abovementioned situations The proposed model is able to circumvent the path dependency problem that can affect the prediction's robustness and also provides useful information for investors and government agencies that have largely based their views on the notion that crude oil markets negatively affect stock market returns. Indeed, the results show that the A Haar Trous Wavelet decomposition method appears to be an important step toward improving accuracy of the smooth signal in detecting key real crude oil volatility features. Additionally, apart from UK and Japanese cases, the responses of the stock market to an oil shock depend on the geographic area for the main source of supply whether it is from the North Sea or from North America (as two oil benchmarks are used, WTI and Brent respectively).

R. Jammazi (✉)
The University of Management and Economic Sciences of Sousse, Sousse, Tunisia
e-mail: jamrania2@yahoo.fr

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_4

Keywords Trivariate BEKK-Markov switching · Wavelet decomposition · Oil shocks · Stock markets

1 Introduction

The stock market movements as contained in the stock price (among other economic indicators) send us some obvious "signals" of a country's economic strength and development. For instance, a bull stock market, i.e. a market which goes up and maintains upward trends, is associated with increasing business investment and vice versa.

However, the majority of Organisation for Economic Co-operation and Development (OECD) countries have become increasingly dependent upon oil over the last century and this is now recognised as the most essential energy source. In 2008, the US was the largest consumer of oil, consuming around 20 million barrels per day, followed by China (7.8) and Japan (4.8) (EIA 2008). The 2007–2008 period marked the fastest price changes in the history of oil. In fact, oil prices rose dramatically to more than 140 dollars per barrel in August 2008 (the record peak), and then sharply dropped to around 30 dollars per barrel in December 2008.[1] This (and also other sequences of very large increases and decreases observed in crude oil prices over the last three decades) will obviously affect companies' earnings very significantly as oil operating costs lead to a remarkable change in stock prices.

Despite the considerable attention that has been paid to the investigation of the relationship between changes in the price of crude oil and stock prices, conclusions on these effects cannot yet be drawn. More than 20 years ago, Jones and Kaul (1996) observed that stock market returns of USA, Canada and Japan respond negatively to oil shocks. However, Huang et al. (1996) found no evidence of the relationship between US stock returns and changes in the price of oil futures. Wei (2003) argued that the decline in stock prices after the 1973/74 oil crisis seems too large to be explained by the rise in oil prices. Chen et al. (1986) in contrast, concluded that oil price changes have no impact on asset pricing. Using structural VAR, Kilian and Park (2009) demonstrate that it is useful to differentiate between three distinct sources of oil shocks in the global market for crude oil before assessing the impact of an oil price shock on aggregate US real stock returns. In particular, they report that only an oil price increase driven by a precautionary demand for oil associated with concerns about future oil supply shortfalls, namely "precautionary demand shocks", negatively affects stock prices. In contrast, shocks to the production of crude oil "oil supply shocks" have no significant impact on the US stock returns. Finally, shocks driven by strong global demand for industrial commodities including crude oil, "aggregate demand shocks", have persistent positive effects on cumulative stock returns within the first year of the expansionary shock.

---

[1] Source: Wikipedia, the free encyclopedia; http://en.wikipedia.org/wiki/Price_of_petroleum.

However, the impact of oil prices on other macroeconomic variables such as inflation, real Gross Domestic Product (GDP) growth rate, unemployment rate and exchange rates, is a matter of great concern for all economies. Hamilton (1983) documents that oil price increases have often been followed by economic recessions in the US since the Second World War. However, Hooker (1996) did not confirm Hamilton's results and argued that the negative relationship between oil prices and output no longer exists when the sample is extended to the 1990s. Lee et al. (1995), Ferderer (1996) and Hamilton (1996) demonstrate for sample periods that include recent years that nonlinear transformations of oil price changes restore that relationship. More recently, several studies have highlighted that economic activity is significantly affected by oil price changes (Kilian (2008) and Cologni and Manera (2008)) among others). Blanchard and Gali (2009) also found that oil price shocks have exhibited a decreased impact on GDP since 1990 for the US and other developed countries. This result can thus be explained by the fact that *"US has become less volatile and more insolent from external shocks, better economic policy, lack of large adverse shocks, or a smaller degree of energy dependence (i.e. more efficient use of energy resources and a larger share of the services sector in the economy)"* (Wu and Cavallo 2009, p. 3).

A number of studies have given special attention to the Multivariate Generalized AutoRegressive Conditional Heteroskedasticity models (*M-GARCH*) as they provide a better understanding of both volatility and co-volatility dynamics for multiple series than the nested univariate model, namely GARCH of Bollerslev (1986). The specifications include the Baba et al. 1987 (BEKK) (Engle and Kroner 1995), constant correlation model (CCC) (Bollerslev (1990), dynamic conditional correlation model (DCC) (Engle 2002) …etc.[2] The *M-GARCH* with the parameterisation BEKK (*BEKK M-GARCH*) model introduced by Engle and Kroner (1995) appears to be an appropriate methodology to reveal much more crucial information on the interaction among a given set of financial time series. Examples of recent studies on this subject include; Agren (2006) who use weekly data on the aggregate stock markets of Japan, Norway, Sweden, the UK and the US to investigate volatility spillovers from oil prices to stock markets within an asymmetric *BEKK* model. He found strong evidence of volatility spillovers for all stock markets with the exception of Sweden where evidence was weak. On the other hand, Aloui et al. (2008) find that changes in crude oil prices have a significant effect on the volatility of the stock market return of six developed countries, namely; US, UK, France, Japan, Germany and Canada using univariate (cross correlation functions) and *BEKK M-GARCH*) approaches.

Several authors have discussed in detail the inadequacy of linear models for capturing asymmetries. Therefore, regime switching models arose as an alternative to standard GARCH models allowing the behaviour of dynamic variables to depend on the state that takes place at any given point in time. The main advantage of the Markov Switching processes, often advocated in the literature, is that they can handle many

[2] For an extensive survey, see Bauwens et al. (2003).

crucial features of time series such as nonlinear phenomena, temporal asymmetries as well as persistence of the macroeconomic times series (Diebold 1986; Hamilton and Susmel 1994; Lamoureux and Lastrapes 1990). Univariate regime switching models were first proposed by Hamilton (1989, 1990) to examine the relation between turning points and changes in regimes. Markov Switching models are utilised to investigate the heteroskedastic behaviour of asset returns (Schwert 1989), the effects of oil prices on US GDP growth (Raymond and Rich 1997)…*inter alia*. Aloui and Jammazi (2009) have used univariate Markov switching *EGARCH* model with constant or time varying transition probabilities to analyse the response of the stock market returns to the oil shocks in UK, France and Japan.

Most studies to date have assumed that shock spillover intensity does not vary over time. To overcome this problem, some authors extend the standard methodology by allowing for regime switches in the volatility and spillover parameters (Beale 2002). Assuming state-dependent conditional correlations, several different Multivariate versions of Markov Switching GARCH models (*M-MSG*) have also been developed. *M-MSG* models are nested within constant conditional correlation (*CCC-GARCH*), time-varying conditional correlation (*DCC-GARCH*) of Engle and Sheppard (2001) and *BEKK-GARCH* of Gray (1996). In order to solve the path dependency problem of the Markov Switching GARCH model, i.e. the conditional variance and conditional covariance will depend on all past information, Gray (1996) suggests a tractable formulation for the conditional variance process by using the conditional expectation of the variance without giving up GARCH terms (the latter was elaborated by Hamilton and Susmel (1994) and Cai (1994) as a first solution to the path dependency problem). Haas et al. (2004), among others, modify Gray's approach to circumvent the path dependency problem. Gray's (1996) bivariate *BEKK MSG* models is perhaps the most applied model in a wide variety of applications such as estimating time-varying optimal hedge ratios (Alizadeh et al. 2008, or Lee and Yoder 2007), understanding the source and the intensity of shock spillover between stock market returns (Beale 2002). Based on Gray's approach, we propose a tractable model, namely the trivariate *BEKK MSG* model, which is more suitable for modeling the relationship between real crude oil price volatility and international real stock market returns.

In addition, using this kind of models represents another major contribution to the literature on the crude oil—stock market relationship. In fact, one limitation of existing work on the analysis of this relationship is that the price of crude oil is often treated as exogenous. However, Kilian (2008) suggests that models relying on exogenous oil price variables have been misleading in recent years. Further, Kilian (2008a, b) argue that "*direct measures of exogenous shocks to the production of crude oil have low explanatory power for the real price of crude oil*" (Kilian 2009, p. 19). Therefore, based on Kilian' arguments, our new class of model again proves to be helpful to understanding the relationship between real crude oil prices and stock market returns.

In particular, this paper analyses the shock and volatility transmission from the crude oil market to the stock market returns of US, UK, Germany, Japan and Canada under the trivariate *BEKK MSG* approach with two common states in the period January 1989 to December 2007. We combine the former with the wavelet decomposition approach, especially the Â Haar Trous Wavelet approach (*Â HTW*) in order to glean a better understanding of crude oil transmission.

Undoubtedly, *GARCH* models worked well to capture the leptokurtosis and volatility clustering generally observed in financial time series but they demonstrate some inaccuracies in terms of changes of time scales (Yalamova 2006). One major advantage afforded by wavelets analysis is its ability to perform local analysis—that is, to analyse a localised sub image area of a larger image (or signal). Therefore, wavelet analysis is capable of revealing aspects of data that other signal analysis techniques (like *GARCH* models) usually miss; aspects like trends, sharp spikes, discontinuities in higher derivatives, self-similarity…etc. Similarly, wavelet analysis can often compress or de-noise a signal without appreciable degradation (Misiti et al. 2008) because it affords a different view of data from that presented by traditional techniques. In their brief history within the signal processing field, wavelets have already proven a very useful tool for data de-noising and deconvolution (separation between two convolved signals namely smooth and detail). In this paper, we restrict our attention to "the *Â HTW* transform", introduced by Murtagh et al. (2004) and designed as well suited for outlier detection in order to decompose the real crude oil returns into six scales and a smooth part. We therefore extract the smooth series in light of the empirical evidence suggesting that the latter contains less noise than the original signal, allowing for more accurate detecting dynamic regime shifts, see Jammazi and Aloui (2009).

In summary, this paper introduces a novel insight for characterizing the relationship between crude oil market and real stock market returns. Firstly, using 6 levels *Â HTW* decomposition, we extract the main information from the real crude oil signal which is designed by the smooth low frequency part of the original series. Secondly, we examine the transmission mechanisms between the desired variables under a trivariate *BEKK MSG* model with common two states that are characterised as low mean high variance regime and high mean low variance regime. Specifically, we allow volatility in the different equity markets to depend purely on shocks and/or volatilities originated from crude oil market.

The rest of the paper is organised as follows: Sect. 2 presents the two econometric methodologies, namely *Â HTW* decomposition method and the trivariate *BEKK MSG* model. Section 3 presents the data and discusses how the smooth fluctuations of the real crude price of oil might be transmitted to the real stock market returns and Sect. 4 concludes.

2 Econometric Methodology

In this section we give a detailed description of the wavelet transform used for the crude oil data decomposition together with the multivariate *BEKK MSG* applied in our analysis.

2.1 *Signal Decomposition Using the Wavelet Method: Â Haar Trous Wavelet (Â* HTW*)*

The *Â HTW* approach was performed according to Murtagh et al. (2004). Below, we briefly recall the basic notions of the discrete wavelet theory; we present the main characteristics of the "â trous" algorithm as an alternative to the Discrete Wavelet Transform *DWT* and finally we discuss the properties of the "Â Haar Trous" wavelet decomposition approach.

2.1.1 Discrete Wavelet Transform

Contrary to the trigonometric functions, wavelets are defined in a finite domain and unlike the Fourier transform they are well-localised with respect to both time and scale. This behaviour ultimately makes them useful to analyse non-stationary signals. The other most important property of the wavelet method is that it can be used to recreate a series without loss of information. Indeed, the wavelet transform techniques split up a signal into a large timescale approximation (coarse approximation) and a collection of "details" at different smaller timescales (finer details). The coarse image preserves the large-scale structure and the mean of the image, whereas the "detail" or wavelet levels complement the coarse level and thus preserve the total image information. The first step of the wavelet de-noising method is the application of filters.

The dilation and the translation of the basis functions at different resolution levels are described by the scaling function φ, the so-called *father wavelet*, (Strang 1989) given by:

$$\phi_{j,k}(t) = 2^{-j/2}\phi\left(2^{-j}t - k\right) \quad \text{or} \quad \varphi(x) = \sum_{k} h_k \times \varphi(2x - k) \tag{1}$$

h_k denotes the low-pass filter coefficients. The low pass filter is a filter that allows only low frequency signals through its output, so it can be used to reduce the amplitude of signals with high frequencies.

Detail levels are generated from the single basic wavelet ψ, the so-called *mother wavelet*:

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right) \quad \text{or} \quad \psi(x) = \sum_{k} g_k \times \varphi(2x - k) \tag{2}$$

where $j = 1 + \cdots + J$ in a J-level decomposition. g_k is called the high-pass (or a band-pass) filter coefficients closely related to the low-pass filter (h_k) mentioned above. The high pass filter does just the opposite, by allowing only frequency components below some threshold.

The father wavelets are used to capture the smooth, low frequency nature of the data, whereas the mother wavelets are used to capture the detailed and high frequency nature of the data. The father wavelet integrates to one, and the mother wavelet integrates to zero (Heil and Walnut 1989). Thus, an original signal *f(t)* in $L^2(R)$ may be expanded approximately using these two basic wavelet functions (φ and ψ):

$$f(t) \approx \sum_j \sum_k \alpha_{j,k} \phi_{j,k}(t) \approx \sum_k s_{J,k} \phi_{J,k}(t) + \sum_k d_{J,k} \phi_{J,k}(t) + \cdots + \sum_k d_{1,k} \phi_{1,k}(t) \\ \approx \sum_k s_{J,k} \phi_{J,k}(t) + \sum_j \sum_k d_{j,k} \psi_{j,k}(t) \tag{3}$$

where $s_{J,k} = \langle f(t),\ \phi_{j,k}(t) \rangle$ and $d_{j,k} = \langle f(t),\ \psi_{j,k}(t) \rangle$ are the wavelet coefficients. The coefficients $s_{J,k}$ and $d_{j,k}$ are the smooth and the detail component coefficients respectively and are given by the projections:

$$s_{J,k} = \int \phi_{J,k} f(t) dt \tag{4}$$

$$d_{J,k} = \int \psi_{J,k} f(t) dt \tag{5}$$

2.1.2 Â Trous Wavelet Transform

A potential drawback of the application of the *DWT* in time-series analysis is that it suffers from a lack of translation invariance. To overcome this problem, some authors (Coifman and Donoho 1995 among others) suggest applying a redundant or non-decimated wavelet transform.[3]

According to Zhang et al. (2001), the advantage of the redundant wavelet transform, i.e. the so-called Trous (with holes) algorithm, lies in the fact that it is shift invariant and it produces smoother approximations by filling the "gap" caused by decimation, i.e., it is non-decimated (it conserves the original dimensions of the series). A redundant algorithm is based on the so-called *autocorrelation shell representation* using dilations and translations of the autocorrelation functions of compactly supported wavelets.[4]

The scaling and the wavelet functions are chosen to satisfy the following equations respectively:

---

[3] A detailed description of the properties of the Â Trous and the Mallat algorithm is given in Mallat (1989) and Shensa (1992).

[4] For more details, see Saito and Beylkin (1992).

$$\frac{1}{2} \times \phi\left(\frac{x}{2}\right) = \sum_{k} h(k)\phi(x-k) \tag{6}$$

$$\frac{1}{2} \times \psi\left(\frac{x}{2}\right) = \sum_{k} g(k)\psi(x-k) \tag{7}$$

where h is a discrete scaling low-pass filter while g is a discrete high-pass filter associated with the wavelet function.

These two functions satisfy the following equation:

$$\frac{1}{2} \times \psi\left(\frac{x}{2}\right) = \phi(x) - \frac{1}{2}\phi\left(\frac{x}{2}\right) \tag{8}$$

Using the filters h and g, we obtain the pyramid algorithm for expanding into the autocorrelation shell. The smoothed and the detailed signals at a given resolution j and at a position t are obtained by these convolutions:

$$s_j(t) = \sum_{l=-\infty}^{+\infty} h(l)s_{j-1}(t + 2^{j-1} \times l) \tag{9}$$

$$d_j(t) = \sum_{l=-\infty}^{+\infty} g(l)s_{j-1}(t+2^{j-1} \times l) \tag{10}$$

where $1 < j < J$, h is a low-pass filter.

A very important property of the autocorrelation shell coefficients is that signals can be directly derived from them Zhang et al. (2001). In each step, the series is convolved with a cubic *B-spline* filter, h, with $2^{j-1} \times l$ zeros inserted between the *B-spline* filter coefficients at level j, hence the name "with holes". The convolution mask in one dimension is 1/16 [1, 4, 6, 4, 1]. Thus, we get a series of smoothed versions s_j with s_0 ($s_0(t) = x(t)$ the finest scale) as the normalized raw series. Given a smoothed signal at two consecutive resolution levels, the detailed signal $d(t)$ at level j, can be derived as:

$$d_j(t) = s_{j-1}(t) - s_j(t) \tag{11}$$

The set $d = \{d_1(t), d_2(t), \ldots d_J(t), s_J(t)\}$ represents the wavelet transform of the signal up to the scale J, and the signal can be expressed as a sum of the wavelet coefficients and the scaling coefficient:

$$x(t) = s_J(t) + \sum_{j=1}^{J} d_{\mathrm{j}}(t) \tag{12}$$

2.1.3 The Â Haar Trous Wavelet Transform (*Â HTW*)

Here, we select Haar wavelet filter to implement the Â Trous wavelet transform. The asymmetry of the wavelet function used makes it a good choice for edge detection, i.e., localised jumps. However, the usual Haar wavelet transform is decimated. Consequently, Murtagh et al. (2004) develop a non-decimated or redundant version of this transform. The non-decimated or redundant algorithm is the Â Trous algorithm with a low-pass filter h = (1/2, 1/2).

The non-decimated Haar algorithm is exactly the same as the Â trous algorithm, except that the low-pass filter h, (1/16…etc.), is replaced by the simple non-symmetric filter h = (1/2, 1/2). By convolving the original signal with the wavelet filter h, we create the wavelet coefficients.

$$s_{j+1} = \frac{1}{2}\left(s_{j,t-2^j} + s_{j,t}\right) \tag{13}$$

Thus, the scaling coefficients at a higher scale can be easily obtained from the scaling coefficients at a lower scale:

$$d_{j+1}(t) = c_j(t) - c_{j+1}(t) \tag{14}$$

2.2 *Wavelet-Multivariate Markov Switching GARCH-BEKK Model*

Several studies on the transmission volatility between different financial variables are based on the estimation of multivariate *BEKK GARCH* models (Saleem 2009; Li and Majerowska 2008; Bachmeier 2008; Malik and Hammoudeh 2007; Agren 2006 among others).

Although these models are parsimonious, they were based on constant shock and volatility transmissions. Multivariate Regime Switching models, which are both time varying and state dependent, are used henceforth to solve this problem. The main advantage of Markov-switching processes, often advocated in the literature, is their ability to take into account features such as nonlinear phenomena, temporal asymmetries as well as persistence of the macroeconomic time series: these features are crucial in the analysis of the dynamic linkage between crude oil prices and stock market returns (Aloui and Jammazi 2009). Hamilton and Susmel (1994) and Cai (1994) were the first to allow for regime-switches in the ARCH process. Gray (1996) extended their methodology to regime switching GARCH-models. In this section, we extend the standard multivariate BEKK-GARCH model of Engle and Kroner (1995) to allow for the presence of regime shifts. Finally, we discuss the trivariate wavelet *BEKK MSG* that we will use in the current analysis in order to study the transmission mechanism of shocks (volatility) originating from crude oil market to equity market returns.

2.2.1 Generalised Regime Switching GARCH Model with Path Dependent Volatility

Following Haas and Mittnik (2008), in this section we derive the multivariate *BEKK MSG* process.

Let us suppose that the joint process for a given number of series is governed by the following set of equations:

$$\begin{aligned} R_t &= \Phi + E_t \\ e_{t,s_t} &= H_{\Delta_t,t}^{1/2} E_t \qquad E_t/\Omega_{t-1} \rightarrow N(0_{M\times 1}, I_{\mathrm{M}}) \end{aligned} \tag{15}$$

Both the return R and the variance H are made regime dependent. Let R_t be the return matrix at time t, modeled as a constant plus a disturbance term. Φ constitutes the constant vector, I_M denotes the identity matrix of dimension M, The transition between the successive states is governed by a first order Markov process $\{\Delta_t\}$ with finite state space $S = \{1, 2, \ldots, k\}$ and a primitive (i.e., irreducible and aperiodic) fixed $k \times k$ transition probability matrix P,

$$P = \begin{bmatrix} p_{11} & \cdots & p_{k1} \\ \cdots & \cdots & \cdots \\ p_{1k} & \cdots & pkk \end{bmatrix} \tag{16}$$

where the transition probabilities are given by

$$p_{ij} = p(\Delta_t = j/\Delta_{t-1} = i), \quad i,j = 1, \ldots, k$$

The regime-dependent covariance matrix H is assumed to follow a *Multivariate Markov Switching GARCH (p, q, k)*) in Vech form as introduced by Bollerslev et al. (1988);

$$h_{jt} = \gamma_{0j} + \sum_{i=1}^{q} \alpha_{ij}\eta_{t-i} + \sum_{i=1}^{p} \beta_{ij}h_{jt-i} j = 1, \ldots, k \tag{17}$$

where $\alpha_i = [\alpha_{i1}^{'}, \ldots, \alpha_{ik}^{'}]^{'}$, $i = 1, \ldots, q$ and $\beta_i = [\beta_{i1}^{'}, \ldots, \beta_{ik}^{'}]^{'}$, $i = 1, \ldots, p$ are parameter matrices of appropriate dimension. The number of the independent element of the regime-dependent conditional covariance matrices H_{jt}, is $N := M(M + 1)/2$. The "squared", $(ee_t^{'})$ in $h_{jt} := vech(\mathrm{H}_{jt})$ and $\eta_t := vech(e_t e_t^{'})$, respectively.

A major disadvantage of using the model defined in (17) is that the positive definiteness of the estimated conditional covariance matrices is not guaranteed (Ding and Engle 2001) Every covariance matrix must be positive definite but for this model it is probably impossible to give general restrictions on parameters to insure a positive definite covariance matrix.

Parameter constraints are required to make the application trustworthy. Such a parameterisation is provided by the Baba et al. (1987) (BEKK) representation of Engle and Kroner (1995) which specifies the conditional volatility as

$$H_{jt} = \gamma_{0j}^{*}\gamma_{0j}^{*'} + \sum_{l=1}^{L}\sum_{i=1}^{q}\alpha_{ij,l}^{*}e_{t-i}e_{t-i}^{'}\alpha_{ij,l}^{*'} + \sum_{l=1}^{L}\sum_{i=1}^{p}\beta_{ij,l}^{*}H_{t-i}\beta_{ij,l}^{*'}j = \{1,\ldots,k\}$$

where γ_{0j}^{*} are k × k lower triangular matrices of state dependent coefficients, L is the lag operator. γ_{0j}^{*}, α_{ij}^{*} and β_{ij}^{*} are state dependent matrices.

By recombining the GARCH model to regime switching and given h_0^2, recursive substitution in a univariate *MS-G* (1,1) model yields Haas et al. (2004):

$$h_{t,s_t}^2 = \sum_{i=0}^{t-1}(\gamma_{s_{t-i}} + \alpha_{s_{t-i}}e_{t-1-i}^2)\prod_{j=0}^{i-1}\beta_{s_{t-j}} + h_0^2\prod_{i=0}^{t-1}\beta_{s_{t-i}} \tag{18}$$

Although the *BEKK* model involves far fewer parameters than the unrestricted *vech* form, the conditional variance as specified in Eq. (18) suffers from the path dependence problem. Indeed, in this formulation, the state dependent conditional variances are a function of the lagged values of the lagged aggregated variances and aggregated error terms (after integrated the unobserved state variable).

To circumvent the path dependency problem, Gray (1996) introduces a recombining method that collapses the conditional variances in each regime by taking the conditional expectation of h_t^2 based on the regime probabilities.[5] As a consequence, the conditional variance and the residual depend only on the current regime, not on the entire past history of the process. Based on the Gray (1996)'s recombining method, in the following section we analyse how this path dependence problem may be resolved in our trivariate MS-G model case.

2.2.2 Circumventing the Path Dependency Problem: Case of a Trivariate Markov Switching BEKK GARCH (Trivariate *BEKK MSG*)

Since three equations complicate the estimation considerably, we have to make some choices in terms of the required number of volatility states and parameters involved in the estimation procedure. We restrict our study to the case of three equations and two states. Thus, the state-dependent crude oil and stock market returns are specified as:

[5] Gray (1996) proposes a recombining method for the univariate Markov Switching volatility model. For a detailed description of the path-dependence problem and its solution for the univariate MS GARCH process case, see Lee and Yoder (2007).

$$\begin{aligned} r_{s,t} &= \mu_{s,s_t} + e_{s,t,s_t} \\ r_{w,t} &= \mu_{w,s_t} + e_{w,t,s_t} \\ r_{b,t} &= \mu_{b,s_t} + e_{b,t,s_t} \end{aligned} \tag{19}$$

where subscribers s, w, and b denote real stock market returns, WTI and Brent real crude oil volatilities (the smooth part), see Eq. (13) respectively, μ is a constant where $\Phi = (\mu_{s,s_t}\mu_{w,s_t}\mu_{b,s_t})'$. e_{s,t,s_t}, e_{w,t,s_t} and e_{b,t,s_t} are state dependent residual terms. The unobserved state variable $s_t = \{1, 2\}$ is interpreted as the market state or regime when the process is at time t, which follows a first-order, 2-dimensional state Markov process.

The conditional variances are specified as:

$$E_{t,s_t}/\psi_{t-1} = \begin{bmatrix} e_{s,t,s_t} \\ e_{w,t,s_t} \\ e_{b,t,s_t} \end{bmatrix} /\psi_{t-1} \rightarrow TN\left(0, H_{t,s_t}\right) \tag{20}$$

TN denotes the trivariate normal. H_{t,s_t} is a state-dependent conditional variance-covariance matrix of each return.

The time-varying 3 × 3 positive definite conditional covariance matrix, H_{t,s_t}, is specified as (where $p = q = 1$):

$$\begin{aligned} H_{t,s_t} = \begin{bmatrix} h^2_{s,t,s_t} & 0 & 0 \\ 0 & h^2_{w,t,s_t} & 0 \\ 0 & 0 & h^2_{b,t,s_t} \end{bmatrix} &= \begin{bmatrix} \gamma_{ss,s_t} & 0 & 0 \\ 0 & \gamma_{ww,s_t} & 0 \\ 0 & 0 & \gamma_{bb,s_t} \end{bmatrix} \begin{bmatrix} \gamma_{ss,s_t} & 0 & 0 \\ 0 & \gamma_{ww,s_t} & 0 \\ 0 & 0 & \gamma_{bb,s_t} \end{bmatrix}' + \begin{bmatrix} \alpha_{ss,s_t} & \alpha_{sw,s_t} & \alpha_{sb,s_t} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}' \\ &\begin{bmatrix} e^2_{ss,t-1} & e_{s,t-1}e_{w,t-1} & e_{s,t-1}e_{b,t-1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{ss,s_t} & \alpha_{sw,s_t} & \alpha_{sb,s_t} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &+ \begin{bmatrix} \beta_{ss,s_t} & \beta_{sw,s_t} & \beta_{sb,s_t} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}' \begin{bmatrix} h^2_{s,t-1} & h_{sw,t-1} & h_{sb,t-1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{ss,s_t} & \beta_{sw,s_t} & \beta_{sb,s_t} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \Gamma_{s_t}\Gamma'_{s_t} + \mathrm{A}_{s_t}\mathrm{E}_{t-1}\mathrm{A}'_{s_t} + \mathrm{B}_{s_t}\mathrm{H}_{t-1}\mathrm{B}'_{s_t} \end{aligned} \tag{21}$$

where Γ_{st} is a *3 × 3* diagonal matrix of state dependent coefficients, A_{st} and B_{st} are 3 × 3 state dependent coefficient matrices restricted to be of 1 × 3 dimension for further simplification.

h_{sw,t,s_t} and h_{sb,t,s_t} are conditional covariance at time t given s_t, and h^2_{s,t,s_t}, h^2_{w,t,s_t} and h^2_{b,t,s_t} are conditional variances at time t given s_t. The matrices Γ_{s_t}, A_{s_t} and B_{s_t} and $E_{t\text{-}1}$ are compact representations of the state-dependent coefficients γ, α, β and e respectively.

We will refer to the model defined by Eq. (21) as a trivariate BEKK Markov-switching *GARCH (1,1;2)* process or, in short triavariate *BEKK-MSG (1,1;2)*. Since we are interested in providing the results related to the shock and volatility transmission only from the crude oil market to the stock market in presence of regime switching, we assume that only $h^2_{s,t}$ follows a *BEKK-MSG (1,1)* process under two volatility states (high volatility and low volatility) and each of $h^2_{w,t}$ and $h^2_{b,t}$ follow a constant.[6] We allow for the vectors of mean and variance parameters to switch across two regimes.

As in the univariate regime switching GARCH model, the recursive nature of the GARCH process makes the basic form of the model intractable due to the dependence of the conditional variance on the entire past history of the data. Indeed, only the first equation i.e., $h^2_{s,t}$, of the proposed trivariate GARCH model, is subject to the path-dependency problem. Hence, it depends directly on the state variable s_t and $h^2_{s,t-1}$, which itself depends on s_{t-1} and $h^2_{s,t-2}$ and so on. The computation of the likelihood function for a sample of length T requires the integration over all 2^T possible (unobserved) regime path, rendering estimation of the model infeasible in practice. This is the well-known path dependency problem in the regime switching literature (Cai 1994; Hamilton and Susmel 1994; Gray 1995, 1996). Furthermore, this problem is present not only in variances and residuals, but also in the covariance between crude oil and stock market returns $h_{sw,t}$ and $h_{sb,t}$.

Using Gray (1996)'s recombining method at time 1, the path-independent conditional variance, residual and covariance for the stock market variance-covariance equation are given, respectively, by:

$$\begin{aligned} h^2_{s,t} &= E\left(r^2_{s,t}|\psi_{t-1}\right) - E\left(r_{s,t}|\psi_{t-1}\right)^2 \\ &= p_{1,t}\left(\mu^2_{s,1} + h^2_{s,t,1}\right) + (1-p_{1t})\left(\mu^2_{s,2} + h^2_{s,t,2}\right) - \left[p_{1t}\mu_{s,1} + (1-p_{1t})\mu_{s,2}\right]^2 \end{aligned} \tag{22}$$

$$\begin{aligned} e_{s,t} &= r_{s,t} - E\left[r_{s,t}|\psi_{t-1}\right] \\ &= r_{s,t} - \left[p_{1t}\mu_{s,1} + (1-p_{1t})\mu_{s,2}\right] \end{aligned} \tag{23}$$

$$\begin{aligned} h_{si,t} &= Cov(r_{s,t}, r_{i,t}|\psi_{t-1}) \\ &= E\left[r_{s,t}r_{i,t}|\psi_{t-1}\right] - E\left[r_{s,t}|\psi_{t-1}\right]E\left[r_{i,t}|\psi_{t-1}\right] \quad i = \{w, b\} \end{aligned} \tag{24}$$

where;

$$E\left[r_{s,t}r_{i,t}|\psi_{t-1}\right] = p_{1t}\left(\mu_{s,1}\mu_{i,1} + h_{si,t,1}\right) + (1-p_{1t})\left(\mu_{s,2}\mu_{i,2} + h_{si,2}\right) \tag{25}$$

$$E\left[r_{s,t}|\psi_{t-1}\right] = p_{1t}\mu_{s,1} + (1-p_{1t})\mu_{s,2} \tag{26}$$

$$E\left[r_{i,t}|\psi_{t-1}\right] = p_{1t}\mu_{i,1} + (1-p_{1t})\mu_{i,2} \tag{27}$$

[6] Henceforth, the conditional covariances $h_{ws,t-1,s_t}$ and $h_{bs,t-1,s_t}$ and the variances $h^2_{w,t-1,s_t}$ and $h^2_{b,t-1,s_t}$ were fixed to be zero.

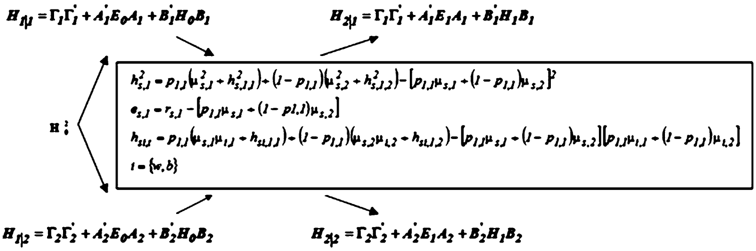


Fig. 1 Path-independent conditional variance of a trivariate BEKK-MSG model

With this definition, the conditional covariance depends only on the current regime, not on the entire past history of the process. The model is then state-independent and tractable even with large samples.

A graphical illustration for the recombining method for *BEKK* Markov Switching model is shown below (Fig. 1).

The regime probability of being in state 1 at time t is:

$$\begin{aligned} p_{1t} &= \Pr(s_t = 1|\psi_{t-1}) \\ &= P\left[\frac{f_{1t-1}p_{1t-1}}{f_{1t-1}p_{1t-1} + f_{2t-1}(1 - p_{1t-1})}\right] + (1 - Q)\left[\frac{f_{2t-1}(1 - p_{1t-1})}{f_{1t-1}p_{1t-1} + f_{2t-1}(1 - p_{1t-1})}\right] \end{aligned} \tag{28}$$

where

$$\begin{aligned} P &= \Pr[s_t = 1|s_{t-1} = 1] \\ Q &= \Pr[s_t = 2|s_{t-1} = 2] \end{aligned} \tag{29}$$

$$f_{st} = f(R_t|s_t = i, \psi_{t-1}) = (2\pi)^{-1}\left|H_{t,i}\right|^{-1/2}\exp\left\{-1/2e'_{t,i}H^{-1}_{t,i}e_{t,i}\right\}, \text{ for } i = \{1, 2\} \tag{30}$$

$R_t = [r_{s,t}r_{w,t}r_{b,t}]'$ is a vector of crude oil and stock market returns at time t. H and e are defined in Eqs. (20) and (21), respectively.

The steady-state probabilities of s_t used as the initial start value for the recursive expression of the regime probability is:

$$\Pr(s_t = 1|\psi_0) = \frac{1 - Q}{2 - P - Q} \tag{31}$$

where P and Q are state transition probabilities assumed to follow a logistic distribution defined as in the following equations;

$$P = \Pr[s_t = 1 | s_{t-1} = 2] = \frac{\exp(p_0)}{1 + \exp(p_0)}$$
$$Q = \Pr[s_t = 2 | s_{t-1} = 2] = \frac{\exp(q_0)}{1 + \exp(q_0)} \tag{32}$$

p_O and q_O denote unconstrained constant terms which have to be estimated along with the regression coefficients' system.

Given the path independent *BEKK MSG* model as described by Lee and Yoder (2007), the unknown parameters that we seek to estimate for our trivariate case model are $\{p_0, q_0, \mu_{s,s_t}, \mu_{\mathrm{w,s_t}}, \mu_{\mathrm{b,s_t}}, \gamma_{ss,s_t}, \gamma_{sw,s_t}, \gamma_{sb,s_t}, \alpha_{ss,s_t}, \alpha_{sw,s_t}, \alpha_{sb,s_t}, \beta_{ss,s_t}, \beta_{sw,s_t}, \beta_{sb,s_t}\}$ for $s_t = \{1, 2\}$. We obtain the estimates parameters by maximising the following log-likelihood function.

$$LL = \sum_{t=1}^{T} \log[p_{1t} f_{1t} + (1 - p_{1t}) f_{2t}] \tag{33}$$

where f_{it} for $i = \{1, 2\}$ is defined as shown in Eq. (30).

3 Methodology Results and Discussions

3.1 Data

Our analysis deals with two variables; (1) real stock returns of five major industrial countries, namely; US (DJIA), UK, (FTSE100), Germany (Dax30), Japan (NIKKEI225) and Canada (TSX) and (2) real prices of two major crude oil products, defined as the US price of West Texas Intermediate Cushing (WTI) and the Europe Brent which are quoted in dollars per barrel. Crude oil prices were extracted from the US Department of Energy (Energy Information Administration), while stock market prices were taken from the International Financial Statistics databases (IFS). All the data are measured on a monthly basis. The use of a monthly frequency is justified by the need to observe common high volatility phases that are expected to be coincident with the ECRI recession dating periods which are also provided in monthly frequency over the investigated period. The sample covers the period from January 1989 to December 2007, for a total of 228 observations. All the data were used in real terms. For each country, real stock returns are defined as the difference between the continuously compounded return on stock price index and the inflation rate given by the log-difference in the consumer price index. Consumer price indices are from OECD databases. On the other hand, the most accurate measure of an oil shock is the real oil price. The world oil prices were therefore deflated by the consumer price index (CPI) of each country. In other words, we take the world price of oil in US $ and divide by the CPI of each country.

This choice of variables may ultimately be crucial for comparison purposes. Indeed, many of the recent studies have shown that net oil prices have predictive content for determining stock market turning points (Aloui and Jammazi 2009). In contrast to some work, we would like to show that the real oil prices are also a useful predictor of turning points in stock markets. Figure 3 (left panel) plot the real equity returns and the smooth part of the real crude oil returns.[7] It is likely that time series include structural changes in the mean during the investigated period. For instance, real DJIA return series increases especially around 1992 and 2007. However, for the other countries, real equity returns experience several jumps throughout most of the period that roughly coincide with the major conventional crises.

The results from Fig. 3 (left panel) provide some preliminary evidence of (roughly) coincidental market volatility switches between real stock returns and the smoothed real crude oil volatility during the study period. In the following sections, we explore this issue further by applying the trivariate wavelet-BEKK MSG model. Let us start with the extraction of the smoothed series for the crude oil volatility index based on the new wavelet decomposition method described above.

3.2 Haar Trous Wavelet Decomposition: Application to the Real Crude Oil Volatility

Oil prices have traditionally been more volatile than many other commodity or asset prices (Regnier 2007). Recently, it has been claimed that "*Wavelet filtering is particularly relevant to volatile and time-varying characteristics of real world time series.*" (Chang and Fan 2008, p. 803).

To verify this, monthly real crude oil price volatilities were used to assess the performance of the $\hat{A}\ HTW$ algorithm in getting a smooth component without losing the underlying characteristics of the respective series. Indeed, the input data consists of the monthly real crude oil price volatility of the West Texas Intermediate Cushing (WTI) and the Europe Brent real oil returns (expressed in \$/bbl) for the period January 1989–December 2007. The real crude oil market volatility R_{it} is taken as the log difference of real crude oil price P:

$$R_{it} = LogP_t - LogP_{t-1}$$

where P_t is the real crude oil price at date t.

The two transformed series are decomposed into their time scale components using $\hat{A}\ HTW$ which is redundant or non-decimated method. The wavelet filter used

[7] We first decompose the original signal (monthly real crude oil returns) using the THW transform. We then extract the smooth part from the signal. We will discuss this in more detail in the following section.

is the discrete low pass filter (*G*) of length, $L = 6$. The sifting processes produce six level details which are captured by *scale 1*, *scale 2*,…, *scale 6* plus the smoothed series (*Smooth*) each containing (the total sample size) 228 samples. At each scale, the corresponding component is reconstituted according to Eqs. (13) and (14). Figure 2 plot the original series (*signal*), the details (*scale 1 to 6*) and the smoothed series (*smooth*) for the real crude oil volatilities of US, UK, Germany, Canada and Japan. The standard deviations (*SD*) of each detail are not uniform across the series but proportional to the *SD* of the underlying signal. Since we use monthly data, the level of details represents the variations within 2^i months horizon which correspond to 4–8, 8–16, 16–32, 32–64 and 64–128 month dynamics, respectively. All the details are listed from the highest to the lowest frequency. The most short-run fluctuations are observed in the two finest components scales 1, and 2 and some in scale 3 which contain the high frequency content, so that they are extremely sensitive to non-smooth data characteristics such as noise, jumps, and spikes in the data. However, scales 4 to 6 depict medium and long-term fluctuations of the series. As the wavelet resolution level increases, the corresponding coefficients become smoother and the smooth trend (the coarsest approximation series) contains the lower frequency movements.

One of the advantages of the wavelet transform is that it can be used to analyse structural break at different time scales (Tommi 2005).

As noted in his article, Hamilton (2005) argues that nine of the last ten recessions during the post- II World War period in the US were preceded by large increases in oil prices. Suppose instead that we believe large oil shocks are followed by sharp recessions. To do so, we first look at the recession history with a particular focus on how each recession is preceded by a specific oil shock.[8] Henceforth, shaded bars in Fig. 2 indicate recessionary periods in months, as identified by Economic Cycle Research Institute (ECRI) from 1989 to 2007 (available upon request). According to ECRI dating, recession periods show some similarities and differences in the growth of business cycles. All the countries experienced six (single or double adjacent) recessions in the period studied (except for UK).[9] These recessions took place in 1990 (the mid-1990s Gulf war), 1994 (the Mexican Peso crisis), 1997 (the East Asian financial crisis), 2000 (economic recession in US), 2004 (Argentine energy crisis) and 2007 (the US mortgage subprime crisis). The 1994 recession in US and Canada lasted longer than in UK, Germany and Japan. However, the 1997 recession was longer in the US and UK. The main difference in the business cycle's growth among these countries concerns the recession in 1990. This recession started earlier in UK, US, Canada but 2 years later in Germany. On the other hand, Japan experienced double recessions during the same period. The recession in early 2000 was long for UK, lasting about 2 years, and shorter for Japan; on the other

[8] It is important to note that we do not attempt to analyse the causality between the crude oil spike volatility and recessions but are just trying to examine graphically the correlation between them at different time scale.

[9] UK experienced only five recessions compared to the other countries.

Fig. 2 Haar A Trous Wavelet decomposition of the real crude oil volatilities. The top panel: the original series (signal) and the smoothed series (smooth). The six panels namely scale 1 to scale 6: the wavelet components (vertical axis represents the amplitude of scaling coefficients (in Hertz). The shaded vertical bars indicate Growth Cycle recessions as dated by ECRI "Economic Cycle Research Institute." The sample period is January 1989 to December 2007, a total of 228 observations

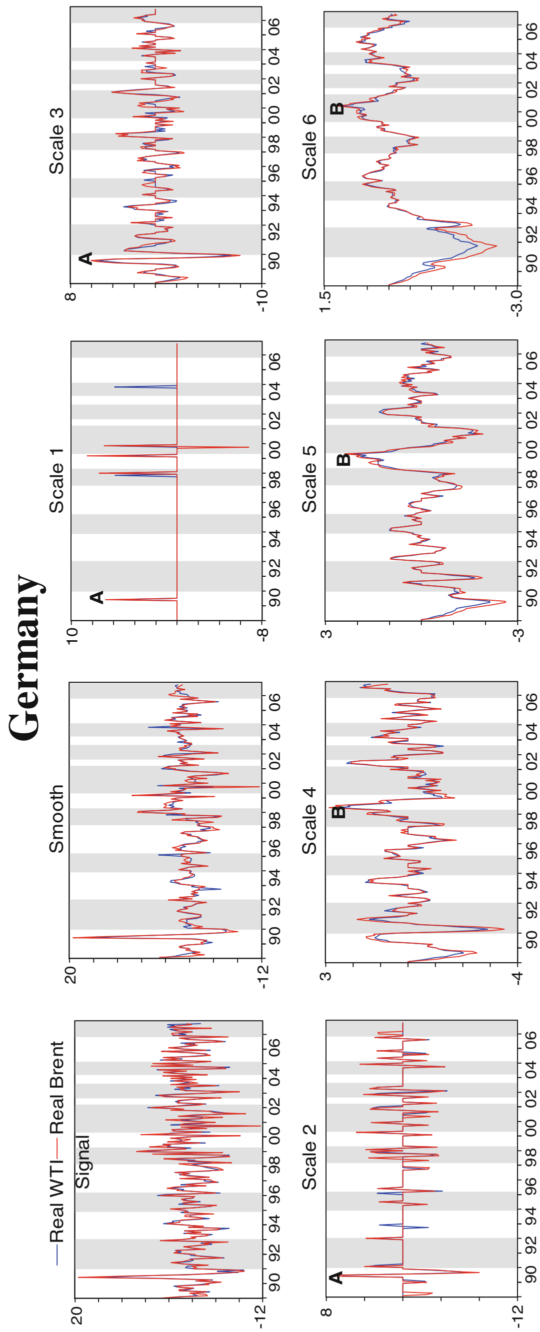


Fig. 2 (continued)

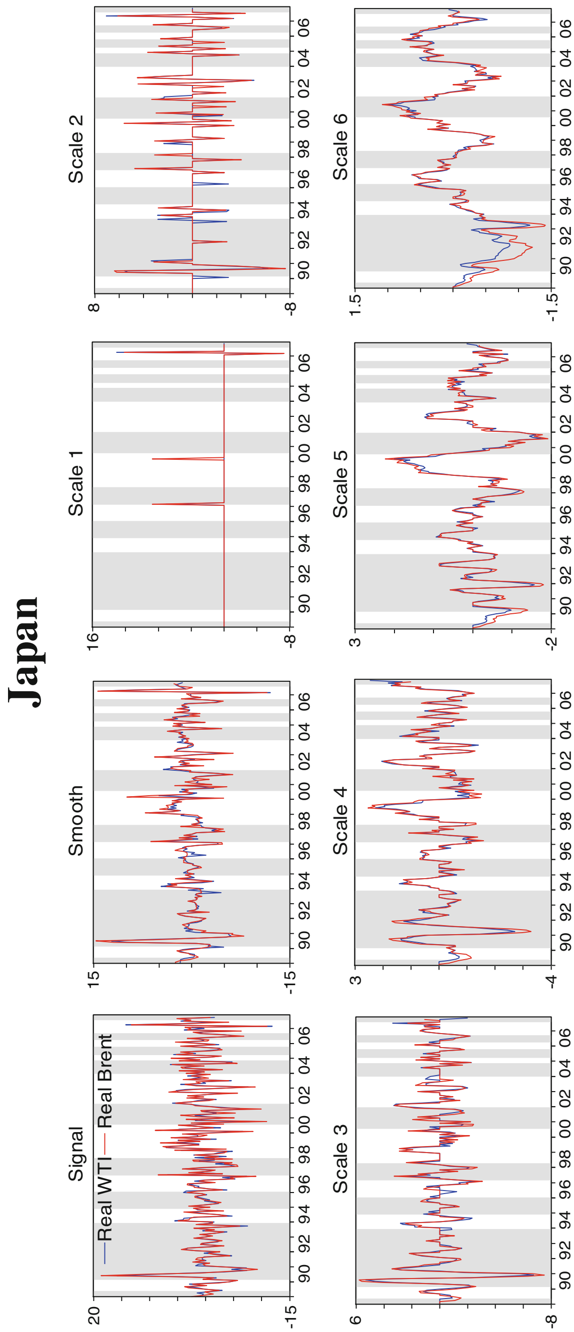


Fig. 2 (continued)

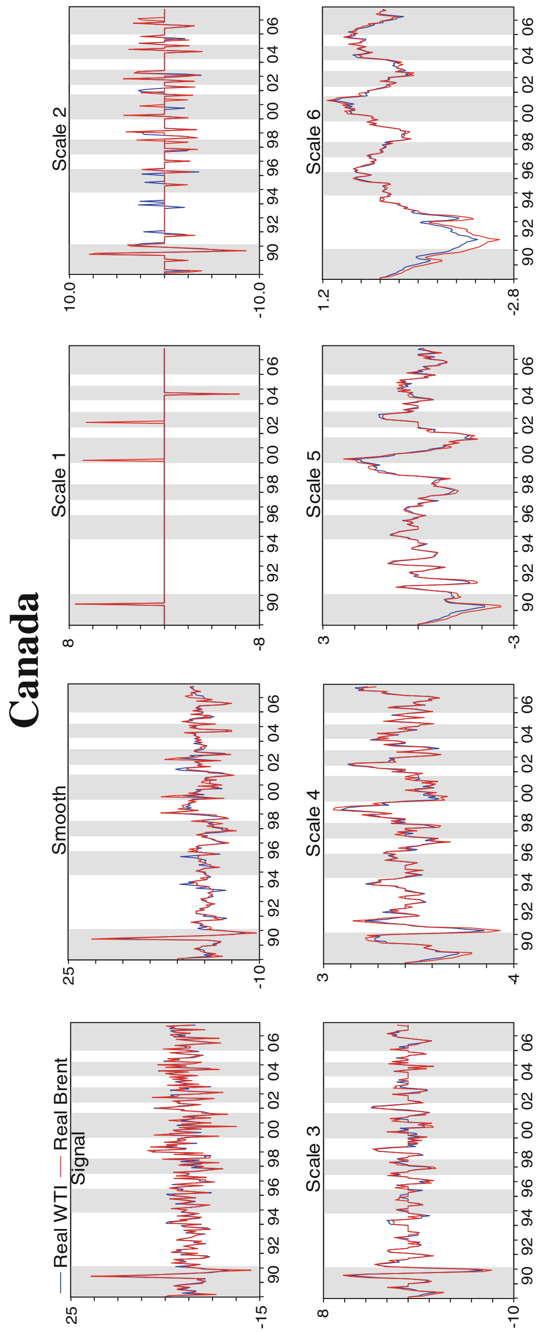


Fig. 2 (continued)

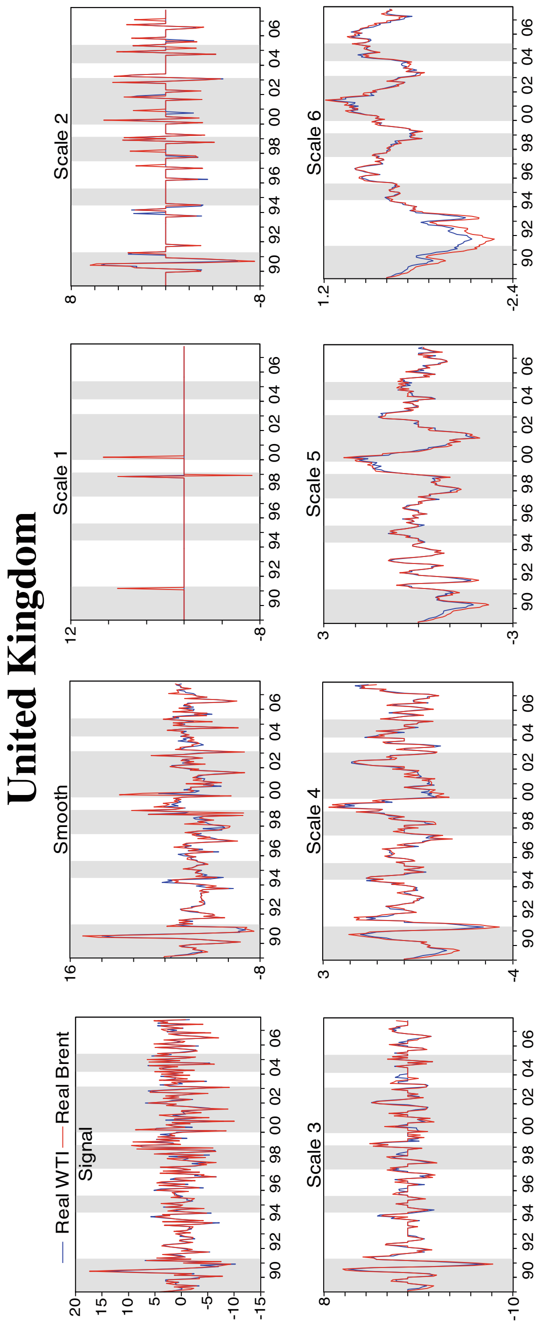
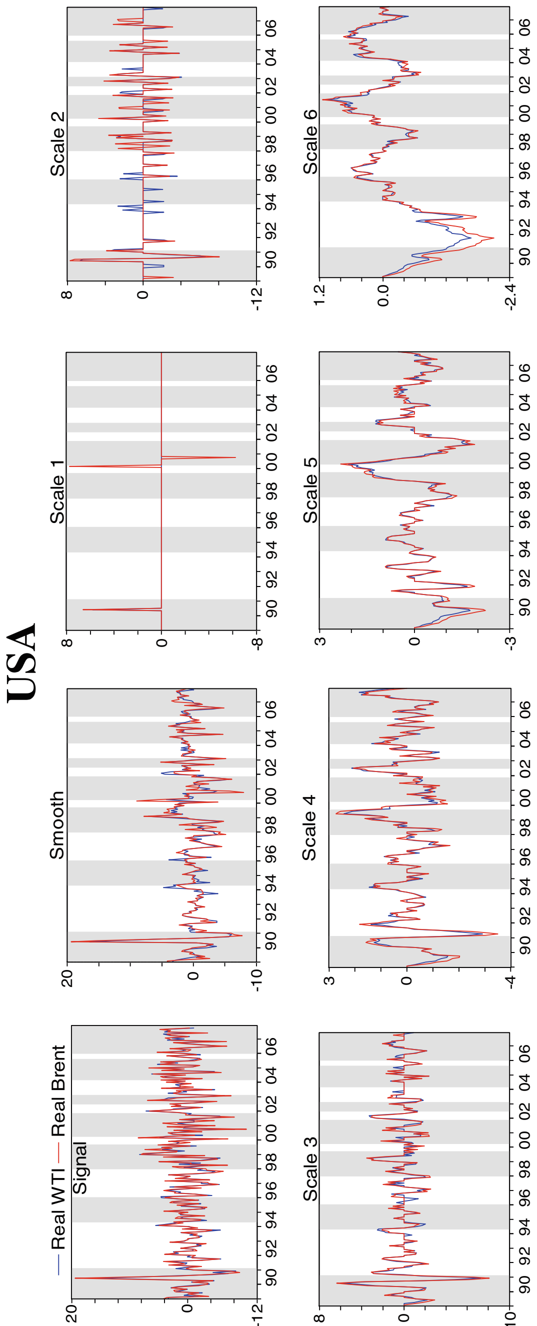


Fig. 2 (continued)



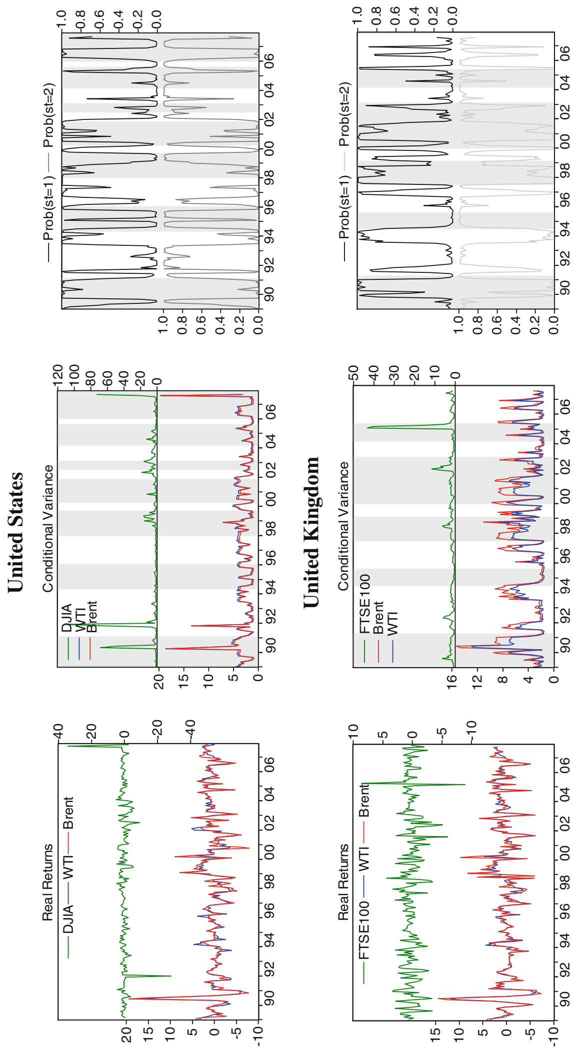


Fig. 3 The *left panel* monthly real stock market returns and the smoothed real crude oil volatilities. The *second panel* the conditional variances obtained from the trivariate RS-BEKK-GARCH model. The *right panel* smoothed probabilities of regime 1 and of regime 2 that the three markets are jointly in regime 1 (high volatility regime) at time *t* and in regime 2 (low volatility regime) at time *t* respectively. The shaded vertical bars indicate Growth Cycle recessions as dated by ECRI "Economic Cycle Research Institute." The sample period is January 1989 to December 2007, a total of 228 observations

hand, two shorter recessions occurred close to each other during the same period for US, Canada and Germany. The 2004 recession started and ended at about the same time while Japan again had two recessions during this period. In 2006, Canada,

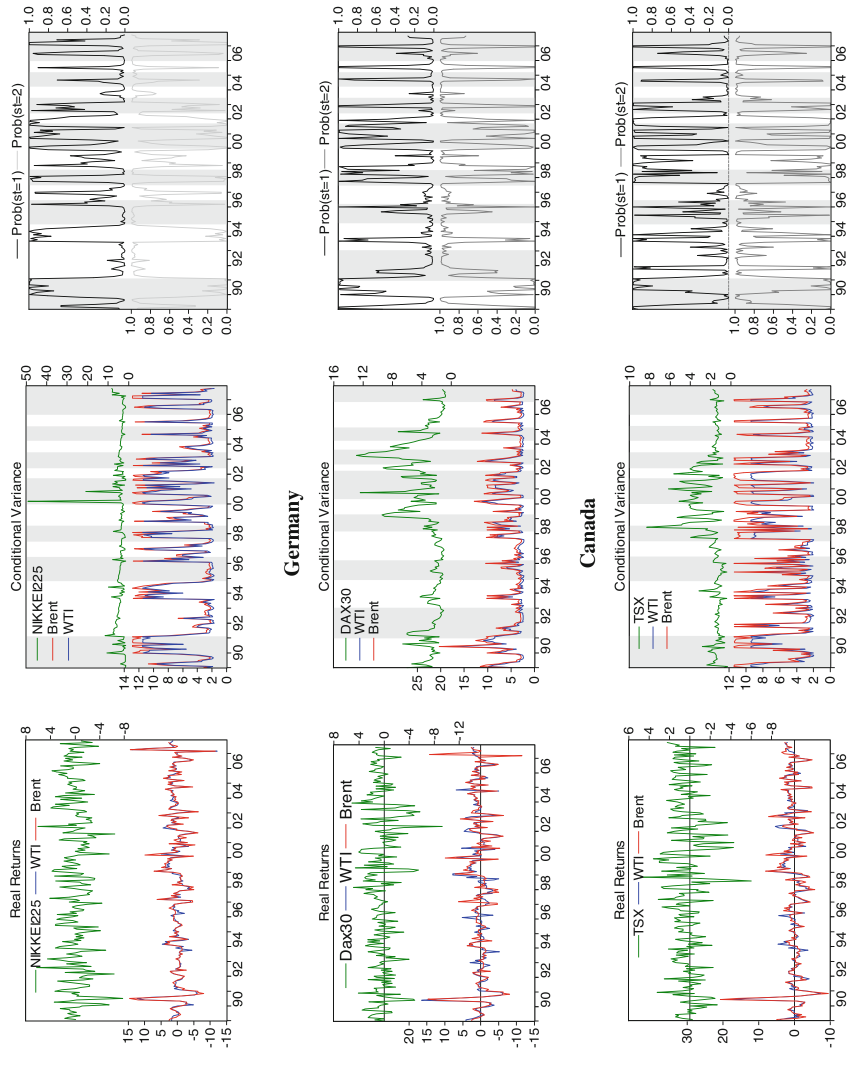


Fig. 3 continued

Germany and Japan sank into a recession at about the same time. However, this latter crisis did not hit UK.

The obtained wavelet coefficients were used to identify characteristics of the time-scale signal (smooth) that were not apparent from the original time domain signal. Therefore, Fig. 2 (scales 1–6) show that crude oil volatility peak detections are easily perceptible in the finest scales (short-term fluctuations of the series) as well as in the coarsest scales (medium and long-term fluctuations of the series). From these plots, it is easy to see which peak features are meaningful at any specific time in world history. For example,[10] in levels 1–3, the wavelets capture well the most intense volatility peak denoted by "A", which has a value of 6 or 7 and occurs in June/July 1990 for all the country cases. Essentially, this huge short-term real crude oil volatility peak leads to the 1990s recession. On the other hand, low frequency waves (scale 4–6) present fewer and thicker spikes with smaller lengths. For instance, wavelet is capable of capturing the long-term real crude oil volatility peak denoted by "B" which has a value of about 2 and occurs in 1999/2000. This followed the early 2000s recession. These plots also highlight the wavelet's strength of detecting pertinent information at varying decomposition levels. It can be seen that this evidence is also supported in the smooth series. Indeed, the studied period began with a huge oil shock in 1990 (Japan has a second largest oil shock which took place at the beginning of 2007). One can observe again that the spike of 1990 seems to be the historical spike at which the global economy can achieve a severe crisis. After this dramatic increase in real crude oil volatility, political controls try to stabilise the oil price trend. The second highest real crude oil volatility, which rises and falls in a distinct series of spikes, was at the beginning of 2000 in almost all the countries. Furthermore, it is unequivocal that there are several instances of coincidence of recessions with crude oil volatility spikes identified by the smooth series. Indeed, the initial spike volatility case was followed by a recession only for Germany and Canada[11] while the latter spike volatility case was followed by a recession for all the economies. The other ECRI recession cases were preceded by rather small oil shocks.

After verifying Hamilton's assumption, we proceed with our analysis by improving further Â THW effectiveness; that is the possibility of noise level reduction while preserving the significant feature of the original signal. Indeed, although the original signal (Fig. 2 (top left panel)) presents several peaks that precede each identified international crisis, unfortunately they are noise contaminated.

It is apparent from the plot of the smooth series (Fig. 2 (top right panel)) that the noise is reduced but the peak height is also reduced slightly. Indeed, the smoothed peaks and original unsmoothed peaks are not perfectly coincident. This is not always the case as the presence of noise can shift the peak by 1–3 sample locations. After undergoing the smoothing algorithm, the peak values are higher in amplitude than the noisy peak, and this agreement is typical of the better quality data. Finally, we could easily argue that the reconstructed signal has a simple and very smooth fluctuation that allows for easy interpretation.

Further probing led to the discovery that each spike in the oil volatility series was matched by transient instabilities in another economic indicator, including stock market returns (Cologni and Manera 2009). Our interest lies in whether oil

[10] This example is only illustrated in the case of Germany. The remaining figures generally report the same behaviour.

[11] A potential explanation of this result is that a prolonged recession occurred at the beginning of 1988 (not included in our dataset) was preceded by successive oil shocks and that conducted to the recession of 1990 for US, UK and Canada.

price changes affect the stock market returns. Figure 3 (left panels) plots real stock returns and the smooth real crude oil returns for each country. The relationships shown in this graph were correlative. Care has thus to be taken since correlation in time does not imply causation. Bearing this in mind, the hypothesis posed was that these recurring spikes of volatility in oil price destabilised the stock market returns.

3.3 Estimation Results of the Multivariate Markov Switching Model

Having the true real crude oil volatility signal in hand, the analysis that follows endeavours to investigate whether switches in this signal have a trend towards higher stock market volatility in the five developed countries. In particular, we assume that high volatility states coincide across the two markets and we use our data set to inquire whether these states coincide with the main international crises.

The estimation of our trivariate *BEKK* MSG (1,1;2) as specified in Eq. (21) already gives us five three-market combinations where each one contains three variables: WTI real returns, Brent real returns and the respective individual real developed-country stock market returns (i.e., US, U.K., Germany, Japan, and Canada). We refer to the crude oil markets as "potential originators" and the stock markets as "potential recipient markets" because we want to explore whether shocks and volatilities originating from these markets are related with shocks and volatilities of the stock markets as in the following pairs of markets[12]:

In order to reduce the computational burden, we allow the triple markets, i.e. the recipient market (the stock market) and the two originator markets (WTI and Brent crude oil markets) to share the same volatility state. In this trivariate formulation, the number of states is six. For instance, for USA, we have the following six primitive states (as for each country case):

$s_t = 1$: DJIA real stock return—low volatility, WTI—low volatility, Brent—low volatility

$s_t = 2$: DJIA real stock return—high volatility, WTI—high volatility, Brent—high volatility

The conditional variance H is specified as a *BEKK* representation where the first element (h^2_{s,s_t}) of the diagonal matrix follows a *BEKK* MSG (1,1;2) process and the two other elements (h^2_{w,s_t} and h^2_{b,s_t}) follow a constant. Regime switching is allowed through the conditional mean intercepts and all the conditional variance parameters.

These choices allows us to refine our aim which consists essentially of finding out whether shocks and/or volatilities originating from crude oil markets are transmitted to stock markets under a jointly "high-high" volatility state or "low-low" volatility

[12] This idea was inspired by that of Edwards and Susmel (2001), who analyse the behaviour of the stock market volatilities for a group of Latin America countries using both univariate and bivariate switching models.

state. Edward and Susmel (2001) call the behaviour under this hypothesis "high volatility synchronisation" which signifies that when the "originator market" is in a high or low volatility state, the "recipient market" is always in the high or low volatility state. Furthermore, we are interested in determining whether these identified transmissions happen around the time of the conventional international crises.

Therefore, it is important to use the best possible model specification. Accordingly, assuming a *BEKK* structure, we consider two different models: (1) a standard Trivariate GARCH model with $p = q = 1$ which we denote MG(1,1) and, (2) our trivariate MSG (1,1;2).

In order to pick the most likely model, Table 1 summarises the critical values of Likelihood Ratio (LR) test, suggested by Garcia and Perron (1996). The log maximum likelihood values for the MMSG (1,1,2) models are higher than for the case where no regime switching is allowed. Notice that the former performs much better than the single regime model. Additionally, one can immediately see that the MMSG ranks better than the MS model according to the SIC, HQC and AIC criteria (not reported here).[13]

The results of estimating the multivariate Markov Switching GARCH model with BEKK parameterisation for each conditional mean and conditional volatility equation are reported in Table 2. Five triple-wise models are estimated and several interesting findings merit attention. It can be seen from the results that the three markets can be separated into two regimes. It is easy to interpret these two regimes. The first regime (labeled $s_t = 1$) indicates that all the real returns are at the same time in a "crash" state with low mean (a_S, a_W, a_B) and high variance (c_{11}, c_{22}, c_{33}). Conversely, regime 2 (labeled $s_t = 2$) captures the behaviour of the real returns in the recovery state with high mean and low variance. These states can differ substantially in durations.

We derived the transition probability matrix for the "originator" and "recipient" markets. It was assumed that the probability law that causes the market to switch among states is given by a K = 2 states Markov chain, P, with a typical element given by $\text{Prob}(s_t = j/s_{t-1} = i) = p_{ij}$. From the estimated transition probabilities P_{11} and P_{22}, we can calculate the duration of being in each regime.[14] In the case of USA, the average expected durations of being in regime 1 and 2 are roughly equal (6.5 months). The expected durations of being in regime 2 for the rest of country cases are about two times higher than those of being in regime 1. Thus, high variance states are less stable for UK, Germany, Japan and Canada. It is expected to persist for as long as the low volatility state in the case of the USA.

One of the study's key objectives is to find out whether the originator and the recipient market states, assumed to be in a joint high-high volatility states, occur around the identified international crises episodes. In other words, we verify

[13] Diagnostic tests for the MG model are available on request.

[14] the average duration of being in state 1 as suggested by Hamilton (1989) can be calculated as: $D_i = (1 - P_{ii})^{-1}$.

Table 1 The likelihood ratio test

	Ln_{MMSG}	Ln_{MS}	LR statistic
USA	−833.7	−898.9	130.4[a]
UK	−750.1	−782.7	65.2[a]
Germany	−842.9	−889.4	93[a]
Japan	−867.8	−886.7	37.8[a]
Canada	−785.7	−833.4	95.4[a]

Note the LR test statistic approximately follows a $\chi 2$ distribution with three degree of freedom. Ln_{MMSG} denotes the log maximum Likelihood value of the Trivariate Markov Switching GARCH-BEKK model and Ln_{MG} designates the log maximum likelihood value of the Multivariate GARCH-BEKK model. [a] denotes significance at the 1 percent level

whether the "volatility synchronisation" between the cycles of stock market and the crude oil market happens around the conventional economic recessions.

To verify this hypothesis graphically, we plot the smoothed probability for the two states $s_t = j$ ($j = 1,2$) in the right panels of Fig. 3. These figures display both the probability that crude oil market and stock markets are jointly in a high-volatility state or state 1 (black line) and the probability that the two markets are jointly in a low-volatility state or state 2 (grey line). The observations are classified following Hamilton's (1989) proposed method for dating regime switches. According to this procedure, an observation belongs to state i if the smoothed probability $\Pr(s_t = i|\psi_t)$ is higher than 0.5.

These figures show that regimes are seen to change frequently although the states are quite persistent. Table 3 compares the ECRI turning points for the five developed countries and the joint high—high volatility periods obtained from our regime switching models. In order to concentrate on the transmission of high volatility from the crude oil market to stock market, in the discussion that follows we focus mostly on the upper line of the bottom panel. As regards the dating results of the joint high—volatility regime, the model is able to delineate all the identified international crises. Additionally, Figures show that around each of the identified ECRI crises, crude oil and stock market jointly experience high volatility states. The common contraction periods differ in length and severity. The duration of the 5 or 6 contractions range from 6 to 27 months for USA, from 5 to 24 months for UK, from 2 to 33 months for Germany, from 3 to 44 months for Canada and from 2 to 23 months for Japan (see Table 3). The longest joint recession probability (a range of two or more successive recessions occurring close to each other) is associated with the 1996 East Asian crisis for USA and UK, the economic recession of 2000 for Canada and Japan and the 1990s Gulf war for Germany. Furthermore, it is obvious that the oil shock of 1990 induces the longest joint recovery period lasting about 3 years for Canada and Japan. In contrast, the oil shock of 2000 triggers the longest common recovery period for USA, UK and Germany.

The estimations of the econometric models are reported in Table 2. we first consider matrix Φ in the mean equation (Eq. 19), captured by the parameters μ_{ij} in Table 2, to see the link in terms of returns across the markets in each triple case.

Table 2 Estimates of the trivariate BEKK-MSG model

Param.	USA	UK	Japan	Germany	Canada
Mean equation					
$\mu_{ss,s_t=1}$	0.43723[c] (2.533)	0.0667 (0.157)	−0.05047 (−0.063)	0.41293 (1.094)	0.37441[a] (1.378)
$\mu_{ww,s_t=1}$	−0.86701[c] (−2.336)	−1.16371 (−0.69)	−0.57928 (−0.123)	−0.47747 (−0.656)	−0.23688 (−0.462)
$\mu_{bb,s_t=1}$	−0.81018[c] (−4.234)	−1.10085 (−0.598)	−0.64003 (−0.127)	−0.7421 (−0.971)	−0.38566 (−0.557)
$\mu_{ss,s_t=2}$	0.8655[c] (4.122)	0.53434[c] (2.786)	0.06613 (0.151)	0.69504[c] (4.57)	0.41688[a] (1.791)
$\mu_{ww,s_t=2}$	0.753669[c] (4.14)	0.71245[a] (1.558)	0.60213 (0.66)	0.64799[c] (2.697)	0.48967[a] (1.747)
$\mu_{bb,s_t=2}$	0.81946[c] (4.234)	0.76233[a] (1.56)	0.67792 (0.683)	0.71545[c] (2.86)	0.54541[a] (1.604)
Variance equation					
$\gamma_{ss,s_t=1}$	1.01169[c] (9.3243)	1.21839[c] (9,802)	1.08123[c] (2.761)	1.09181[c] (7.126)	0.72372[c] (5,453)
$\gamma_{ww,s_t=1}$	1.38159[c] (14.069)	1.58813[c] (3,002)	1.84192[c] (8.559)	1.70612[c] (7.316)	1.75198[c] (13.123)
$\gamma_{bb,s_t=1}$	1.41483[c] (12.920)	1.71815[c] (3,153)	1.89504[c] (11.273)	1.7876[c] (7.357)	1.83818[c] (12.053)
$\gamma_{ss,s_t=2}$	0.82333[c] (5.2642)	0.79844[c] (3,956)	0.64524[c] (3.164)	0.24226 (0.7963)	0.65601 (1.226)
$\gamma_{ww,s_t=2}$	0.96959[c] (17.407)	1.10553[c] (11,223)	1.14041[c] (3.760)	1.20845[c] (16.091)	1.18165[c] (12.373)
$\gamma_{bb,s_t=2}$	1.0292[c] (15.981)	1.13841[c] (10,396)	1.19702[c] (3.110)	1.27353[c] (15.761)	1.23767[c] (11.523)
$\alpha_{ss,s_t=1}$	−0.08768 (−1.287)	0.04765 (0.053)	0.35916 (1.032)	−0.16367[a] (−1.472)	−0.4012[c] (−3.828)
$\alpha_{sw,s_t=1}$	0.10004 (0.638)	−0.05232 (−0.804)	−1.1248 (−1.151)	0.56081[a] (1.776)	−0.50581[c] (−2.642)
$\alpha_{sb,s_t=1}$	−0.08387 (−0.603)	0.01302 (0.074)	1.05857 (1.113)	−0.57145[c] (−2.108)	0.32392[a] (1.894)
$\beta_{ss,s_t=1}$	−0.09402 (−0.093)	−0.39355 (−0.132)	0.0000 (0.000)	0.60485[c] (3.616)	0.67594[c] (5.661)
$\beta_{sw,s_t=1}$	−4.84532 (−1.104)	−1.15252 (−0.166)	−0.61242 (−0.001)	−0.03723 (−0.045)	1.12084 (0.07)
$\beta_{sb,s_t=1}$	5.96062 (1.198)	1.01243 (0.163)	−0.02393 (0.000)	0.21019 (0.298)	−0.81552 (−0.064)
$\alpha_{ss,s_t=2}$	0.8655[c] (4.122)	−0.67316[c] (−2.327)	0.05621 (0.647)	0.21307[c] (2.617)	0.23979[c] (3.5347)
$\alpha_{sw,s_t=2}$	0.89084[a] (1.795)	0.02,779 (0.048)	−0.34,526 (−1.185)	−0.11,484 (−1.296)	0.16,314 (1.116)
$\alpha_{sb,s_t=2}$	−0.27,778 (−0.739)	−0.06,521 (−0.303)	0.22,148 (0.981)	0.15,617[a] (1.81)	−0.07,771 (−0.58)

(continued)

Table 2 (continued)

Param.	USA	UK	Japan	Germany	Canada
$\beta_{ss,s_t=2}$	0.19,619[c] (3.125)	0.55,125[c] (2.855)	0.98,334[c] (64.439)	0.76,459[c] (10.22)	0.95,601[c] (74.63)
$\beta_{sw,s_t=2}$	4.99,591[a] (1.703)	−1.02,577 (−0.155)	3.86,618 (0.026)	0.68,817 (0.186)	−1.28,347 (−0.005)
$\beta_{sb,s_t=2}$	−7.89,867[c] (−2.389)	0.71,428 (0.14)	−2.97,633 (−0.032)	0.1982 (0.055)	0.84,935 (0.004)
Transition probabilities					
P_{11}	0.84,493	0.78,607	0.74,283	0.66,801	0.71,520
P_{22}	0.84,671	0.87,895	0.85,757	0.85,954	0.81,371
Residuals diagnostics					
Log-L	−833.757	−750.196	−867.844	−842.985	−785.767
SIC	−923.123	−839.488	−957.283	−932.351	−875.132
HQC	−889.507	−805.918	−923.62	−898.735	−841.516
AIC	−866.757	−783.196	−900.844	−875.985	−818.767

Notes The regime dependent covariance matrices H evolves according to a trivariate RS-GARCH (1,1) equation with a BEKK representation. The diagonal elements "μ" in matrix Φ represent the constant mean coefficients. While the diagonal elements "γ" in matrix Γ represent the constant variance coefficients. Elements "α" in matrix A captures own and cross-market ARCH effects. Elements "β" in matrix B measure own and cross-market GARCH effects. Subscribers: s, w, and b denote real stock market returns, WTI and Brent real crude oil returns. Student-*t* statistics of parameters are reported in parentheses. [a, b, c] denote statistical significance at 10, 5 and 1 %

The diagonal parameters $\mu_{11,st=2}$, $\mu_{22,st=2}$ and $\mu_{33,st=2}$ for all the modeled triples equations are positively significant (except for Japan) and approximately equal during expansion phases, suggesting that financial markets and crude oil markets tend to become more stable and predictable during an expansion regime. For instance, the average mean of the real DJIA return is 0.69 % while for the real crude oil returns are 0.64 and 0.71 % (respectively for the WTI and the Brent). In contrast, during high volatility states, these diagonal parameters are significant only for USA and Canada (for Canada, only one of the three parameters is significant; $\mu_{11,st=1}$). However, it is shown that while stock market returns appear to be positive, crude oil markets are characterised by negative returns during recession states. This can demonstrate that high volatility regime in crude oil markets are on average more severe, whereas American and Canadian stock markets seem to be more resistant to an economic slowdown. The Japanese case clearly distinguishes itself from the remaining countries. It shows no significant effects on the means of any of the parameters studied either during recessions or during expansions phases.

Results from the constant parameters of the variance equations show that all the intercept terms except $\gamma_{11,st=2}$ for Germany and Canada, are positively significant. However, the amplitude of these parameters is reduced slightly when volatilities switch simultaneously from state 1 to state 2. Interestingly, we observe again that

Table 3 Reference and estimated recession periods extracted from the trivariate MS-GARCH model

USA	1.1989M01–1991M02 (26 months)	1.1989M04–1989M09 (5 months)	a.1989M10–1989M12 (3 months)
	2.1994M05–1996M01 (21 months)	1990M01–1991M05 (17 months)	1991M06–1991M09 (4 months)
	3.1998M01–1999M09 (21 months)	1991M10–1991M12 (3 months)	1992M01–1993M08 (20 months)
	4.2000M04–2001M11 (20 months)	2.1993M09–1994M10 (14 months)	b.1994M11–1995M04 (6 months)
	2002M07–2003M02 (8 months)	1995M05–1995M06 (2 months)	1995M07–1996M01 (7 months)
	5.2004M03–2005M08 (18 months)	1996M02–1996M06 (5 months)	1996M07–1996M11 (3 months)
	6.2006M01–2007M12 (24 months)	3.1996M12–1997M07 (8 months)	c. 1997M08 (1 month)
		1997M09–1999M03 (19 months)	1999M04–1999M12 (9 months)
		4.2000M01–2000M05 (5 months)	d.2000M06–2000M09 (4 months)
		2000M10–2001M01 (4 months)	2001M02 (1 month)
		2001M03–2002M04 (14 months)	2002M05–2003M08 (16 months)
		2003M09 (1 month)	2003M10–2005M06 (21 months)
		5.2005M05–2005M10 (6 months)	e.2005M 11–2006M06 (8 months)
		6.2006M07–2007M02 (8 months)	f.2007M03–2007M10
		2007M11–2007M12 (2 months)	
UK	1.1989M01–1991M04 (28 months)	1.1990M03–1990M05 (3 months)	a.1990M06–1990M07 (2 months)
	2.1994M07–1995M08 (14 months)	1990M08–1992M03 (20 months)	1992M04–1993M11 (20 months)
	3.1997M07–1999M02 (20 months)	2.1993M12–1994M11 (12 months)	b. 1994M12–1997M02 (28 months)
	4.2000M01–2003M02 (38 months)	3.1997M03–1997M07 (5 months)	c. 1997M08–1997M11 (4 months)
	5.2004M03–2005M05 (15 months)	1997M12–1999M02 (15 months)	1999M03–1999M06 (4 months)
		1999M07–1999M10 (4 months)	1999M11–2000M03 (5 months)
		4.2000M04–2000M06 (3 months)	d.2000M07–2000M11 (5 months)

(continued)

Table 3 (continued)

		2000M12–2002M01 (14 months)	2002M02–2003M02 (13 months)
		2003M03–2003M04 (2 months)	2003M05–2005M09 (29 months)
		5.2005M10–2005M11 (2 months)	e.2005M12–2006M08 (9 months)
		2006M09–2006M10 (2 months)	2006M11–2007M03 (5 months)
		2007M04 (1 month)	2007M05–2007M12 (8 months)
Germany	1.1991M01–1993M01 (25 months)	1.1989M03–1989M07 (5 months)	a.1989M08–1990M03 (8 months)
	2.1994M12–1996M03 (16 months)	1990M04–1990M06 (3 months)	1990M07 (1 month)
	3.1998M03–1999M04 (14 months)	1990M08–1991M05 (10 months)	1991M06–1991M09 (4 months)
	4.2000M05–2002M03 (23 months)	1991M10–1992M12 15 months)	1993M01–1993M11 (11 months)
	2002M09–2003M08 (12 months)	2.1994M12–1995M02 (3 months)	b.1995M03–1995M11 (9 months)
	5.2004M04–2005M02 (11 months)	1995M12 (1 month)	1996M01–1996M03 (3 months)
	6.2006M 11–2007M12 (14 months)	1996M04 (1 month)	1996M05–1997M11 (19 months)
		1997M12–1998M03 (4 months)	1998M04–1998M06 (3 months)
		3.1998M07–1998M09 (3 months)	c.1989M10 (1 month)
		1998M11–1999M01 (3 months)	1998M02–1999M03 (15 months)
		4.2000M04–2000M07 (4 months)	d.2000M08–2000M11 (4 months)
		2000M12–2001M01 (2 months)	2001M02 (1 month)
		2001M03–2001M07 (5 months)	2001M08–2001M09 (2 months)
		2001M10–2001M12 (2 months)	2002M01 (1 month)
		2002M02–2002M03 (2 months)	2002M04–2003M01 (10 months)
		2003M02–2003M04 (3 months)	2003M05–2004M11 (19 months)
		5.2004M12–2005M01 (2 months)	e.2005M02–2005M09 (8 months)
		2005M10–2005M11 (2 months)	2005M12–2006M02 (3 months)
		6.2007M03–2007M04 (2 months)	f.2007M05–2007M12 (8 months)

(continued)

Table 3 (continued)

Canada	1.1989M01–1991M02 (26 months)	1.1989M04–1989M08 (5 months)	a.1989M09–1990M07 (11 months)
	2.1994M11–1996M06 (20 months)	1990M08–1991M06 (11 months)	1991M07–1991M12 (6 months)
	3.1997M07–1998M07 (13 months)	1992M01–1992M03 (3 months)	1992M04–1993M11 (20 months)
	4.2000M01–2001M09 (21 months)	1993M12–1994M02 (3 months)	1994M03–1994M04 (2 months)
	2002M06–2003M06 (13 months)	2.1994M05–1994M06 (2 months)	b.1994M07–1995M08 (14 months)
	5.2004M04–2005M03 (12 months)	1995M09 (1 month)	1995M10–1996M02 (5 months)
	6.2006M01–2007M12 (24 months)	1996M03–1996M04 (2 months)	1996M05–1996M07 (3 months)
		1996M08 (1 month)	1996M09–1997M11 (3 months)
		3.1997M12–1998M07 (8 months)	c. 1998M08/M10 (2 months)
		1998M09 (1 month)	1999M06 (1 month)
		1998M11–1999M05 (7 months)	1999M11–2000M02 (4 months)
Japan	1.1989M01–1989M05 (5 months)	1.1989M05–1989M06 (2 months)	a.1989M07–1990M01 (7 months)
	1990M03–1993M12 (46 months)	1990M02–1991M04 (15 months)	1991M05–1993M10 (30 months)
	2.1994M12–1996M01 (14 months)	2.1993M11–1994M10 (12 months)	b.1994M11–1996M04 (18 months)
	3.1997M03–1998M04 (14 months)	1996M05 (1 month)	1996M06–1997M01 (8 months)
	4.2000M08–2001M12 (17 months)	3.1997M02–1997M03 (2 months)	c.1997M04–1997M11 (8 months)
	5.2004M01–2004M11 (11 months)	1997M12–1998M04 (5 months)	1998M05–1998M12 (8 months)
	2005M04–2005M10 (7 months)	1999M01/09 (2 months)	1999M02–1999M08 (7 months)
	6.2006M04–2006M09 (6 months)	2000M02–2000M05 (4 months)	1999M10–2000M01 (4 months)
	2007M08–2007M12 (5 months)	4.2000M11–2001M06 (8 months)	2000M06–2000M10 (5 months)
		2001M09–2002M05 (9 months)	d.2001M07–2001M08 (2 months)
		2002M10 (1 month)	2002M06–2002M09 (4 months)
		2002M12–2003M04 (5 months)	2002M11 (1 month)

(continued)

Table 3 (continued)

		5.2004M11 (1 month)	2003M05–2004M10 (18 months)
		2005M09 (1 month)	e.2004M12–2005M08 (9 months)
		6.2006M08–2006M10 (3 months)	2005M10–2006M07 (10 months)
		2007M04–2007M06 (3 months)	f.2006M11–2007M03 (5 months)
		2007M08 (1 month)	2007M07 (1 month)
			2007M09–2007M12 (4 months)

Note: *Growth rate cycle peak and trough dates from 1989 to 2007 (source: Economic Cycle Research Institute (ECRI)). Figures in parentheses indicate the average length of the period in month

the volatility of crude oil returns is lengthened more than the volatility of the stock market returns in both states. Thus, high crude oil market volatilities have the potential to damage the conditions of economic growth much more and so these volatilities might be the primary cause of financial market turbulence.

To demonstrate the stock market's response to crude oil market movement, Table 2 shows the estimated interaction parameters between the degrees of turbulence or stability emanating from real crude oil volatility series to real stock market returns.

As a result, we find that almost two stock markets utilised in our analysis are affected by news (i.e. shocks) and volatility generated from their own markets, namely Dax30 and TSX during joint recession state. However, almost all the markets are affected by news (except for Japan) and volatility generated from their own markets during the joint expansion state.

Table 2 provides results from estimating the model using equity markets and WTI, Brent crude oil markets subscribed by the letters s, w and b respectively.

The results apparently indicate that FTSE 100 and NIKKEI 225 stock market returns do not receive significant shocks/volatility originating from crude oil markets during either joint high volatility state or joint low volatility state.

Therefore the biggest danger to financial stability does not seem to have come from high increases in crude oil market volatility.

As shown in the second panel of Fig. 3, excepting the abnormal increase (during early 2000 and 2005 for Japan and UK respectively),[15] UK and Japanese stock market volatilities remain static over all the period despite the presence of large spikes in the volatility of crude oil markets. Henceforth, UK and Japanese equity

15 Britain and Japanese stock market volatility saw an unprecedented rise of about 50 % (in 2005 and 2000 respectively) followed by rapid reversals. These meteoric rises may not be explained by any change in oil (or fundamentals), which barely changed during this period but may be indicative of explosive bubbles (e.g. the UK housing market bubble of 2004–2005).

market returns are not interrelated during the last 20 years in spite of the heavy dependency on oil.[16] This may indicate the important role that improvement in energy efficiency plays in reducing oil shock transmission to the volatility of the stock market. Indeed, according to the data of IEA (2009), UK and Japan have had the lowest primary energy intensities of any countries since the 1970s oil shock, indicating a higher efficiency than the other developed countries. Together, high volatility states in stock markets may be affected by diverse factors other than oil shocks such as interest rates or exchange rates (Apergis and Miller 2009).

The recessionary WTI (Brent) oil price shocks are positively (negatively) and significantly transmitted to the high volatility state of the German Dax 30 stock market. Then this transmission intensity switches to the joint recovery state and becomes negative (positive) and insignificant (significant) with 5 times lower amplitude. The finding for Canada can be interpreted in a similar way as for Germany with a difference in the amplitude and the sign of the coefficients $\alpha_{sw,st=1}$ and $\alpha_{sb,st=2}$ where the oil shock transmission switches from negative (positive) and significant during simultaneous high volatility state to positive (negative) and insignificant with a 3 times lower amplitude during simultaneous low volatility state. However, there is no evidence of volatility transmission running from the crude oil market to stock market.

This finding suggests that recessionary "external oil shocks"[17] (WTI) affect the German and Canadian (Brent) stock markets by increasing their volatilities. On the other hand, reaching the expansion regime, the underlying shocks negatively affect the stock market volatility and their transmission intensities become much less pronounced or even insignificant. In contrast, the opposite happens for "domestic oil shocks". Indeed, they stabilise the underlying stock markets by decreasing their volatilities during the joint recessionary state. This may highlight the decreased role that hedging policy efficiency plays in order to neutralise any potential oil price impact (particularly "external oil shocks") on the volatility of the stock market.

[16] Japan imports all of its oil. It is considered the third largest oil consumer in the world (behind US and China) and the second largest net importer of oil (behind US) in spite of its limited domestic oil reserves and production. UK is largest producer of oil and natural gas in the European Union but it cannot produce enough oil to meet its domestic demand (EIA 2008).

[17] Brent oil is, by definition, produced from Europe (UK), Africa and the Middle East (Brent North Sea crude). However, WTI oil is produced from North America (North America crude such as Canada). In what follows, we denote WTI oil shock as "External oil shock", i.e. extra-North sea oil shock, for European countries like Germany and as "Domestic oil shock" for American countries like Canada. In the same way, we denote Brent oil shock by "External oil shock", i.e. extra-American oil shock (North America as well as South America), for American countries and as "Domestic oil shock" for European countries.

In 2006, Germany is the fourth largest net-oil importing country (it imported 2.483 million barrels of crude oil per day to meet most of its oil needs). It was dependent on external oil sources even in peacetime. The top three sources of German crude oil imports were Russia (34 %), Norway (16 %) and UK (12 %) (Hsing 2007). Furthermore, Canada is both an exporter and importer of crude oil. From Stats Canada for 2005, domestic crude accounts for only about 45 % of Canada's oil consumption. Imports represent the remaining 55 %, mostly coming from North Sea Countries (UK and Norway) or the Middle East (Iraq, Saudi Arabia…etc.).

Decision makers are advised to drive domestic oil production and seek renewable energy technologies in order to reduce its reliance on foreign oil.

It should be emphasised, as shown in Fig. 3 (second panels) for Canadian and German cases, that these transmissions were concentrated during the 1999–2004 period of severe worldwide economic contractions (the bursting of the equity bubble of 1990, the US terrorist attack and the Enron scandals in 2001, the Argentine energy crisis, the Iraq disarmament crisis). They were opposite and weaker than those observed before and after these crises periods. Indeed, as clearly illustrated in these figures, the conditional variances of TSX and Dax30 varied dramatically over the 2000–2003 period which coincides with the sharp increases in oil volatility. Together, as previously demonstrated in Sect. 3.2, these respective low frequency components of crude oil volatility shock take longer period to stabilise. Moreover, especially in the case of Canada, real Brent is more volatile and therefore far more vulnerable to the real TSX than do the real WTI.

The US stock market response differs systematically from that of other oil-importing countries. Table 3 shows that the crude oil market does not transmit any signals (shock or volatility) to the DJIA stock market return during a common recession state. The significant coefficient on $\alpha_{12,st=1}$ shows that shocks of WTI arising during simultaneous low volatility states are transmitted positively and significantly to the DJIA stock market. There is also evidence of positive (negative) volatility transmission from WTI (Brent) oil market ($\beta_{12,st=2}$ and $\beta_{13,st=2}$) to the US stock market during those same periods. In addition, the DJIA stock market volatility is very sensitive to volatility coming from crude oil returns (4.9 and 7.8), underlying the major role that crude oil plays in this country as the largest oil importer.[18]

The positive transmission of the WTI's shock/volatility to the expansion phase of the USA stock market may underline the latter's greater vulnerability to shocks/volatilities from American sources of crude oil prices[19] than from the North Sea crude oil prices, but not to the point of leading to a stock market crash. In fact, with the declining production volumes of the Brent fields, more of the North Sea crude oil supply is being absorbed locally and less is available for sale to the USA. US dependence on the Brent crude fell sharply; this sudden change can be explained mainly by the rapid increase in oil demand by high growth countries particularly China and India,[20] the so-called "US Middle East

[18] According to US energy information Administration 2008, USA is the world's largest net importer of crude oil. It imported 10,984 thousand barrels per day, followed by Japan (4652) and China (3858).

[19] In 2000, North and South American countries particularly Canada (17.8 %), Mexico (14.2 %), Venezuela (14 %) supplied much more crude oil to the USA. However, Middle East countries (Saudi Arabia and Iraq) provide less than 23 % of USA oil imports, 25 % comes from African countries (Nigeria, Angola and Algeria), and less than 3 % from European countries (UK, Norway). (http://import-export.suite101.com/article.cfm/usa_oil_imports_by_country_2007).

[20] In 2008, Chinese crude oil imports, largely concentrated in the volatile Middle East, was roughly 4 times higher than in 1978 (Leung 2010).

oil independence”[21] (Kraemer 2006), as well as by the improvements made in energy efficiency by the US policy to reduce the inflationary effects of oil shocks. As a result, the decreasing US dependence on Brent crude may help make the stock market more resilient to the disruption of Brent supplies.

It can be concluded from these findings that the increased dependence on American crude oil supplies and the decreased dependence on North Sea crude oil supplies (the most unstable countries in the world) may be welcomed in the stock market.

The economic intuition for our main findings is most easily explained with reference to the second panel of Fig. 3. In this panel, the crude oil variances vary considerably over time and low spikes (state 2) are associated with very moderate investments in stocks (see the period 1999–2004). In contrast sharp spikes (state 1) are associated especially with small stock and reduced allocations (the two subsequent high volatility periods occurred in 1990 and the other one in 2007). Because regimes are persistent, short-horizon investors clearly attempt to time the market by reducing (increasing) the allocation to the riskiest assets when investment opportunities are poor (good) based on the information offered by the crude oil market volatility.

As there is no spillover effect between the stock market and crude oil market for USA during the joint high volatility state, there is limited potential for making riskless excess profit on the US stock market in much less time based on information from WTI, for example. Except for these periods, volatility in US equity markets remained generally low.

4 Summary and Concluding Remarks

In this paper, we use monthly stock market prices and two crude oil data (WTI and Brent) for a group of five developed countries (USA, UK, Germany, Japan and Canada) to quantify the magnitude and time-varying nature of volatility spillovers running from the crude oil market to the equity markets (DJIA, FTSE100, Dax30, NIKKEI225 and TSX).

With the objective of finding the most efficient way to model the behaviour of crude oil price volatilities, we use wavelet filtering, particularly Trous Haar wavelet decomposition method, as it has already proved it can provide a better insight into the dynamics of financial time series.

Moreover, most studies assume that the relationship between variables (especially asset returns) is generated by a linear process with stable coefficients so the predictive power of state variables does not vary over time. However, there is

[21] Indeed, many American politicians (President George W. Bush, among others) had worked toward US energy independence in order to reduce US imports of oil and other foreign sources of energy (see also “US energy independence” article from Wikipedia, http://en.wikipedia.org/wiki/United_States_energy_independence).

mounting empirical evidence that spillover parameters follow a more complicated process with multiple “regimes”, each of which is associated with a very different distribution of asset returns. The restricted trivariate *BEKK MSG* model used in our analysis is quite general and allows means, variances and parameters of shock/ volatility transmission to vary across states. Hence assuming that the two variables are in common states, the stock market return can vary across states in response to a shock or volatility originating from the crude oil market.

The results show that the *Â HTW* decomposition method appears to be an important step towards obtaining more accurate results. Indeed, we find that it seems to be very useful in detecting break-points, which implies that crude oil shock intensity varies significantly through time. Further, the resulting signals are smooth and give us a better approximation or reconstruction of the original signal. We also improve accuracy of this variable in detecting key real crude oil volatility features.

On the other hand, the trivariate *BEKK-MSG* estimations suggest that there are quite close connections between the joint equity and crude oil high volatility state and international recessions. Additionally, apart from UK and Japanese cases, the responses of the stock market to an oil shock depend on the geographic area for the main source of supply, be it from the North Sea or from North America (as we take two oil benchmarks, WTI and Brent respectively). Then, for Germany and Canada, external oil sources contribute more to causing a stock market crash even though these countries import less oil from abroad (Western America for Germany and Europe for Canada. However, oil shocks originating from Eurasian or European countries (North America) appear to be far less vulnerable.

The results for the US stock market volatility response to the crude oil shock and to volatility are different. Indeed, WTI crude oil volatility (American sources of oil) increases the DJIA stock market volatility, whereas the latters exhibit the inverse reaction to Brent crude oil. The US stock appears to be more resilient to crude oil shocks since even they exist they do not lead to a potential stock market crash.

However, Japanese and Britain equity markets do not show any reaction to shocks and/or volatilities coming from crude oil market.

Our results might be of interest to:

(1) investors; results show that the current crude oil market state is a persistent bear state with more attractive assets than in a bull crude oil market state.
(2) Monetary policy makers; the results obtained suggest that there are divergences between the hedging performance of WTI and Brent. For example, the presence of a positive transmission of the temporary WTI oil price shocks to the recessionary stock market phase highlights that the hedging policy in Germany is less efficient to neutralise the WTI oil price effect on the volatility of the German Dax30 stock market. Reaching the expansion phase, the opposite occurs but shocks take longer to stabilise. Here, monetary policy may play a more active role as a

(3) Energy policy makers; since German stock market may be more vulnerable to a WTI shock than a Brent shock (the inverse case for Canada), the government should import little to no oil from the main production countries of WTI crude but diversify sources and promote incentives for developing alternative energy sources (both in industrial and household sectors); this would reduce dependence on any one area outside the Brent crude main source countries. Conversely, the results for the US case can be attributed to the successful efforts of American policy makers to promote efficient energy since it depends mostly on WTI.

References

Agren M (2006) Does oil price uncertainty transmit to stock market ? department of economics. Working Paper 23, Uppsala University

Alizadeh AH, Nomikos NK, Pouliasis PK (2008) A Markov regime switching approach for hedging energy commodities. J Bank Finance 32:1970–1983

Aloui C, Jammazi R (2009) The effects of crude oil shocks on stock market shifts behaviour: a regime switching approach. Energy Econ 31:789–799

Aloui C, Jammazy R, Dhakhlaoui I (2008) Crude oil volatility and stock market returns. J Energy Markets 1:69–96

Apergis N, Miller SM (2009) Do structural oil-market shocks affect stock prices? Energy Econ 31 (4):569–575

Baba Y, Engle R, Kraft D, Kroner K (1987) Multivariate simultaneous generalized ARCH. University of California, San Diego (unpublished manuscript)

Bachmeier L (2008) Monetary policy and the transmission of oil shocks. J Macroecon 30:1738–1755

Bauwens L, Laurent S, Rombouts JVK (2003) Multivariate GARCH models: a survey. CORE Discussion Paper 2003/31

Beale L (2002) Volatility spillover effects in european equity markets: evidence from a regime switching model. Department of Financial Economics, Ghent University. http://fetew.rug.ac.be/fineco/personal/baele/papers/spillovers.htm

Blanchard OJ, Gali J (2009) The macroeconomic effects of oil shocks: why are the 2000s so different from the 1970s? In: Gali J, Gertler M (eds) International dimensions of monetary policy, University of Chicago Press, Chicago, pp 373–428

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econom 31:307–327

Bollerslev T, Engle R, Wooldridge M (1988) A capital asset pricing model with time-varying covariances. J Polit Econ 96(1):116–131

Bollerslev T (1990) Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. Rev Econ Stat 72:498–505

Cai J (1994) A markov model of unconditional variance in ARCH. J Bus Econ Stat 12:309–316

Chang PC, Fan CY (2008) Hybrid system integrating a wavelet and TSK fuzzy rules. IEEE Trans Syst Man Cybern—Part C: Appl Rev 38(6):802–815

Chen NF, Roll R, Ross SA (1986) Economic forces and the stock market. J Bus 59:383–403

Coifman RR, Donoho DL (1995) Translation-invariant de-noising. In: Antoniades A (ed) Wavelets and statistics, springer lecture notes. Springer, New York

Cologni A, Manera M (2008) Oil prices, inflation and interest rates in a structural cointegrated VAR model for the G-7 countries. Energy Econ 30:856–888

Cologni A, Manera M (2009) The asymmetric effects of oil shocks on output growth: a Markov-Switching analysis for G7 countries. Econ Model 26:1–29
Diebold FX (1986) Modeling the persistence on conditional variance: a comment. Econom Rev 5:51–56
Ding Z, Engle R (2001) Large scale conditional covariance modeling, estimation and testing. Acad Econ Pap 29:157–184
Edwards S, Susmel R (2001) Volatility dependence and contagion in emerging equity markets. J Dev Econ 66:505–532
Energy Information Administration (EIA, 2008) http://tonto.eia.doe.gov/country/index.cfm
Engle RF (2002) Dynamic conditional correlation—a simple class of multivariate GARCH models. J Bus Econ Stat 20:339–350
Engle R, Kroner FK (1995) Multivariate simultaneous generalized ARCH. Econom Rev 11:122–150
Engle R, Sheppard K (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. UCSD Discussion Paper 2001–2015.
Ferderer JP (1996) Oil Price Volatility and the Macroeconomy. J Macroecon 18:1–26
Garcia R, Perron P (1996) An analysis of the real interest rate under regime shifts. Rev Econ Stat 78:111–125
Gray SF (1995) An analysis of conditional regime-switching models. Working Paper, Duke University
Gray SF (1996) Modeling the conditional distribution of interest rates as a regime-switching process. J Financ Econ 42:27–62
Haas M, Mittnik S, Paolella MS (2004) Mixed normal conditional heteroskedasticity. J Financ Econom 2:211–250
Haas M, Mittnik S (2008) Multivariate regime switching GARCH with an application to international stock Markets. Working Paper Series 2008/08, Center for Financial studies
Hamilton JD (1983) Oil and the macroeconomy since World War II. J Polit Econ 91:228–248
Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57(2):357–384
Hamilton JD (1990) Analysis of time series subject to regime changes. J Econom 45:39–70
Hamilton JD (2005) Oil and the macroeconomy. Prepared for: Palgrave Dictionary of Economics
Hamilton JD (1996) This is what happened to the oil price-macroeconomy relationship. J Monetary Econom 38(2):215–220
Hamilton JD, Susmel R (1994) Autoregressive conditional heteroskedasticity and changes in regime. J Econom 64:307–333
Heil CE, Walnut DF (1989) Continuous and discrete wavelet transforms. Soc Ind Appl Math Rev 31:628–666
Hooker MA (1996) What happened to the oil price-macroeconomy relationship? J Monetary Econ 38:195–213
Hsing Y (2007) Impacts of higher crude oil prices and changing macroeconomic conditions on output growth in Germany. Int Res Jo Finance Econ 11:1450–2887
Huang RD, Masulis RW, Stoll HR (1996) Energy shocks and financial markets. J Futures Markets 16:1–27
International Energy Agency (IEA, 2009) Progress with implementing energy efficiency policies in the G8. www.iea.org/G8/docs/Efficiency_progress_g8july09.pdf
Jones C, Kaul G (1996) Oil and stock markets. J Finance 51:453–491
Kraemer DT (2006) Addicted to oil: strategic implications of american oil policy. Carlisle Papers in Security Strategy. Strategic Studies Institute United States Army War College
Kilian L (2008) Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. http://www-personal.umich.edu/~lkilian/aer061308final.pdf
Kilian L (2008b) Exogenous oil supply shocks: how big are they and how much do they matter for the U.S. economy? Rev Econ Stat 90:216–240
Kilian L (2008c) The economic effects of energy price shocks. J Econ Lit 46(4):871–909

Kilian L (2009) Oil price volatility: origins and effects. world trade organization, Economic Research and Statistics Division

Kilian L, Park C (2009) The impact of oil price shocks on the U.S. stock market. Int Econ Rev 50 (4):1267–1287

Lamoureux CG, Lastrapes WD (1990) Persistence in variance, structural change and the GARCH model. J Bus Econ Stat 8:225–234

Lee HT, Yoder JK (2007) A bivariate Markov regime switching GARCH approach to estimate the time varying minimum variance hedge ratio. Appl Econ 39(10):253–1265

Lee K, Shwan N, Ratti RA (1995) Oil shocks and the macroeconomy: the role of price variability. Energy J 16:39–56

Leung GCK (2010) China's oil use, 1990-2008. Energy policy 38(2):932–944

Li H, Majerowska E (2008) Testing stock market linkages for Poland and Hungary: a multivariate GARCH approach. Res Int Bus Finance 22:247–266

Malik F, Hammoudeh S (2007) Shocks and volatility transmission in the oil, US and Gulf equity markets. Int Rev Econ Finance 16:357–368

Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11:674–693

Misiti M, Misiti Y, Openheim G, Poggi JM (2008) Wavelet analysis: what can wavelet analysis do? the mathworks. http://www.d.umn.edu/~skatsev/Phys5053/WaveletHelp.pdf

Murtagh F, Starck JL, Renaud O (2004) On neuro-wavelet modeling. Decis Support Syst; Spec Issue Data Min Financ Dec Making 37:475–484

Raymond JE, Rich RW (1997) Oil and the macroeconomy: a markov state-switching approach. J Money Credit Bank 29:193–213

Regnier E (2007) Oil and energy price volatility. Energy Econ 29:405–427

Saito N, Beylkin G (1992) Multiresolution representations using the auto-correlation functions of compactly supported wavelets. IEEE Trans Signal Process 41(12):3584–3590

Saleem K (2009) International linkage of the Russian market and the Russian financial crisis: a multivariate GARCH analysis. Res Int Bus Finance 23:243–256

Schwert GW (1989) Why does stock market volatility change over time? J Finance 44(5):1115–1153

Shensa MJ (1992) The discrete wavelet transform: wedding the á trous and Mallat algorithms. IEEE Trans Signal Process 10:2463–2482

Strang G (1989) Wavelets and dilation equations: a brief introduction. Soc Ind Appl Math Rev 31:614–627

Tommi AV (2005) A wavelet analysis of scaling laws and long memory in stock market volatility. Reasearch Discussion Paper from Bank of Englend Number 27

Yalamova R (2006) Wavelet test of multifractality of Asia-Pacific index price series. Asian Acad Manage J Acc Finance 1:63–83

Wei C (2003) Energy, the stock market, and the putty-clay investment model. Am Econ Rev 93 (1):311–323

Wu T, Cavallo M (2009) Measuring oil-price shocks using market-based information. Working Papers 0905, Federal Reserve Bank of Dallas

Zhang B-L, Coggins R, Jabri MA, Dersch D, Flower B (2001) Multiresolution forecasting for futures trading using wavelet decompositions. IEEE Trans Neural Netw 12(4):765–775

Forcing Variables in the Dynamics of Risk Spillovers in Oil-Related CDS Sectors, Equity, Bond and Oil Markets and Volatility Market Risks

Shawkat Hammoudeh and Ramazan Sari

Abstract This study examines migration and cascading of credit default swaps (CDS) risks among four oil-related sectors -autos, chemical, oil and natural gas production, and utility—in two models. Model 1 encompasses fundamental variables, and Model 2 includes market risks. The key finding of the study suggests that replacing the two financial fundamental variables (the 10-year Treasury bond rate and the S&P 500 index) of Model 1 with the two market risk variables (the S&P VIX and the Oil VIX) of Model 2 reduce the long- and short-run risk migration and cascading in the second model for both the full sample and the subperiod. The CDS and VIX indices both reflect fear and risk on their own. Among the four oil-related CDS spreads, the chemical and auto spreads are the most responsive to the other credit and market risks and the fundamentals in the long-run, while those of utility and oil and natural gas sectors are not responsive. The recent quantitative easing in the United States adds to spikes in the levels of the chemical CDS and the S&P 500 index in Model 1, and to the S&P VIX and default risk spread in Model 2. Implications for model builders and policy makers are also discussed.

Keywords CDS · Oil-related sectors · S&P 500 index · S&P VIX · Oil VIX

Corresponding author. This paper was written while the second author (R. Sari) was a visiting professor at Middle East Technical University-Northern Cyprus Campus.

S. Hammoudeh (✉)
Department of Economics and International Business, Lebow College of Business,
Drexel University, 3200 Market Street, Philadelphia, PA 19104, USA
e-mail: hammousm@drexel.edu

R. Sari
Department of Business Administration, Middle East Technical University,
06800 Ankara, Turkey
e-mail: rsari@metu.edu.tr

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_5

1 Introduction

Oil is the most wildly traded commodity and one of the most volatile commodities in the world. It plays a pivotal role in the modern economy since its impacts dominate many economic sectors including the oil-related sectors: automobile, chemical, oil and natural gas, and utility. Given the high volatility of this commodity, companies that deal with oil, whether as an output, a fuel or a feedstock, have opted to protect themselves by buying counterparty risk protection contracts against volatility and default events. These companies in those oil-related sectors buy credit default swaps (CDS) to protect themselves from credit risks related to events that impact the oil markets and the overall economy.

The CDSs for these oil-related sectors are pertinent measures of expected credit risk and fear in these sectors, which is relevant information on movements of oil prices and changes in the business cycles. Each of these sector CDSs may also relate to or reflect fear in other oil-related sectors, the stock market, and the government and private bond markets. It will be interesting to discern risk migration and the lead/lag causal relationships between these sector CDS indices and changes in the oil, bond and stock markets. The oil credit risk which is represented by these oil-related CDSs may also have directional relationships with credit risks of the expected volatility in the stock and bond markets. The credit risks of the equity and bond markets are measured by the CBOE volatility equity index VIX and the credit risk spread which is the difference between the Baa bond rate and the 10-year Treasury bond rate.

Similar to the rest of CDS sector indices, the oil-related CDS indices for the automobile, chemical, oil and natural gas production, and utility sectors are highly liquid, standardized credit securities that trade at a very small bid-ask spread. The CDS indices can be efficient at processing information on evolving risks in various sectors of the economy (see Norden and Weber 2004; Greatrex 2008, among others). The magnitude of the oil and oil-related sector credit spreads gauges the default risk exposure of the firms in the oil-related sectors. A widening of a CDS spread in response to certain oil or credit events indicates an increase in the level of credit risk in the pertinent sector, while a narrowing spread shows a decrease in the credit risk.

Several studies examine CDS indices for specific major sectors of the U.S. economy but not for the oil-related sectors. Berndt et al. (2008) assess the variations in the risk premium that forms a major component of the CDS spread for the U.S. corporate debt at the firm level in three sectors: broadcasting and entertainment, health care, and oil and gas for the period 2000–2004. Raunig and Scheicher (2009) compare the market pricing of the default risk of banks and non-banks before and after the 2008 financial crisis, using monthly data. Using the decomposition of the CDS premia (or market prices) divided as the expected loss and the risk premium, their results demonstrate that the CDS traders had drastically changed their judgments on the riskiness of banks after the crisis by viewing these financial institutions as at least as risky as the other firms. Hammoudeh and Sari (2011) employ the

Autoregressive Distributed Lag (ARDL) approach to uncover the relationship between the financial CDS spread indices of the banking, financial services and insurance sectors and short- and long-term Treasury securities and the S&P 500 index. However, those authors do no account of other measures of financial stress and credit risks such as the default risk spreads and the expected volatility risk. More recently, Stanton and Wallace (2011) examine the relevance of the ABX.HE indices, which track CDSs on the US sub-prime residential mortgage-backed securities (RMBS), to the mortgage default rates during the financial crisis. Their results cast doubts on the suitability of the prices of the AAA ABX.HE index CDS as valuation benchmarks. Hammoudeh et al. (2013) examine the CDS spread indices for three financial-sectors, banking, financial services and insurance- in the short- and long-run and find the individual dynamic adjustments to the equilibrium to be different for those sectors.

To our knowledge, no published research has examined the CDS sector indices using the ARDL approach to figure out the relationship between the forcing variables in the four oil-related CDS sector indices and changes in the oil, equity and bond markets, equity VIX, oil VIX and default risk spread. The advantage of ARDL is that it allows one to define equations individually for all cointegrating vectors even if the variables have a mixed order of integration. Thus, the ARDL approach helps to define the forcing variables. The objective of this paper is to explore the lead/lag relationships between the risk and fundamental variables and examine risk migration between the market and credit risks in the oil-related sector CDS indices. We seek to fill the gap and complement the previous studies on sector CDS indices by focusing on the four oil-related sector CDS indices and their forcing variables, be they the fundamental variables or the market risk variables. These oil-related sectors are among the top S&P sectors.

The findings of our study underscore the relative importance of the CDS risks of the cyclical chemical and auto sectors over those of the utility and oil and natural gas production sectors. They also highlight the relative significance of the financial and oil fundamentals over the volatility market risk measures such as the equity VIX, oil VIX and default risk spreads.

This paper is organized as follows. Following this introduction, Sect. 2 presents the descriptive statistics of the series, and Sect. 3 gives a summary of the relevant literature. Section 4 discusses the methodology and the results, and Sect. 5 concludes.

2 Data and Descriptive Statistics

The data series include the closings for the CDS sector indices for the auto, chemical, oil and natural gas production and utility sectors, the 3-month West Texas Intermediate (WTI) crude oil futures price, the 10-year Treasury bond rate (DGS10), the default risk spread (DFR) and the measures of expected equity, and oil volatility indices, VIX and OVX respectively. The data on the four CDS series were obtained from DataStream, and on WTI were sourced from the database of Energy Information Administration (EIA). Moreover, the data on the default risk spread and the 10-year T bond rate were accessed from the database of the Federal Reserve Bank of Saint Louis, and on the S&P 500 VIX and the oil OVX were obtained from CBOE's Market Statistics Summary Data. All variables are expressed in the logarithmic form. The four oil–related CDS indices, particularly the auto and chemical indices, register a jump around January 2009, displaying greater default fear near the end of 2008 and beginning of 2009 (Fig. 1). This default fear is also reflected in the S&P VIX and the oil VIX around that time. Correspondingly, there is a dip in the S&P 500 index and the WTI.

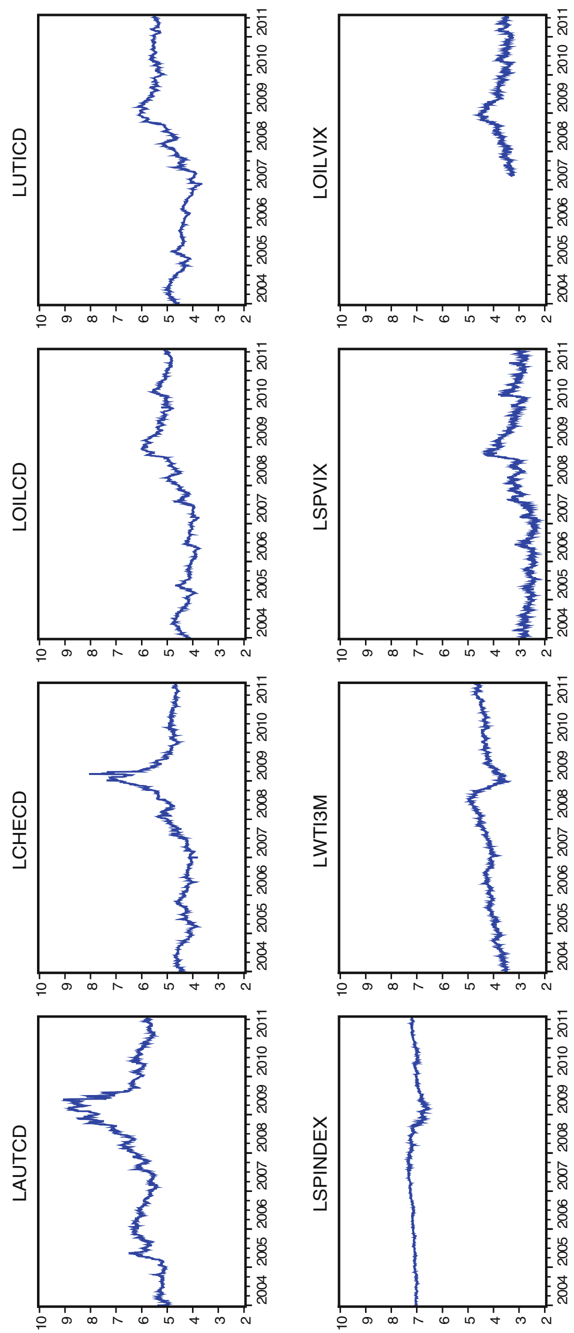
The expected market volatility indices are for the equity market (S&P VIX) and the oil market (oil VIX).[1] It's worth noting that the series for the oil VIX started on May 10, 2007. We examine these series for the full period January 2, 2004 to July 13, 2011. Thus, the series do not include the oil VIX over the whole sample. However, the subperiod June 1, 2009 to July 13, 2011 spanning the recovery period after the 2008/2009 Great Recession includes the oil VIX. The dummy variables *QE1* and *QE2* represent quantitative easing for the second half of 2009 and first half of 2010, respectively.

CDS sector indices, which are based on the most liquid 5-year term, are equally weighted and reflect an average mid-spread calculation of the given index's constituents. These proprietary indices are rebalanced every six months to better reflect liquidity in the CDS market. The identification of the CDS sector indices follows the DJ/FTSE Industry Classification Benchmark (ICB) supersectors as their basis and reflects the price performance of a basket of corporate 5-year CDSs within a given sector. As stated, the data for the CDS sector indices are available from 2004 only. The years 2004–2007 of the full sample for the CDS market were rapid growth years. However, the years 2008–2009 were troubling years for this market, which experienced fiscal stimulus packages and two monetary quantitative easings. The economic recovery years of 2009–2011 make up our subperiod.

As indicated above, the equity VIX is an index which measures *expectations* of volatility of the S&P 500 index and typically moves in an adverse direction to the latter. That is, an increase in equity VIX is associated with a decrease in the S&P 500 index to reflect fears in the equity market. VIX assembles risk information on events related to more than the stock market. In fact, the equity VIX increased by more than 30 % in the week following the major earthquake with a magnitude of 9.0 in Eastern Japan on March 11, 2011. This index has sentiment extremes where the range (30–32) signals excessive bearishness in the stock market that foreshadows bullish reversals, while (16–18) signals excessive bullishness that presages bearish reversals.

[1] Each volatility index series has a given number of reference entities at a fixed coupon. The coupon is determined prior to the onset of each index series, and is the current spread of the underlying reference entities that equate the value of the index to par value (100 %) at the time of calculation. The levels of the indices are calculated at the end of each business day at around 5:15 pm.

Fig. 1 Historical evolution of CDS, VIXs and oil price. *Notes* All the oil–related CDS indices, particularly the auto and chemical indices, have a spike around January 2009. This is also reflected by the S&P VIX and oil VIX. All variables are in a logarithmic form



Oil VIX, ticker OVX, is the CBOE crude oil ETF volatility index which measures the market's expectations of 30-day volatility of crude oil prices by applying the VIX methodology to the United States Oil Fund (USO) options spanning a wide range of strike prices. Its range since its inception on May 10, 2007 is 25.42–100.42. Unlike the S&P VIX, which typically rises when there is panic in the U.S. stock market and equity prices fall, the OVX goes up as oil prices, which incorporate a fear component, also increase. When there is uncertainty in the crude oil market, both oil prices and the OVX are more likely to rise in tandem because the tail risk is to the upside rather than the downside. Thus, the oil VIX is positively correlated with oil prices because higher risk levels will increase oil prices rather than discount them. There are those who believe that oil VIX can predict oil prices (Jagerson 2008).

Finally, the default risk spread is the difference between the corporate BAA bond rate and the 10-year Treasury bond rate, which measures the rises and falls in corporate credit risk in anticipation of recessions and booms, respectively.[2] A rise in the default risk spread presages a decline in economic activity and vice versa.

The descriptive statistics for the first logarithmic (L) differences of the series are presented in Table 1. The average percentage change over the sample period is the highest for the WTI futures price, followed by the CDS indices for the utility and oil/natural gas production sectors. It is negative for the 10-year Treasury bond rate (DGS10). In sum, the average return for the futures oil price is much higher than for any of the CDS sector indices.

The highest percentage change volatility as measured by the standard deviation is for the S&P VIX, followed by the 10-year Treasury bond rate (DGS10) over the full period. The lowest volatility is for the utility CDS index, followed by the WTI oil futures price. It is interesting to note that the volatilities for the S&P VIX, the 10-year Treasury rate, and the CDS indices of auto and chemicals in Table 1 are similar on the relatively high side, while those for CDS indices for utility, oil and gas production and the oil futures price are close on the low side. The volatility for oil VIX and default risk comes in the middle of the spectrum. In sum, the volatilities of the four oil-related CDS indices are dissimilar.

The series of the auto and chemicals CDS indices and the 10-year Treasury rates are skewed to the left over the full period, suggesting that the mass of the distributions for these three series is concentrated on the right of the figure, and have a few extremely low values in the distributions. This means the spreads for the returns of these series are bunched up on the high end of the spread scale. In comparison, the utility, oil and gas CDS indices, S&P VIX, oil VIX, and DFR are skewed to the right.

The kurtosis results indicate that the distributions of the series are more leptokurtic (peaked with fat tails) for the returns or first log differences over the full period. The Jarque-Bera statistics reject the null hypothesis of a normal distribution for all the series during the full period. This result is consistent with the statistics for skewness and kurtosis for most speculative assets.

[2] The BAA rate has more default risk than AAA which is almost close to zero default.

Table 1 Descriptive statistics-full period

Difference (Δ)	ΔLAUTCD	ΔLCHECD	ΔLDFR	ΔLDGS10	ΔLOILCD	ΔLOILVIX	ΔLSPVIX	ΔLUTICD	ΔLWTI3 M
Mean	0.000214	9.63E-05	0.000214	−0.00069	0.000456	0.000198	3.28E-05	0.000474	0.000559
Median	−0.00034	−0.0004	0	0	−0.00069	−0.00211	−0.00406	−0.00076	0
Maximum	0.46414	0.866491	0.38	0.24	0.370979	0.288702	0.496008	0.16291	0.164137
Minimum	−1.41405	−1.26928	−0.14	−0.51	−0.34725	−0.33336	−0.35059	−0.15323	−0.12827
Std. Dev.	0.056322	0.054146	0.033364	0.062467	0.025779	0.046921	0.064489	0.022824	0.025226
Skewness	−8.0411	−4.60205	2.210684	−0.26261	0.680903	0.589905	0.655089	0.579923	0.019871
Kurtosis	215.3242	207.807	25.29372	6.533463	43.95346	10.56904	7.541675	9.463333	7.688131
Jarque-Bera	3710340	3436003	42271.69	1044.293	137401.5	2662.709	1828.433	3528.647	1798.709
Probability	0	0	0	0	0	0	0	0	0
Observations	1964	1962	1964	1964	1964	1089	1964	1964	1964

Notes Full period is January 2, 2004-July 13, 2011. The variables are the first logarithmic differences. Δ*LAUTCD*, Δ*LCHECD*, Δ*LOILCD* and Δ*LUTICD* are the CDS sector spreads for the auto, chemical, oil and natural gas production sectors, respectively. Δ*LOILVIX* and Δ*LSPVIX* are the expected option volatility index for oil and S&P 500 index, respectively. Δ*LDFR* is the default risk spread, Δ*LWTI3M* is the three-month WTI futures price, and Δ*LDGS10* is the 10-year Treasury bond rate

Table 2 Unit root tests for Models 1 and 2 (full period)

	ADF	PP	Model 1	Model 2
LAUTCD	I(1)	I(1)	√	√
LCHECD	I(1)	I(1)	√	√
LOILCD	I(1)	I(1)	√	√
LUTICD	I(1)	I(1)	√	√
LOILVIX	N/A	N/A		
LSPVIX	I(1)	I(0)		√
LDGS10	I(1)	I(1)	√	
LDFR	I(1)	I(1)		√
LWTI3M	I(1)	I(1)	√	√
LSPINDEX	I(1)	I(1)	√	

Notes ADF stands for augmented Dickey-Full test while PP represents the Phillips-Parron test. Number 1 refers to I(1), while 0 indicates I(0). The symbol √ refers to the variables included in each model. The unit root test for the subperiod for the models indicates more I(0) than I(1) and they are available upon request. N/A means that the variable is not applicable to this model

The ADF and Phillips-Perron (PP) unit root tests for the intercept and the intercept plus trend were calculated for all variables over the full period and the subperiod. The unit root results are mixed for the subperiod, indicating that some variables are I(1) while others are I(0). A summary of these results is reported in Table 2 for the full period for both models. It can be contended that the VIX's are implied option volatility indices, and thus are proxies for option prices. This explains why VIX has a unit root behavior. The same logic applies for the CDS indices. Due to limited space, the results for the subperiod are not presented but are available on request.

These results warrant the use of the ARDL approach. We also run the Johansen-Juselius cointegration technique for the model with the same order of integration.[3]

3 Review of the Literature

Data series on the CDS sector index spreads start in 2004. Therefore, the literature on these credit risks is still quite sparse, particularly studies that examine the 2008 financial and 2010 sovereign debt crises. As indicated above, the available studies deal with financial sectors' CDS indices (see Stanton and Wallace 2011; Hammoudeh et al. 2013, among others). Clearly, there is a substantial scope for contributions in this area, particularly on oil-related sector CDSs.

[3] We will not report the results for the Johansen-Juselius approach due to the lack of space. Those results are available from the authors.

The recent CDS literature examines the difference between the spread in the cash/asset market and the CDS credit market, known as the "basis" (Das and Hanouna 2006). Longstaff et al. (2005) examine the basis using an approach that extracts the corporate bond-implied CDS spreads. When comparing it with the actual market CDS spreads, they find the corporate bond-implied CDS spreads to be higher.

Berndt et al. (2008) investigate the variations in the credit risk premium that comprises a major component of the CDS spread for three sectors, namely broadcasting and entertainment, health care, and oil and gas. They find striking differences in the spread variations between these sectors. Zhang et al. (2009) use an approach that identifies the volatility and jump risks of individual firms from high frequency stock prices to explain the CDS premium. Their empirical results suggest that the volatility risk alone predict 48 % of the variations in CDS spread levels, whereas the jump risk alone forecasts 19 %. After controlling for credit ratings, macroeconomic conditions, and firm balance sheet information, they predict 73 % of the total variations. Simulation results suggest that the high frequency-based volatility measures can help explain the credit spreads above and beyond what is already captured by the true leverage ratio.

Other studies have examined the relationships between the equity and credit markets using time series instead of cross section data, as in the cases discussed above. Bystrom (2006) examines the properties of the Dow Jones iTraxx index, which is an index of CDS securities on 12 European reference entities. He finds that the CDS spreads are significantly autocorrelated in the seven sectors comprising the iTraxx index, and are also significantly negatively related to the contemporaneous stock returns in all sectors, except for energy, consumers, and financials.

Fung et al. (2008) study the relationship between the stock market and high yield and investment grades and the CDS markets in the United States and find that the lead/lag relationship between them depends on the credit quality of the underlying reference entity. Forte and Lovreta (2008) examine the relationship between company-level CDS and stock market-implied credit spreads (ICS) in recent years. They find the relationship to be stronger, and the probability that the stock market leads in the price discovery to be higher at lower credit quality levels. However, the probability of CDS spreads leading in the price discovery rises with increases in the frequency of the severity of credit downturns.

Zhu (2006) discovers a long-run (cointegrating) relationship between credit risk in the corporate bond market and the CDS market, although a substantial deviation from the theoretical parity relationship can arise in the short-run. The VECM analysis suggests that the deviation is largely due to the higher responsiveness of CDS premia to changes in the credit conditions. Norden and Weber (2009) examine the relationships between CDS, bond and stock markets during the 2000–2002 period. They investigate monthly, weekly and daily lead-lag relationships using VAR/VEC models, and find that stock returns lead the CDS and bond spread changes. They also find that the CDS spread changes Granger-cause the bond spread changes for a higher number of firms than the reverse. They contend that the CDS market is more sensitive to the stock market than the bond market, and that this sensitivity increases for the lower credit quality. Finally, they find that the CDS market contributes more to price discovery than the bond market, with this result being stronger for the U.S. than for European firms.

On the informational content of VIX, Luo and Zhang (2010) extended this volatility index to other maturities and constructed daily VIX term structure data, proposing a simple two-factor stochastic volatility framework for VIX. Their results indicate that the framework captures both the time series dynamics of VIX and the rich cross-sectional shape of the term structure. Consistent with previous studies, it has been found that VIX contains more information than historical volatility.

Becker et al. (2009) examine two issues pertinent to the informational content of the VIX implied volatility index. One relates to whether it subsumes information on how historical jump activity contributed to the price volatility, and the other one relates to whether VIX reflects any incremental information pertaining to future jump activity relative to model-based forecasts. It is found that VIX both subsumes information linked to past jump contributions to total volatility and reflects incremental information relevant to future jumps.

In a related study, Figuerola-Ferretti and Paraskevopoulos (2010) consider the cointegration and the lead in the price discovery process between credit risk, as represented by CDS spreads, and market risk embedded in the equity VIX. They find that CDS and VIX are cointergated and that VIX has a clear lead over the CDS market in the price discovery process, implying that CDS adjusts to market risk when there is temporary mispricing from the long-run equilibrium. They find that there are long-term arbitrage relationships between VIX and CDS for most companies, implying that excess returns may be earned using "pairs trading" strategies.[4]

Fernandes et al. (2009) examine the time series properties of daily equity VIX. Their results suggest that VIX display long-range dependence. They confirm the evidence in the literature that there is a strong negative relationship between VIX and S&P500 index returns, as well as a positive contemporaneous link with the volume of the S&P500 index. Moreover, they demonstrate that equity VIX tends to decline as the long-run oil price increases, reflecting the high demand from oil in recent years, as well as the recent trend of shorting energy prices in the hedge fund industry.

Gogineni (2010) examines the impact of changes in daily oil price on the equity returns of a wide array of industries. He finds that stock returns both of industries that depend heavily on oil and those that use little oil are sensitive to changes in oil price. The latter industries are impacted because their main customers are affected by oil prices. The results also demonstrate that the sensitivity of industries' returns to the oil price changes depends on the cost-side as well as the demand-side dependence on oil.

In this study, we will examine the counterparty credit risks embedded in the oil-related sectors, and their relations to market risks including the expected option

[4] The pairs trade or pair trading is a market neutral trading strategy which enables traders to profit from virtually any market conditions: uptrend, downtrend, or sideways movement. One pairs trade would be to short the outperforming asset and to long the underperforming one, betting that the "spread" between the two assets would eventually converge.

volatility in the stock (equity VIX), and oil (oil VIX) markets, when fundamental variables such as the S&P 500 index, WTI oil futures price and the 10-year Treasury bond rate are accounted for. This analysis will provide room to examine the migration of risks in the different sectors and markets. The near bankruptcy of GM attests to the importance of such a risk-related examination.

We will also study the dynamic dependent-forcing relationships between these markets in the recovery period that followed the 2008/2009 financial crisis. Thus, our approach contrasts with the previous literature, which focused on the firm level, by examining forcing-dependent variable relationships at the sector level. We will use the ARDL approach which has flexibility to the degree of integration of these highly mixed and diversified variables.

4 The ARDL Procedure and Results

Technically, the ARDL approach is a multiple step procedure (Pesaran and Pesaran 2009). In the first step, the bounds testing procedure is utilized to test the presence of cointegration among the variables to identify the long-run relationship(s) between a dependent variable and its forcing variables (independent variables). In the second step, the ARDL models are constructed based on the results obtained in the first step. The short-run dynamics are estimated in the third step. To use the bounds test procedure, we estimate the following regressions for the first model (Model 1) which consists of the four oil-related sectors CDS spreads, the S&P 500 index, the oil futures price, and the 10-year Treasury bond rate, as well as two dummies *QE1* and *QE2* representing quantitative easing over the full period.

$$\begin{aligned}\Delta \ln LAUTCD_t = a_{0A} &+ \sum_{i=1}^{n} b_{iA}\Delta \ln LAUTCD_{t-i} + \sum_{i=1}^{n} c_{iA}\Delta \ln LCHECD_{t-i} + \sum_{i=1}^{n} d_{iA}\Delta \ln LOILCD_{t-i} \\ &+ \sum_{i=1}^{n} e_{iA}\Delta \ln LUTICD_{t-i} + \sum_{i=1}^{n} f_{iA}\Delta \ln LDGS10_{t-i} + \sum_{i=1}^{n} g_{iA}\Delta \ln LWTI3M_{t-i} \\ &+ \sum_{i=1}^{n} h_{iA}\Delta \ln SPINDEX_{t-i} + \lambda_{1A}\ln LAUTCD_{t-1} + \lambda_{2A}\ln LCHECD_{t-1} \\ &+ \lambda_{3A}\ln LOILCD_{t-1} + \lambda_{4A}\ln LUTICD_{t-1} + \lambda_{5A}\ln LDGS10_{t-1} \\ &+ \lambda_{6A}\ln LWTI3M_{t-1} + \lambda_{7A}\ln SPINDEX_{t-1} + \gamma_{1A}QE1_t + \gamma_{2A}QE2_t + \varepsilon_{tA}\end{aligned} \tag{1}$$

$$\begin{aligned}\Delta \ln LCHECD_t = a_{0C} &+ \sum_{i=1}^{n} b_{iC}\Delta \ln LAUTCD_{t-i} + \sum_{i=1}^{n} c_{iC}\Delta \ln LCHECD_{t-i} + \sum_{i=1}^{n} d_{iC}\Delta \ln LOILCD_{t-i} \\ &+ \sum_{i=1}^{n} e_{iC}\Delta \ln LUTICD_{t-i} + \sum_{i=1}^{n} f_{iC}\Delta \ln LDGS10_{t-i} + \sum_{i=1}^{n} g_{iC}\Delta \ln LWTI3M_{t-i} \\ &+ \sum_{i=1}^{n} h_{iC}\Delta \ln SPINDEX_{t-i} + \lambda_{1C}\ln LAUTCD_{t-1} + \lambda_{2C}\ln LCHECD_{t-1} \\ &+ \lambda_{3C}\ln LOILCD_{t-1} + \lambda_{4C}\ln LUTICD_{t-1} + \lambda_{5C}\ln LDGS10_{t-1} \\ &+ \lambda_{6C}\ln LWTI3M_{t-1} + \lambda_{7C}\ln SPINDEX_{t-1} + \gamma_{1C}QE1_t + \gamma_{2C}QE2_t + \varepsilon_{tC}\end{aligned} \tag{2}$$

$$\begin{aligned}\Delta \ln LOILCD_t = a_{0O} &+ \sum_{i=1}^{n} b_{iO}\Delta \ln LAUTCD_{t-i} + \sum_{i=1}^{n} c_{iO}\Delta \ln LCHECD_{t-i} + \sum_{i=1}^{n} d_{iO}\Delta \ln LOILCD_{t-i}\\ &+ \sum_{i=1}^{n} e_{iO}\Delta \ln LUTICD_{t-i} + \sum_{i=1}^{n} f_{iO}\Delta \ln LDGS10_{t-i} + \sum_{i=1}^{n} g_{iO}\Delta \ln LWTI3M_{t-i}\\ &+ \sum_{i=1}^{n} h_{iO}\Delta \ln SPINDEX_{t-i} + \lambda_{1O} \ln LAUTCD_{t-1} + \lambda_{2O} \ln LCHECD_{t-1}\\ &+ \lambda_{3O} \ln LOILCD_{t-1} + \lambda_{4O} \ln LUTICD_{t-1} + \lambda_{5O} \ln LDGS10_{t-1}\\ &+ \lambda_{6O} \ln LWTI3M_{t-1} + \lambda_{7O} \ln SPINDEX_{t-1} + \gamma_{1O} QE1_t + \gamma_{2O} QE2_t + \varepsilon_{tO}\end{aligned} \tag{3}$$

$$\begin{aligned}\Delta \ln LUTICD_t = a_{0U} &+ \sum_{i=1}^{n} b_{iU}\Delta \ln LAUTCD_{t-i} + \sum_{i=1}^{n} c_{iU}\Delta \ln LCHECD_{t-i} + \sum_{i=1}^{n} d_{iU}\Delta \ln LOILCD_{t-i}\\ &+ \sum_{i=1}^{n} e_{iU}\Delta \ln LUTICD_{t-i} + \sum_{i=1}^{n} f_{iU}\Delta \ln LDGS10_{t-i} + \sum_{i=1}^{n} g_{iU}\Delta \ln LWTI3M_{t-i}\\ &+ \sum_{i=1}^{n} h_{iU}\Delta \ln SPINDEX_{t-i} + \lambda_{1U} \ln LAUTCD_{t-1} + \lambda_{2U} \ln LCHECD_{t-1}\\ &+ \lambda_{3U} \ln LOILCD_{t-1} + \lambda_{4U} \ln LUTICD_{t-1} + \lambda_{5U} \ln LDGS10_{t-1}\\ &+ \lambda_{6U} \ln LWTI3M_{t-1} + \lambda_{7U} \ln SPINDEX_{t-1} + \gamma_{1U} QE1_t + \gamma_{2U} QE2_t + \varepsilon_{tU}\end{aligned} \tag{4}$$

$$\begin{aligned}\Delta \ln LDGS10_t = a_{0D} &+ \sum_{i=1}^{n} b_{iD}\Delta \ln LAUTCD_{t-i} + \sum_{i=1}^{n} c_{iD}\Delta \ln LCHECD_{t-i} + \sum_{i=1}^{n} d_{iD}\Delta \ln LOILCD_{t-i}\\ &+ \sum_{i=1}^{n} e_{iD}\Delta \ln LUTICD_{t-i} + \sum_{i=1}^{n} f_{iD}\Delta \ln LDGS10_{t-i} + \sum_{i=1}^{n} g_{iD}\Delta \ln LWTI3M_{t-i}\\ &+ \sum_{i=1}^{n} h_{iD}\Delta \ln SPINDEX_{t-i} + \lambda_{1D} \ln LAUTCD_{t-1} + \lambda_{2D} \ln LCHECD_{t-1}\\ &+ \lambda_{3D} \ln LOILCD_{t-1} + \lambda_{4D} \ln LUTICD_{t-1} + \lambda_{5D} \ln LDGS10_{t-1}\\ &+ \lambda_{6D} \ln LWTI3M_{t-1} + \lambda_{7D} \ln SPINDEX_{t-1} + \gamma_{1D} QE1_t + \gamma_{2D} QE2_t + \varepsilon_{tD}\end{aligned} \tag{5}$$

$$\begin{aligned}\Delta \ln LWTI3M_t = a_{0W} &+ \sum_{i=1}^{n} b_{iW}\Delta \ln LAUTCD_{t-i} + \sum_{i=1}^{n} c_{iW}\Delta \ln LCHECD_{t-i} + \sum_{i=1}^{n} d_{iW}\Delta \ln LOILCD_{t-i}\\ &+ \sum_{i=1}^{n} e_{iW}\Delta \ln LUTICD_{t-i} + \sum_{i=1}^{n} f_{iW}\Delta \ln LDGS10_{t-i} + \sum_{i=1}^{n} g_{iW}\Delta \ln LWTI3M_{t-i}\\ &+ \sum_{i=1}^{n} h_{iW}\Delta \ln SPINDEX_{t-i} + \lambda_{1W} \ln LAUTCD_{t-1} + \lambda_{2W} \ln LCHECD_{t-1}\\ &+ \lambda_{3W} \ln LOILCD_{t-1} + \lambda_{4W} \ln LUTICD_{t-1} + \lambda_{5W} \ln LDGS10_{t-1}\\ &+ \lambda_{6W} \ln LWTI3M_{t-1} + \lambda_{7W} \ln SPINDEX_{t-1} + \gamma_{1W} QE1_t + \gamma_{2W} QE2_t + \varepsilon_{tW}\end{aligned} \tag{6}$$

$$\begin{aligned}\Delta \ln SPINDEX_t = a_{0S} &+ \sum_{i=1}^{n} b_{iS}\Delta \ln LAUTCD_{t-i} + \sum_{i=1}^{n} c_{iS}\Delta \ln LCHECD_{t-i} + \sum_{i=1}^{n} d_{iS}\Delta \ln LOILCD_{t-i}\\ &+ \sum_{i=1}^{n} e_{iS}\Delta \ln LUTICD_{t-i} + \sum_{i=1}^{n} f_{iS}\Delta \ln LDGS10_{t-i} + \sum_{i=1}^{n} g_{iS}\Delta \ln LWTI3M_{t-i}\\ &+ \sum_{i=1}^{n} h_{iS}\Delta \ln SPINDEX_{t-i} + \lambda_{1S} \ln LAUTCD_{t-1} + \lambda_{2S} \ln LCHECD_{t-1}\\ &+ \lambda_{3S} \ln LOILCD_{t-1} + \lambda_{4S} \ln LUTICD_{t-1} + \lambda_{5S} \ln LDGS10_{t-1}\\ &+ \lambda_{6S} \ln LWTI3M_{t-1} + \lambda_{7S} \ln SPINDEX_{t-1} + \gamma_{1S} QE1_t + \gamma_{2S} QE2_t + \varepsilon_{tS}.\end{aligned} \tag{7}$$

The coefficients *b, c, d, e, f,* g and *h* are the short-run coefficients for the respective variables, while the λs are the long-run coefficients of the ARDL model. The null hypothesis of no cointegration is that $\lambda_{1j} = \lambda_{2j} = \lambda_{3j} = \lambda_{4j} = \lambda_{5j} = \lambda_{6j} = \lambda_{7j} = 0$, where j represents one of the seven variables. We will construct the second model (Model 2) by replacing LDGS10 and SPINDEX by LSPVIX and LDFR over the full sample and replacing LSPVIX with the oil VIX over the subsample.

To determine the appropriate lag length for the bounds testing procedure, we utilize various criteria with two dummy variables. The Final Prediction Error (FPE), Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), Hannan-Quinn Information Criterion (H-Q), and the Sequential Modified LR Test Statistic are the common criteria used to determine the lag lengths. These criteria yielded mixed results. We used the mostly suggested lags for the models. In the case of conflict, we utilized the lag length suggested by AIC due to the preservation of the degrees of freedom.

The results of the bounds testing procedure are estimated for the two models in the full period January 2, 2004 to July 13, 2011 and the 2009 recovery subperiod June 1, 2009 to July 13, 2011. Model 1, which is comprised of seven variables including the four oil-related credit risks and three fundamental variables, focuses on the dependent-forcing variable relationships for the oil-related CDS sector indices and the fundamentals: the 10-year Treasury bond rate, the oil futures price and the S&P 500 index. As indicated above, Model 2 concentrates on the oil credit risks' relationships with the measures of market risks including the equity VIX, and the oil VIX as well as the default risk spread. The second model strives to examine the migration of risks between the oil credit risks and the market risks.

4.1 Model 1

4.1.1 Cointegration in Model 1

In Model 1 for the full period, there are five significant cointegration hypotheses for the dependent variables, suggesting the presence of five long-run relationships that bind the seven variables included in this model. The two important dependent variables that do not have a significant cointegration hypothesis are the CDS index for the oil and natural gas production sector, and the CDS index for the utility sector (see Table 3).

For the equation with the auto sector CDS index as the dependent variable, the cointegration hypothesis in this model is

$F(LAUTCD_t \mid LCHECD_t, LOILCD_t, LUTICD_t, LDGS10_t, LWTI3M_t, LSPINDEX_t)$ which yields significant F-statistics for all variables in the model, whether they are oil- related CDS indices, or the fundamental variables such as the 10-year Treasury bond rate, the oil price and the stock market. Based on Eq. (1), the most specific hypothesis is $\lambda_{1A} = \lambda_{2A} = \lambda_{3A} = \lambda_{4A} = \lambda_{5A} = \lambda_{6A} = \lambda_{7A} = 0$. Thus, the variable on the left-had side of "|" indicates the dependent variable, while those on the right-hand side

Table 3 Model 1's Bounds-Testing procedure results (full period)

Cointegration hypotheses	F-statistics
$F(LAUTCD_t \mid LCHECD_t, LOILCD_t, LUTICD_t, LDGS10_t, LWTI3M_t, LSPINDEX_t)$	4.1860*
$F(LCHECD_t \mid LAUTCD_t, LOILCD_t, LUTICD_t, LDGS10_t, LWTI3M_t, LSPINDEX_t)$	2.9141**
$F(LOILCD_t \mid LAUTCD_t, LCHECD_t, LUTICD_t, LDGS10_t, LWTI3M_t, LSPINDEX_t)$	1.6291
$F(LUTICD_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LDGS10_t, LWTI3M_t, LSPINDEX_t)$	1.6461
$F(LDGS10_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LWTI3M_t, LSPINDEX_t)$	4.4824*
$F(LWTI3M_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LDGS10_t, LSPINDEX_t)$	2.6193**
$F(LSPINDEX_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LDGS10_t, LWTI3M_t)$	2.2558***

Notes *, ** and *** represent significance at the 1, 5 and 10 % significance levels, respectively. This table indicates that there are five significant cointegration hypotheses according to the ARDL approach. In comparison, the Johansen procedure gives six cointegrating equations. The null hypothesis of no cointegration is that $\lambda_{1j} = \lambda_{2j} = \lambda_{3j} = \lambda_{4j} = \lambda_{5j} = \lambda_{6j} = \lambda_{7j} = 0$, where j represents one of the seven variables as in Eqs. (1–7). The variable on the left-hand side of "|" indicates the dependent variable, while those on the right-hand side are the potential *forcing variables*

are the potential *forcing variables*. This result indicates a long-run equilibrium relationship between the credit risk in the utility sector and the three other oil-related credit risks, the 10-year Treasury rate and the S&P500 index. All right-hand side variables are the *forcing variables* of the left-hand side auto sector CDS index. This finding demonstrates the degree of sensitivity of the auto credit risk to oil-related credit risks and oil/financial variables. The auto sector has forward and backward linkages with the rest of the economy and is highly sensitive to the business cycle.

The cointegration hypothesis for the equation with the chemical CDS as the dependent variable is

$F(LCHECD_t \mid LAUTCD_t, LOILCD_t, LUTICD_t, LDGS10_t, LWTI3M_t, LSPINDEX_t)$.

This hypothesis also attests that all other oil-related CDS indices and the fundamental variables are also forcing variables, as is the case with the auto CDS index. The chemical sector is cyclical and can be negatively related to oil which it uses as a feedstock.

When the dependent variable is the benchmark 10-year Treasury rate, the F-statistic for its cointegration hypothesis

$F(LDGS10_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LWTI3M_t, LSPINDEX_t)$

is also significant, underscoring the importance of the CDS, oil and stock market variables to this benchmark of the bond market. The same significant result holds for the oil and S&P 500 index equations.

However, no significant findings are reported for the equations of the CDS index for oil and gas and for utility. Parity between natural gas and oil prices has been weakening, particularly after the development of the technique that cracks gas shale, leading to the multiplication of natural gas reserves. The utility sector includes natural monopolies regulated by local authorities, which may have something to do with failing to have the other variables as its forcing variables.

4.1.2 Estimation of Long-Run and Short-Run Relationships in Model 1

The next step in the ARDL procedure is to estimate the coefficients of the long-run relationships using the following ARDL(x,y,z,l,m,n,s) models. The models are determined by the bounds testing procedure. The long-run relationships are given by:

$$\begin{aligned} \ln LAUTCD_t = a_1 &+ \sum_{i=1}^{x} \alpha_{i1} \ln LAUTCD_{t-i} + \sum_{i=0}^{y} \beta_{i1} \ln LCHECD_{t-i} + \sum_{i=0}^{z} \gamma_{i1} \ln LOILCD_{t-i} \\ &+ \sum_{i=0}^{l} \delta_{i1} \ln LUTICD_{t-i} + \sum_{i=0}^{m} \phi_{i1} \ln LDGS10_{t-i} + \sum_{i=0}^{n} \vartheta_{i1} \ln LWTI3M_{t-i} \\ &+ \sum_{i=0}^{s} \varphi_{i1} \ln LSPINDEX_{t-i} + \zeta_{i1} QE1 + \eta_{i1} QE2 + \varepsilon_{1t} \end{aligned} \tag{8}$$

$$\begin{aligned} \ln LCHECD_t = a_2 &+ \sum_{i=0}^{x} \alpha_{i2} \ln LAUTCD_{t-i} + \sum_{i=1}^{y} \beta_{i2} \ln LCHECD_{t-i} + \sum_{i=0}^{z} \gamma_{i2} \ln LOILCD_{t-i} \\ &+ \sum_{i=0}^{l} \delta_{i2} \ln LUTICD_{t-i} + \sum_{i=0}^{m} \phi_{i2} \ln LDGS10_{t-i} + \sum_{i=0}^{n} \vartheta_{i2} \ln LWTI3M_{t-i} \\ &+ \sum_{i=0}^{s} \varphi_{i2} \ln LSPINDEX_{t-i} + \zeta_{i2} QE1 + \eta_{i2} QE2 + \varepsilon_{2t} \end{aligned} \tag{9}$$

$$\begin{aligned} \ln LDGS10_t = a_3 &+ \sum_{i=0}^{x} \alpha_{i3} \ln LAUTCD_{t-i} + \sum_{i=0}^{y} \beta_{i3} \ln LCHECD_{t-i} + \sum_{i=0}^{z} \gamma_{i3} \ln LOILCD_{t-i} \\ &+ \sum_{i=0}^{l} \delta_{i3} \ln LUTICD_{t-i} + \sum_{i=1}^{m} \phi_{i3} \ln LDGS10_{t-i} + \sum_{i=0}^{n} \vartheta_{i3} \ln LWTI3M_{t-i} \\ &+ \sum_{i=0}^{s} \varphi_{i3} \ln LSPINDEX_{t-i} + \zeta_{i3} QE1 + \eta_{i3} QE2 + \varepsilon_{3t} \end{aligned} \tag{10}$$

$$\begin{aligned} \ln LWTI3M_t = a_4 &+ \sum_{i=0}^{x} \alpha_{i4} \ln LAUTCD_{t-i} + \sum_{i=0}^{y} \beta_{i4} \ln LCHECD_{t-i} + \sum_{i=0}^{z} \gamma_{i4} \ln LOILCD_{t-i} \\ &+ \sum_{i=0}^{l} \delta_{i4} \ln LUTICD_{t-i} + \sum_{i=0}^{m} \phi_{i4} \ln LDGS10_{t-i} + \sum_{i=1}^{n} \vartheta_{i4} \ln LWTI3M_{t-i} \\ &+ \sum_{i=0}^{s} \varphi_{i4} \ln LSPINDEX_{t-i} + \zeta_{i4} QE1 + \eta_{i4} QE2 + \varepsilon_{4t} \end{aligned} \tag{11}$$

$$\begin{aligned} \ln LSPINDEX_t = a_5 &+ \sum_{i=0}^{x} \alpha_{i5} \ln LAUTCD_{t-i} + \sum_{i=0}^{y} \beta_{i5} \ln LCHECD_{t-i} + \sum_{i=0}^{z} \gamma_{i5} \ln LOILCD_{t-i} \\ &+ \sum_{i=0}^{l} \delta_{i5} \ln LUTICD_{t-i} + \sum_{i=0}^{m} \phi_{i5} \ln LDGS10_{t-i} + \sum_{i=0}^{n} \vartheta_{i5} \ln LWTI3M_{t-i} \\ &+ \sum_{i=1}^{s} \varphi_{i5} \ln LSPINDEX_{t-i} + \zeta_{i5} QE1 + \eta_{i5} QE2 + \varepsilon_{5t} \end{aligned} \tag{12}$$

where *QE1* and *QE2* stand for quantitative easing for the two six months: one in the second half of 2009, and the other for the first half of 2010, respectively. For the ARDL models, we use a maximum lag order of 4, which can be considered as sufficiently high, given the fact that we use daily data.

The long-run results for the auto sector CDS spread show a significant relationship with only the CDS risk for the chemical sector. An increase in the chemical market risk leads to a spike in the risk in the auto sectors (Table 4). This sector is highly dependent on oil as a feedstock and is cyclical. Therefore, the auto and the chemical sectors have a common factor, namely the business cycle. However, the CDS spreads of the oil and natural gas production and utility sectors have no significant influence on the auto CDS spread in the long-run. In terms of the fundamentals factors strength in both the oil and stock market gives rise to less credit risk in this sector. When commodity and stock markets are improving, there is less need for traders and investors to buy protection against credit risk.

The short-run error-correction representation for the auto CDS spread shows more significant relationships than in the long-run. There are significant relationships with CDS spreads of the chemical, oil and natural gas production and utility sectors. Thus, in this representation there is a migration of credit risk from other oil-related sectors to the auto sector. There are also significant relationships with the three fundamental variables—the oil futures price, the S&P 500 index and the 10-year Treasury bond rate. These relationships are negative, suggesting that when the business cycles strengthen, the CDS spread drop in the auto sector. In this framework, neither *QE1* nor *QE2* has an impact on increasing the auto CDS spread in the short-run (Table 5).

Table 4 Estimated long-run coefficients of Model 1 (full period)

Forcing variable	Dependent variable				
	LAUTCD (ARDL(4, 0, 1, 1, 2, 2, 2))	LCHECD (ARDL(3, 0, 1, 1, 2, 2, 4))	LDGS10 (ARDL(3, 3, 2, 1, 1, 0, 4))	LWTI3M (ARDL(4, 2, 3, 0, 0, 2, 3))	LSPINDEX (ARDL(4, 4, 3, 2, 1, 4, 1))
LAUTCD		0.3337*	−0.0180	0.2778*	−0.0775***
LCHECD	1.3783*		0.2169	−0.2906***	0.0787
LOILCD	−1.0332	1.1954***	−2.1175*	−0.1283	0.1464
LUTICD	−0.4042	−0.1629	0.8770	0.4954***	−0.3878**
LDGS10	−0.1845	0.0535		−0.1524	0.0077
LWTI3M	1.2490*	−0.3882	−0.2967		0.1677**
LSPINDEX	−3.4845*	1.1136	0.8585	2.7122*	
C	26.6461*	−8.5243	3.4647	−16.5553*	7.6477*
QE1	−0.0857	−0.3178***	0.1613	0.1385	0.1018
QE2	−0.1986	−0.1405	−0.4989	−0.0721	0.2136**

All the ARDL(.) models are based on AIC. The asterisks *, ** and *** represent significance at the 1 %, 5 % and 10 % levels, respectively

Table 5 Error-correction representations Model 1 (full period)

Regressor	Dependent variables				
	ΔLAUTCD	ΔLCHECD	ΔLDGS10	ΔLWTI3M	ΔLSPINDEX
ΔLAUTCD		0.0080**	−0.1136*	0.0044*	−0.0316*
ΔLAUTCD1	0.1413*		0.0433***		
ΔLAUTCD2	−0.0960*				
ΔLAUTCD3	0.0784*				
ΔLCHECD	0.0214*		−0.0352	0.0297*	−0.0016
ΔLCHECD1		0.0262	0.0250	0.0214**	−0.0123**
ΔLCHECD2		−0.1074*	0.0504**	0.0320*	
ΔLOILCD	0.2786*	0.2245*	−0.1938*	−0.0020	−0.0571*
ΔLOILCD1					0.0284**
ΔLOILCD2					0.0130
ΔLOILCD3					0.0216***
ΔLUTICD	0.4036*	0.4413*	−0.1928**	0.0828*	−0.0973*
ΔLUTICD1				−0.1297*	
ΔLDGS10	−0.0862*	−0.0310		0.0050	0.0529*
ΔLDGS101	0.0707*	−0.0389***	0.0001	0.0479*	0.0012
ΔLDGS102			−0.0857*		0.0145*
ΔLWTI3M	0.0114	0.1331*	−0.0056		−0.0060
ΔLWTI3M1	−0.0956**	0.1339*		−0.0326	−0.0176
ΔLWTI3M2				0.0097	0.0068
ΔLWTI3M3				0.0608*	0.0240**
ΔLSPINDEX	−0.5843*	−0.0326	1.2654*	−0.0239	
ΔLSPINDEX1	−0.4985*	−0.2070**	−0.3295*	0.4258*	−0.2076*
ΔLSPINDEX2		−0.1903**	−0.0830	0.2057*	−0.1184*
ΔLSPINDEX3		−0.2124**	−0.2660*		0.0478**
QE1	−0.0013	−0.0076***	0.0030	0.0022	0.0016
QE2	−0.0031	−0.0034	−0.0094	−0.0011	0.0033**
ecm(−1)	−0.0156	−0.0239*	−0.0188*	−0.0159*	−0.0155*

Notes For the ARDL models see Table 4. The asterisks *, ** and *** represent significance at the 1%, 5% and 10% levels, respectively. Δ stands for the first difference

The behavior of the chemical sector's CDS spread is somewhat different from that of the auto sector in the short- and long-run. The results show that the CDS credit risk migrates from both the auto and oil and natural gas production sectors to the chemical sector in the long-run. As in the auto CDS framework, there is no significant directional relationship between the chemical CDS spread and the risk-free 10 year Treasury bond market in the chemical CDS framework. Interestingly, the chemical CDS spread is positively sensitive to *QE1*, which is not the case for the auto CDS spread.

In the short-run, all sector CDS spreads and fundamental variables have an influence over the chemical CDS spread. The financial fundamental factors (excluding the oil price) have an inverse directional relationship with the chemical CS spread. On the other hand, surges in the oil price lead to significant increases in the chemical CDS spread, and this is clearer than in the case of the auto CDS spread. Rises in the CDS spreads of all sectors also increase the CDS spread in the chemical sector.

In the long-run, the relationships between the oil price and the credit risks and the fundamental variables are less potent than in the short-run. In the long-run, the CDS spreads for the chemical, auto and utility sectors have differential impacts on the oil futures prices, with the auto and utility sector having a positive impact while the chemical having a negative relationship. This is somewhat surprising because auto and chemical sectors are cyclical while the utility sector is defensive. The fundamental variables also have varying impacts, with the S&P 500 index sharing a common business cycle with the "financialized" oil while the 10-year bond rate moves counter cyclically with the oil price.

For oil in the short-run over the full sample, rises in the CDS risk spread for the chemical and auto sectors cause the oil futures price to increase. The auto and chemical sectors use oil as a feedstock and are also highly cyclical, which may imply that during the expanding phase of the business cycle these sectors experience an increase in their risk protection in the form of higher option prices. The oil price includes a fear premium component, which possibly picks up spikes in fears in those cyclical oil-related sectors. On the other hand, increases in the CDS risks for the utility sector have mixed effects on oil futures, increasing the oil futures in the current period and reducing it in the previous period. As indicated previously, this differential cycling impact is probably linked to the nature of the utility sector, which is regulated by state governments and is considered a defensive sector at times of recessions.

When it comes to the financial fundamentals and the oil price in the long-run, there is a common factor that commoves the oil futures price with those variables, namely the strength of the business cycle and the overall economy. The result is also consistent with the notion of the "financialization" of oil. This finding should explain the positive relationship between the S&P500 index and the 10-year bond rate and the oil price.

As expected, the S&P 500 index as a dependent variable has fewer directional relationships with the oil-related CDS spreads and the fundamental variables in the long-run over the full sample. Increases in the auto and utility CDS spreads move this major stock index negatively, with the CDS spread of the other two sectors having no effects. This is due to the non-oil- related sectors that are included in the index. Among the fundamental variables, only the oil price positively co-moves the stock index. Interestingly, *QE2* but not *QE1* adds to the spike in the level of the S&P 500 index in this sector CDS framework.

In the short run, similarly to the oil price, the S&P 500 index has many directional relationships with the oil-related sector CDSs and the fundamental variables. In terms of the relationships with the sector CDS indices, increases in the credit risk spread of the chemical, auto and utility sectors reduce the S&P 500 index as more traders and investors purchase options to hedge against the rising credit risk in these oil-sensitive sectors. On the other hand, rising risk in the oil and natural gas production sector leads to a higher S&P 500 index. As for the short-run links with the fundamentals, this stock index responds positively to both the oil price and the 10-year Treasury bond rate. Thus, the relationship between the S&P 500 index and the oil price is positive and bidirectional. As in the long run, *QE2* and not *QE1* adds to increases in the stock index. Among the other fundamentals, only the S&P 500 index is significant in influencing the long-run interest rate benchmark and the relationship is negative, suggesting that a rising stock market index in this framework imply a lower 10-year bond rate.

The risk-free 10-year Treasury bond rate has a greater directional relationship with the sector CDS spreads and the other fundamental factors than the oil price and the S&P 500 index. In the long-run, this benchmark has a positive directional relationship with only the CDS risk in the oil and natural production sector. The unilateral causal relationship may reflect changes in the credit risk in the oil sector on inflationary expectations which also capture another measure of risk.

In the short run, the risk-free interest rate has fewer relations than the oil price and the stock index. Spikes in the chemical and delayed auto CDS spreads lead to a higher Treasury bond rate. On the other hand, increases in risk spreads in the utility and oil production and gas sector dampen the risk-free interest rate benchmark.

4.1.3 The Recovery Subperiod for Model 1

Model 1 is also estimated for the recovery subperiod which spans the period June 1, 2009 to July 13, 2011. There are five cointegration hypotheses for the dependent variables that are significant in Model 1 as in the previous model but the dependent variables that are significant have changed somewhat (see Table 6). In this model, the hypothesis for the utility CDS index becomes significant, while that for the S&P 500 index loses its significance. However, the CDS for oil and natural gas production remains exogenous in both periods for Model 1.

In the long-run relationship for the auto CDS spread, the significance of the variables has changed somewhat in the subperiod. The chemical CDS spread remains a significant forcing variable for the auto CDS spread (Table 7). But unlike the case of the full period, in the subperiod the CDS spread for the oil and natural gas production has become significant as a forcing variable for the auto CDS spread, while the oil price becomes insignificant. This is probably a sign of an increase in the long-run credit risk. The stock market also remains a significant forcing variable. In the short-run error correction representation, the dependent-forcing variable relations for the auto CDS spread basically remain the same. A notable difference is the change in the role of the oil and natural gas production

Table 6 Bounds-testing procedure results for Model 1 (subperiod)

Cointegration hypotheses	F-statistics
F($LAUTCD_t$ | $LCHECD_t$, $LOILCD_t$, $LUTICD_t$, $LDGS10_t$, $LWTI3M_t$, $LSPINDEX_t$)	2.3685***
F($LCHECD_t$ |$LAUTCD_t$, $LOILCD_t$, $LUTICD_t$, $LDGS10_t$, $LWTI3M_t$, $LSPINDEX_t$)	2.1802***
F($LOILCD_t$ |$LAUTCD_t$, $LCHECD_t$, $LUTICD_t$, $LDGS10_t$, $LWTI3M_t$, $LSPINDEX_t$)	1.2801
F($LUTICD_t$ |$LAUTCD_t$, $LCHECD_t$, $LOILCD_t$, $LDGS10_t$, $LWTI3M_t$, $LSPINDEX_t$)	3.7145*
F($LDGS10_t$ |$LAUTCD_t$, $LCHECD_t$, $LOILCD_t$, $LUTICD_t$, $LWTI3M_t$, $LSPINDEX_t$)	2.7710**
F($LWTI3M_t$ |$LAUTCD_t$, $LCHECD_t$, $LOILCD_t$, $LUTICD_t$, $LDGS10_t$, $LSPINDEX_t$)	2.7916**
F($LSPINDEX_t$ |$LAUTCD_t$, $LCHECD_t$, $LOILCD_t$, $LUTICD_t$, $LDGS10_t$, $LWTI3M_t$)	1.9665

Notes *, ** and *** represent significance at the 1 %, 5 % and 10 % levels, respectively

Table 7 Estimated long-run coefficients of Model 1 (subperiod)

Forcing variable	Dependent variable				
	LAUTCD (ARDL(2, 4, 0, 2, 1, 1, 3))	LCHECD (ARDL(4, 4, 3, 4, 2, 4, 2))	LUTICD (ARDL(4, 1, 2, 2, 0, 4, 2))	LDGS10 (ARDL(2, 0, 0, 1, 2, 0, 1))	LWTI3M (ARDL(2, 0, 0, 0, 3, 1, 2))
LAUTCD		0.0308	−0.1160	−0.2773	0.1890**
LCHECD	1.1891*		0.6907*	−2.1811	−0.2172
LOILCD	−0.6224**	−0.1608	0.2307***	−5.9542***	−0.1019
LUTICD	0.4155	0.5469		5.1377	0.2350
LDGS10	0.1170	0.0111	−0.0788***		0.0007
LWTI3M	0.5298	0.6353	−0.0941	−1.9527	
LSPINDEX	−3.3799*	−2.0130***	0.5254	−7.6963	1.6456*
C	22.3582*	13.7948*	−1.3465	80.4024	−8.0683*

All the ARDL(.) models are based on AIC. The asterisks *, ** and *** represent significance at the 1 %, 5 % and 10 % levels, respectively

CDS from a credit risk-elevation forcing variable in the full period to a dampening forcing variable in the subperiod. The oil price has also become a risk dampening risk forcing variable under the subperiod probably because of its steep decline as a result of the Great Recession.

The long-run relationship for the chemical CDS spread has weakened in the subperiod. Only the S&P 500 index is the significant forcing variable and has a cooling effect on the chemical CDS in this subperiod. Whereas the current effect elevates the credit risk, the delayed effect is risk-cooling (Table 8). Among the fundamentals, only the S&P 500 index maintains a consistent risk-dampening effect on the chemical CDS.

There is no long-run equation for the utility CDS spread in the full period. In the subperiod, the CDS spreads of both oil and natural gas production are risk-elevating forcing variables for the utility CDS spread. The only fundamental variables forcing

Table 8 Error-correction representations of Model 1 (subperiod)

Regressor	Dependent variables				
	ΔLAUTCD	ΔLCHECD	ΔLUTICD	ΔLDGS10	ΔLWTI3M
ΔLAUTCD		0.0392*	0.0421*	−0.2095*	−0.0268***
ΔLAUTCD1	0.3936*	0.0098	−0.0244***	0.1065**	
ΔLAUTCD2		−0.0260***			
ΔLAUTCD3		0.0335**			
ΔLCHECD	0.4222*		0.4227*	−0.3542*	0.1018**
ΔCHECD1	−0.0793	0.0795***	−0.0561		−0.1428*
ΔLCHECD2	0.3333*	0.0511			0.1106*
ΔLCHECD3	−0.1546***	−0.0746***			
ΔLOILCD	−0.0458**	0.1223*	0.1112*	−0.0856*	−0.0072
ΔLOILCD1		−0.0105			
ΔLOILCD2		0.0813*			
ΔLUTICD	0.5074*	0.4758*		0.0738***	0.0166
ΔLUTICD1	−0.2448***	−0.0396	0.1674*		
ΔLUTICD2		−0.0383	0.0209		
ΔLUTICD3		−0.0833***	0.0932*		
ΔLDGS10	−0.1053*	−0.0233**	−0.0035***		0.0000
ΔLDGS101		−0.0164		−0.0893**	
ΔLWTI3M	−0.1746***	0.0611***	0.0254	−0.0281	0.0687***
ΔLWTI3M1		0.0235	−0.0150		
ΔLWTI3M2		−0.0787**	0.0509***		
ΔLWTI3M3		−0.0814*	0.0789*		
ΔLSPINDEX	−0.4423**	−0.3233*	−0.2543*	2.1695*	0.1893**
ΔLSPINDEX1	0.4764**	−0.1987**	−0.1136***		0.7348*
ΔLSPINDEX2	0.4157**				
ecm(−1)	−0.0736*	−0.0288*	−0.0439*	−0.0144***	−0.0708*

Notes For the ARDL models see Table 7. The asterisks *, ** and *** represent significance at the 1%, 5% and 10% levels, respectively. Δ stands for the first difference

the utility CDS spread is the 10-year Treasury bond rate, causing dampening of this spread. In the short-run error correcting representation, the utility CDS spread has multiple relationships in the subperiod. All the CDS spreads for all the sectors as well as the three fundamentals influence the utility CDS spread.

Finally, in this subperiod all three financial and oil fundamentals have fewer long- and short-run representations whether in terms of the other sectors' CDS spreads or the other fundamentals. This could be due to the persistently high uncertainty in the markets.

4.2 Model 2

While this model has seven variables as in Model 1, the two financial fundamental variables in the previous model, specifically the S&P 500 index and the 10-year Treasury bond rate, are replaced with two measures of market and credit risks, namely the S&P VIX index and the default risk spread. Thus, the aim of this model is to examine the spillovers among the CDS credit risks for the four oil-related sectors and with the market risks of the stock and bond markets. The default risk spread is a much better measure of credit risk than the Merrill Lynch option volatility estimate (MOVE), which interrelates little with market and credit risks other than the S&P VIX. As in the previous model, we will examine the results for the full period and the subperiod in Model 2.

4.2.1 Cointegration in Model 2

This model has four cointegration relationships, compared to five in the previous model (Table 9). This suggests that replacing the financial variables with two market risk variables in Model 2 reduces the long-run relationships with the CDS credit risks of the four oil-related sectors. The significant cointegration hypotheses are found for the CDS indices of the auto and chemical sectors, the S&P VIX and the default risk variables. In the model, the CDSs of the utility sector and the oil and natural gas production are not significant like in the previous model. For the equation with the auto sector CDS index as the dependent variable, the cointegration hypothesis in this model is:

$F(LAUTCD_t \mid LCHECD_t, LOILCD_t, LUTICD_t, LSPVIX_t, LWTI3M_t, LDFR_t)$.

The other three cointegration hypotheses for the dependent variables $LCHECD_t$, $LSPVIX_t$ and $LDFR_t$ in the model are similar to the above cointegration hypothesis but alternate their dependent variables. The S&P VIX detects fears in all markets, while the default risk spread presages changes in economic activity where it spikes if it predicts a recession and dips if it forecasts a boom.

Table 9 Bounds-Testing procedure results for Model 2 (full period)

Cointegration hypotheses	F-statistics
$F(LAUTCD_t \mid LCHECD_t, LOILCD_t, LUTICD_t, LSPVIX_t, LWTI3M_t, LDFR_t)$	4.0210**
$F(LCHECD_t \mid LAUTCD_t, LOILCD_t, LUTICD_t, LSPVIX_t, LWTI3M_t, LDFR_t)$	3.6697**
$F(LOILCD_t \mid LAUTCD_t, LCHECD_t, LUTICD_t, LSPVIX_t, LWTI3M_t, LDFR_t)$	2.8175
$F(LUTICD_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LSPVIX_t, LWTI3M_t, LDFR_t)$	1.8562
$F(LSPVIX_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LWTI3M_t, LDFR_t)$	3.2510***
$F(LWTI3M_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LSPVIX_t, LDFR_t)$	1.2529
$F(LDFR_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LSPVIX_t, LWTI3M_t)$	3.6072***

Notes *, ** and *** represent significance at the 1 %, 5 % and 10 % levels, respectively. This table indicates that there are five significant cointegration hypotheses according to the ARDL approach

4.2.2 Estimation of Long-Run and Short-Run Relationships in Model 2

The next step in the ARDL procedure is to estimate the coefficients of the long-run relationship using the ARDL(x,y,z,l,m,n,s) specifications for Model 2. This model is determined by the bounds testing procedure. The long-run results for the dependent variable, the auto CDS, indicate significant relationships with both the CDS risks for the chemical and oil and natural sectors as well as with the oil price and the default risk spread (Table 10). Surprisingly, the auto CDS has no long-run directional relationship with the S&P VIX.

The short-run error-correction representation for the auto CDS spread in Model 2 shows fewer significant relationships than Model 1. There are significant relationships with the other three CDS spreads, the S&P VIX and the default risk spread in the short-run (Table 11). Thus, there is a migration of market risk to the auto credit risk in this equation. In this framework, as in Model 1, neither *QE1* nor *QE2* has an impact on increasing the auto CDS spread in the short-run.

The chemical CDS credit risk has no long-run relations with the other sectors' CDSs, VIX or default risk in this model. This is surprising given the similarity between the oil and chemical sectors. But in the short-run, the chemical CDS has a dependence relationship with the oil and natural gas production, utility CDSs and the oil price but not with the S&P VIX or the default risk spread in Model 2. There is no migration from the market risks to the credit risk in the chemical sector.

The estimate of the long-run relationship for the S&P VIX index has four long-run forcing variables, including specifically the CDS of the oil and natural gas production and the default risk spread as well as *QE2*. Thus, in this framework there

Table 10 Estimated long-run coefficients of Model 2 (full period)

Forcing variable	Dependent variable			
	LAUTCD (ARDL (4, 0, 1, 1, 2, 0, 0))	LCHECD (ARDL (3, 0, 3, 1, 0, 2, 0))	LSPVIX (ARDL (3, 0, 2, 4, 2, 0, 4))	LDFR (ARDL (4, 2, 0, 0, 1, 4, 4))
LAUTCD		0.1067	−0.0383	0.0916
LCHECD	0.6252***		0.1172	0.5253*
LOILCD	−1.4855**	0.6509	0.3067***	0.3760
LUTICD	0.3109	−0.3242	−0.0149	0.0114
LSPVIX	0.1781	0.3578		1.2847*
LWTI3M	0.4294***	−0.0731	0.2375*	0.1942
LDFR	0.7622*	0.2329	0.1556**	
C	4.3131**	1.3131	−0.1186	−6.7468*
QE1	0.1550	−0.2231	−0.0364	−0.6641*
QE2	−0.3636	−0.0940	−0.2443**	−0.0863

All ARDL(.) models are based on AIC. The asterisks *, ** and *** represent significance at the 1 %, 5 % and 10 % levels, respectively

Table 11 Error-correction representations of Model 2 (full period)

Regressor	Dependent variables			
	ΔLAUTCD	ΔLCHECD	ΔLSPVIX	ΔLDFR
ΔLAUTCD		0.0030	0.1156*	0.0015
ΔLAUTCD1	0.1390*		−0.0465***	
ΔLAUTCD2	−0.1020*			
ΔLAUTCD3	0.0897*			
ΔLCHECD	0.0105***		0.0082	0.0084**
ΔLCHECD1		0.0319		
ΔLCHECD2		−0.1107*		
ΔLOILCD	0.3293*	0.2616*	0.3848*	0.1210*
ΔLOILCD1		0.0310	−0.1179***	
ΔLOILCD2		0.1017**	−0.1183**	
ΔLOILCD3			−0.1056***	
ΔLUTICD	0.4777*	0.4542*	0.4887*	0.1864*
ΔLUTICD1			0.1203	0.0772**
ΔLUTICD2				0.0408
ΔLUTICD3				−0.0778**
ΔLSPVIX	0.0935*	0.0100		0.0767*
ΔLSPVIX1	0.0299		−0.1589*	−0.0299**
ΔLSPVIX2			−0.0980*	
ΔLWTI3M	0.0072	0.1037**	0.0166*	−0.0208
ΔLWTI3M1		0.1310*		−0.0296
ΔLWTI3M2				−0.0490***
ΔLWTI3M3				−0.0558**
ΔLDFR	0.0127*	0.0065	0.2855*	
ΔLDFR1			−0.1390*	0.0662*
ΔLDFR2			0.0585	0.0800*
ΔLDFR3			0.0636	0.1109*
QE1	0.0026	−0.0062	−0.0025	−0.0106*
QE2	−0.0061	−0.0026	−0.0171**	−0.0014
ecm(−1)	−0.0167*	−0.0280*	−0.0699*	−0.0160*

Notes For the ARDL models see Table 10. The asterisks *, ** and *** represent significance at the 1%, 5% and 10% levels, respectively. Δ stands for the first difference is a spillover from the oil and natural gas credit risk to the market risk S&P VIX, and a market to market risk migration between the VIX and the default risk spread. In the short-run, the S&P VIX has more significant relationships with the credit and market risks than in the long-run, including relationships with the utility and oil and natural gas CDSs, the default risk spread and *QE2*.

Finally, the long-run and short-run results for the dependent variable default risk spread indicate that there is a directional relationship between this market risk and its forcing variables the S&P VIX, and also with the oil and gas production CDS, and *QE1*. In the short-run, there is a relationship with the CDSs of the chemical, oil and gas production sectors, the oil price and *QE1*.

4.2.3 The Recovery Subperiod for Model 2

The S&P VIX in Model 2 of the full period is replaced in this subperiod with the oil VIX which has available data from May 17, 2007. The purpose of modeling this subperiod is to understand the directional relationships between the oil VIX and the oil-related sector CDSs which should focus on credit and market migration and cascading with the oil-related sectors. The sample size is dictated by the length of the oil VIX series which influenced the estimation in this subperiod.

Model 2 has one cointegrating hypothesis for this subperiod, compared to five for Model 1 over the recovery subperiod (Table 12). In this subperiod, there is no long-run relationship between these market risks and credit risks in the oil related sectors (Table 13).

In the short-run, the dependent variable default spread risk, has four forcing variables: Oil VIX, chemical CDS and auto CDS (Table 14).

Table 12 Bounds-testing procedure results of Model 2 (subperiod)

Cointegration hypotheses	F-statistics
$F(LAUTCD_t \mid LCHECD_t, LOILCD_t, LUTICD_t, LOILVIX_t, LWTI3M_t, LDFR_t)$	2.4257
$F(LCHECD_t \mid LAUTCD_t, LOILCD_t, LUTICD_t, LOILVIX_t, LWTI3M_t, LDFR_t)$	2.1203
$F(LOILCD_t \mid LAUTCD_t, LCHECD_t, LUTICD_t, LOILVIX_t, LWTI3M_t, LDFR_t)$	1.2500
$F(LUTICD_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LOILVIX_t, LWTI3M_t, LDFR_t)$	3.0590
$F(LOILVIX_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LWTI3M_t, LDFR_t)$	2.7870
$F(LWTI3M_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LOILVIX_t, LDFR_t)$	1.6205
$F(LDFR_t \mid LAUTCD_t, LCHECD_t, LOILCD_t, LUTICD_t, LOILVIX_t, LWTI3M_t)$	4.2180**

Notes *** represent significance at the 10 % levels, respectively

Table 13 Estimated long-run coefficients of Model 2 with OILVIX (subperiod)

Forcing variable	Dependent variable
	LDFR ARDL(3, 1, 4, 3, 0, 0, 0)
LOILVIX	0.2376
LCHECD	1.1920***
LAUTCD	−0.3820
LOILCD	0.4479
LUTICD	0.8066
LWTI3M	−0.0590
CC	−7.8820**

The ARDL(.) is based on AIC. The asterisks ** and *** represent significance at the 5 % and 10 % levels, respectively

Table 14 Error-correction representation of Model 2 with OILVIX (subperiod)

Regressor	Dependent variable
	ΔLDFR
ΔLDFR1	0.1144*
ΔLDFR2	0.0942**
ΔLOILVIX	0.0652**
ΔLCHECD	0.2562*
ΔLCHECD1	0.0744
ΔLCHECD2	0.0280
ΔLCHECD3	0.1478**
ΔLAUTCD	−0.0099
ΔLAUTCD1	0.0078
ΔLAUTCD2	−0.0647**
ΔLOILCD	0.0138
ΔLUTICD	0.0248
ΔLWTI3M	−0.0018
ecm(−1)	−0.0308*

Notes For the ARDL model see Table 13. The asterisks * and ** represent statistical significance at the 1 % and 5 % levels, respectively. Δ stands for the first difference

5 Conclusions

Given the pivotal role that the oil price plays in the economy and financial markets and the rising risk and uncertainty, this study analyzes primarily the dependent-forcing variable relationships for the insurance protections for four oil-related sectors- auto, chemical, oil and natural gas production, and utility, as well as their relations with the other risk and fundamental variables. The oil-related sectors are considered among the largest S&P sectors. The first two sectors are highly cyclical and follow the business cycle, while the last sector is defensive and could fare better even during a contractionay phase of the business cycle. Buyers of debt protection pay a premium or a spread which increases during periods of high risk.

This study examines migration and cascading of credit risks among those sectors, while controlling for the oil and financial fundamentals and market risks. It also examines the 2009/2010 impact of quantitative easing on market and CDS credit risks. The analysis is performed within the framework of two models estimated over the full sample period and a subperiod that spans the 2009 recovery following the 2007/2008 recession. Model 1 examines the relationships for the four oil-related credit risks, given the presence of three fundamental variables: the WTI oil futures price, the 10-year Treasury bond rate and the S&P 500 index. Model 2 examines the four oil-related credit risks and controls for two market risks represented by the S&P VIX and the default risk spread. Moreover, in this model the S&P VIX of the full sample is replaced with the oil VIX in the 2009 recovery subperiod due to the availability of data in May 2007.

The main finding of the study indicates that replacing the two financial fundamental variables with the two market risks reduces the long- and short-run risk migration and cascading in the second model relative to the first model in both the full sample and the subperiod. This finding underscores the importance of including fundamental variables when are modelling and examining credit risks that also reflect risk. The oil price has more and varying credit risk spillover effects in the short-run than in the long-run; it has a dampening effect on the auto CDS spread and a heightening effect on the chemical CDS spread. A surging10-year Treasury bond rate has a dampening effect in the short-run, suggesting that the position of monetary policy matters for oil-related CDS risks. The S&P 500 index moves in the opposite direction to those CDS risks, benefiting from greater liquidity.

Among the two market risk variables, the default risk spread which foreshadows changes in economic activity has a stronger lead/lag relationship with the other types of the credit and market risks than the S&P VIX index which gauges fears in the overall economy. When the S&P VIX is replaced with the oil VIX in the subperiod of Model 2, the directional relations weaken significantly, abating the migration and cascading of different risks among the risk variables.

The long- and short-run relationships among the oil-related credit risks and with the other variables are more diverse. The cyclical chemical and auto CDS spreads are the most responsive of these sector CDS and the oil and natural gas production and utility CDS spreads are least responsive.

The recent quantitative easing *QE1* and *QE2* has limited impact affecting mainly the financial variables. In Model 1, the impact is on the chemical CDS and the S&P 500 index, while in Model 2 it affects the S&P VIX and the default risk spread. Therefore, QE does not seem to contribute considerably to the oil-related credit risks.

References

Becker R, Clements AE, McClelland A (2009) The jump component of S&P 500 volatility and the VIX index. J Banking Finan 33(6):1033–1038

Berndt A, Douglas R, Duffie D, Ferguson M, Schranz D (2008) Measuring default risk premium from default swap rates and EDFs. Working paper 173, bank of international settlements. http://www.andrew.cmu.edu/user/aberndt/Beetal08.pdf. Accessed 25 April 2011

Bystrom H (2006) Credit default swaps and equity prices: the iTraxx CDS index market. Finan Anal J 62:65–76

Das SR, Hanouna P (2006) Credit default swap spreads. J investment Manag 4:93–105

Fernandes M, Medeiros MC, Scharth M (2009) Modeling and predicting the CBOE market volatility index. http://webspace.qmul.ac.uk/mfernandes/Papers/Vix.pdf

Figuerola-Ferretti I, Paraskevopoulos I (2010) Pairing market with credit risk. http://ssrn.com/abstract=1553863

Forte S, Lovreta L (2008) Credit risk discovery in the stock and CDS markets: who, when and why leads. http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2008-athens/Lovreta.pdf

Fung HG, Sierra GE, Yau J, Zhang G (2008) Are the U.S. stock market and credit default swap market related? Evidence from the CDX indices. J Altern Invest 11:43–61

Gogineni S (2010) Oil and stock market: an industry level analysis. Finan Rev 45(4):891–1123

Greatrex C (2008) The credit default swap market's determinants, efficiency and relationship to the stock market. ETD collection for Fordham University. Paper AAI3301438. http://fordham.bepress.com/dissertations/AAI3301438

Hammoudeh S, Nandha M, Yuan Y (2013) Dynamics of CDS spread indices of US financial sectors. Appl Econom 45:213–223

Hammoudeh S, Sari R (2011) Financial CDS, stock market and interest rates: which determine which? North Am J Econom Finan 22(3):257–276

Jagerson J (2008) Using the oil VIX (OVX) to forecast energy prices. http://www.learningmarkets.com/using-the-oil-vix-ovx-to-forecast-energy-prices/. Accessed 3 Oct 2011

Longstaff F, Mithal S, Neis E (2005) Corporate yield spreads: default risk or liquidity? New evidence from the credit default swaps market. J Finan 60:2213–2253

Luo X, Zhang JE (2010) The term structure of VIX. http://hdl.handle.net/10722/127838

Norden L, Weber M (2009) The co-movement of credit default swap, bond and stock markets: an empirical analysis. Eur Finan Manag 15:529–562

Norden L, Weber M (2004) Informational efficiency of credit default swaps and stock market. J Bank Finan 28:2813–2843. Also available as discussion paper DP 4250, Center for Economic Policy Research. London, United Kingdom. www.cepr.org/pubs/dps/DP4250.asp

Pesaran MH, Pesaran B (2009) Time series econometrics using Microfit 5.0. Oxford University Press, New York

Raunig B, Scheicher M (2009) Are the banks different? Evidence from CDS market, working paper 152, Oesterreichische National Bank. https://www.bis.org/bcbs/events/cbrworkshop09/raunigscheicher.pdf. Accessed 2 Oct 2011

Stanton R, Wallace N (2011) The Bear's lair: index credit default swaps and the subprime mortgage crisis. https://fisher.osu.edu/blogs/efa2011/files/REF_2_2.pdf. Accessed 2 Oct 2011

Zhang B, Zhou H, Zhu H (2009) Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. Rev Finan Stud 22(12):5099–5131

Zhu H (2006) An empirical comparison of credit spreads between the bond market and the credit default swap market. J Finan Serv Res 29:211–235

Oil Futures Market: A Dynamic Model of Hedging and Speculation

Giulio Cifarelli and Giovanna Paladino

Abstract This paper develops a non linear model for oil futures prices which accounts for pressures due to hedging and speculative activities. The corresponding spot market is assumed to maintain a long term equilibrium relationship with the futures prices in line with the presence of an arbitrage led time varying basis. The model combines an error correction relationship for the cash returns and a non linear parameterization of the corresponding futures returns with a bivariate CCC-TGARCH representation of the conditional variances. The dynamic interaction between spot and futures returns in the oil market has been investigated over the 1990–2010 time period. We have found clear evidence of the activity of hedgers and speculators on the futures markets and the role of the latter is far from negligible. Finally, in order to capture the consequences of the growing impact of financial flows on commodity market pricing, a two-state regime switching model for futures returns has been implemented. The empirical findings indicate that hedging and speculative behavior change across the two regimes, which we associate with low and high return volatility, according to a distinctive pattern.

Keywords Spot and futures markets · Dynamic hedging · Speculation · Non linear GARCH · Markov regime switching

G. Cifarelli (✉)
Dipartimento di Scienze Economiche, University of Florence, via delle Pandette 9, 50127 Florence, Italy
e-mail: giulio.cifarelli@unifi.it

G. Paladino
Economics Department, LUISS University and Intesa Sanpaolo, Piazza San Carlo 156, 10156 Turin, Italy
e-mail: gpaladino@luiss.it

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_6

1 Introduction

This research focuses on hedging and speculation with futures contracts. Futures trading involves an exchange of contracts between people with opposing views of the market and/or with a different degree of risk aversion. The risk is shifted from a party that desires less risk to a party that is willing to accept it in exchange for an expected profit.[1]

In the real world there is no clear separation between agents since hedgers and speculators do not play a strictly uniform role. It may well be that typical hedgers, such as commercial firms, take a certain view on the market and speculate on price direction; alternatively, speculators can find it profitable to engage in hedging activities (see Stulz 1996; Irwin et al. 2009). Consequently it is misleading to consider hedgers as pure risk-averse agents and speculators as risk-seekers. In this paper the futures' demand function avoids this simplistic division.

Speculators, mainly non commercial firms or private investors, are essential for the smooth functioning of commodity markets as they assure liquidity and assume the risks discarded by hedgers in order to earn profits stemming from the expected price changes. Speculators intervene directly in the futures market where transaction costs are low and no physical delivery is involved.

The literature on commodity market speculation has followed two main strands. A direct approach attempting to micro model simultaneously speculative and hedging behavior, and an indirect approach analyzing the excess co-movement of commodity prices that ascribes this evidence to 'herding' behavior.

As for the direct approach, an important paper by Johnson (1960) suggests that hedging and speculation in futures markets are interrelated. Speculation is mainly attributed to traders' expectations on future price changes that bring about an increase/decrease of the optimal hedging ratio in a short hedging context. Ward and Fletcher (1971) generalize Johnson's approach to both long and short hedging and find that speculation is associated with optimal futures positions (short or long) that are in excess of the 100 % hedging level.

As stated above, the indirect approach focuses on the presence of excess co-movement of returns (with respect to a component explained by fundamentals) of unrelated commodities (Pyndick and Rotemberg 1990). Subsequent research—see among others Cashin et al. (1999), Ai et al. (2006), and Lescaroux (2009)—challenged the excess co-movement hypothesis on empirical and methodological grounds. The results are mixed and could indeed depend on the selection of the estimation techniques and/or of the information set (Le Pen and Sévi 2010).

Finally, the disclosure of data on the Commitments of Traders Reports, provided by the Commodity Futures Trading Commission, has recently produced papers that try to assess the impact of speculation on commodity prices, measuring speculative positions in terms of open interest. The weekly open interest of each commodity is

---

1 Fagan and Gencay (2008) find that hedgers and speculators are often counterparties, since they tend to take opposing positions. Their respective long positions exhibit a strong negative correlation.

broken down, according to the purposes of traders. The empirical results, however, are mixed (Fagan and Gencay 2008).

The analysis of hedging is less variegated. Stein (1961) and McKinnon (1967) model hedgers' behavior consistently with the minimization of the variance of the return of a portfolio constructed with cash and futures contracts. The optimal cover ratio (the Minimum Variance Hedge ratio or MVH) is defined as the percentage of cash position or value matched by futures contracts that minimizes the variance of the hedged portfolio. The MVH strategy pays no attention to the hedged portfolio expected return. Subsequent improvements include strategies based on hedged portfolio return mean and variance expected utility maximization (Cecchetti et al. 1988), minimization of the extended mean-Gini coefficient (Kolb and Okunev 1992), or based on the Generalised Semivariance (Lien and Tse 2000). It has been shown, however, that if futures prices are martingale processes and if spot and futures returns are jointly normal then the optimal hedge ratio converges to the MVH ratio.

Given the stochastic nature of futures and spot prices, the hedge ratio is unlikely to be constant. Static OLS hedge ratio estimation recognizes that the correlation between the futures and spot prices is less than perfect (Figlewski 1984), but imposes the restriction of a constant correlation between spot and futures price rates of change. As such, it could lead to sub-optimal hedging decisions in periods of high basis volatility. The properties of the joint distribution of the returns have prompted the implementation of GARCH techniques in studies which find that optimal hedge ratios are time dependent and that dynamic hedging reduces in-sample portfolio variance substantially more than static hedging. They are based on the estimation of bivariate conditional variance models of varying complexity (see, among others, Kroner and Sultan (1993), Chan and Young (2006), who incorporate a jump component in a bivariate GARCH, and Lee and Yoder (2007), who implement a Markov switching GARCH).

This paper investigates some complex dynamic properties of cash and futures prices—typically disregarded in literature. We develop a plausible model of oil market hedger and speculator short-run reaction to expected returns and volatility shifts. The empirical findings corroborate our a priori hypotheses and provide innovative insights into the impact on futures pricing of the interaction between hedging and speculation across volatility regimes. In line with Tokic (2011) we find evidence of a change in the attitude of hedgers in periods of high volatility, when their comprehension of the market declines. They adopt a destabilizing feedback trading behavior, reducing their optimal hedge ratio, and becoming more active on the futures market.

This paper contributes to the current debate as follows:

a. Using a complex non linear CCC-TGARCH approach we model explicitly the reaction of hedgers and speculators to volatility shifts in the oil market. In this way the literature is extended by adding a dynamic component to the standard optimal hedge ratio computation.
b. A two-state Markov switching procedure is used to model the impact of changes in the behavior of the oil market, changes due to bullish/bearish reactions to futures price changes and/or to shifts in risk aversion brought about by return

volatility changes. We thus identify a financial pattern that seems to play an important role in recent oil market pricing.

c. We model and assess empirically the relative impact of speculative versus hedging drivers on oil futures pricing, and investigate whether periods of high futures return volatility may be associated with a more intense speculative activity.

Following a discussion of the properties of a dynamic model of hedging and speculation (Sect. 2), the paper outlines the main features of the non linear multivariate CCC-TGARCH used in the empirical investigation (Sect. 3), sets forth the estimates for the oil market (Sect. 4), and presents a Markov switching framework in which the drivers of futures returns are assumed to switch between two different processes that are dictated by the state of the market (Sect. 5). Section 6 provides the main conclusions.

2 A Dynamic Model of Hedging and Speculation

Hedging transactions are intended to reduce the risk of unwanted future cash price changes to an acceptable level by offsetting spot market trades with trades of the opposite sign in the corresponding futures market. Thus, if current cash and futures prices are positively correlated, the financial loss in one market will be compensated by the gains obtained from holding the opposite position in the other market.

In more detail, let $r_{c,t} = \Delta \log C_t = \Delta c_t$ and $r_{f,t} = \Delta \log F_t = \Delta f_t$, where C_t is the oil cash (spot) price and F_t is the price of the corresponding futures contract. An investor who takes a long (short) position of one unit in the cash market will hedge by taking a short (long) position of β units in the corresponding futures market, which he will buy (sell) back when he sells (buys) the cash. The hedge ratio β can be seen as the proportion of the long (short) cash position that is covered by futures sales (purchases).[2]

The revenue of this hedging position (or portfolio), i.e. the hedger's return $r_{H,t}$, is given by

$$r_{H,t} = r_{c,t} - \beta r_{f,t} \tag{1}$$

The variance of this portfolio is given by

$$\sigma^2_{r_H,t} = \sigma^2_{r_c,t} + \beta^2 \sigma^2_{r_f,t} - 2\beta \sigma_{r_c,t} \sigma_{r_f,t} \rho_{r_c r_f,t} \tag{2}$$

where $\sigma^2_{r_c,t}$ is the variance of $r_{c,t}$, $\sigma^2_{r_f,t}$ is the variance of $r_{f,t}$, and $\rho_{r_c r_f,t}$ is the correlation between $r_{c,t}$ and $r_{f,t}$.

The optimum hedge ratio β^* is derived from the first order condition of the hedging portfolio variance minimization and reads as

[2] The hedge ratio is also defined as the ratio between the number of futures and cash contracts.

$$\beta^* = \frac{\sigma_{r_c,t}\sigma_{r_f,t}\rho_{r_c r_f,t}}{\sigma^2_{r_f,t}} \tag{3}$$

The optimum hedge ratio depends on both the covariance between changes in futures and cash prices, $\sigma_{r_c r_f,t} = \sigma_{r_c,t}\sigma_{r_f,t}\rho_{r_c r_f,t}$, and the variance of futures price changes.

We extend the standard hedging model by introducing a dynamic component. The expected utility of hedgers is assumed to be an inverse function of the expected variability of their optimally hedged position. The variance of this position (or portfolio) can be defined, replacing the optimal hedge ratio β^* in Eq. (2) with its determinants set out in Eq. (3), as

$$\sigma^2_{r_H,t} = \sigma^2_{r_c,t} - \frac{\left(\sigma_{r_c r_{f,t}}\right)^2}{\sigma^2_{r_{f,t}}} = \sigma^2_{r_c,t}\left(1 - \rho^2_{r_c r_f,t}\right) \tag{4}$$

where $\rho_{r_c r_f,t} = {\sigma_{r_c r_f,t}}\big/{\sigma_{r_f,t}\sigma_{r_c,t}}$

An increase in the minimum portfolio variance may be due to a rise in the variability of cash price changes and/or to a decrease in the correlation between cash and futures price changes.

The demand of futures contracts of an hedger wishing to minimize the variance of his optimal portfolio is defined as

$$D_t^H = a_0 + b^H\sigma^2_{r_c,t}\left(1 - \rho^2_{r_c r_f,t}\right) \tag{5}$$

where a_0 is a constant term and b^H measures the demand sensitivity to the variability of the return of the optimally hedged position. We can thus reasonably assume that b^H is positive if consumers' hedging prevails, since consumers, concerned about cash price increases, will demand more futures contracts whenever the portfolio variance increases. Conversely, b^H will be negative if producers' hedging prevails, since producers, worried about possible cash price decreases, will supply more (i.e. demand fewer) contracts if the variability of their hedged position rises.

The demand for futures contracts of a speculator is defined as

$$D_t^S = c_0 + d^S E_t r_{f,t+1} - e^S\sigma^2_{r_f,t} \tag{6}$$

where c_0 is a constant term. d^S is assumed to be always positive because of the positive impact on oil speculation of an increase in expected futures returns, whereas e^S can be either positive or negative, depending on speculator reaction to risk. We assume that $e^S < 0$ for risk lover and $e^S > 0$ for risk averse agents.

It is generally accepted that futures trading is a zero sum game. Thus we can assume that the net demands of both agents are balanced on a daily basis or, equivalently, that the demands of hedgers and speculators add up to 1

$$D_t^H + D_t^S = 1 \tag{7}$$

By substituting Eqs. (5) and (6) in Eq. (7) and readjusting terms, we obtain the following expression for the expected futures return

$$E_t r_{f,t+1} = \frac{1}{d^S}\left(1 - a_0 - c_0 - b^H \sigma^2_{r_c,t}\left(1 - \rho^2_{r_c r_f,t}\right) + e^S \sigma^2_{r_f,t}\right)$$

Since $r_{f,t+1} = E_t r_{f,t+1} + u_{r_f,t+1}$, we obtain the following testable short run relationship

$$r_{f,t+1} = e_0 - \left(b^H/d^S\right)\sigma^2_{r_c,t}\left(1 - \rho^2_{r_c r_f,t}\right) + \left(e^S/d^S\right)\sigma^2_{r_f,t} + u_{r_f,t+1} \tag{8}$$

where $e_0 = (1 - a_0 - c_0)/d^S$. Equation (8) relates futures returns to their own volatility and to the variability of the optimally hedged portfolio. The short-run dynamics of this relationship is in line with the stylized facts detected in the paper by Fagan and Gencay (2008), where the negative correlation between futures returns and hedger net long positions supports the idea that large speculators are net buyers in rising markets, while large hedgers are net sellers. This behavior is encompassed by our (more general) model, when it contemplates the case of hedgers being net sellers—when b^H is negative—and futures returns rising.

3 A Bivariate Non Linear CCC-TGARCH Representation

We focus on futures prices since commodity prices are typically discovered in futures markets and price changes are passed on from futures to cash markets (Garbade and Silber 1983). Economic theory, however, suggests that the prices of cash assets and of the corresponding futures contracts are jointly determined (Stein 1961). Our empirical estimation thus includes a relationship that describes the behavior of cash returns, along with a futures returns relationship, and analyzes their covariance. Over the longer term equilibrium, prices are ultimately determined in the cash market as all commodity futures prices at delivery date converge to the cash price (plus or minus a constant). This behavior justifies the existence of a cointegration relationship between futures and cash prices and the use of an error correction parameterization of the conditional mean equation for $r_{c,t}$, where cash prices adjust to futures prices (the forcing variable) in line with the adopted framework of price discovery.[3] In the long run the relation between cash and futures prices holds and accounts for the presence of an identified basis or convenience yield.

[3] On this point see Figuerola-Ferretti and Gonzalo (2010). They successfully apply a VECM approach to cash and futures commodity returns where cash prices adjust to futures prices, in line with the Garbade and Silber (1983) framework of price discovery.

A non linear bivariate GARCH model for futures and spot returns is thus estimated. The conditional mean of the futures returns is modeled by Eq. (8′), while the conditional mean of the cash returns, Eq. (9), is parameterized by an autoregressive error correction structure and the conditional second moments are quantified by a bivariate CCC-TGARCH(1,1).

$$r_{c,t} = a_0 + \sum_{j=1}^{n} a_j r_{c,t-j} + \sum_{k=1}^{m} b_k r_{f,t-k} + \varepsilon_1 (f_{t-1} - d_0 - d_1 c_{t-1}) + u_{r_c,t} \tag{9}$$

$$r_{f,t} = e_0 - \left(b^H/d^S\right)\left(h^2_{r_c,t-1} - h^2_{r_c r_f,t-1}\Big/h^2_{r_f,t-1}\right) + \left(e^S/d^S\right)h^2_{r_f,t-1} + u_{r_f,t} \tag{8′}$$

$$u_t = \begin{bmatrix} u_{r_c,t} \\ u_{r_f,t} \end{bmatrix}$$

$$u_t | \Omega_{t-1} N(0, H_t)$$

$$\begin{aligned}
H_t &= \Delta_t R \Delta_t \\
R &= \begin{bmatrix} 1 & \rho_{r_c r_f} \\ \rho_{r_c r_f} & 1 \end{bmatrix} \quad \Delta_t = \begin{bmatrix} h_{r_c,t} & 0 \\ 0 & h_{r_f,t} \end{bmatrix} \\
h^2_{r_c,t} &= \varpi_c + \alpha_c h^2_{r_c,t-1} + \beta_c u^2_{r_c,t-1}; \\
h^2_{r_f,t} &= \varpi_f + \alpha_f h^2_{r_f,t-1} + \beta_f u^2_{r_f,t-1} + \gamma_f S_{t-1} u^2_{r_f,t-1} \\
S_{t-1} &= \begin{cases} 1 & if \quad u_{r_f,t-1} < 0 \\ 0 & if \quad u_{r_f,t-1} \geq 0 \end{cases}
\end{aligned}$$

4 The Empirical Behavior of the Oil Market

Our daily data span the 3 January 1990–26 January 2010 time period. All contracts are traded on the NYMEX (New York Mercantile Exchange), refer to West Texas Intermediate oil price, and are taken from Datastream. Futures prices relate to a continuation contract. Both spot (C_t) and futures prices (F_t) are expressed in US dollars. Futures prices correspond to the highly liquid 1 month (nearest to delivery) futures contract.[4] Returns are computed as first differences of the log of the price levels, their summary statistics are presented in Table 1.

Average daily returns and standard deviations are small but not negligible. Both cash and futures returns have mildly skewed and significantly leptokurtic distributions. The departure from normality is confirmed by the size of the corresponding Jarque Bera (J.B.) test statistics whereas the presence of volatility clustering

[4] The futures contract expires on the third business day prior to the 25th calendar day of the month preceding the delivery month. If the 25th calendar day of the month is a non-business day, trading ceases on the third business day prior to the business day preceding the 25th calendar day.

Table 1 Descriptive statistics

Return	Mean	St. dev.	Sk.	Kurt.	J.B.	$Q_x^2(1)$	$Q_x^2(6)$
Oil futures	0.000270	0.0250	−0.95	17.52	67710.6	73.16 [0.00]	369.33 [0.00]
Oil cash	0.000240	0.0240	−1.23	24.63	1333762.2	147.36 [0.00]	672.60 [0.00]

Daily sample from 3 January 1990 to 26 January 2010 (5,325 observations)
Notes Sk. skewness; *Kurt.* kurtosis; *J.B.* Jarque Bera test statistic; $Q_x^2(k)$ Ljung Box Q-statistic for kth order serial correlation of the squared variable x^2; probability levels are in square brackets

supports the choice of a GARCH parameterization of the conditional second moments. As expected, the logarithms of the prices of the cash and futures contracts are always I(1) and their first differences I(0).[5]

Table 2 presents parsimonious estimates of the conditional mean and variance equations of the bivariate non linear CCC-TGARCH(1,1) system set forth in Sect. 3. The overall quality of fit is satisfactory. The estimated parameters are significantly different from zero and the conditional heteroskedasticity of the residuals is captured by our GARCH parameterization.[6] The usual misspecification tests suggest that the standardized residuals v_{z_t}, $z_t = r_{c,t}, r_{f,t}$, are symmetric and well behaved; the J.T.A. test statistics are not significant, $E[v_{z_t}] = 0$, $E\left[v_{z_t}^2\right] = 1$, and $v_{z_t}^2$ is serially uncorrelated. Considering that d^S is positive by construction and that the sign of the coefficient ratios b^H/d^S and e^S/d^S will depend upon the sign of b^H and e^S, the futures return mean Eq. (8′) provides some useful information on market drivers:

(i) coefficient b^H estimates are positive reflecting the predominance of consumer agents on the market. This result is also in line with the effects of hedging pressure, where futures prices increase when hedgers trade short and decrease when hedgers are long[7];
(ii) the absolute value of the ratio between speculative and hedging factors $e^S\sigma_{r_f,t}^2 \Big/ b^H\sigma_{r_c,t}^2\left(1-\rho_{r_c r_f,t}^2\right)$—see the SPEC index set forth in Table 2—measures the relative impact of different sources of risk on futures returns using a "level of importance" criterion.[8] It is higher than 1 which suggests that speculators are more reactive than hedgers;

[5] The unit root test statistics are not reported due to lack of space.

[6] The t-ratios reported in the tables are based on the robust quasi-maximum likelihood estimation procedure of Bollerslev and Wooldridge (1992) since the J.B. test statistics reject the null of normality of the standardized residuals.

[7] See Chang (1985) and Bessembinder (1992).

[8] For a definition of this measure, see Achen (1982, pp. 72–73).

Table 2 Bivariate non linear CCC-TGARCH(1,1) full sample estimates

Conditional means n = 1, m = 1			Conditional Variances			
	$r_{c,t}$	$r_{f,t}$				
a_0	−0.014 (−84.75)		ϖ_c	1.2E − 05 (38.38)	ϖ_f	1.6E − 05 (47.89)
a_1	−0.223 (−26.98)		α_c	0.895 (891.37)	α_f	0.876 (701.98)
			β_c	0.083 (83.99)	β_f	0.088 (58.37)
b_1	0.245 (32.34)				γ_f	0.020 (7.43)
			$\rho_{r_c r_f}$	0.761 (227.70)		
ε_1	0.071 (87.24)					
d_0	–					
d_1	0.956 (1492.95)					
e_0		1.1E – 04 (0.65)				
$\left(b^H/d^S\right)$		3.968 (5.13)				
$\left(e^S/d^S\right)$		2.019 (6.58)				
Residual diagnostics						
$v_{r_c t} = u_{r_c t}\Big/\sqrt{h^2_{r_c t}}$		$v_{r_f t} = u_{r_f t}\Big/\sqrt{h^2_{r_f t}}$				
$E(v_{r_c t})$	−0.002 [0.85]	$E(v_{r_f t})$	−0.003 [0.80]			
$E\left(v^2_{r_c t}\right)$	1.000	$E\left(v^2_{r_f t}\right)$	1.000			
Sk.	−0.583	Sk.	−0.542			
Kurt.	7.045	Kurt.	5.924			
$Q^2_x(1)$	0.160 [0.69]	$Q^2_x(1)$	0.068 [0.79]			
$Q^2_x(6)$	11.691 [0.07]	$Q^2_x(6)$	1.654 [0.95]			
J.T.A.	0.073 [0.97]	J.T.A.	1.321 [0.265]			
J.B.	11116.0 [0.00]	J.B.	7906.65 [0.00]			
SPEC	1.1649					
LLF	27560.79					

Notes Sk. skewness; *Kurt.* excess kurtosis; $Q^2_x(k)$ Ljung Box Q-statistic for kth order serial correlation of the squared variable x^2 ; *J.T.A.* joint Wald test of the null hypothesis of no asymmetry distributed as χ^2 with 3 degrees of freedom (Engle and Ng 1993); *SPEC* speculative to hedging factors ratio defined as the absolute value of $e^S\sigma^2_{r_f,t}\Big/b^H\sigma^2_{r_c,t}\left(1-\rho^2_{r_c r_f,t}\right)$; *LLF* log likelihood function

(iii) speculators are risk averse since the corresponding e^S coefficient estimates are positive; this finding may be due to the size of the volatility shocks. This will be further investigated in the next section as it could be affected by futures pricing regime shifts.

In principle, the dynamic specification of our model might introduce distortive effects in the estimation of the optimal hedge ratio β^*, which reduce its effectiveness. We have thus performed the standard comparison of its hedging performance with the performance of a naive portfolio hedge ratio ($\beta = 1$) and of an OLS hedge ratio, obtained from the futures return coefficient estimate of a regression of cash returns on a constant and on futures returns. An artificial daily portfolio is introduced where an investor is assumed to buy (sell) one unit of the cash asset and to sell (buy) β units of the corresponding futures contract. The unconditional portfolio return standard deviations are computed over the whole sample and are set forth in Footnote 9.[9] The naive hedge portfolio is clearly outperformed by the optimal hedge portfolios. Moreover, our CCC-TGARCH model provides the minimum risk hedge, which supports the validity of our parameterization.

5 Hedging, Speculation, and Futures Pricing Regime Shifts

Sarno and Valente (2000) and Alizadeh and Nomikos (2004) analyzed the changes in the relationship between futures and spot stock index returns using a Markov switching model set out originally by Hamilton (1994). This technique is used here in order to analyze the shifts over two regimes in hedging and speculative behavior.

Using the full sample estimates of the conditional second moments obtained in the previous section, Eq. (8′) is adapted to a two-state Markov switching framework in which the drivers of futures returns are assumed to switch between two different processes, dictated by the state of the market.[10]

[9]

CCC-TGARCH estimates		OLS estimates		Naive
Optimal hedge ratio β^*	St. dev. of the optimal hedge portfolio	Optimal hedge ratio β^*	St. dev. of the optimal hedge portfolio	St. dev. of the naive portfolio
0.75	0.016309	0.70	0.016416	0.018018

[10] In order to eliminate the potential errors-in-variables distortions due to the use of a two-step procedure, we follow Pagan (1984). We replace the conditional variances by the fitted value of a regression of the futures (cash) return conditional variance on a constant, on its own lagged values (up to two lags), on the lagged values (up to two lags) of the conditional variance of the cash (futures) returns and on the one period lagged cash rate of return. The estimated coefficients are consistent, whereas the corresponding standard errors may underestimate their true values. However, this potential bias does not affect the SPEC index, which is consistent.

Equation (8′) is thus rewritten as

$$r_{f,t} = e_{0s_t} - \left(b^H_{s_t}\middle/d^S_{s_t}\right)\left(h^2_{r_c,t-1} - h^2_{r_c r_f,t-1}/h^2_{r_f,t-1}\right) + \left(e^S_{s_t}\middle/d^S_{s_t}\right)h^2_{r_f,t-1} + u_{r_f,s_t t} \quad (10)$$

where $u_{r_f,s_t t} \sim N\left(0, \sigma^2_{s_t}\right)$, and the unobserved random variable s_t indicates the state of the market.

The value of the current regime s_t is assumed to depend on the state of the previous period only, s_{t-1}, and the transition probability $P\{s_t = j|s_{t-1} = i\} = p_{ij}$ gives the probability that state i will be followed by state j. In the two state case $p_{11} + p_{12} = 1$ and $p_{22} + p_{21} = 1$, and the corresponding transition matrix is given by

$$P = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix} \quad (11)$$

The joint probability of $r_{f,t}$ and s_t is then given by the product

$$p\left(r_{f,t}, s_t = j|Y_{t-1}, \psi\right) = f\left(r_{f,t}\middle|s_t = j; Y_{t-1}, \psi\right)P(s_t = j|Y_{t-1}, \psi) \quad j = 1, 2 \quad (12)$$

where Y_{t-1} is the information set that includes all past information on the population parameters and $\psi = \left(e_{0s_t}, \left(b^H_{s_t}\middle/d^S_{s_t}\right), \left(e^S_{s_t}\middle/d^S_{s_t}\right), \sigma^2_{s_t}\right)$ is the vector of parameters to be estimated. $f(.)$ is the density of $r_{f,t}$, conditional on the random variable s_t, and $P(.)$ is the conditional probability that s_t will take the value j.

Following Hamilton (1994, Chap. 22), the density distribution of $r_{f,t}$ for the two-state case is

$$g\left(r_{f,t}\middle|Y_{t-1}, \psi\right) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{\frac{-u^2_{r_f,1t}}{2\sigma_1^2}\right\} + \frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left\{\frac{-u^2_{r_f,2t}}{2\sigma_2^2}\right\} \quad (13)$$

where $u_{r_f,s_t t}$ is the residual of Eq. (10). If the unobserved state variable s_t is i.i.d. maximum likelihood estimates of the parameters in ψ are obtained maximizing the following log likelihood function with respect to the unknown parameters

$$\mathrm{L}(\psi) = \sum_{t=1}^{T} \log g\left(r_{f,t}\middle|Y_{t-1}, \psi\right) \quad (14)$$

where T is the total number of sample observations.

In this paper the identification process of the nature of the regimes, essential for the interpretation of a Markov switching model, relies on the estimates of Eq. (10) and on the analysis of the behavior over time of the state probabilities. Table 3 sets out the estimates of Eq. (10). The quality of the fit is highly satisfactory; the relevant coefficients change across regimes and are significantly different from zero.

Table 3 Markov switching regime estimates of Eq. (10)

	Oil	
	$s_t = 1$	$s_t = 2$
p_{s_t,not,s_c}	0.009 (6.17)	0.074 (7.52)
e_{0s_t}	−0.000 (−0.14)	−0.004 (−2.04)
$\left(b^H_{s_t} / d^S_{s_t}\right)$	2.783 (2.73)	5.505 (3.50)
$\left(e^S_{s_t} / d^S_{s_t}\right)$	2.632 (6.19)	2.356 (4.05)
$\sigma^2_{s_t}$	0.018 (96.27)	0.053 (61.36)
Number of days in regime s_t [a]	110	13
SPEC	0.583	1.699
Optimal hedge ratio β^*	0.746	0.006
LLF	12637.924	

Note [a] average expected duration of being in state s_t

The regime (state) 2 variance is three times larger than that of regime (state) 1. The probability of switching from a low variance to a high variance state p_{12} is much lower than the probability of switching from a high variance to a low variance state p_{21}. Indeed the transition probabilities are $p_{12} = 0.009$ and $p_{21} = 0.074$ and indicate that the average expected duration of being in state 1 is close to 110 working days (about 5 months) and the average expected duration of being in state 2 is of 13 working days.[11]

A relevant difference in hedging and speculation can be easily detected. A risk averse speculative behavior in state 1 is maintained despite the increase in volatility across regimes; speculators only slightly decrease their demand for futures contracts whenever the volatility rises. What really stands out is the change in hedgers' behavior; they increase their demand in state 2 and switch to a speculative attitude (the SPEC index trebles from 0.58 to 1.70 and β^* collapses from 0.746 to 0.006 because of a sharp drop of the covariance between cash and futures returns).[12]

The upper graph of Fig. 1 presents the behavior over the sample of the time t probability that the market is in regime 1. The lower graph sets out the rate of return of the corresponding futures contract. Visual inspection suggests that regime 1 may be associated with periods in which return variability is low (and thus regime 2 with

[11] The average expected duration of being in state 1 is computed according to Hamilton (1989) as $\sum_{i=1}^{\infty} i p_{11}^{i-1}(1-p_{11}) = (1-p_{11})^{-1} = (p_{12})^{-1}$. If we posit that regime 1 (2) at time t holds if the probability of being in state 1 based on data through t is larger (smaller) than 0.5, the oil market is in the low volatility regime 1 for 4,777 days and in the high volatility regime 2 for 466 days.

[12] We detect therefore an overall increase in the share of agents that follow a speculative rationale and a corresponding decrease in the number of standard risk minimizing investors.

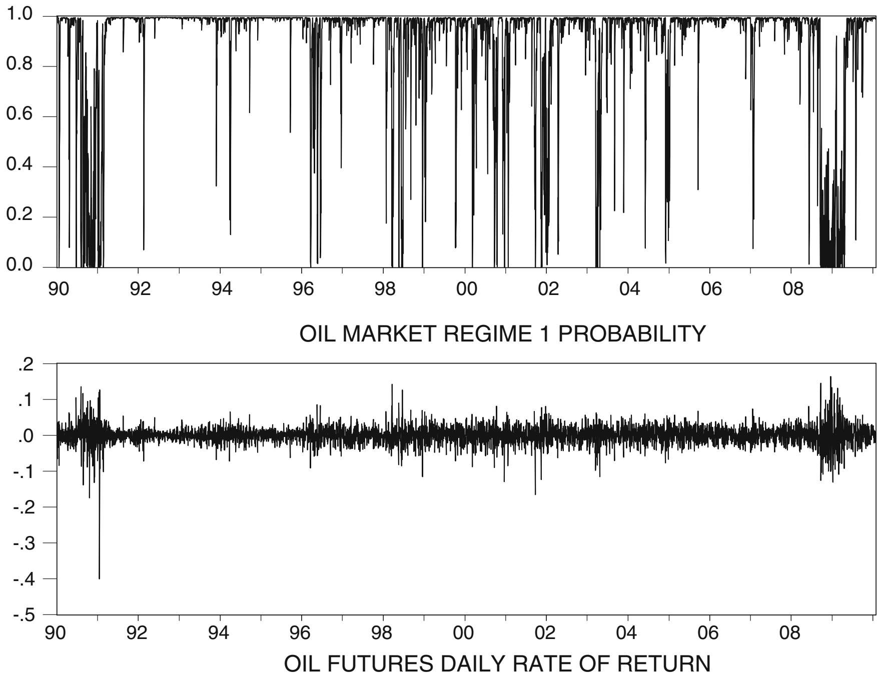


Fig. 1 Oil market regime 1 probability

periods in which it is high).[13] Since regime 2 essentially corresponds to the turbulence associated with the Gulf war of 1990 and with the oil price turmoil of 2008, our empirical evidence is in line with the interpretation of the recent price dynamics provided by Tokic (2011), who maintains that the institutional investors' recent inroads in the oil market, motivated by a desire to diversify their portfolios and/or hedge inflation, destabilized the interaction between commercial participants and liquidity-providing speculators. More precisely, in periods of severe price turbulence, such as the 2008 price upswing, well informed commercial hedgers lose their informational advantage and are misled by unexplained price shifts into significantly reducing their short positions, thus engaging in positive feedback trading.

[13] According to the standard ADF unit root tests the time t regime 1 probability time series is I(0). The correlation coefficients between the regime 1 probability and the daily oil futures' rate of return and standard deviation are, respectively, 0.035 (2.52) and −0.756 (−83.68), where the t-ratios are in parentheses. Regime 1 is to be associated with both low futures return variability and, to a lesser extent, with positive futures price rates of change (i.e. possibly with a bullish market), and regime 2 with high return variability and negative futures price rates of change (i.e. with a bearish market).

6 Conclusions

This paper has examined the dynamic behavior of futures returns in the oil market. The interaction between hedgers and speculators is modelled using a highly non linear parameterization where hedgers react to deviations from the minimum variance of the hedged portfolio, and speculators respond to standard expected risk returns considerations. The relationship between expected spot and futures returns and time varying volatilities is estimated using a non linear in mean CCC-TGARCH approach. The results point to the suitability of this choice because of the quality of the fit and of the sound meaning of the parameter estimates. In spite of the growing role of speculation, over the 1990–2010 sample period, hedgers play a dominant role since futures returns dynamics are mostly associated with the variability of the hedged portfolio, especially in the frequent low volatility periods. We account for the impact of financial integration of the commodity markets by allowing the demand of futures to be dependent upon the "state of the market" via a Markov regime switching approach. Both visual inspection and correlation analysis suggest that regime 1 is associated with periods in which return variability is low and regime 2 with periods in which it is high. Optimal hedging ratios computed in each state are smaller in the high volatility regime. The differences across regimes in hedging and speculative behavior are distinctive. The impact on futures returns of the ratio of speculative to hedging drivers seems to be strong, when market volatility is high.

References

Achen C (1982) Interpreting and using regression. Sage Publications, Thousand Oaks
Ai C, Chatrath A, Song F (2006) On the comovement of commodity prices. Am J Agric Econ 88:574–588
Alizadeh A, Nomikos N (2004) A Markov regime switching approach for hedging stock indices. J Futures Markets 24:649–674
Bessembinder H (1992) Systematic risk, hedging pressure and risk premiums in futures markets. Rev Finan Stud 5:637–667
Bollerslev T, Wooldridge JM (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometr Rev 11:143–172
Cashin P, McDermott CJ, Scott A (1999) The myth of comoving commodity prices. IMF working paper 169
Cecchetti SG, Cumby RE, Figlewski S (1988) Estimation of the optimal futures hedge. Rev Econ Stat 70:623–630
Chan W, Young D (2006) Jumping hedges: an examination of movements in copper spot and futures markets. J Futures Markets 26:169–188
Chang B (1985) Returns to speculators and the theory of normal backwardation. J Finan 4:193–208
Engle RF, Ng VK (1993) Measuring and testing the impact of news on volatility. J Finan 48:1749–1778
Fagan S, Gencay R (2008) Liquidity-induced dynamics in futures markets. MPRA working paper 6677 January
Figlewski S (1984) Hedging performance and basis risk in stock index futures. J Finan 39:657–669

Figuerola-Ferretti I, Gonzalo J (2010) Modelling and measuring price discovery in commodity markets. J Econometr 158:95–107
Garbade KD, Silber WL (1983) Price movements and price discovery in futures and cash markets. Rev Econ Stat 65:289–297
Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57:357–384
Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton
Irwin SH, Sanders DR, Merrin RP (2009) Devil or angel? The role of speculation in the recent commodity price boom (and bust). J Agric Appl Econ 41:377–391
Johnson LL (1960) The theory of hedging and speculation in commodity futures. Rev Econ Stud 27:139–151
Kolb RW, Okunev J (1992) An empirical evaluation of the extended Mean-Gini coefficient for futures hedging. J Futures Markets 12:177–186
Kroner KF, Sultan J (1993) Time-varying distribution and dynamic hedging with foreign currency futures. J Finan Quant Anal 28:535–551
Lee TH, Yoder J (2007) Optimal hedging with a regime-switching time-varying correlation Garch model. J Futures Markets 27:495–516
Le Pen Y, Sévi B (2010) Revisiting the excess co-movements of commodity prices in a data-rich environment. Memo Université de Nantes, Nantes
Lescaroux F (2009) On the excess co-movement of commodity prices—a note about the role of fundamental factors in short-run dynamics. Energy Policy 37:3906–3913
Lien D, Tse YK (2000) Hedging downside risk with future contracts. Appl Finan Econ 10:163–170
McKinnon RL (1967) Futures markets, buffer stocks and income stability for primary producers. J Polit Econ 73:844–861
Pagan A (1984) Econometric issues in the analysis of regressions with generated regressors. Int Econ Rev 25:221–247
Pyndick RS, Rotemberg JJ (1990) The excess co-movement of commodity prices. Econ J 100:1173–1189
Sarno L, Valente G (2000) The cost of carry model and regime shifts in stock index futures markets: an empirical investigation. J Futures Markets 20:603–624
Stein JL (1961) The simultaneous determination of spot and futures prices. Am Econ Rev 51:1012–1025
Stulz RM (1996) Rethinking risk management. J Appl Corp Finan 9:8–24
Tokic D (2011) Rational destabilizing speculation, positive feedback trading, and the oil bubble of 2008. Energy Policy 39:2051–2061
Ward RW, Fletcher LB (1971) From hedging to pure speculation: a micro model of futures and cash market positions. Am J Agric Econ 53:71–78

Evaluating the Empirical Performance of Alternative Econometric Models for Oil Price Forecasting

Andrea Bastianin, Matteo Manera, Anil Markandya and Elisa Scarpa

Abstract The empirical literature is very far from any consensus about the appropriate model for oil price forecasting. Several specifications have been proposed: some concentrate on the relationship between spot and futures prices ("financial" models), while others assign a key role to economic fundamentals ("structural" models). In this work we systematically test and evaluate the ability of several alternative econometric specifications to capture the dynamics of oil prices. Moreover, we propose a new class of models which combines the relevant aspects of financial and structural specifications ("mixed" models). We evaluate the forecasting performance of each class of models using different measures of forecast accuracy. We also analyse the effects of different data frequencies on the coefficient estimates and forecasts of each selected specification. Our empirical findings suggest that financial models are to be preferred to time series models. Both financial and time series models are better than mixed and structural models. Although the random walk model is not statistically outperformed by any of the alternative models, our empirical results suggest that theoretically well-grounded financial models are valid instruments for producing accurate forecasts of the WTI spot price.

Keywords Oil price · Forecasting · Econometric models · Forecast evaluation

A. Bastianin
Department of Economics, Management, and Quantitative Methods, University of Milan, Milan, Italy

A. Bastianin · M. Manera
Fondazione Eni Enrico Mattei, Milan, Italy

M. Manera (✉)
Department of Economics, Management and Statistics,
University of Milan-Bicocca, Via Bicocca degli Arcimboldi, 8 - Building U7,
20126 Milan, Italy
e-mail: matteo.manera@unimib.it

A. Markandya
BC3 Basque Centre for Climate Change, Bilbao, Spain

E. Scarpa
Edison Trading, Milan, Italy

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_7

1 Introduction

The relevance of oil in the world economy is undisputable. The world oil production in 2009 amounted to 82,165 thousand barrels per day (tbd). OPEC countries produced 33,363 tbd (40.6 % of the world oil production) in 2009, while OECD countries and Europe (25 countries) were responsible of 19,427 tbd (23.6 %) and 2,187 tbd (2.7 %), respectively. In January 2010 world oil stocks were estimated at 1,191,066 million barrels. If OPEC countries alone hold 80.2 % of world oil reserves, OECD and European countries can directly count only on 7 and 0.8 %, respectively. Moreover, world oil consumption in 2009 was measured in 85,006 tbd, 59.6 % of which originates from the OECD countries (Eni 2010). The impact of oil on the financial markets is at least equally important. The NYMEX average daily open interest volume (OIV)[1] on oil futures and options contracts, which was equal to 634,549 contracts during the period 2002–2005, increased to 1,255,986 contracts during the period 2006–2010 (Commodity Futures Trading Commission 2010).

Moreover, the peculiar nature of oil price dynamics has attracted the attention of many researchers in recent years. As an example, in Fig. 1 we report the behaviour of the WTI spot price over the period January 1986–December 2005. From an inspection of this graph, it is easy to verify that both level and volatility of the WTI spot price are highly sensitive to specific economic and geo-political events. For instance, the small price fluctuations of the years 1986–1990 are the result of the OPEC's production quotas repeated adjustments. The 1990 sharp increase in WTI spot price is obviously due to the Gulf war. The remarkable price falls of the period 1997–1998 coincide with the pronounced slowdown of Asian economic growth. The reduction in OPEC's production quotas of 1999 has been followed immediately by a sharp price increase. Finally, if the price decreases in 2001 are related to terrorist attack of 11 September, the reduction of the WTI spot price levels recorded in the period 2002–2005 are again justified by falling OPEC production quotas and spare capacity.

The more recent evolution of the WTI spot price shows that forecasting the price of crude oil is very challenging. In August 2005 oil price has risen to over US$ 60 per barrel (pb), while one year later it has topped out at the record level of US$ 77.05 pb. Experts have again attributed the spike in oil price to a variety of economic and geo-political factors, including the North Korean crisis, the Israel-Lebanon conflict, the Iranian nuclear threat and the decline in US oil reserves. At the end of the summer 2006, the WTI oil price has begun to decrease and reached the level of US$ 56.82 pb on 20 October 2006. In the meantime, OPEC has announced production cuts to stop the sliding price. On 16 January 2007 prices have been even lower: US$ 51.21 pb for the WTI spot price and US$ 51.34 for the first position of the NYMEX oil futures contract.

[1] Open interest volume is measured as the sum of all long contracts (or, equivalently, as the sum of all short contracts) held by market participants at the end of a trading day. It is a proxy for the flow of money into the oil futures and options market.

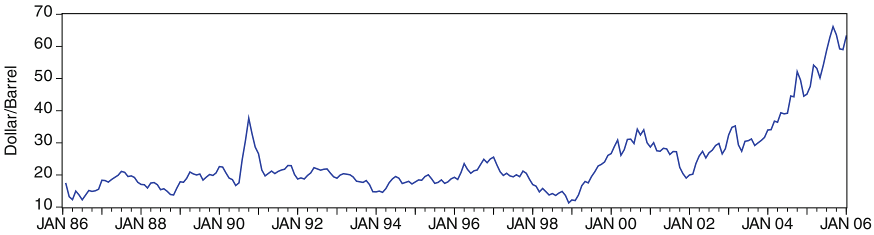


Fig. 1 WTI spot price for the period January 1986–December 2005 (monthly data)

Given the relevance of oil in the world economy and the peculiar characteristics of the oil price time series, it is not surprising that considerable effort has been devoted to the development of different types of econometric models for oil price forecasting.

Several specifications have been proposed in the economic literature. Some are based on financial theory and concentrate on the relationship between spot and futures prices ("financial" models). Others assign a key role to variables explaining the characteristics of the physical oil market ("structural" models). These two main groups of models have often been compared to standard time series models, such as the random walk and the first-order autoregressive model, which are simple and, differently from financial and structural models, do not rely on additional explanatory variables.

It should be noticed that many econometric models for oil price forecasting available in the literature are single-equation, linear reduced forms. Two recent noticeable exceptions are represented by Moshiri and Foroutan (2006) and Dees et al. (2007). The first study uses a single-equation, non-linear artificial neural network model to forecast daily crude oil futures prices over the period 4 April 1983–13 January 2003. The second contribution discusses a multiple-equation, linear model of the world oil market which specifies oil demand, oil supply for non-OPEC producers, as well as a price rule including market conditions and OPEC behaviour. The forecasting performance of this model is assessed on quarterly data over the period 1995–2000.

The empirical literature is very far from any consensus about the appropriate model for oil price forecasting that should be implemented. Findings vary across models, time periods and data frequencies. This study provides fresh new evidence to bear on the following key question: does a best performing model for oil price forecasting really exist, or aren't accurate oil price forecasts anything more than a mere illusion?

Relative to the previous literature, this work is novel in several respects. First of all, in this contribution we test and systematically evaluate the ability of several alternative econometric specifications proposed in the literature to capture the dynamics of oil prices. We have chosen to concentrate our investigation on single-equation and multiple-equations linear reduced forms, since models of this type are the most widely used in the literature and by the practitioners. In this respect, our study complements the empirical findings presented in Moshiri and Foroutan (2006), which are focused on the forecasting performance of a single non-linear model.

Second, this study analyses the effects of different data frequencies (daily, weekly, monthly and quarterly) on the coefficient estimates and forecasts obtained using each selected econometric specification. The factors which potentially affect the goodness of fit and forecasting performance of an econometric model are numerous, the most important being sample period and data frequency. The fact that no unanimous conclusions could be drawn by previous studies on the forecasting performance of similar models may depend upon, among other things, the particular data frequency used in each investigation.

Third, we compare different models at different data frequencies on a common sample and common data. For this purpose, we have constructed specific data sets which enable us to evaluate different types of econometric specifications involving different explanatory variables on the same sample period. Within our composite data base, the WTI spot oil price as well as the majority of the explanatory variables are recorded at different frequencies.

Fourth, we evaluate the forecasting performance of each selected model using one step-ahead forecasts, as well as different measures of forecast accuracy based on symmetric and asymmetric loss functions. At the same time, we present formal statistical procedures for comparing the predictive ability of the models estimated.

Lastly, we propose a new class of models, namely the mixed models, which combine the relevant aspects of the financial and structural specifications proposed in the literature.

The chapter is organized as follows. In Sect. 2 we briefly review the existing empirical literature related to oil price forecasting. Section 3 presents and describes the data collected for the empirical analysis. In Sect. 4 the empirical results obtained by forecasting oil prices with alternative econometric models are discussed. The performance of each model is analysed using different measures of forecasting ability and graphical evaluation "within" each class of models (i.e. financial, structural, time series and mixed models). Section 5 summarizes the forecasting performance of the alternative specifications, with particular emphasis on "between"-class analogies and differences. Some conclusions and directions for future research are presented in Sect. 5.

2 The Existing Literature on Oil Price Forecasting

The literature on oil price forecasting has focused on two main classes of linear, single-equation, reduced-form econometric models. The first group ("financial" models) includes models which are directly inspired by financial economic theory and based on the efficient market hypothesis (EMH), while models belonging to the second class ("structural" models) consider the effects of oil market agents and real variables on oil prices.[2] Both financial and structural models often use pure time series specifications for benchmarking.[3]

2.1 Financial Models

In general, financial models for oil price forecasting examine the relationship between the oil spot price at time t (S_t) and the oil futures price at time t with maturity T (F_t), analyzing, in particular, whether futures prices are unbiased and efficient predictors of spot prices. The reference model is:

$$S_{t+1} = \beta_0 + \beta_1 F_t + \varepsilon_{t+1} \tag{1}$$

where the joint null hypothesis of unbiasedness ($\beta_0 = 0$ and $\beta_1 = 1$) should not be rejected, and no autocorrelation should be found in the error terms (efficiency). A rejection of the joint null hypothesis on the coefficients β_0 and β_1 is usually rationalised by the literature in terms of the presence of a time-varying risk premium.

A sub-group of models, which are also based on financial theory but have been less investigated, exploits the following spot-futures price arbitrage relationship:

$$F_t = S_t e^{(r+\omega-\delta)(T-t)} \tag{2}$$

where r is the interest rate, ω is the cost of storage and δ is the convenience yield.[4]

Samii (1992) attempts at unifying the two approaches described in Eqs. (1) and (2) by introducing a model where the spot price is a function of the futures price and the interest rate. Using both daily (20 September 1991–15 July 1992) and monthly (January 1984–June 1992) data on WTI spot price and futures prices with 3- and 6-month maturity, he concludes that the role played by the interest rate is unclear and that, although the correlation between spot and futures prices is very high, it is not possible to identify which is the driving variable.

---

[2] As pointed out in the Introduction and at the beginning of Sect. 2, the models analysed in this study are linear, single-equation, reduced-forms. In this context, we use the term "structural model" to identify a specification whose explanatory variables capture the real and strategic (as opposed to financial) aspects of the oil market.

[3] Interesting exceptions are Pindyck (1999) and Radchenko (2005), who propose alternative forecasting models in a pure time series framework. See Sect. 2.2 for details.

[4] The arbitrage relationship (2) means that the futures price must be equal to the cost of financing the purchase of the spot asset today and holding it until the futures maturity date (which includes the borrowing cost for the initial purchase, or interest rate, and any storage cost), once the continuous dividend yield paid out by the underlying asset (i.e. the convenience yield) has been taken into account. See, among others, Clewlow and Strickland (2000) and Geman (2005) for details on the arbitrage relationship (2) for energy commodities.

An overall comparison of financial and time series models is offered by Zeng and Swanson (1998), who evaluate the in-sample and out-of-sample performance of several specifications. The authors use a daily dataset over the period 4 January 1990–31 October 1991 and specify a random walk, an autoregressive model and two alternative Error Correction models (ECM, see Engle and Granger 1987), each with a different definition of long-run equilibrium. The deviation from the equilibrium level which characterizes the first ECM is equal to the difference between the futures price tomorrow and the futures price today, i.e. the so-called "price spread". In the second ECM, the error correction term recalls the relationship between spot and futures prices, which involves the cost of storage and the convenience yield, as reported in Eq. (2). The predictive performance of each model is evaluated using several formal and informal criteria. The empirical evidence shows that the ECM specifications outperform the others. In particular, the ECM based on the cost-of-storage theory performs better than the ECM which specifies the error correction term as the spot-futures price spread.

Bopp and Lady (1991) investigate the performance of lagged futures and spot oil prices as explanatory variables in forecasting the oil spot price. Using monthly data on spot and futures prices for heating oil during the period December 1980–October 1988, they find empirical support to the cost-of-storage theory.[5] The authors also compare a random walk against the reference financial model. In this case, the empirical evidence suggests that both models perform equally well.

Serletis (1991) analyses daily data on 1-month futures price (as a proxy for the spot price) and 2-month futures price (quoted at NYMEX) for heating oil, unleaded gasoline and crude oil, relative to the period 1 July 1983–31 August 1988 (the time series of gasoline starts on 14 March 1985). He argues that the presence of a time-varying premium worsens the forecasting ability of futures prices.

In the empirical literature on oil prices there is no unanimous consensus about the validity of EMH. For instance, Green and Mork (1991) offer evidence against the validity of unbiasedness and EMH, analysing monthly prices on Mideast Light and African Light/North Sea crude oils over the period 1978–1985. Nevertheless, the authors notice that, if the subsample 1981–1985 is considered, EMH is supported by the data, because of the different market conditions characterizing the two time periods.

The unreliability of unbiasedness and EMH is also pointed out by Moosa and Al-Loughani (1994), who analyse WTI monthly data covering the period January 1986–July 1990. The authors exploit cointegration between the series on spot price and 3- and 6-month futures contracts using an ECM, and show that futures prices are neither unbiased nor efficient. Moosa and Al-Loughani apply a GARCH-in-mean model to take into account the time-varying structure of the risk premium.

[5] Two different spot prices are considered, namely the national average price reported by the Energy Information Administration (EIA) in the Monthly Energy Review, and the New York Harbour ex-shore price, while the futures contract is quoted at NYMEX.

Gulen (1998) asserts the validity of EMH by introducing the posted oil price as an additional explanatory variable in the econometric specification. In particular, using monthly data on WTI (spot price and 1-, 3- and 6-month futures prices) for the period March 1983–October 1995, he verifies the explanatory power of the posted price by using both futures and posted prices as independent variables. Empirical evidence from this study suggests that futures prices outperform the posted price, although the latter has some predictive content in the short horizon.

Morana's analysis (2001), based on daily data from 2 November 1982 to 21 January 1999, confirms that the Brent forward price can be an unbiased predictor of the future spot price, but in more than 50 % of the cases the sign of the changes in oil price cannot be accurately predicted. He compares a financial model with a random walk specification and shows that, when considering a short horizon, both specifications are biased.

Chernenko et al. (2004) test the EMH by focusing on the price spread relationship:

$$S_{t+T} - S_t = \beta_0 + \beta_1 (F_t - S_t) + \varepsilon_{t+1} \tag{3}$$

Analysing monthly data on WTI for the period April 1989–December 2003, the authors compare model (3) with a random walk specification and find that the empirical performance of the two models is very similar, confirming the validity of EMH.

The same model (3) is tested by Chinn et al. (2005) with a monthly dataset on WTI spot price and 3-, 6- and 12-month futures prices covering the period January 1999–October 2004. The empirical findings are, in this case, supportive of unbiasedness and EMH.

Another interesting application of financial models to the oil spot-futures price relationship is proposed by Abosedra (2005), who compares the forecasting ability of the futures price in model (3) with a naïve forecast of the spot price. Specifically, assuming that the WTI spot price can be approximated by a random walk with no drift, he forecasts the daily 1-month-ahead price using the previous trading day's spot price and constructs the naïve monthly predictor as a simple average of the daily forecasts. Using data for the period January 1991–December 2001, he finds that both the futures price and the naïve forecast are unbiased and efficient predictors for the spot price. The investigation of the relationship between the forecast errors of the two predictors allows the author to conclude that the futures price is a semi-strongly efficient predictor, i.e. the forecast error of the futures price cannot be improved by any information embedded in the naïve forecast.

2.2 Structural Models

Structural models, that is models based on economic fundamentals, emphasise the importance of explanatory variables describing the peculiar characteristics of the oil market. Some examples are offered by variables which are strategic for the oil market (e.g. industrial and government oil inventory levels), "real" variables (e.g. oil consumption and production), and variables accounting for the role played by OPEC in the international oil market.

Kaufmann (1995) models the real import price of oil using as structural explanatory variables the world oil demand, the level of OECD oil stocks, OPEC productive capacity, as well as OPEC and US capacity utilisation (defined as the ratio between oil production and productive capacity). The author also accounts for the strategic behaviour of OPEC and the 1974 oil shock with specific dummy variables. His analysis exploits an annual dataset for the period 1954–1989. Regression results show that his specification is successful in capturing oil price variations between 1956 and 1989, that is the coefficients of the structural variables are significant and the model explains a high percentage of the oil price changes within the sample period.

More recently, Kaufmann (2004) and Dees et al. (2007) specify a different forecasting model on a quarterly dataset. In particular, the first paper refers to the period 1986–2000, while the second contribution considers the sample 1984–2002. In these studies the authors pay particular attention to OPEC behaviour, using as structural regressors the OPEC quota (defined as the quantity of oil to be produced by OPEC members), OPEC overproduction (i.e. the quantity of oil produced which exceeds the OPEC quota), capacity utilisation and the ratio between OECD oil stocks and OECD oil demand. Using an ECM, the authors show that OPEC is able to influence real oil prices, while their econometric specification is able to produce accurate in-sample static and dynamic forecasts.

A number of authors introduce the role of the relative oil inventory level (defined as the deviation of oil inventories from their normal level) as an additional determinant of oil prices, for this variable is supposed to summarize the link between oil demand and production. In general, two kinds of oil stocks can be considered, namely industrial and governmental. The relative level of industrial oil stocks (*RIS*) is calculated as the difference between the actual level (*IS*) and the normal level of industrial oil stocks (*IS**), the latter corresponding to the industrial oil inventories de-seasonalised and de-trended. Since the government oil stocks tend to be constant in the short-run, the relative level of government oil stocks (*RGS*) can be obtained by simply removing the trend component.

Ye et al. (2002, 2005, 2007) develop three different models based on the oil relative inventory level to forecast the WTI spot price. In their 2002 paper, the authors build up a model on a monthly dataset for the period January 1992–February 2001, where oil prices are explained in terms of the relative industrial oil stocks level and of a variable describing an oil stock level lower than normal. Ye et al. (2005) present a basic monthly model of WTI spot prices which uses, as explanatory variables, three lags of the relative industrial oil stock level, the lagged dependent variable, a set of dummies accounting for the terrorist attack of 11 September 2001 (*D01*) and a "leverage" (i.e. step) dummy equal to one from 1999 onwards (*S99*) and zero before 1999, aimed at picking a structural change of the

OPEC behaviour in the oil market.[6] The authors compare this specification with: (i) an autoregressive model which includes AR(1) and AR(12) terms and dummies *D01* and *S99*; (ii) a structural model where the oil spot price is a function of the 1-month lag of the industrial oil inventories, the deviation of industrial oil stocks from the previous year's level, the 1-month lag of the oil spot price, as well as the dummy variables *D*01 and *S*99. Each model is estimated over the period 1992–2003. The basic model outperforms the other two specifications: in particular, the time series model is unable to capture oil price variability. The performance of each model is evaluated by calculating out-of-sample forecasts for the period 2000–2003. The forecasting accuracy of the two structural models depends on the presence of oil price troughs and peaks within the sample period. When considering 3-month-ahead forecasts, the basic model exhibits a higher forecasting performance in presence of oil price peaks, while the second structural specification outperforms the basic model in presence of oil price troughs. On the basis of this last evidence, Ye et al. (2007), using the same dataset, take into account the asymmetric transmission of oil stock changes to oil prices. The authors define a low (*LIS*) and a high (*HIS*) relative industrial oil stock level as follows:

$$\begin{cases} LIS_t = RIS_t + \sigma_{IS} & \text{if} \quad RIS_t < -\sigma_{IS} \\ LIS_t = 0 & \text{otherwise} \end{cases}$$
$$\begin{cases} HIS_t = RIS_t - \sigma_{IS} & \text{if} \quad RIS_t < \sigma_{IS} \\ HIS_t = 0 & \text{otherwise} \end{cases} \tag{4}$$

where σ_{IS} indicates the standard deviation of the industrial oil stock level.

The estimated model is:

$$S_t = \alpha_0 + \alpha_1 S_{t-1} + \sum_{j=0}^{5} \psi_j D01_{jt} + \lambda S99_t + \sum_{i=0}^{k} \beta_i RIS_{t-i} + \sum_{i=0}^{k} \left(\gamma_i LIS_{t-i} + \delta_i LIS^2_{t-i}\right)$$
$$+ \sum_{i=0}^{k} \left(\phi_i HIS_{t-i} + \varphi_i HIS^2_{t-i}\right) + \varepsilon_t \tag{5}$$

which shows a more accurate forecasting performance than the linear specification proposed by Ye et al. (2005).

[6] The oil price increases, characterizing the 90s, came to a rapid end in 1997 and 1998 when the impact of the economic crisis in Asia was either ignored or severely underestimated by OPEC who increased its quota by 10 % January 1, 1998. The combination of lower consumption and higher OPEC production sent prices into a downward spiral. In response, OPEC cut quotas by 1.25 million b/d in April and another 1.335 million in July. Price continued down through December 1998. Prices began to recover in early 1999 and OPEC reduced production another 1.719 million barrels in April. Not all of the quotas were observed but between early 1998 and the middle of 1999 OPEC production dropped by about 3 million barrels per day and was sufficient to move prices above $25 per barrel.

Following Ye et al. (2002), Merino and Ortiz (2005) specify an ECM with the percentage of relative industrial oil stocks and "speculation" (defined as the log-run positions held by non-commercials of oil, gasoline and heating oil in the NYMEX futures market) as explanatory variables. Evidence from January 1992 to June 2004 demonstrates that speculation can significantly improve the inventory model proposed by Ye et al., especially in the last part of the sample.

Zamani (2004) proposes a forecasting model based on a quarterly dataset for the period 1988–2004 and specifies an ECM with the following independent variables: OPEC quota, OPEC overproduction, *RIS*, *RGS*, non-OECD oil demand and a dummy for the last two quarters of 1990, which accounts for the Iraq war. The accuracy of the in-sample dynamic forecasts is indicative of the model's capability of capturing the oil price evolution.

In the pure time series framework, two models, which are particularly useful for forecasting oil prices in the long-run, are proposed by Pindyck (1999) and Radchenko (2005). The data used by the authors cover the period 1870–1996 and refer to nominal oil prices deflated by wholesale prices expressed in US dollars (base year is 1967). Pindyck (1999) specifies the following model:

$$\begin{cases} S_t = \rho S_{t-1} + (\beta_1 + \phi_{1t}) + (\beta_2 + \phi_{2t})t + \beta_3 t^2 + \varepsilon_t \\ \phi_{1t} = \alpha_1 \phi_{1,t-1} + \upsilon_{1t} \\ \phi_{2t} = \alpha_2 \phi_{2,t-1} + \upsilon_{2t} \end{cases} \tag{6}$$

where $\phi_{1t}$and ϕ_{2t} are unobservable state variables. He estimates the model with a Kalman filter and compares its forecasting ability with the following specification:

$$S_t = \rho S_{t-1} + \beta_1 + \beta_2 t + \beta_3 t^2 + \varepsilon_t \tag{7}$$

on the full dataset and three sub-samples, namely 1870–1970, 1970–1980 and 1870–1981. Model (6) offers a better explanation of the fluctuations of oil prices, while specification (7) produces more accurate forecasts.

Radchenko (2005) extends Pindyck's model, allowing the error terms to follow an autoregressive process:

$$\begin{cases} S_t = \rho S_{t-1} + \beta_1 + \phi_{1t} + \phi_{2t} t + \varepsilon_t \\ \phi_{1t} = \alpha_1 \phi_{1,t-1} + \upsilon_{1t} \\ \phi_{2t} = \alpha_2 \phi_{2,t-1} + \upsilon_{2t} \\ \varepsilon_t = \varphi \varepsilon_{t-1} + u_t \end{cases} \tag{8}$$

The forecasting horizons are 1986–2011, 1981–2011, 1976–2011 and 1971–2011. Overall, the empirical findings confirm Pindyck's results, although the model is unable to account for OPEC behaviour, leading to unreasonable price declines. Nevertheless, the author suggests that forecasting results can be improved significantly by combining specification (8) with a random walk and an autoregressive model, which can be considered a proxy for future OPEC behaviour.

3 Data and Methods

3.1 Data

We have constructed four different datasets, with the following frequencies: daily, weekly, monthly and quarterly. Prices refer to WTI crude oil spot price (S_t) and WTI crude oil futures prices contracts with 1-, 2-, 3- and 4-month maturity ($F_{t,1} - F_{t,4}$), as reported by EIA. Weekly, monthly and quarterly data have been obtained by aggregating daily observations with simple arithmetic means, taking into account that the futures contract rolls over on the third business day prior to the 25th calendar day of the month preceding the delivery month. The sample covers the period 2 January 1986–31 December 2005 (see Fig. 1).

Due to the limited availability of structural variables at high frequencies, the daily and weekly datasets include observations on the WTI prices only. Therefore, we have concentrated our analysis on financial and time series models at daily and weekly frequencies, whereas we have estimated the structural specifications using monthly and quarterly data.

The monthly dataset includes observations over the period January 1988–August 2005 for the following variables: OECD industrial crude oil stocks (RIS); oil demand in the OECD countries (*OD*); the world crude oil production (*WP*); the commodity price index (*PPI*), with June 1982 as basis. All variables are expressed in million barrels per day (mbd) and are obtained from EIA, with the single exception of *PPI*, which is from the Bureau of Labor Statistics.

The quarterly data range from the first quarter of 1993 to the third quarter of 2005 and refer to the following variables: total oil demand, computed (*TOTD*) as the sum of the OECD (*OOD*) and non-OECD (*NOOD*) oil demand, RIS, and the OPEC (*OP*) crude oil production.

Moreover, both the monthly and quarterly dataset include a variable labelled as NCLP, that is a measure of long position held by non-commercial derivative traders. Commercial and non-commercial are the labels currently used by the U.S. Commodity Futures Trading Commission (CFTC) to categorize traders. Commercial traders (commonly called hedgers) are futures market participants whose line of business is in the related cash market, while non-commercial traders (commonly called speculators) are participants whose main line of business is unrelated to the cash market. The complete list of the variables employed in the empirical analysis is summarized in Table 1.

Table 1 Complete list of variables used in the empirical analysis

Variable	Sample	Frequency	Source	Acronym
WTI spot price	2/1/1986–31/12/2005	D, W, M, Q	EIA	*S*
WTI futures price contract $i = 1, \ldots, 4$	2/1/1986–31/12/2005	D, W, M, Q	EIA	F_i
Non-commercial long positions	3/1995–8/2005 Q1/1995–Q42005	M, Q	CFTC	*NCLP*
OECD oil consumption	1/1988–8/2005	M	EIA	*OD*
OECD industrial oil stocks	1/1988–8/2005 Q1/1993–Q3/2005	M, Q	IEA	*RIS*
World oil production	1/1988–8/2005	M	EIA	*WP*
Commodity price index	1/1988–8/2005	M	BLS	*PPI*
OECD oil demand	Q1/1993–Q3/2005	Q	IEA	*OOD*
Non-OECD countries oil demand	Q1/1993–Q3/2005	Q	IEA	*NOOD*
Total oil demand	Q1/1993–Q3/2005	Q	Computed as: *OOD* + *NOOD*	*TOTD*
OPEC oil production	Q1/1993–Q3/2005	Q	EIA	*OP*

Notes D daily frequency; *W* weekly frequency; *M* monthly frequency; *Q* quarterly frequency; *Qi* ith quarter, $i = 1, 2, 3, 4$; *EIA* Energy Information Administration; *CFTC* U.S. Commodity Futures Commission; *BLS* Bureau of Labor Statistics; *IEA* International Energy Agency

3.2 Models

We have evaluated the forecasting performance of different econometric models available in the existing literature, which can be subsumed under the two main classes described in Sect. 2, that of financial and that of structural models. We also propose a new class of models which combine the relevant aspects of financial and structural models (i.e. mixed models), and are based on the assumption that the interaction between financial and macroeconomic variables can improve the accuracy of oil price forecasts. Financial, structural and mixed models are confronted with pure time series specifications. As already noted, due to data constraints, structural and mixed forecast are produced only with monthly and quarterly data.

Irrespective of the sampling frequency of the data, all variables, with the only exception of RIS, have been transformed into logarithms. We denote the logarithm of a variable with lower-case letters (i.e. $x_t = \log X_t$). Moreover, we use Δ to indicate the difference operator (i.e. $\Delta^k x_t = x_t - x_{t-k}$).

3.2.1 Time Series Models

When evaluating a set of competing forecasts it is important to define a benchmark model; in the case of the price of oil the Random Walk (RW) represents a natural choice:

$$s_t = s_{t-1} + e_t \tag{9}$$

where e_t is a white noise error. The *RW* model is also known as "no-change forecast", since it is assumed that the best predictor for the oil price tomorrow is the oil price today.

The second time series model we consider is also a *RW*, but in this case we add a drift term (RWD):

$$s_t = \delta + s_{t-1} + e_t \tag{10}$$

The strength of these models, that explicitly impose a unit root behaviour for s_t, is their simplicity in both the estimation and forecasting stages. Actually, while the RW model does not need to be estimated, the RWD requires just to compute the OLS estimate of the sample average of Δs_t. Finally, we note that the usefulness of random walk models as benchmarks stems from the fact that they often out-perform more complex alternatives (Zeng and Swanson 1998).

3.2.2 Financial Models

In Sect. 3 we have pointed out that, irrespective of the frequency considered, the WTI spot price and the four WTI futures prices involved in the empirical analysis are $I(1)$.[7] Moreover, the WTI spot price and each WTI futures price are cointegrated, that is there exists a stationary, long-run equilibrium relationship between the WTI spot price and the WTI futures price at different maturities. Interestingly, these statistical findings can be explained by standard economic theory and used to build a forecasting models for the spot price of oil. In particular, the cost-of-carry model posits that the futures price of storable commodities, such as crude oil, depends on the spot price as well as on the cost of holding the commodity until the delivery date. This cost, known as the cost-of-carry, includes both the storage and the opportunity costs of awaiting future delivery (see Pindyck 2001, for a survey). Assuming that investors can trade simultaneously in the spot and futures markets, we can write the (log) cost-of-carry model as:

$$f_{t,i} - s_t = d_t + Q_t \tag{11}$$

[7] The results of the unit root tests, which are available from the authors upon request, are omitted to save space.

where the term on the left-hand side is knows as the "basis", d_t is the (log) cost-of-carry and Q_t is an adjustment term accounting for the marking-to-market feature of futures markets. As shown by Brenner and Kroner (1995), if we are willing to assume that the log-spot price follows a random walk with drift and that investors are rational, we can use Eq. (11) to derive the set of financial models:

$$s_t = \alpha + \beta f_{t,i} + \varepsilon_t \tag{12}$$

where α subsumes the terms on the right-hand side of Eq. (12) and ε_t is an uncorrelated error term. Notice that we can derive a joint test of hypotheses; in fact testing if $(\alpha\ \beta)' = (0\ 1)'$ is both a test of the optimality of $f_{t,i}$ as a predictor for s_t and a test of EMH (i.e. if new information is immediately incorporated into spot prices, then, on average, the futures price should be equal to the spot price).

These considerations form the basis for deriving the operational versions of financial models which are used to produce a second set of forecasts. All these models exploit the cointegrating relation between spot and futures prices. We consider four bivariate Vector Error Correction Models (VECM), denoted as FUT1–FUT4, which exploit the information content of futures contracts with different maturities:

$$\Delta s_t = \beta_{0i} + \beta_{1i}\Delta s_{t-1} + \beta_{2i}\Delta f_{t-1,i} + \gamma_{si}\left(s_{t-1} - b_{0i} - b_{1i}f_{t-1,i} - b_{2i}t\right) + e_{t,i} \tag{13}$$

$$\Delta f_{t,i} = \alpha_{0i} + \alpha_{1i}\Delta f_{t-1,I} + a_2\Delta s_{t-1} + \gamma_{fi}\left(s_{t-1} - b_{0i} - b_{1i}f_{t-1,i} - b_{2i}t\right) + u_{t,i} \tag{14}$$

for $i = 1, \ldots, 4$.

The fifth financial model is a multivariate VECM and is denoted as FUT(1,4):

$$\Delta s_t = \beta_0 + \beta_1\Delta s_{t-1} + \Sigma_{i=1}^{4}\beta_{2i}\Delta f_{t-1,i} + \Sigma_{i=1}^{4}\gamma_{s,i}\left(s_{t-1} - b_{0i} - f_{t-1,i} - b_{2i}t\right) + e_{t,i} \tag{15}$$

$$\Delta f_{t,i} = \alpha_{0i} + \Sigma_{i=1}^{4}\alpha_{1i}\Delta f_{t-1,i} + \alpha_{2,i}\Delta s_{t-1} + \Sigma_{i=1}^{4}\gamma_{fi}\left(s_{t-1} - b_{0i} - f_{t-1,i} - b_{2i}t\right) + u_{t,i} \tag{16}$$

for $i = 1, \ldots, 4$.

There are two main differences between this specification and models FUT1–FUT4. First, FUT(1,4) jointly models the relation between the spot price and the term structure of futures. Second, we impose restrictions on the cointegrating parameters in order to treat futures as unbiased predictors of the spot price. Finally, we also consider a sixth financial model, namely AVG(1,4), which uses the sample average of futures prices $\bar{f}_t = (1/4)\Sigma_{i=1}^{4}f_{t,i}$. As in model (15)-(16), the intuition for taking the simple average is to exploit the information content of the term structure of future prices. The model can be written as models FUT1–FUT4, with $\bar{f}_t$ in place of $f_{t,i}$.

The lag order of all models has been selected according to well established information criteria, as well as a set of Lagrange Multiplier tests for residuals autocorrelation. Estimation and inference of VECMs is carried out following the Johansen's (1995) approach to vector cointegration.[8]

3.2.3 Structural and Mixed Models

Structural and mixed models have been estimated only for monthly and quarterly frequencies, due to the lack of data on the structural variables at higher frequencies.

For monthly data, we propose two different specifications. In the basic mixed model (MIX) the WTI spot price is regressed on the non-commercial long positions (*nclp*), OPEC consumption (*od*), the relative inventory industrial level (*RIS*), a step dummy for 1999 (*S99*), which accounts for a structural change of the OPEC's behaviour in the international oil market, and the world oil production (*wp*):

$$s_t = \alpha + \beta nlcp_t + \gamma od_t + \delta RIS_t + \lambda S99_t + \phi wp_t + \varepsilon_t \tag{17}$$

The structural specification (STR) considers as explanatory variables the relative oil inventory level (RIS), the commodity price index (*ppi*), the OECD oil demand (*od*), the step dummy *S*99 and a set of dummy variables capturing the effects of 11 September 2001 (*D*01):

$$s_t = \alpha + \beta RIS_t + \delta ppi_t + \varphi od_t + \lambda S99_t + \gamma D01_t + \varepsilon_t \tag{18}$$

On quarterly data we estimate the following two different types of models:

$$s_t = \alpha + \beta RIS_t + \gamma totd_t + \delta nclp_t + \varepsilon_t \tag{19}$$

$$s_t = \alpha + \beta RIS_t + \gamma totd_t + \delta op_t + \varepsilon_t \tag{20}$$

where $totd_t$ denotes oil demand and op_t is OPEC production. Specification (19) is a mixed model, model (20) is purely structural.

Although oil demand might be naturally thought as endogenous when used as explanatory variable for oil price, in our case endogeneity of oil demand is not a issue, for the previous models are estimated in VECM form. Moreover, it is worth pointing out that for monthly, as well as for quarterly, data seasonality in oil demand and industrial oil stocks has been removed by regressing oil demand and industrial oil stocks on a set of monthly dummies.

[8] The estimation results for all models, which have been omitted to save space, are available from the authors upon request.

3.3 Forecast Evaluation

The estimation period for time series and financial models runs from January 1986 up to December 2003, while the interval from January 2004 to December 2005 is used for forecast evaluation. Structural and mixed models have been estimated on the sample January 1993–December 2003, and monthly (quarterly) forecasts have been produced for the period January (first quarter, Q1) 2004–August (fourth quarter, Q4) 2005.

All models have been selected and estimated once on the estimation sample; then one-step ahead forecasts have been produced by keeping the estimated parameters fixed.

The number of observations used to evaluate the forecasting performance of different models is determined by the sampling frequency of the data: for daily, weekly, monthly and quarterly the number of predictions is 329, 123, 20 and 8, respectively.

Before discussing our forecast evaluation framework, it is worth introducing some notation. We use $h_{i,t}$ to denote forecast from model i, the corresponding forecast error is $u_{i,t}$ and $L_{i,t}(u_{i,t})$ is a loss function. If not needed, we drop both model and time subscripts.

Our forecast evaluation strategy relies on the family of flexible loss functions put forth by Elliott et al. (2005):

$$L(u;\rho,\phi) = [\phi + (1-2\phi)I(u<0)]\,|u|^{\rho} \tag{21}$$

where $I(.)$ is the indicator function. The shape of the loss function is determined by two parameters: $\rho > 0$ and $0 < \phi < 1$; the loss is asymmetric whenever $\phi \neq 0.5$. More precisely, over-forecasting is costlier than under-forecasting for $\phi < 0.5$; on the contrary, when $\phi > 0.5$ positive forecast errors (under-prediction) are more heavily weighted than negative forecast errors (over-prediction). As shown in Fig. 2, special cases of the loss include: the quad–quad loss for $\rho = 2$ and the lin–lin loss for $\rho = 1$. Moreover, we get the mean absolute error (MAE) loss for $\rho = 1$ and $\phi = 0.5$ and the mean square error (MSE) loss for $\rho = 2$ and $\phi = 0.5$.

When evaluating forecasts from different models we will focus on quaq–quad losses ($\rho = 2$) with three different values for the asymmetry parameter $\phi = (0.2, 0.5, 0.8)$.

The values chosen for the parameters of the loss function allow for a greater flexibility than the traditional model-ranking approach based on symmetric losses, such as the MSE. There are several reasons for considering a flexible loss function. First, given that the shape of the loss function often influences the ranking of models, an asymmetric flexible loss function allows to evaluate forecasts taking into account the degree of aversion of the decision maker with respect to under- and over-prediction. Second, in order to consistently evaluate the prediction ability of models, forecasts producers and users should have the same loss function. On the contrary, when the loss function of the forecaster does not coincide with that of the

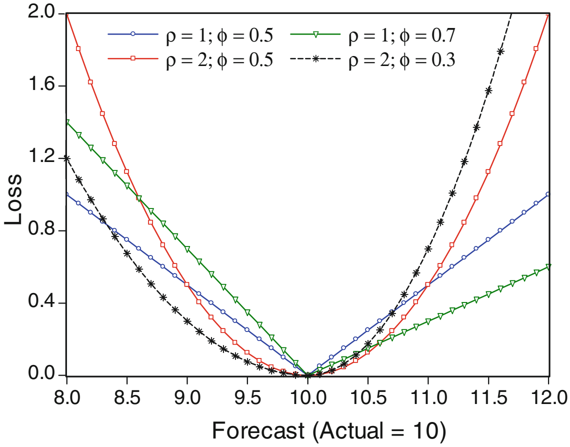


Fig. 2 Generalized loss function.
Notes The generalized loss function refers to Elliott et al. (2005); Forecasts are shown on the horizontal axis; The actual value is equal to 10; Over-prediction, $u < 0$, (under-prediction, $u > 0$) occurs to the *right* (*left*) of the actual value; The graph shows four different loss functions: the mean absolute error (MAE) loss for $\rho = 1$ and $\phi = 0.5$ (*circles*), the mean squared error (MSE) loss for $\rho = 2$ and $\phi = 0.5$ (*squares*), the asymmetric lin–lin (piecewise linear) loss for $\rho = 1$ and $\phi = 0.7$ (*triangles*), and the asymmetric quad–quad loss for $\rho = 2$ and $\phi = 0.3$ (*stars*); The function is defined for $\rho > 0$ and $0 < \phi < 1$; Over-prediction is costlier than under-prediction when $\phi < 0.5$

user, the optimality of the forecast can be judged only with respect to the producer's loss function. Therefore, unless the user knows the form of the forecaster's loss function, the evaluation of forecast optimality implies also a test of the functional form of the loss function (see Elliott et al. 2005, 2008). Third, there is evidence that loss functions of some decision makers are asymmetric (Elliott et al. 2005, 2008; Patton and Timmermann 2007). For instance, Auffhammer (2007) estimates the asymmetry parameter of the flexible loss function using the annual forecasts of the United States Energy Information Administration. In the case of the world price of oil, for both the lin–lin and quad–quad losses, the asymmetry parameter, ϕ, is very close one, suggesting that over-predictions are considered much less costly than under-predictions.

In this study, forecasts evaluation goes one step beyond that of a simple model ranking. As a matter of fact, in order to compare the forecast performance of each specification (at any sampling frequency and for any shape of the loss function), we run the test for equal predictive ability proposed by Diebold and Mariano (1995). The test statistic is based on the loss differential, $d_{iRW,t} = L_{i,t} - L_{RW,t}$, where the subscript attached to the second loss function indicates that the i-th model is evaluated against the random walk (RW). Under the null hypothesis, H_0: $E(d_{iRW})$, the Diebold-Mariano test statistic is asymptotically Gaussian. Given that the number of available forecasts produced by our models is, in at least two cases, insufficient in order to guarantee the validity of asymptotic results, we implement the Diebold and Mariano test corrected for small samples, where the appropriate p-values are computed using the moving block bootstrap of Künsch (1989).[9]

4 Empirical Results

We start the evaluation of forecasts with an heuristic model comparison based on the Approximate Bayesian Model Averaging (ABMA). ABMA is a method to combine forecasts that delivers a set of weights that are functions of the Schwarz Information Criterion (see Garratt et al. 2003).

Results are shown in Fig. 3. Irrespective of the sampling frequency of the data, the largest ABMA weights are always associated with models RW and RWD. While this finding is expected, given the parsimony of RW and RWD, nonetheless it is interesting to notice that, at daily and weekly sampling frequencies, ABMA would be essentially equivalent to assign equal weights to each model. Focusing on models for monthly and quarterly data (and keeping in mind the small size of the forecasting sample), we can confirm some of the previous results. In particular, the most heavily weighted models are, once again, RW (first), RWD (second) and AVG (1,4) (third), while the lowest (approximate) posterior probability is assigned to FUT(1,4). The success of the AVG(1,4) model is due to its ability to summarize the whole term structure of futures with two equations only. On the contrary, the multivariate FUT(1,4) model involves five equations and some coefficient restrictions that might not be supported by the data in the forecasting sample. As for the MIX and STR models, they appear on the bottom end of this ranking, with the sum of their weights not larger than that associated to the third best model, which in turn belongs to the financial class. In summary, our empirical results do not suggest a single winning option, however they clearly indicate the presence of a hierarchical order among the different classes of models, which can be summarized as: time series (first), financial (second), mixed (third), structural (fourth).

There are many ways to test for forecast optimality. One simple approach is to analyze the properties of forecast errors. In particular, it is well known that forecast errors from optimal forecasts should have zero mean. If forecast errors follow a Gaussian white noise process, as it should be for one-step ahead errors, then a standard t-test is the obvious diagnostic tool. However, due to the limited number of observations, we implement a finite-sample corrected t-test by relying on bootstrap standard errors and p-values obtained with the moving block bootstrap of Künsch (1989). Results are shown in Table 2, where the statistic OUR, which measures the incidence of over- and under-forecasts (i.e. an entry larger than unity suggests that the *i*-th model produces more negative forecast errors than positive forecast errors), is also presented.

[9] Details on this procedure and a small Monte Carlo study of its performance are available from the authors upon request.

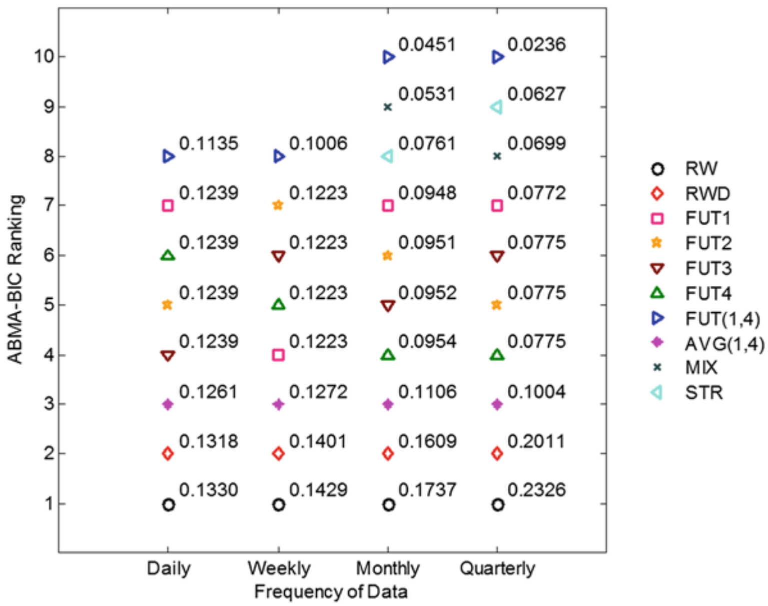


Fig. 3 Ranking of models using ABMA weights. *Notes* Models RW and RWD are described in Sect. 3.2.1 (Eqs. (9) and (10)); Models FUT1—FUT4 are described in Sect. 3.2.2 (Eqs. (13) and (14)); Models FUT(1,4) and AVG(1,4) are described in Sect. 3.2.2 (Eqs. (15) and (16)); Models MIX and STR are described in Sect. 3.2.3 (Eqs. (17—20))

None of the models for daily data presents a statistically significant bias. As for weekly forecasts, only the RW and FUT(1,4) models show a positive and statistically significant bias. Interestingly, for data sampled at weekly frequency all models produce more under-forecasts than over-forecasts; this result holds also for models that at daily frequency present a value of OUR > 1.

At monthly and quarterly frequency, OUR is always below unity, suggesting that all models tend to over-forecast. However, in both cases the class of financial models is the only producing unbiased forecasts and the one with OUR closer to unity (at least at monthly frequency). This finding can be explained by referring to the cost-of-carry model and its relationship with EMH. Comparing the size of biases at monthly frequency, we can compile the following model ranking: financial (first), structural (second), time series (third), mixed (fourth).

Figure 4 shows the rankings and the magnitude of the flexible loss functions associated to different models. In panel (a) the MSE ranking is reported. The set of points with the label “overall” on the x-axis represent the ranking of models obtained by summing the loss function over all forecast horizon. First, we can notice that the loss differential across models are not very large in magnitude, suggesting that it will be very hard to identify a best option. Second, when the performance of models across sampling frequencies is compared, we can see that the magnitude of the losses increases. Third, in the majority of cases bivariate

Table 2 Bias of forecast errors and ratio of over- to under-predictions

	Daily		Weekly		Monthly		Quarterly	
	Bias	Over/Under	Bias	Over/Under	Bias	Over/Under	Bias	Over/Under
RW	0.0526	0.8156	0.2852	0.6400	1.5572	0.5385	3.7256	0.1429
	(0.4259)		(0.0935)		(0.0510)		(0.0006)	
RWD	0.0448	0.8380	0.2631	0.6400	1.4313	0.6667	3.2794	0.3333
	(0.5049)		(0.1214)		(0.0723)		(0.0043)	
FUT1	−0.0549	1.0309	0.4225	0.6622	0.6692	0.8182	2.0701	0.3333
	(1.0000)		(0.0437)		(0.2939)		(0.0835)	
FUT2	−0.2264	1.3500	0.1667	0.6849	0.5635	0.8182	2.0883	0.3333
	(1.0000)		(0.4290)		(0.4049)		(0.0687)	
FUT3	−0.2132	1.3333	0.0451	0.7083	0.3434	0.8182	1.8374	0.3333
	(1.0000)		(0.8311)		(0.6182)		(0.0690)	
FUT4	−0.2057	1.3333	0.0230	0.7571	0.2068	0.8182	1.5554	0.3333
	(1.0000)		(0.9154)		(0.7581)		(0.1230)	
FUT (1,4)	−0.0412	1.0061	0.4469	0.5570	0.5353	0.8182	−0.1200	0.3333
	(1.0000)		(0.0318)		(0.4376)		(1.0000)	
AVG (1,4)	−0.2775	1.4191	−0.0183	0.7083	0.3776	0.8182	1.7585	0.3333
	(1.0000)		(1.0000)		(0.5783)		(0.0778)	
MIX					2.4991	0.5385	2.8809	0.1429
					(0.0030)		(0.0407)	
STR					1.0648	0.6667	3.4798	0.1429
					(0.0728)		(0.0014)	

Notes Even columns from 2 to 8 report the bias of the forecast errors; Bootstrap p-values in round brackets denote the probability of accepting the null hypothesis of a forecast bias equal to zero; Bootstrap p-values have been calculated on 9,999 moving block bootstrap samples; The length of blocks, b, is set according to the rule $b = \text{floor}(4(H/100)^{2/9})$; Odds columns from 3 to 9 show the relative occurrence of negative and positive forecast errors; An entry lower than one indicates that there are more positive forecast errors than negative forecast errors and that the model tends to under-forecast the spot price; An entry greater than one suggests that the model tends to over-forecast the spot price

financial models make in the first positions. The performance of structural and mixed model changes according to the sampling frequency of the data.

When the loss function becomes asymmetric (see panels (b) and (c)), the only models that have a good and consistent global performance are, once again, those belonging to the financial class. They are outperformed by time series models only when over-forecasting is costlier than under-forecasting. In this case there are

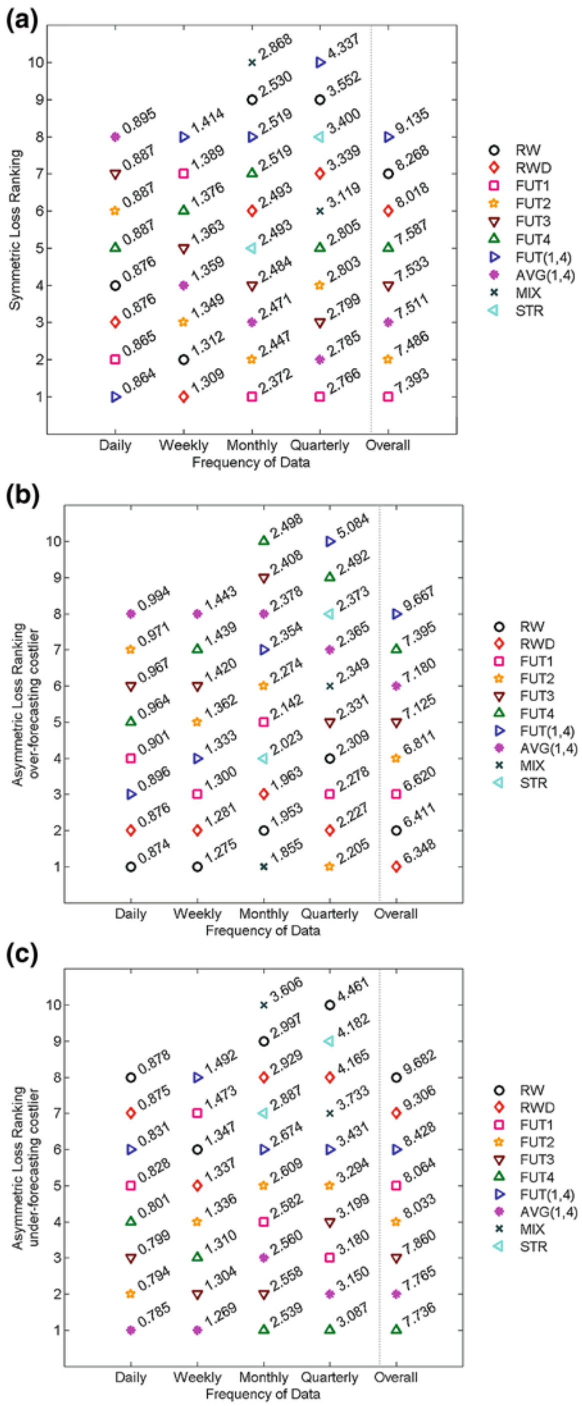


Fig. 4 Ranking of models using the generalized loss function. Notes See Notes of Fig. 3; Panel **a** reports the ranking based on MSE; Panel **b** reports the ranking based on the asymmetric loss function, under the assumption that over-forecasting is costlier; Panel **c** reports the ranking based on the asymmetric loss function, under the assumption that under-forecasting is costlier

Table 3 Diebold-Mariano test

	Daily			Weekly			Monthly			Quarterly		
	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.8$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.8$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.8$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.8$
RWD	6.9076	−0.5789	−9.1184	3.2281	−1.3613	−5.9011	0.4146	−1.8669	−3.3557	−1.2125	−3.1140	−3.6084
	(0.0000)	(1.0000)	(1.0000)	(0.0067)	(1.0000)	(1.0000)	(0.7097)	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)
FUT1	3.2063	−0.8291	−2.4350	0.6289	2.0764	2.6188	0.9616	−0.8743	−1.9359	−0.0660	−1.4913	−1.8125
	(0.0009)	(1.0000)	(1.0000)	(0.5533)	(0.0447)	(0.0093)	(0.3628)	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)
FUT2	7.7983	0.9841	−6.2729	2.8041	1.1876	−0.2357	1.2729	−0.4074	−1.8436	−0.2887	−1.8150	−2.1601
	(0.0000)	(0.3172)	(1.0000)	(0.0230)	(0.2739)	(1.0000)	(0.2592)	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)
FUT3	7.8468	1.0686	−6.8173	3.6598	1.7014	−1.0974	1.4250	−0.1844	−1.8110	0.0483	−1.6078	−2.1715
	(0.0000)	(0.2838)	(1.0000)	(0.0110)	(0.0887)	(1.0000)	(0.2136)	(1.0000)	(1.0000)	(0.9695)	(1.0000)	(1.0000)
FUT4	7.9135	1.0999	−6.9774	3.9062	2.0717	−0.9820	1.4828	−0.0406	−1.7293	0.3183	−1.4107	−2.0956
	(0.0000)	(0.2642)	(1.0000)	(0.0062)	(0.0411)	(1.0000)	(0.2129)	(1.0000)	(1.0000)	(0.7529)	(1.0000)	(1.0000)
FUT(1,4)	2.6237	−0.9382	−2.2516	1.4656	2.1452	2.2910	1.3884	−0.0434	−1.1652	1.0430	0.5027	−0.9792
	(0.0092)	(1.0000)	(1.0000)	(0.1554)	(0.0568)	(0.0331)	(0.2181)	(1.0000)	(1.0000)	(0.4128)	(0.5785)	(1.0000)
AVG(1,4)	8.1890	1.5558	−6.3827	3.8532	1.3745	−1.6460	1.3702	−0.2487	−1.7929	0.1141	−1.5665	−2.1426
	(0.0000)	(0.1218)	(1.0000)	(0.0040)	(0.1724)	(1.0000)	(0.2343)	(1.0000)	(1.0000)	(0.8943)	(1.0000)	(1.0000)
MIX							−0.3199	1.3912	2.3679	0.1125	−1.0136	−1.2795
							(1.0000)	(0.1931)	(0.0779)	(0.9016)	(1.0000)	(1.0000)
STR							0.3409	−0.2419	−0.6435	0.3186	−0.5600	−0.7745
							(0.7310)	(1.0000)	(1.0000)	(0.7513)	(1.0000)	(1.0000)

Notes Entries report the calculated Diebold and Mariano statistic; Bootstrap p-values in round barkets denote the probability of accepting the null hypothesis of a zero loss differential; Bootstrap p-values have been calculated on 9,999 moving block bootstrap samples; The length of blocks, b, is set according to the rule b = floor(4(H/100)2/9) interesting exceptions: the mixed model applied to monthly data delivers the lowest loss, while FUT2 is the best option in the case of quarterly data.

In summary, the ranking of models seems to suggest that, irrespective of the shape of the loss function, the class of financial models is to be preferred to time series models. Both financial and time series models are, in turn, better than mixed and structural models.

Finally, we use the Diebold and Mariano test to evaluate if the loss differentials of RWD, financial, structural and mixed models are not statistically significant when the RW model is used as a benchmark. Results reported in Table 3 are not conclusive, since the loss differential seems to be statistically insignificant in the large majority of cases. Although the RW model is not statistically outperformed by any of the alternative models, the empirical findings seem to suggest that theoretically well-grounded financial models are valid instruments for producing accurate forecasts of the WTI spot price.

5 Conclusions

In this paper, we have tested and systematically evaluated the ability of several alternative econometric specifications proposed in the literature to capture the dynamics of oil prices. We have concentrated our investigation on single- as well as multiple-equation, linear reduced forms, since models of this type are the most widely used in the academic literature and by the practitioners.

We have also analysed the effects of different data frequencies (daily, weekly, monthly and quarterly) on the coefficient estimates and forecasts obtained using each selected econometric specification. We have evaluated the forecasting performance of each selected model using static forecasts, as well as different measures of forecast errors.

Finally, we have proposed a new class of models, namely “mixed” models, which combine the relevant aspects of the financial and structural specifications proposed in the literature.

The empirical findings of this study can be summarized as follows. According to an heuristic model comparison based on the ABMA, a hierarchical order among the different classes of models can be found: time series (first), financial (second), mixed (third), structural (fourth). The finite-sample corrected t-test for the null hypothesis of zero-mean forecast errors, and the statistic OUR, show that none of the models for daily data presents a statistically significant bias. For data sampled at weekly frequency all models produce more under-forecasts than over-forecasts. At monthly and quarterly frequency, OUR is always below unity, suggesting that all models tend to over-forecast. However, in both cases the class of financial models is the only producing unbiased forecasts and the one with OUR closer to unity (at least at monthly frequency). Comparing the size of biases at monthly frequency, the following model ranking emerges: financial (first), structural (second), time series (third), mixed (fourth). The ranking of models seems to suggest that, irrespective of the shape of the loss function, the class of financial models is to be preferred to time series models. Both financial and time series models are, in turn, better than mixed and structural models. The Diebold and Mariano test is inconclusive, since the loss differentials seem to be statistically insignificant in the large majority of cases. Although the random walk model is not statistically outperformed by any of the alternative models, the empirical findings seem to suggest that theoretically well-grounded financial models are valid instruments for producing accurate forecasts of the WTI spot price.

Acknowledgments This work has been presented at the First FEEM Conference on the Economics of Sustainable Development held at the Fondazione Eni Enrico Mattei (FEEM), Milan, January 25–25, 2007. The authors would like to thank Carlo Carraro, Giliola Frey, Marzio Galeotti, Alessandro Lanza, Michael McAleer and Yves Smeers for insightful discussion, as well as seminar participants at the University of Bath, FEEM and the University of Milan-Bicocca for helpful comments. The authors are grateful to Chiara Longo for her research assistance related to an older working paper version of this contribution. Andrea Bastianin gratefully acknowledges the following research grant: "Dote Ricercatori: FSE, Regione Lombardia".

References

Abosedra S (2005) Futures versus univariate forecast of crude oil prices. OPEC Rev 29:231–241
Auffhammer M (2007) The rationality of EIA forecasts under symmetric and asymmetric loss. Res Energy Econ 29:102–121
Bopp AE, Lady GM (1991) A comparison of petroleum futures versus spot prices as predictors of prices in the future. Energy Econ 13:274–282
Brenner R, Kroner K (1995) Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets. J Financ Quant Anal 30:23–42
Chernenko SV, Schwarz KB, Wright JH (2004) The information content of forward and futures prices: market expectations and the price of risk. International Finance Discussion Papers 808. Board of Governors of the Federal Reserve System, US
Chinn MD, LeBlanch M, Coibion O (2005) The predictive content of energy futures: an update on petroleum, natural gas, heating oil and gasoline. NBER Working Paper 11033. National Bureau of Economic Research, Cambridge
Clewlow L, Strickland C (2000) Energy derivatives: pricing and risk management. Lacima Publications, London
Commodity Futures Trading Commission (2010) The commitments of traders report. http://www.cftc.gov/marketreports/commitmentsoftraders/index.htm. Accessed 26 Apr 2012
Dees S, Karadeloglou P, Kaufmann RK, Sanchez M (2007) Modelling the world oil market: assessment of a quarterly econometric model. Energy Policy 35:178–191
Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13:253–263
Engle RF, Granger CWJ (1987) Co-integration and error correction: representation, estimation and testing. Econometrica 55:251–276
Elliott G, Komunjer I, Timmermann A (2005) Estimation and testing of forecast rationality under flexible loss. Rev Econ Stud 72:1107–1125
Elliott G, Komunjer I, Timmermann A (2008) Biases in macroeconomic forecasts: irrationality or asymmetric loss? J Eur Econ Assoc 6:122–157
Eni (2010) World oil and gas review. Eni, Rome
Garratt A, Lee K, Pesaran HM, Shin Y (2003) Forecast uncertainties in macroeconomic modeling: an application to the UK economy. J Am Stat Assoc 98:829–838

Geman H (2005) Commodities and commodity derivatives. Wiley, Chichester

Green SL, Mork KA (1991) Towards efficiency in the crude oil market. J Appl Econ 6:45–66

Gulen SG (1998) Efficiency in the crude oil futures markets. J Energy Finance Develop 3:13–21

Johansen S (1995) Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press, Oxford

Kaufmann RK (1995) A model of the world oil market for project LINK: integrating economics, geology, and politics. Econ Model 12:165–178

Kaufmann RK (2004) Does OPEC matter? An econometric analysis of oil prices. Energy J 25:67–91

Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. Ann of Stat 17:1217–1241

Merino A, Ortiz A (2005) Explaining the so-called price premium in oil markets. OPEC Rev 29:133–152

Moosa IA, Al-Loughani NE (1994) Unbiasedness and time varying risk premia in the crude oil futures market. Energy Econ 16:99–105

Morana C (2001) A semiparametric approach to short-term oil price forecasting. Energy Econ 23:325–338

Moshiri S, Foroutan F (2006) Forecasting nonlinear crude oil futures prices. Energy J 27:81–95

Patton AJ, Timmermann A (2007) Testing forecast optimality under unknown loss. J Am Stat Assoc 102:1172–1184

Pindyck RS (1999) The long-run evolution of energy prices. Energy J 20:1–27

Pindyck RS (2001) The dynamics of commodity spot and futures markets: a primer. Energy J 22:1–29

Radchenko S (2005) The long-run forecasting of energy prices using the model of shifting trend. Working Paper. University of North Carolina at Charlotte

Samii MV (1992) Oil futures and spot markets. OPEC Rev 4:409–417

Serletis A (1991) Rational expectations, risk and efficiency in energy futures markets. Energy Econ 13:111–115

Ye M, Zyren J, Shore J (2002) Forecasting crude oil spot price using OECD petroleum inventory levels. Int Adv Econ Res 8:324–334

Ye M, Zyren J, Shore J (2005) A monthly crude oil spot price forecasting model using relative inventories. Int J Forecast 21:491–501

Ye M, Zyren J, Shore J (2007) Forecasting short-run crude oil price using high and low inventory variables. Energy Policy 34:2736–2743

Zamani M (2004) An econometrics forecasting model of short term oil spot price. In: Proceedings of the 6th IAEE European conference

Zeng T, Swanson NR (1998) Predictive evaluation of econometric forecasting models in commodity future markets. Stud in Nonlinear Dyn Econ 2:159–177

Part III
Electricity Markets

Commodity Price Interaction: CO_2 Allowances, Fuel Sources and Electricity

Mara Madaleno, Carlos Pinho and Cláudia Ribeiro

Abstract This work anlyses the relationship between the returns for carbon, electricity and fossil fuel price (coal, oil and natural gas), focusing on the impacts of emissions trading via a Vector Error Autoregressive Correction Model (VECM) for both German and French markets. Results show that the effect of carbon depends on the energy mix of the country under analysis but that it is not the only factor. Less carbon coercion takes place in the European Energy Exchange (EEX) and innovations in carbon are not strongly reflected in electricity prices. Also, market power affects the correct transfer of prices, thus limiting cost increases.

Keywords CO_2 emission allowance trading · Environmental management · Spot prices · European union · Energy mix impact

1 Introduction

The European Union Emission Trading System (EU ETS) officially began on 1st January 2005 following the 2003/87/EC directive. It is one of the largest multinational emission trading schemes in the world and a major pillar of the EU climate policy created in the ambit of the Kyoto Protocol[1] which aims to cut 1990 levels of CO_2 emissions by 8 %.

[1] Signatories of the Kyoto Protocol in 1997 decided to reduce greenhouse gases (namely CO_2) by limiting quantified emissions; Under the treaty, industrialized countries agreed to reduce their 1990 levels of greenhouse gas emissions by at least 5 % until 2012.

M. Madaleno (✉) · C. Pinho
GOVCOPP—Unidade de Investigação em Governança, Competitividade e Políticas Públicas, DEGEI - Departamento de Economia, Gestão e Engenharia Industrial, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
e-mail: maramadaleno@ua.pt

C. Ribeiro
CEF.UP/Faculdade de Economia da Universidade do Porto,
Rua Dr. Roberto Frias,
4200-464 Porto, Portugal

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_8

The EU ETS sets a ceiling on emissions from the most energy-intensive industrial sectors and introduced the emissions market. Large European CO_2 emitting installations receive permits from their government to emit tonnes of CO_2, and their equivalent is traded on the spot and derivatives markets (mostly options and futures) whenever targets are met at the scheduled time (Mansanet-Bataller et al. 2007).

The energy sector is clearly on the front line of climate change as it is responsible for 60 % of global greenhouse gas emissions and much of the regional and urban air pollution (World Energy Council 2010). Moreover, air quality is a major concern in the urban environment as 50 % of the world's population lives in a city. Emission trading is a market-based scheme aimed at improving the environment and it allows parties to buy or sell both permits for emissions and credits for reductions in the emission of certain pollutants (Dellink et al. 2010). Electricity generation is the main polluting activity in the energy sector and it has been opened up to competition due to the liberalisation of the electricity market in Europe. Electricity is produced from various primary energy sources including nuclear, coal, oil, gas and renewable energies. A country's energy mix is determined by the proportion of the different primary energy sources used in electricity generation. It varies from one European country to another as a result of differences in energy policies as well as geographical and geological features. Electricity prices are therefore determined by the cost of fossil fuels, the impact of environmental policies, and also by climatic factors (Mohammadi 2009).

Carbon allowances are currently traded in electricity exchanges throughout Europe and their price is a result of supply and demand (Benz and Trück 2009). In general, CO_2 production depends on a number of factors such as weather, fuel prices and economic growth (Springer 2003; Mansanet-Bataller et al. 2007; Alberola et al. 2008; Chevallier 2012; Creti et al. 2012).

In this chapter, we intend to extend previous analyses of electricity prices, fuel prices and carbon interactions in at least five ways: (1) our period of analysis is from 2009 to 2012 (Phase II); (2) we broaden previous works to the German and French markets. These countries were selected for the following reasons: the German electricity market is one of the biggest by number of participants and generation capacity, and has strong connections with the rest of the European countries (Madaleno and Pinho 2011a, b); allowances have been traded since 2005 in both markets; the German market is completely open to competition, while the French market is still characterised by monopolistic behaviours; both appear to behave coherently (Silva and Soares 2008; Pinho and Madaleno 2011b); although France and Germany are already geographically close to each other, they formed a regional market in January 2010; (3) We include other fuel prices such as oil due to the energy mix that distinguishes the markets under analysis, and we provide a VECM model with five endogenous variables; (4) we give a clear answer on how the EU ETS has affected the electricity generation sector by addressing countries' heterogeneity (for both short and medium term interactions); (5) finally, we include temperature dummies.

Empirical findings show that in the period under analysis, the European emission allowances market failed to compel electricity producers to reduce their emissions and invest in cleaner technologies whose efficiency depends on the energy mix of the country under analysis. Policies related to the coal industry have therefore a marginal influence on electricity prices.

The remainder of the chapter is organised as follows. Section 2 presents the functioning of the carbon market and its determinants, before showing the data used and its statistical properties in Sect. 3. Section 4 provides the methodology, empirical analysis, results, and policy recommendations, and Sect. 5 concludes.

2 How the Carbon Allowances Market Works and What Affects It

The EU-ETS is the first large scale CO_2 emission trading system in the world. It has been organised in three phases with a pilot phase (Phase I) going from 2005 to 2007,[2] Phase II going from 2008 to 2012 and Phase III of arrangement from 2013 until 2020. The EU ETS is set to expire in 2020 if no other international climate agreement is reached (Creti et al. 2012). Any company wishing to participate in the emission allowances market must open an account in the registry of the country of origin, where allocations are stipulated along with each company's the purchases and sales.

The EU ETS covers more than 11,000 industrial installations in 25 countries; each participating country proposes their National Allocation Plan (NAP) including caps on greenhouse gas emissions for power plants and other sources, which must subsequently be approved by the European Commission. The NAP of each member state determines the total quantity of CO_2 allowances granted per year for each company and for a specific commitment period (each Phase[3]). Allowances are allocated free of charge in the first stage.[4] Thereafter, additional allowances must be purchased directly from the market when required.

---

[2] Considered the trial phase when administrative and regulatory bodies were put on-line.

[3] During each of these Phases, allowances delivery is made on a yearly basis and follows a precise calendar: on February 28 of year N, European operators receive their allocation for the commitment year N; March 31 of year N is the deadline for the submission of the verified emissions report during year N − 1, from each installation to the European Commission; April 30 of year N is the deadline for the restitution of quotas utilized by operators during year N − 1; May 15 of year N corresponds to the deadline of the official publication by the European Commission of verified emissions for all installations covered by the EU ETS during year N − 1 (European Commission reports).

[4] This will be limited for Phase III (beginning in 2013), where allowances will not be issued completely free of charge (Friends of the Earth 2010). The allocation of allowances will be made primarily by auction, but until 2020, some allowances will continue to be allocated free of charge to the industrial sector in particular to reduce the costs to facilities in areas considered to be exposed to significant competition, especially from third countries. According to the DG Clima, this decision establishes the rules, including benchmarks for emissions of greenhouse gas emissions, but it is the responsibility of member states to calculate the number of allowances that will be provided free of charge to these areas each year.

The purpose of the EU ETS is primarily to reduce emissions by promoting low carbon technologies and energy efficiency among CO_2 emitting plants and to establish a market price for allowances. European polluters will therefore be aware of the environmental consequences of their polluting activities. As such, installations need to surrender as many allowances during this period as the amount of carbon dioxide emitted during the reference year. The EU ETS is a cap-and-trade scheme; the overall level of emissions is capped up to this limit, and installations short (in excess) of allowances (emissions rights) with respect to their individual allocation level may purchase (sell) allowances on the spot market in order to meet their compliance requirement in the EU ETS (Alberola et al. 2008). Installations that do not meet their target in Phase II must pay a penalty of 100 €/ton of CO_2, up from 40 €/ton of CO_2 in Phase I.

At the start of Phase I, major emitters were allocated an initial amount of permits and were free to trade them on the market. A similar new supply was given every year to the same sources. However, the early environmental benefits were limited because of concerns among member states of over-allocation (Ellerman et al. 2010) and the implementation of banking restrictions between 2007 and 2008; as a result, carbon spot and futures prices of maturity fell to zero levels in December 2007 (Alberola and Chevallier 2009). This first experience also highlighted the need for reliable verified emissions data, harmonised monitoring and reporting rules, as well as concentrated their attention on the first Phase, despite the fact that this was a learning period which revealed the weaknesses of the scheme.

Academics had investigated carbon price patterns in 2005–2007 discussing both their determinants (Alberola et al. 2008; Mansanet-Bataller et al. 2007) and stochastic behavior to forecast trends (Paolella and Taschini 2008; Seifert et al. 2008). Ferkingstad et al. (2011) study the dynamics of price information flow among weekly Nordic and German electricity prices and oil, gas, coal, wind power in Germany and Nordic water reservoir levels but did not take the price of allowances into account. Creti et al. (2012) try to shed light on the determinants of carbon futures prices in Phase II by testing whether energy prices and indicators of economic activity still hold for this phase and evolve toward a stable long-run relationship; they used daily futures contracts from 2005 until 2010 in their cointegration testing. These authors did not include weather variables arguing that the literature thus far only shows that their impact on carbon prices is indirect and captured by sudden shocks in energy demand.

Phase II brought more clarity. The audited figures for each installation were disclosed publicly and installations that had initially received a substantial surplus were subsequently given much less. Supply and demand of allowances was adjusted through exchanges and over-the-counter transactions based on price levels and institutional characteristics of the (Creti et al. 2012).

Economic theory teaches us that carbon price is a marginal cost and that carbon permits have an opportunity cost equal to their market price. Thus, it is to be expected that the price of carbon will be an additional increment to the short-term fuel costs of power generation and must therefore be included in the price of electricity. However, the aggregate effect of carbon prices will depend on the technology mix across the whole of the EU and firms' pricing behaviors. Moreover, electricity prices that reflect the cost of CO_2 are needed to encourage investment in clean generation, demand-side response and the adoption of efficient end-use technologies. The increase of CO_2 in the atmosphere caused by the rampant use of fossil fuels has negative impacts on natural systems and is a main contributor to climate change. Coal and oil should thus be replaced with renewable alternatives which do not emit CO_2. Accordingly, trading allowances for the emission of CO_2 gives value to reducing emissions and has resulted in a market with an asset value worth tens of billions of euros annually.

However, trading CO_2 is different from more traditional commodities. First, whereas producers in this market may hold emission allowances to reduce the costs of adjusting production over time or to avoid stock outs, assets in financial markets can be used for insurance, hedging and speculation. Second, the emissions of sellers are expected to be lower than their allowance, so the unused allowances are bought by those who emit more than their allocated amount. The carbon credit system strives to reduce emissions by encouraging countries to honour their emission quotas and offer incentives to stay below them (Prabhakant and Tiwari 2009; Bhardwaj and Wadadekar 2010). Third, the value of a stock is based on the expected profit of the firm that distributes the shares, while the price of emission allowances is determined by the balance between supply and demand (Benz and Trück 2009). Fourth, while the annual quantity of allocated emission allowances is limited and specified by the EU-Directive for all trading periods, it is the firm that decides whether to issue additional shares and thus fosters the stock's liquidity. Fifth, unlike other markets, emissions trading schemes create a commodity which has one sole producer and supplier, i.e. the government is the only source of allowances and emissions permits. Moreover, there are no apparent production and storage costs. Finally, allowances have a limited validity.

Literature has found evidence that a change in carbon prices is closely linked to the power price (Convery and Redmond 2007). Moreover, German wholesale power prices were found to be closely related to European Union Allowances price change (Zachmann and von Hirschhausen 2008). Also, previous authors analysed CO_2 spot price behaviour (Benz and Trück 2009; Paolella and Taschini 2008; Seifert et al. 2008; Daskalakis et al. 2009) and CO_2 futures markets (Uhrig-Homburg and Wagner 2006, 2008; Wei et al. 2008).

Through Vector Autoregressive (VAR) analysis, long-term and short-term dynamics of electricity, gas and coal prices and the price of carbon permits were studied in the Finnish market (Honkatukia et al. 2007). Similar structural approaches were used to analyse the English electricity market, this time excluding the price of coal and including temperature and dummies as exogenous variables (Bunn and Fezzi 2007).

Previous authors using an autoregressive distributed lag model concluded that other determinants of fossil fuel used in Swedish electricity generation probably diminished the effects of the EU ETS (Widerberg and Wräke 2009). Reasons for the less than 100 % pass-through of CO_2 costs into firm and industry were attributed to demand responses, market structure, and competition from non-fossil fuel generators (Sijm et al. 2006). Among other variables, prices of European Union Allowances (EUA) are also influenced by coal and natural gas prices (Mansanet-Bataller et al. 2007). Moreover, significant interactions are found between European Union Allowances prices and input fuel prices (Bunn and Fezzi 2007). Our results reveal that electricity prices have null short-term responses to CO_2 price shocks, although the response increases over time. This conclusion is the inverse of others taken elsewhere (Fell 2008) using daily data for NordPool for 2005–2008 under a VECM methodology although not using oil prices, but including reservoir levels. For the US market and using VECM, Mohammadi (2009) concludes that there is only a significant long-term relation between electricity and coal, and while the role of oil prices is significant, that of natural gas is statistically weak.

The different results obtained in studies not only reflect distinct approaches but also the fact that the countries surveyed have very diverse energy mixes. The absence of a unanimous response to the problem of the effect of the EU ETS on the price of electricity (Reinaud 2007) is therefore due mainly to the coexistence of various electricity markets in Europe and the heterogeneity of energy mixes. Furthermore, as these studies did not cover any more than the period 2005–2006, on the demand side, carbon prices are impacted by energy prices because they reflect the producing process of the utilities regulated by the EU ETS.[5]

3 Data and Statistical Properties

Electricity prices were obtained from the electricity stock exchanges of Powernext (FR) for France, and European Energy Exchange (EEX) for Germany. We focus on the French and German electricity markets where the major fuel sources are gas, coal and oil (Ferkingstad et al. 2011). The German electricity data collected starts in June 2000 and the French data in November 2001. CO_2 only started to be traded after the liberalisation of electricity markets, namely October 2005 in Germany and April 2005 in France.

Weekly day-ahead (base load price—the day's arithmetic 24 h average) electricity prices (in €/MWh) were obtained by means of the price on the last trading day in the week. Due to data restrictions and the misbehaviour of carbon markets until 2009 our period of analysis is from January 2, 2009 to July 6, 2012. Moreover, Chevallier (2012) identifies three breakpoints in carbon spot series.[6] Our results would not necessarily say much about price information flow between the weekly price levels if we chose the Sunday price, and using weekly average spot prices might have induced additional correlation into the series or differenced price series.

[5] For more details on the relationship between coal, energy prices and fuel switching behaviour, institutional decisions and weather events between 2007 and 2009 see Chevallier (2012).

[6] These were May 28, 2007; December 30, 2008; and February 11, 2009.

Daily data can avoid additional complications induced by averaging but results obtained when we performed this analysis proved to be less reliable.[7]

The carbon spot price of the respective stock exchange expressed in € per ton was used; in other words, the Bluenext carbon spot European Union Allowances price was used for France, and the EEX-EU CO_2 emissions allowances price was used for the German market. Furthermore, we collected data on exchange rates so that all electricity prices and carbon were in the same denomination as other primary energy fuels used (gas, oil and coal), i.e. we converted all prices to US dollars to control the impact of exchange rates. Monthly exchange rates were collected from the "Bank of Portugal"[8] covering the corresponding sample periods.

For crude oil, we use weekly spot prices of the London Brent Crude Oil Index, one of Europe's benchmarks for crude. Weekly spot prices set on Brent are denominated in US dollars per barrel but transformed into Euros. Brent is a North Sea deposit; as its oil is representative of the crudes produced in this region, it has the best characteristics to match other energy variables traded in Continental Europe (Chevallier 2012). For coal data, we take the Antwerp/Rotterdam/Amsterdam (ARA) coal price which is denominated in US dollars per Gigajoule. Weekly prices on natural gas are those reported in the Zeebrügge Hub where data is denominated in €/MWh. We expect this market to be more important for electricity price formation as it is closer to the German market (Ferkingstad et al. 2011) which is the most liquid gas trading market in Europe. As argued by Chevallier (2012) this market has a major influence on the price that consumers pay for their gas in Europe and therefore constitutes a good proxy. Data descriptive statistics are presented in Table 1. All time series have been log-transformed into returns.

As evidenced by the data, mean returns for all electricity spot markets are positive. The Jarque-Bera statistic indicates that the distribution of returns for all samples has fat tails and sharper peaks (kurtosis) than the normal distribution (kurtosis being higher for natural gas and carbon prices). Skewness, which measures the degree of a distribution's asymmetry, is also very different from zero, and is negative for carbon, natural gas, oil and Powernext electricity returns. Results for skewness and kurtosis are not shown here but are available on request.

Moreover, volatility is high for all markets and there are no significant differences between the average wholesale electricity returns in the two markets. Powernext relies heavily on nuclear power, followed by hydro, and given the results obtained here we are able to confirm the finding that the mix of generation technology has an impact on the standard deviation of market prices (Wolak 1998). Wolak (1998) finds that prices in markets dominated by fossil fuel or thermal technology tend to be much more volatile than prices in markets dominated by hydroelectric capacity. According to the standard deviation obtained, which we use

[7] Results will be provided upon request.

[8] http://www.bportugal.pt/pt-PT/Estatisticas/PublicacoesEstatisticas/BolEstatistico/Paginas/BoletimEstatistico.aspx.

Table 1 Descriptive statistics for energy and carbon returns

	elect_PN	elect_EEX	Gas	Coal	Oil	CO_2_PN	CO_2_EEX
Mean	0.002	0.002	0.001	0.002	0.006	−0.003	0.033
Median	0.003	0.007	0.007	−0.001	0.005	0.001	0.002
Maximum	0.641	1.070	0.342	0.130	0.157	0.256	6.420
Minimum	−0.860	−1.780	−0.515	−0.116	−0.115	−0.233	−0.217
Std. Dev.	0.197	0.231	0.098	0.034	0.039	0.061	0.477
Jarque-Bera test	86.657	3470.0	183.342	9.976	16.706	64.862	236000.0
Probability	0.001	0.001	0.001	0.015	0.004	0.001	0.001
Observations	184	184	184	184	184	184	184

EEX stands for European Energy Exchange in Germany. PN stands for Powernext in France. The period analysed goes from January 2009 to July 2012 for both EEX and Powernext. The gas, coal and oil series descriptive statistics results are for the same period of electricity and carbon series analysed for both markets. *Std. Dev*. standard deviation; Jarque-Bera test statistic; Probability values are those associated to the Jarque-Bera test; *elect.* electricity return; CO_2 carbon return

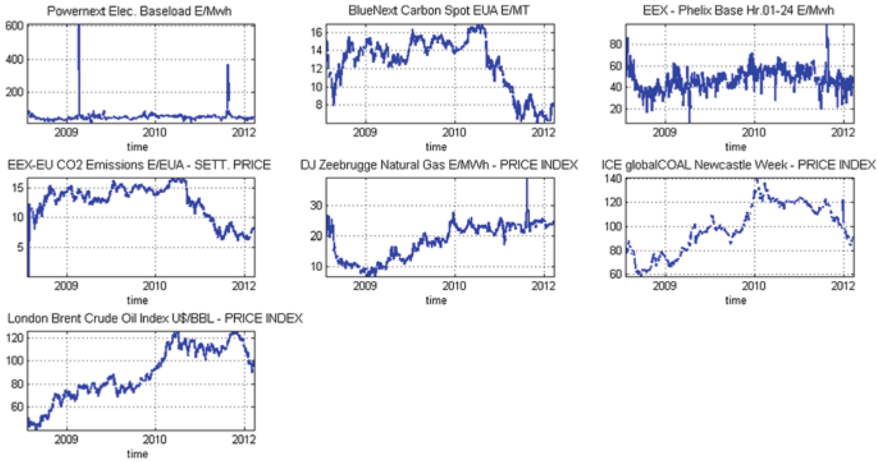


Fig. 1 Weekly price dynamics plots for electricity, gas, coal, carbon and oil

as our volatility proxy, EEX presents higher volatility of both electricity prices and allowances.

Volatility increases costs for emitters and they prefer stable and predictable carbon prices. In the carbon markets, there are generally two types of risk that participants may want to transfer: carbon price volatility and carbon default risk (the risk that offset projects may not achieve some or all of their carbon reductions). Both types of risk would be found in a system with a high proportion of offsets and volatile carbon prices.

Since 2005, electricity prices have been affected by two major changes: an increase in fossil fuel prices and natural gas in particular, and the introduction of CO_2 allowances, itself boosted by increasing gas prices. The two factors have resulted in higher market prices—and costs—for energy intensive users. Figure 1 shows that electricity contract prices have varying volatilities; they are most volatile than of all energy markets, whereas CO_2 volatility is very similar among markets. The price of coal rose sharply in 2010 and only decreased at the end of 2011.

As stated previously, CO_2 emission allowances have a limited validity as they expire after each commitment period. However, the decision to allow banking[9] from the pilot phase (2005–2007) into the first Kyoto commitment period was left to the individual EU member states (whereas Germany decided against allowing it, France permitted it in the initial stage). An intertemporal ban in banking meant all licences became invalid at the end of 2007 and environmental institutions had to

[9] Banking occurs when the right to emit carbon can be saved for future use, i.e. we can use a 2007 allowance in 2008. On the other hand, borrowing means that current emissions are extended against future abatement, i.e., we can borrow permits from future allocations for use in the current period (using 2008 allowances in 2007). Both banking and borrowing were forbidden between phase I and II.

issue companies with new allowances. Therefore, Phase I spot prices for carbon went down to zero by the end of Phase I due to banking restrictions implemented between 2007 and 2008 (Alberola and Chevallier 2009). This induced an excessive supply of allowances on the market which in turn led to a fall in the carbon price initiating a convergence towards zero in January 2007. Moreover, two structural breaks were also identified in the literature (Alberola et al. 2008) in 2005–2007 and three have since been explored by Chevallier (2012).

The second year in Phase II of the EU ETS, 2009, started with a fall in European Union Allowances prices; this followed the decline that had begun towards the end of 2008 due to the widening of the financial crisis and it stoked fear among market participants of a reoccurrence of the problems at the end of Phase I when allowances were being sold to improve companies' balance sheets (see European Commission reports). As a result we excluded 2008 from our analysis.

Carbon and coal prices seem to follow opposite paths. The price dynamic is consistent with the intuition that when the demand for carbon permits increases, the coal price decreases. They will increase when the relative price of coal decreases because a coal-fired power station is more carbon-intensive than a gas-fired station However, there seems to be a downward trend in both from 2010 onwards with some evident peaks with respect to coal.

The electricity markets under analysis differ in their underlying production structure. The recommendations throughout "green markets" are showing some evolution with respect to hydro and wind. Renewables are still not the main production source for both countries. According to Eurostat data, Germany generated 10 % of its electricity from renewable sources in 2005 and France 10.98 %. In 2008, the figures rose to 14.63 and 14.07 % for Germany and France respectively, followed by another increase for with in 2010 to 16.9 % compared with just 14.45 % for France. This demonstrates the huge effort being made in Germany.

At this stage it is interesting to notice the differences in the energy mix among countries. For example, France has a large nuclear and hydro production (Pinho and Madaleno 2011a). Of the EU-15 countries, France is expected to be a relative winner in the EU emission trading due to its large proportion of nuclear energy[10] (Pinho and Madaleno 2011a). The percentage of nuclear in EEX is also high, and is followed by coal (see Table 2). Germany clearly switched from coal to natural gas and wind, while France is still relying on nuclear. The German EEX market is the largest market in Europe, dominated by coal (47 %), nuclear power (23 %), gas (17 %), hydro and increasing wind power production (Ferkingstad et al. 2011).

Reducing the concentration in the electricity industry was another of the main objectives of the EU Directives: "increasing competition to reduce market power". Table 3 presents the percentage share of the largest generator for the markets under analysis between 1999 and 2010.

As demonstrated by the data, the French market has the highest level of generators concentration but the concentration in both markets was lower in 2010 than

[10] We were unable to include nuclear, wind or even hydro production due to lack of available data .

Table 2 Percentage of electricity production by fuel source in Germany and France

	Germany		France	
Fuel source/year	1998	2008	1998	2008
Hard coal	27.56	19.56	6.22	4.24
Petroleum	1.15	1.35	2.28	1.02
Natural gas	9.76	11.91	0.97	3.80
Nuclear	29.03	23.30	75.92	76.29
Hydro	3.88	4.23	13.04	11.95
Wind	0.82	6.37	0.00	0.99

Figures are in percentages computed as: (type of fuel used to produce electricity/total gross electricity generated) * 100. Total gross electricity generation (GWh) covers gross electricity generation in all types of power plants. The gross electricity generation at plant level is defined as the electricity measured at the outlet of the main transformers, i.e. it includes the consumption of electricity in the plant auxiliaries and in transformers. The gross electricity generation in power stations burning hard coal (GWh), in power stations burning natural gas (GWh), in nuclear power plants (GWh) and in wind turbines (GWh) are measured as above Gross electricity generation in power stations burning petroleum (GWh) products cover hydrocarbons like motor spirit, gas oil, kerosene, etc. produced in oil refineries or in some rare cases obtained without refining Hydroelectricity covers potential and kinetic energy of water converted into electricity in hydroelectric plants (GWh), also expressed as gross generation. *Data comes from* http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

in 2000. Nevertheless, the high levels of concentration create scope for market power and therefore they influence spot prices, which could induce environmental costs being transferred erroneously to electricity prices (Pinho and Madaleno 2011a).

The correlation matrix between European Union Allowances price markets is also studied for the estimation period. Results are presented in Table 4.

Table 4 shows that European Union Allowances markets have positive pairwise correlations (except between carbon and all the other fuel sources in both markets); although this implies interactions between electricity prices and fuel prices, they are not so strong as initially expected. Higher correlations are observed between gas and electricity, coal and oil, as well as between gas and oil.

Similar to Chevallier (2012) we also considered the broad European temperatures index[11] to be a suitable exogenous variable that drives energy and allowances prices, and therefore included it as a dummy in our model. Weather conditions are expected to affect the price path of carbon by influencing energy demand. In cold winters, more heating is needed and this requires extra power extra power generation. On the other hand, hot summers lead to a greater consumption of air-conditioning, also raising electricity production. However, the fuel used in response to

[11] See http://www.weatherindices.com/index. Moreover, due to data limitations and lack of availability for the countries considered here we do not consider other potential weather events such as wind.

Table 3 Percentage share of the largest generator of electricity

Market/year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Germany	28.1	34.0	29.0	28.0	32.0	28.4	30.0	31.0	30.0	30.0	26.0	28.4
France	93.8	90.2	90.0	90.0	89.5	90.2	89.1	88.7	88.0	87.3	87.3	86.5

Data comes from http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

Table 4 Correlations among weekly returns for electricity, carbon, gas, coal and oil

Powernext—France						EEX—Germany					
Variable	Elect.	Gas	Coal	Oil	CO_2	Variable	Elect.	Gas	Coal	Oil	CO_2
Elect.	1					Elect	1				
Gas	0.529	1				Gas	0.299	1			
Coal	0.024	0.129	1			Coal	−0.059	0.129	1		
Oil	0.063	0.141	0.276	1		Oil	0.029	0.141	0.276	1	
CO_2	−0.006	−0.073	−0.079	−0.095	1.000	CO_2	−0.023	−0.020	−0.052	−0.039	1

The period analysed for both markets is from January 2009 to July 2012. Figures presented are in absolute terms. Elect. stands for electricity returns, CO_2 for European union allowances (carbon) returns the demand for increased production is not always the lowest CO_2 emitting source; this means more CO_2 allowances are required, which will be reflected in prices. The national business-climate index used was computed by Metnext (the average daily temperature of the regions that compose a country weighted by their population). CDC Climate Research has extended this methodology by creating the European temperatures index (expressed in degrees Celsius), which is equal to the average of the national temperature indices for 18 European countries (including France and Germany), weighted by the weight of each country in the total volume of distributed allowances. For our analysis we define two dummy variables: one to capture the influence of cold temperatures and the other to capture the influence of hot temperatures.[12]

4 Model and Empirical Results

The descriptive statistics provided above indicate that energy series and carbon prices are non-stationary. This implies that any particular price measured over time will not be tied to its historical mean. Moreover, electricity, carbon and fuel prices are not expected to be independent from each other, whereas similar economic forces are expected to influence each market.

In order to address stationarity, the Augmented Dickey-Fuller test (ADF) was used (null hypothesis: non-stationarity of the tested time series), assuming a constant, a constant and a trend and none, for all series (in logs and log first differences) under analysis. The presence of a unit root for all the series after differencing one time is rejected (except for natural gas assuming a constant and a trend). Overall, the series are integrated of order one, I(1), or first-difference stationary, and we conduct the model analysis in logarithmic first differences (returns).[13]

We also tested for cointegration using Engle and Granger (1987) cointegration tests but do not present results in order to save space.[14] Tests performed indicate the existence of 1–2 cointegrating vectors depending on the market under analysis, and the null hypothesis of no cointegration is rejected.

For the empirical estimations, we define $y_t^T = \left(Log_{elec}, Log_{gas}, Log_{coal}, Log_{oil}, Log_{carbon}\right)$, the vector of the log prices of electricity, gas, coal, oil and carbon emission permits. Exogenous variables considered were the lagged values of endogenous variables and the two dummies used for hot and cold extreme temperatures; the Vector Autoregressive Model (VAR) lag order selection indicated

[12] The dummy that captures the influence of cold temperatures equals one when the temperatures index in a given month is −1.97 °C below decennial seasonal averages and that of the influence of hot temperatures equals one when the temperatures index in a given month is 1.47 °C above decennial seasonal averages.

[13] Results are not provided here but are available on request.

[14] We test for the number of cointegrating vectors using the trace test introduced in Johansen (1992) and the Max-eigenvalue test.

two lags for both markets selected by both LR (sequential modified Likelihood Ratio test statistic) and AIC (Akaike information criteria).[15]

The vector autoregressive (VAR) model (Hamilton 1994) is a standard and useful tool of econometrics and multivariate time series analysis. To explain the model, consider that endogenous variables y_t and exogenous variables x_t are observed random vectors depending on time t = 1, 2, … The main idea of this model is that endogenous variables depend linearly on their p previous values and also the current value of the exogenous variables. For now, we consider a VAR with p-lags (when p is long enough to ensure absence of autocorrelation):

$$y_t = v + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \delta x_t + \varepsilon_t \tag{4.1}$$

where y_t is a n × 1 vector of variables, v is a n × 1 vector of parameters, $A_1, \ldots, A_p$ are n × n matrices of parameters, δ is a coefficients matrix of size n × d and ε_t is a n × 1 vector of disturbances, with mean 0, covariance matrix ∑, and i.i.d. is normal over time. In this case, n stands for the number of endogenous variables and d for the number of exogenous variables (x_t).

From the econometric literature, we know that any VAR(p) can be rewritten as a Vector Error Correction Model (VECM) when the stability condition is not satisfied. In fact, all variables must have the same order of integration. If all variables are stationary, I(0), we can easily use the VAR specification. If not, or if the variables are non-stationary, I(k), k ≥ 1 we can do two things: If the variables are not cointegrated, they must be differenced k times in order to obtain a VAR; but if the variables are cointegrated, we may use a vector error correction model (VECM).

Here we define the VECM of order p as:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + \Phi x_t + \varepsilon_t \quad t \in Z \tag{4.2}$$

where y_t is a n × 1 random vector, $y_t \sim$ CI(1) meaning y_t sequence is a VAR(p) process cointegrated of order 1; $\Pi, \Gamma_1, \ldots, \Gamma_{p-1}$ are n × n fixed coefficient matrices and ε_t is a n × 1 white noise Gaussian process. In the present setting, we have a VECM with p = 2 for both Powernext and EEX. The Π matrix has a rank $r \leq n$ and $\Pi = \alpha\beta^T$. The n × r, α, matrix is called the loading matrix. The r × n, β, matrix is called the cointegration matrix. The columns of β, β_i are such that $\beta_i^T y_t$ is stable, and are cointegrating vectors. When we find the rank of cointegration for the VECM, y_t, we find the rank of Π, the number of cointegrating vectors β_i (if more than one, otherwise just one vector). Hence, βy_{t-1} can be regarded as an error correction element, with α then being a speed of adjustment vector. Given that we have defined y_t as being the vector of endogenous containing the log prices, Δy_t will be the vector containing log first differences (or else, returns). δ is a coefficients matrix

[15] Schwartz criteria was also used and given the difference of the selected lag structure and the need to keep the VAR model parsimonious, we ran the χ^2 lag exclusion test.

of size n × d associated with the d × 1 vector x_t that represents the two temperature dummies or exogenous variables. Notice that here n = 5 and d = 2.

Response of $y_{j,t+s}$ to a one-time impulse in $y_{i,t}$ is described by impulse-response functions, with all the other variables held fixed. They can be used to produce the time path of the dependent variables in the VAR to shocks from all the explanatory variables. If the system of equations is stable, any shock should decline to zero, whereas an unstable system would produce an explosive time path. In order to save space we omit the presentation of the VECM estimates.

Figure 2 displays the impulse response functions for all series in the France Powernext, namely the responses of each series to a shock in each series. The horizontal axis represents the up to 9-week responses of all series caused by an impulse (a one-time-only shock) in one of the series (column headers show the impulses and row headers the responses). The responses are normalised so that they can be compared with each other.

Each series response to its own shock shows to be positive, significant and strong in the short term. All series responses to shocks in oil prices seem to be positive, except for electricity in EEX (see Fig. 3), but they do not last across the entire time horizon considered (9 weeks). For Powernext, electricity response to carbon and gas appears to be positive in the short term but negative for coal. With respect to oil shocks, electricity only responds negatively in the 2-week period. Coal responses are generally positive, and natural gas seems to show a positive response to a shock in carbon, while negative for oil between 1 and 2 weeks. Moreover, oil response to coal is found to be sharply decreasing for a 1-week period. Electricity seems to indicate a negative response to carbon prices with a delay of approximately 1 week; the first impact is positive but not strong.

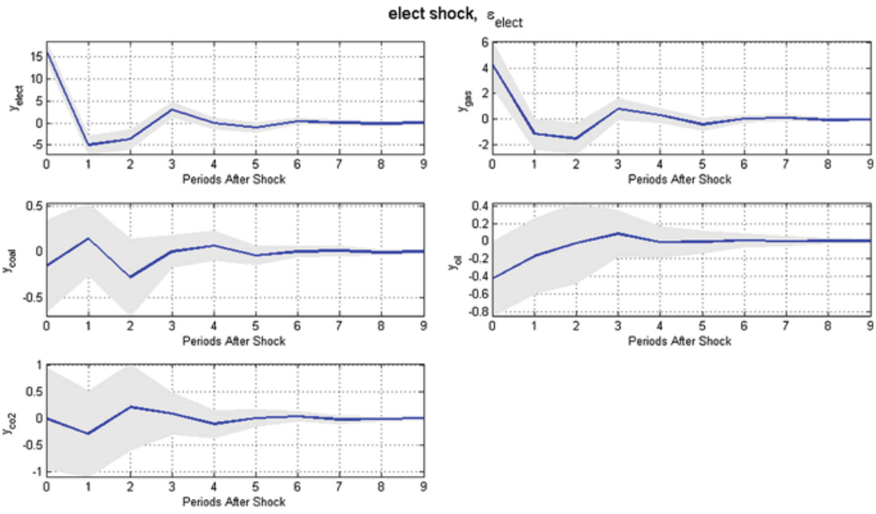


Fig. 2 Impulse response plot for Powernext. Each *column* shows the up to 9-week responses in all series to a one-time-only shock in the series listed in the column header

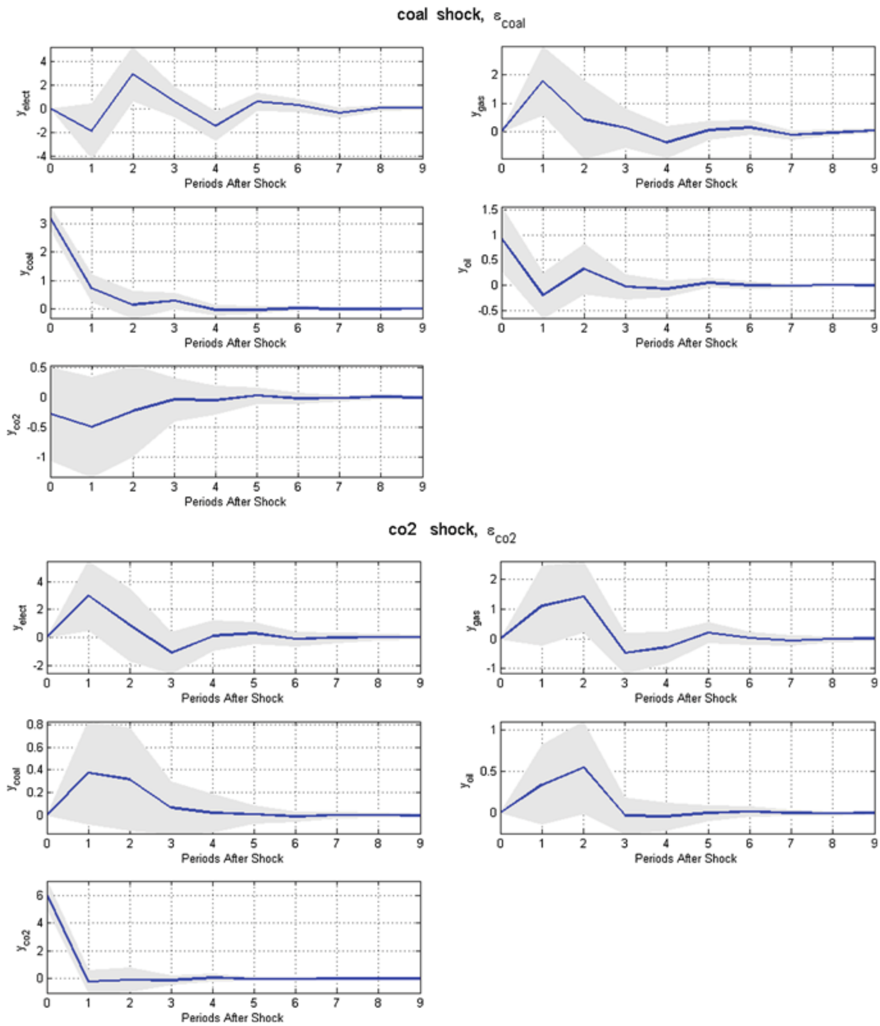


Fig. 2 continued

In addition, CO_2 response to fuel prices proved to be almost negligible in Powernext, although positive for electricity price shocks in EEX. An impact of electricity in natural gas is negative in both markets, but is positive for coal and for oil only after a stable period of 2 weeks. Natural gas seems to react positively in the short term, turning out to decrease after that; oil response to natural gas is negative and persistent until 2 weeks. The response of coal to natural gas disappears after 3 weeks, but coal reacts negatively to oil price shocks. In fact, oil shows the strongest response of all the relevant fuels to CO_2 prices in the short term even though it remains minimal over time. It was observed that whereas electricity prices

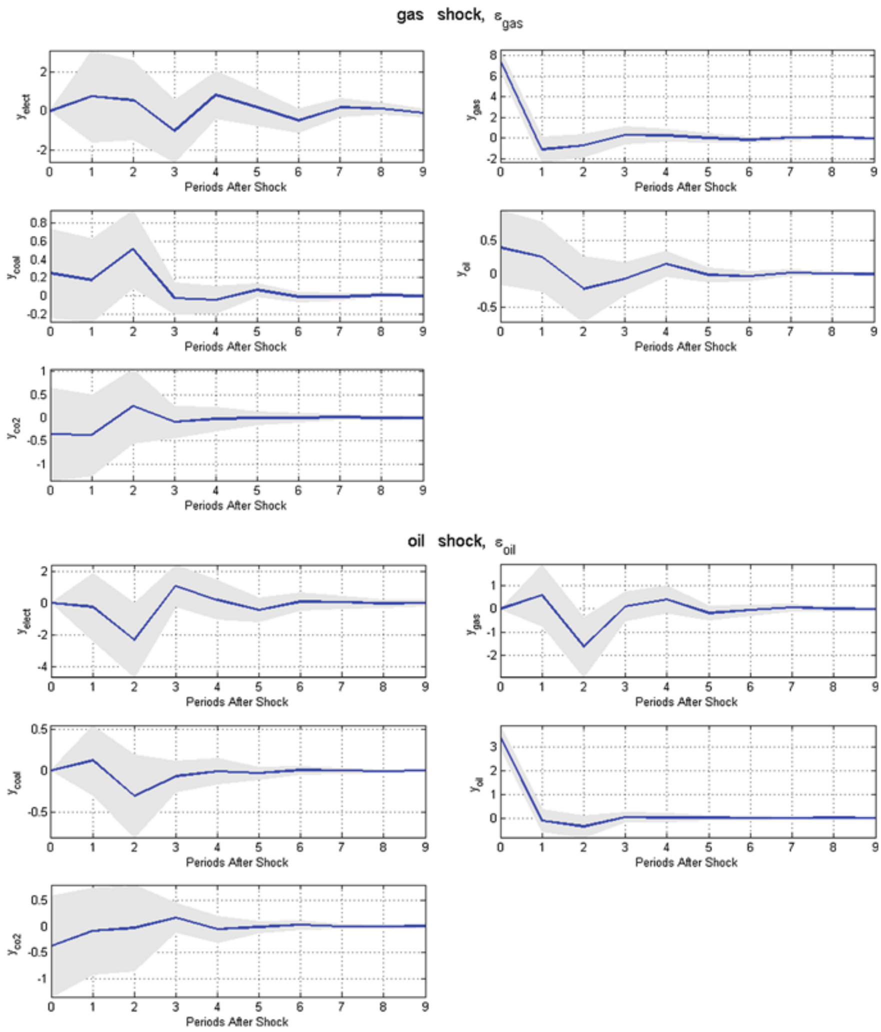


Fig. 2 continued

appear to react negatively after a shock in EEX, they react positively in Powernext, and after compensation in the following periods the response to CO_2 weakens. Moreover, natural gas seems not to respond significantly to European Union Allowances shocks.

Both carbon and gas shocks on electricity prices seem to produce a similar effect in the first week but the gas price shock is completely absorbed after a 3-week period, whether or not the shock in carbon price is persistent and unstable until 4-week, implying a significant marginal effect. This can be explained by the fact that the gas market is relatively mature.

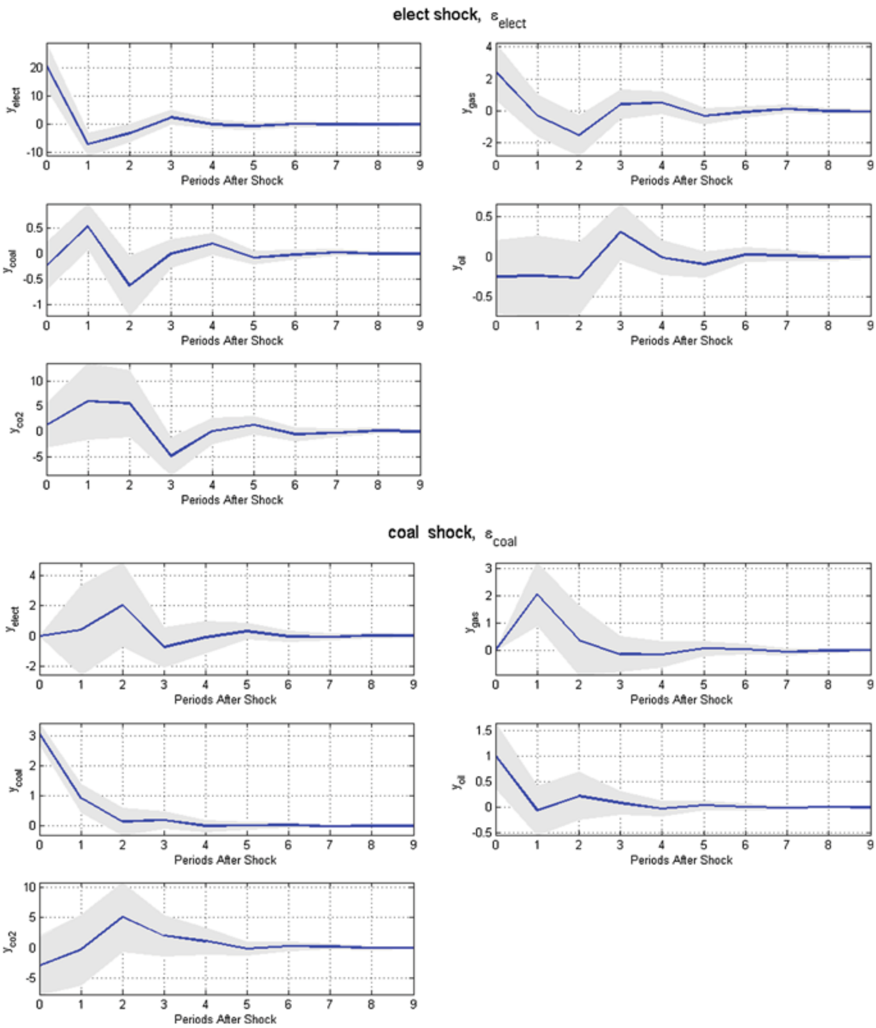


Fig. 3 Impulse response plot for EEX. Each *column* shows the up to 9-week responses in all series to a one-time-only shock in the series listed in the column header

In the German case, gas prices do not seem to be significantly affected by a shock in carbon prices and yet a gas price shock seems to affect both electricity and carbon prices positively. A possible reason is that in the EEX a significant quantity of electricity (around 11.9 %) is produced by gas-fired power stations (see Table 2) and the main initiative, in order to fulfil the Kyoto target, has been to switch from coal to gas, which occurred when we compare the values from 1998 to 2008. Switching becomes more expensive if gas prices are high, and this is reflected in higher carbon prices.

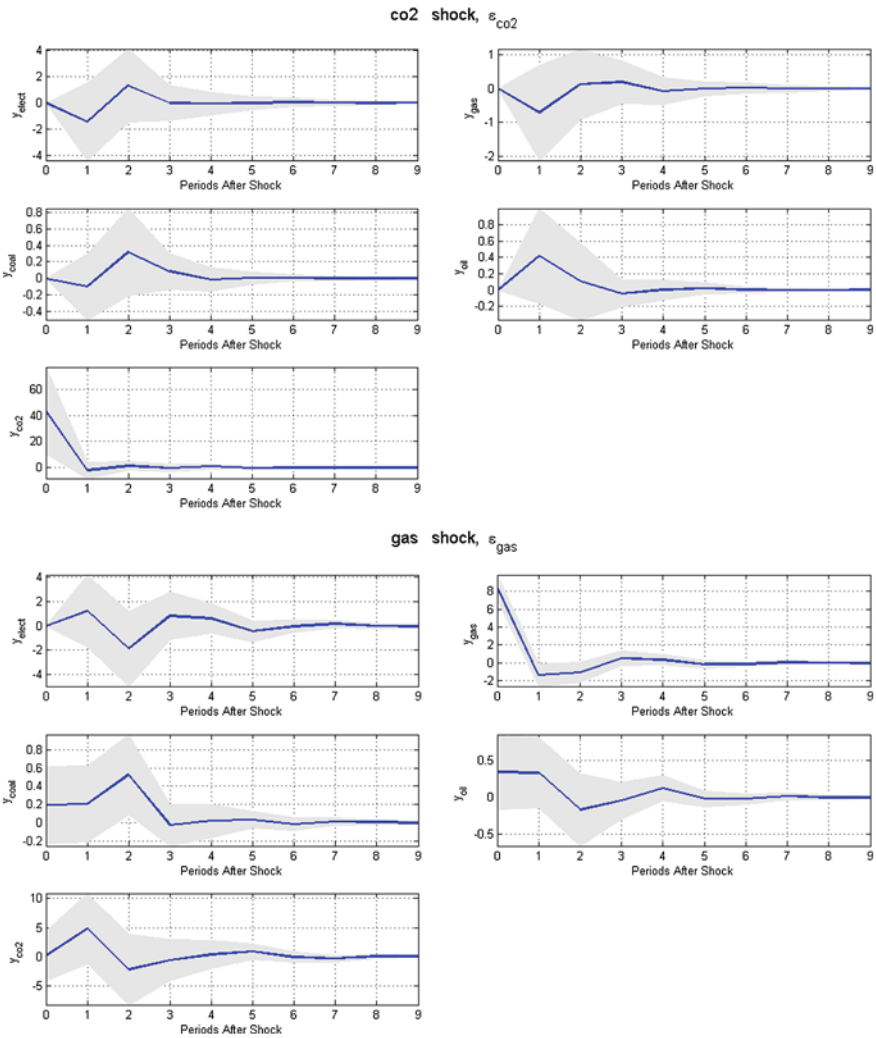


Fig. 3 continued

Despite using impulse response functions, variance decomposition (VD) is useful for examining the effects of shocks on the dependent variables. It determines how much of the forecast error variance for any variable in a system is explained by innovations to each explanatory variable over a series of time horizons. The result will depend on the order in which the equations are estimated in the model and here the selected order was: electricity, natural gas, coal, oil and EU ETS carbon.

Variance decomposition results are provided in Figs. 4 and 5, and Tables 5 and 6, for EEX and Powernext, respectively. Coefficients of the VD can be interpreted as the price of elasticity; this implies for example that a 1 % gas price rise would, in

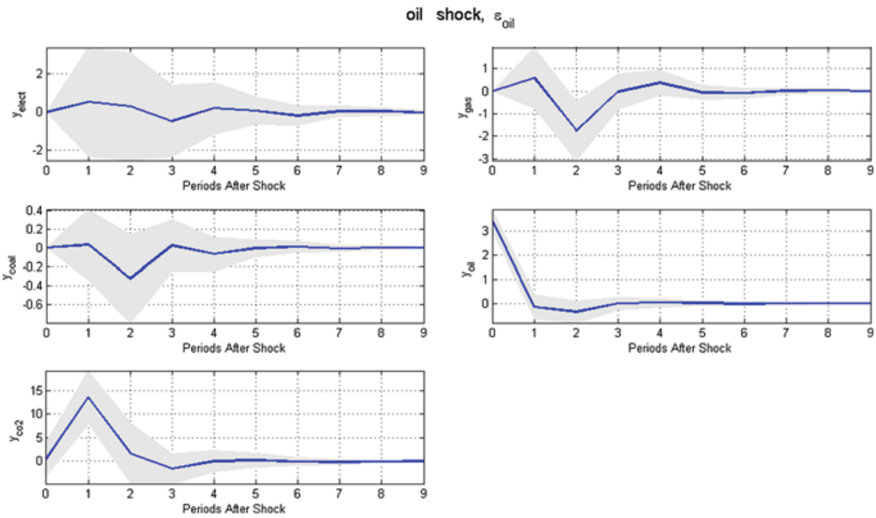


Fig. 3 continued

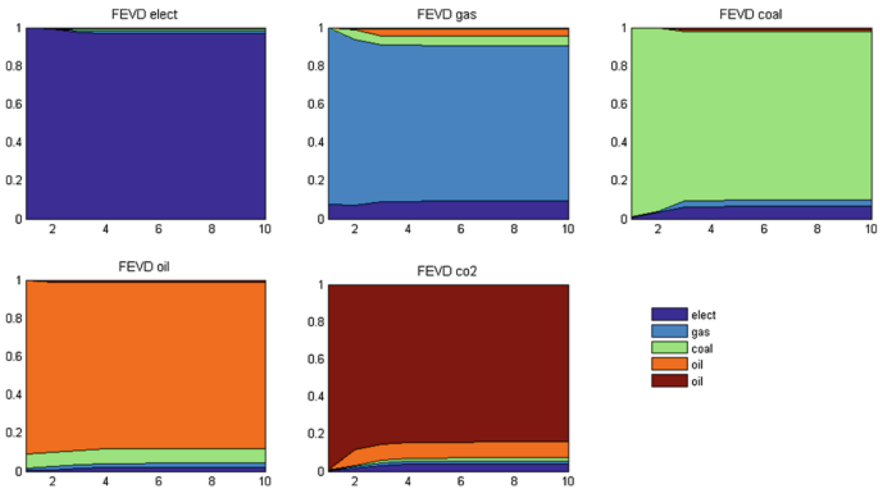


Fig. 4 Forecast error variance decomposition plot for the EEX market. FEVD stands for forecast error variance decomposition of electricity (elect.), gas, coal, oil and carbon (CO_2). The period analysed is January 2009 to July 2012 for the EEX market (corresponding to 148 observations). Values are plotted in relative (%) units. The results of the likelihood ratio (LR) test for lag length in the VAR for EEX (German market) favour the selection of two lags

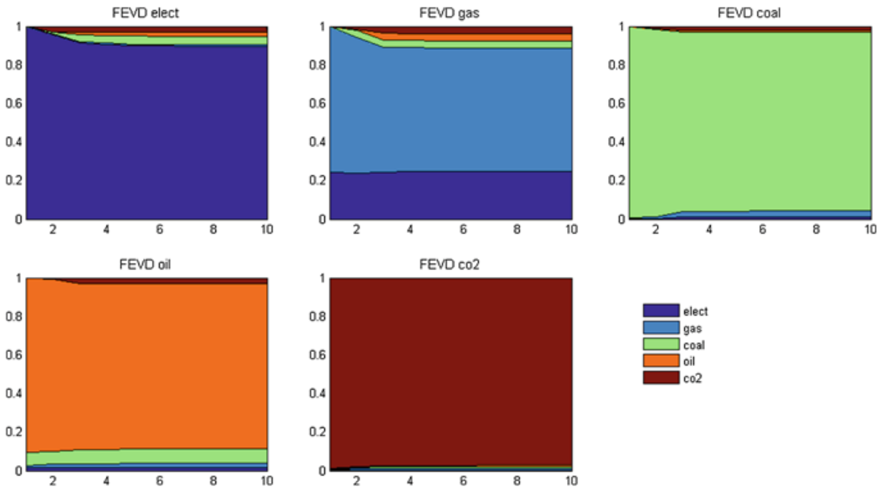


Fig. 5 Forecast error variance decomposition plot for the Powernext market. FEVD stands for forecast error variance decomposition of electricity (elect.), gas, coal, oil and carbon (CO_2). The period analysed is January 2009 to July 2012 for the Powernext market (corresponding to 184 observations). Values are presented in relative (%) units. The results of the likelihood ratio (LR) test and Akaike information criteria (AIC) for lag length in the VAR for Powernext (French market) favour the selection of two lags

equilibrium, be associated with a 1.2 % electricity price rise in the EEX market for a 5 week period (see Table 5).[16] Furthermore, since all coefficients are significant, all price variables are important to define the equilibrium vector.

For the German market, gas, coal and carbon prices may be considered the source of randomness, that represents the main driver of electricity. However, the coal price is the main driver of the source of randomness. Innovations in gas, electricity and carbon play a negligible role in explaining oil prices but the short-term effect increases over time.

Innovation effects in the carbon market to electricity and other fuel markets are null in the short term but the effect increase over time, and are stronger in oil, coal and electricity markets, in this order. Electricity and natural gas explain more uncertainty in coal prices in long horizons.

As we can also see, oil price seems to be mostly explained by coal prices, among the variables considered here (around 7.7 % for all periods). In sum, shocks in the German electricity, gas, coal and oil markets alone are not strong enough to influence the behaviour of the carbon price traded, the impact of which should be explained by factors other than those analysed here. Moreover, none of the fuels and carbon shocks seem to have a short-term effect on electricity, and carbon does

[16] Endogenous lagged variables were transformed into their natural logarithms to reduce variability, and thus we obtain elasticity values directly from parameter estimates.

Table 5 Forecast error variance decomposition (FEVD) for the EEX market

FEVD of: (variance due to a … shock)	Period/weeks	Elect.	Gas	Coal	Oil	CO_2
Elect.	1	100	7.6	0.6	0.5	0.1
	5	97.0	9.5	6.7	2.1	4.1
	10	96.9	9.6	6.8	2.2	4.1
	20	96.9	9.6	6.8	2.2	4.1
Gas	1	0.0	92.4	0.4	0.9	0.0
	5	1.2	81.2	3.1	2.0	1.3
	10	1.2	81.1	3.1	2.0	1.3
	20	1.2	81.1	3.1	2.0	1.3
Coal	1	0.0	0.0	99.0	7.7	0.4
	5	0.9	4.8	88.2	7.7	1.7
	10	1.0	4.8	88.2	7.7	1.7
	20	1.0	4.8	88.2	7.7	1.7
Oil	1	0.0	0.0	0.0	90.9	0.0
	5	0.1	3.9	1.0	86.8	8.5
	10	0.1	3.9	1.0	86.7	8.5
	20	0.1	3.9	1.0	86.7	8.5
CO_2	1	0.0	0.0	0.0	0.0	99.5
	5	0.7	0.6	1.0	1.3	84.5
	10	0.7	0.6	1.0	1.3	84.3
	20	0.7	0.6	1.0	1.3	84.3

FEVD stands for Forecast Error Variance Decomposition of electricity (elect.), gas, coal, oil and carbon (CO_2). The period analysed goes from January 2009 until July 2012 for the EEX market (corresponding to 148 observations). Values are presented in relative (%) units. The results of the Likelihood ratio (LR) test for lag length in the VAR for EEX (German market), favour the selection of two lags

not seem to be affected by gas, oil and electricity for the 1-week period in the German EEX market.

Our results for the short term (considering 1-week period) in the EEX market can be summarised as follows: gas shocks do not affect electricity and carbon; coal does not affect electricity and gas; oil is not the source of randomness for electricity, gas, coal and carbon; and carbon has a null impact on electricity, gas, coal and oil. Although electricity seems to have a negligible impact on carbon, the effect is null and vice versa.

Turning our attention to the Powernext market, we see that the oil price uncertainty in the French market is explained in the long term mainly by coal prices (7.7 %) and by carbon (3.0 %). However, in France the carbon price uncertainty is mostly explained by coal prices, 1 % for longer periods, followed by natural gas prices and oil. Since natural gas has only residual usage in this market (3.80 % in

Table 6 Forecast error variance decomposition (FEVD) for the French market

FEVD of: (variance due to a … shock)	Period/ months	Elect.	Gas	Coal	Oil	CO_2
Elect.	1	100.0	24.5	0.2	1.4	0.0
	5	90.0	24.8	1.1	1.6	0.4
	10	89.7	24.9	1.1	1.6	0.4
	20	89.7	24.9	1.1	1.6	0.4
Gas	1	0.0	75.5	0.6	1.2	0.3
	5	0.7	63.9	3.1	2.2	0.9
	10	0.8	63.7	3.1	2.2	0.9
	20	0.8	63.7	3.1	2.2	0.9
Coal	1	0.0	0.0	99.2	6.7	0.2
	5	4.2	3.8	92.8	7.4	1.0
	10	4.3	3.9	92.7	7.4	1.0
	20	4.3	3.9	92.7	7.4	1.0
Oil	1	0.0	0.0	0.0	90.7	0.4
	5	1.9	3.6	1.0	85.9	0.5
	10	2.0	3.7	1.0	85.8	0.5
	20	2.0	3.7	1.0	85.8	0.5
CO_2	1	0.0	0.0	0.0	0.0	99.1
	5	3.1	3.9	2.1	3.0	97.2
	10	3.2	3.9	2.1	3.0	97.2
	20	3.2	3.9	2.1	3.0	97.2

FEVD stands for forecast error variance decomposition of electricity (elect.), gas, coal, oil and carbon (CO_2). The period analysed is from January 2009 to July 2012 for the Powernext market (corresponding to 184 observations). Values are presented in relative (%) units. The results of the Likelihood ratio (LR) test and Akaike information criteria (AIC) for lag length in the VAR for Powernext (French market) favour the selection of two lags

2008), innovations in natural gas prices explain only a small percentage of both short and medium/long term carbon prices, which is even more evident for oil (1.35 % in 2008). As also observed here, expanded nuclear power generation could limit increases in electricity prices (Kara et al. 2008; Pinho and Madaleno 2011a) more than in Germany. In France, gas and carbon shocks are the biggest sources of randomness for electricity prices.

While oil and electricity are the major sources of randomness that drive the carbon market for EEX (about 8.5 and 4.1 %, respectively), this is the case of coal and gas in France (1.0 and 0.9 % respectively). Table 6 seems to indicate that coal and carbon are the major sources of randomness for electricity prices for Powernext (4.3 and 3.2 %, respectively), unlike EEX where it is gas and coal (1.2 and 1.0 %, respectively).

As electricity generation in the French market relies mainly on nuclear (77.17 % in 2007), innovations in carbon have an almost negligible impact on electricity prices (Table 6—3.2 %), though it is still higher than that of gas and oil. In fact, from the two markets under analysis, the results of forecast error variance decomposition for the German market seem to indicate that electricity prices hardly react to fuel price and carbon shocks (1.2 % for gas, 1 % for coal, 0.1 % for oil and 0.7 % for CO_2), which confirms the relationship between production source, market structure and electricity price response. These results are consistent with the fact that there has been a large increase in the use of wind for electricity production in the German market in recent years, but this will be addressed in future research due to the current unavailability of data to include this source.

Carbon is not contemporaneous for either market, meaning that 1-week returns (Tables 5 and 6 present a 0 % value for that period) is affected by other energy market. Therefore there are pressures from external factors not captured by the model.

Results reveal the absence of a unified energy market and, contrary to previous literature (Mohammadi 2009), it seems policies related to the coal industry continue to have a marginal influence on electricity, although the impact depends on the country's energy mix (for a more complete analysis see also Pinho and Madaleno 2011a).

On the power generation side, the price of gas affects operating choices more than the price of coal. High gas price encourages a greater use of coal; if everything else remains constant this should increase the demand for CO_2 allowances as coal emits twice the CO_2 content of natural gas. Therefore, if fossil fuels become more expensive, prices of EU ETS are likely to decrease or rise less than otherwise. Moreover, another hypothesis can be explored in this setting. Relationships between energy prices imply the possibility of substitution among the different forms of energy (results would obviously depend upon the country's energy mix).

Additionally, a more competitive market for electricity implies that spot market prices may respond promptly to price changes in input fuel source markets. The French market is the one that most deviates from the desired competition degree. In the EEX, a carbon innovation is reflected less in electricity prices. More recently, sharper increases in the price of allowances have led to speculation that electricity producers might have manipulated the allowance market so as to raise the allowance price, which then triggers an electricity price rise. If producers act as price takers, raising prices artificially is not easy. Since all of them benefit from a price increase, they might collude to manipulate the market and a reduction in market power would be the only solution to reduce speculation.

Moreover, it cannot be assumed that profits from trading in secondary carbon markets finance climate mitigation completely: an increasing number of participants in the carbon market participate to profit from speculation.[17] This trading of the same carbon allowance or carbon derivative takes place mainly among financial speculators who profit from speculating on the volatility of the price of carbon, and not because they are subject to emission reduction targets or have an interest in

[17] World Bank Carbon Finance Unit (2010): State and Trends of the Carbon Market 2010.

climate mitigation. Increased involvement of speculative actors with no interest in cost-effective implementation of greenhouse gas emission reduction targets may hinder the carbon market achieving its original objective. The motivations of the increasing number of speculative participants in the trading of carbon are opposed to the motivations of those trading to manage their cost of compliance with an emissions target. Participants whose trading is motivated by speculation will use their trading power to generate, exploit and profit from price volatility, as speculators profit from unpredictable price movements. Moreover, linking trading schemes that operate in jurisdictions where the enforcement capacity differs significantly will provide further ground for trading in "subprime" carbon derivatives in particular, given that much of the trading activity in carbon offsets takes place over-the-counter.

Even though the EU-Directive on trade of CO_2 allowances is a promising step, much more needs to be done to reach the ideal system. First, national governments in the EU allocate the CO_2 allowances in different ways; some are more generous than others and there is a natural influence of lobbying. Second, outside the EU there is no such system of allocation so that CO_2 intensive industries outside the EU have no incentive to economise on their CO_2 emissions. In that case, cooperation between EU and non-EU companies could result in additional allowances. Production technologies for electricity differ greatly in their CO_2 emissions and it proves difficult to reduce the aggregate level of emissions by governmental directives.

It can also be questioned whether allowances price act as reliable price signals for companies to invest in less CO_2 intensive production technologies. If a company uses these desirable technologies, it may not be awarded allowances in the future so that it cannot sell these and gain additional profits. Thus, the net benefit from switching to a technology without CO_2 emissions is dubious. Moreover, reducing the use of CO_2 intensive technologies would foster the debate on the use of nuclear energy.

5 Conclusions

In this chapter, we analyse the relationships between electricity prices, primary energies prices used in electricity generation and the price of carbon dioxide emission permits in France and Germany using a VECM model. The difference in responses to carbon constraints in the electricity generation sector were accounted and allowed us to of the EU ETS given the energy mix heterogeneity of both countries for the Phase II period.

We were able to show that the impact of carbon constraints on energy markets depends on the countries' energy mix. This allows us to conclude that it is not always producers in countries using predominantly fossil fuels, which are great carbon emitters, that undertake more carbon coercion; results indicate that they do not necessarily include the price of emission permits in their electricity generation and cost functions (EEX). Using other sources of electricity production like wind might have helped us obtain more useful results to explain this. We also found that oil and electricity are the major sources of randomness that drive the carbon market for EEX, but not vice versa given that it is coal that most impacts oil, and gas that most impacts electricity. Furthermore, natural gas is significantly affected by electricity in both the short and the long term. We also found that coal and gas have the biggest impact on electricity prices.

Coal and gas are the major sources of randomness for carbon in France; however, coal is mostly affected by gas, and gas by electricity in this market. Whereas carbon is the major source of randomness for electricity and gas in Powernext, this is the case for coal and oil in EEX. For Powernext, we also found that coal and carbon have the biggest impact on electricity prices. Also, coal is mostly used as a power source in EEX and explains carbon better in this market than in Powernext. However, carbon explains coal more in Powernext than in EEX.

Hitherto, it has been understood that policies related to the coal industry have a marginal influence on electricity prices. Empirical results seem to show that policies towards clean air still do not imply a rise in the cost of coal and electricity production, but we have also seen that the coal market is the major source of randomness for oil prices in both France and Germany. Throughout the period analysed, the efficiency of the European market for emission allowances was therefore unable to compel electricity producers to eliminate their emissions and invest in cleaner technologies, whereas the desired effects also depended on policies pursued for distributing allowances.

Given that CO_2 markets are relatively new markets, we could improve the quality of results by repeating the analysis some years from now because more data becomes available as markets evolve. In addition, it would be productive to use daily data which is currently impossible due to data restrictions. Moreover, portfolio analysis using these different commodities from a trader's point of view could offer valuable insights into necessary strategies for these markets.

References

Alberola E, Chevallier J, Chèze B (2008) Price drivers and structural breaks in European carbon prices 2005–2007. Energy Policy 36(2):787–797

Alberola E, Chevallier J (2009) European carbon prices and banking restrictions: evidence from Phase I (2005–2007). Energy J 30(3):107–136

Benz E, Trück S (2009) Modelling the price dynamics of CO_2 emission allowances. Energy Econ 31(1):4–15

Bhardwaj M, Wadadekar A (2010) Carbon credit for environmental management. NSE Newsletter 1–4

Bunn DW, Fezzi C (2007) Interaction of European carbon trading and energy prices. FEEM working paper, No. 63

Chevallier J (2012) CO_2 price fundamentals. In: Econometric analysis of carbon markets: the European union emissions trading scheme and the clean development mechanism. pp 19–54. http://www.springer.com/978-94-007-2411-2. Accessed September 2012

Convery FJ, Redmond L (2007) Market and price developments in the European union emissions trading scheme. Rev Environ Econ Policy 1(1):88–111

Creti A, Jouvet P-A, Mignon V (2012) Carbon price drivers: Phase I versus Phase II equilibrium? Energy Econ 34:327–334
Daskalakis G, Psychoyios D, Markellos R (2009) Modeling CO_2 emission allowance prices and derivatives: evidence from the European trading scheme. J Bank Finan 33(7):1230–1241
Dellink R, Jamet S, Chateau J, Duval R (2010) Towards global carbon pricing: direct and indirect linking of carbon markets. OECD environmental working paper, No. 20, OECD Publishing, 1–39
Engle RF, Granger CWJ (1987) Co-integration and Error Correction: Representation, Estimation, and Testing. Econometrica, 55(2):251–276
Ellerman D, Convery F, De Perthuis C (eds) (2010) Pricing carbon: the European union emissions trading scheme. Cambridge University Press, Cambridge, United Kingdom
Fell H (2008) EU-ETS and Nordic electricity: a CVAR analysis. RFF DP 08-31. Discussion paper
Ferkingstad E, Loland A, Wilhelmsen M (2011) Causal modelling and inference for electricity markets. Energy Econ 33(3):404–412
Friends of the Earth (2010) Friends of the Earth Europe annual review 2010 for the people | for the planet | for the future. Brussels, Belgium. http://www.foeeurope.org/sites/default/files/publications/FoEE_annual_review_2010.pdf
Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton
Honkatukia J, Mälkönen V, Perrels A (2007) Impacts of the European emission trade system on Finnish wholesale electricity prices. VAAT Discussion paper: 2007-405; Government Institute for Economic Research, Helsinki
Johansen S (1992) Testing weak exogeneity and the order of cointegration in UK money demand data. J Policy Model 14(3):313–334
Kara M, Syri S, Lehtilä A, Helynen S et al (2008) The impacts of EU CO_2 emissions trading on electricity markets and electricity consumers in Finland. Energy Econ 30(2):193–211
Madaleno M, Pinho C (2011a) Risk premia in CO_2 allowances: spot and futures prices in the EEX market. Manag Environ Qual: Int J 22(5):550–565
Madaleno M, Pinho C (2011b) Hedging performance and multiscale relationships in the German electricity spot and futures markets. J Risk Finan Manag 2:26–62
Mansanet-Bataller M, Pardo A, Valor E (2007) CO_2 Prices Energy and Weather. Energy J 28 (3):67–86
Mohammadi H (2009) Electricity prices and fuel costs: long-run relations and short-run dynamics. Energy Econ 31(3):503–509
Paolella M, Taschini L (2008) An econometric analysis of emission allowance prices. J Bank Finan 32(10):2022–2032
Pinho C, Madaleno M (2011a) CO_2 emission allowances and other fuel markets interaction. Environ Econ Policy Stud 13(3):259–281
Pinho C, Madaleno M (2011b) Multiscale analysis of European electricity markets comovements. Tékhne Polyt Stud Rev IX(15):33–57
Prabhakant P, Tiwari G (2009) Evaluation of carbon credits earned by energy security in India. Int J Low Carbon Technol 4(1):42–51
Reinaud J (2007) CO_2 allowances and electricity price interaction: impact on industry's electricity purchasing strategies in Europe. IEA information paper
Seifert J, Uhrig-Homburg M, Wagner M (2008) Dynamic behavior of CO_2 spot prices. J Environ Econ Manag 56(2):180–194
Sijm J, Chen Y, Donkelaar M, Hers JS, Scheepers MJ (2006) CO_2 price dynamics: a follow-up analysis of the implications of EU emission trading for the price of electricity. ECN report, No: ECN-C-06-015
Silva P, Soares I (2008) EU spot prices and industry structure: assessing electricity market integration. Int J Energy Sector Manag 2(3):340–350
Springer U (2003) The market for tradable GHG permits under the Kyoto protocol: a survey of model studies. Energy Econ 25:527–551
Uhrig-Homburg M, Wagner M (2006) Success chances and optimal design of derivatives on CO_2 emission certificates. Working paper, University of Karlsruhe

Uhrig-Homburg M, Wagner M (2008) Derivative instruments in the EU emissions trading scheme: an early market perspective. Energy Environ 19(5):635–655

Wei YM, Liu LC, Fan Y, Wu G (2008) China energy report 2008: carbon emission research. Science Press, Beijing

Widerberg A, Wräke M (2009) The impact of the EU emissions trading system on CO_2 intensity in electricity generation. Working papers in economics, University of Gothenburg

Wolak FA (1998) Market design and price behavior in restructured electricity markets: an international comparison. Stanford University, Working paper, 1–64

World Bank Carbon Finance Unit (2010) Carbon finance for sustainable development annual report 2010. Washington, DC. www.carbonfinance.org, https://wbcarbonfinance.org/docs/64897_World_Bank_web_lower_Res..pdf

World Energy Council (2010) Survey of Energy Resources. London, United Kingdom. ISBN: 978 0 946121 021

Zachmann G, von Hirschhausen C (2008) First evidence of asymmetric cost pass-through of EU emissions allowances: examining wholesale electricity prices in Germany. Econ Lett 99 (3):465–469

An Overview of Electricity Price Regimes in the U.S. Wholesale Markets

José G. Dias and Sofia B. Ramos

Abstract The U.S. electricity market is organized in several deregulated regional markets. In this paper we specify a multi-regime switching model to study price dynamics of electricity in the U.S. markets. Our results show that electricity prices from the West and East coasts have different regime dynamics with the latter prices switching more frequently between regimes. Additionally, our methodology suggests that electricity prices are better parameterized by four regimes: the base regime with low volatility; a spike up and a reverse regime both with high volatility and short duration; finally, a fourth one has extremely high volatility. This latter regime describes West coast prices during the California electricity crisis, but East coast prices are also frequently in that regime. We find evidence of price synchronization in the lowest and highest volatility regimes, i.e., prices from the East and West coasts tend to be in the same regimes at the same time.

Keywords U.S. electricity markets · Deregulation · Electricity prices · Regime-switching models · Volatility

1 Introduction

The electricity business activity can be roughly characterized by three sectors: generation, transmission, and distribution, which were usually tied within a utility. Generation is the process of generating electric energy from other forms of energy

Financial support from Fundação para a Ciência e Tecnologia is greatly acknowledged (PTDC/EGE-GES/103223/2008).

J.G. Dias · S.B. Ramos
Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL), Lisbon, Portugal

S.B. Ramos (✉)
Department of Finance, ISCTE – Instituto Universitário de Lisboa, Business Research Unit (BRU-UNIDE), Lisbon, Portugal
e-mail: sofia.ramos@iscte.pt

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_9

such as hydro energy, fossil fuels, harnessing wind, solar, or through nuclear fission. After being generated, electricity is distributed through high-voltage, high-capacity transmission lines to local regions, where it is consumed. When the electricity reaches the local destination of consumption, it is transformed into a lower voltage and sent through local distribution wires to end-use consumers.

In the U.S., for many years each of these segments was investor-owned but state-regulated or owned by the local municipality. But the 1980s saw the introduction of a wave of deregulatory reforms that reached the electricity sector. Reforms were implemented with the argument that competitiveness would rise and benefit consumers by lowering prices in both the short and long runs.

The establishment of a competitive wholesale electricity market, i.e., a market where competing agents offer and buy electricity was a key element of the deregulation process. While wholesale pricing used to be of the exclusive domain of large retail suppliers, a market in a competitive framework should open up to new participants such as generators, retailers, or financial intermediaries or end-users. To reach this goal, the Federal Energy Regulatory Commission (FERC), the regulatory agency, introduced rules such open access to transmission service tariffs and on the availability of transmission service of networks. Moreover, transmission owners had to provide access to their networks at cost-based prices to end discriminatory practices against unaffiliated generators.

Market power and the potential upsurge of prices are a major issue in the market design of wholesale markets. As will be explained in more detail below, the physical features of electricity favor imperfect competition, and ultimately deregulation could have adverse effects by increasing prices for end-users. Knittel and Roberts (2005) refer that when regulated prices were set by state public utility commissions in order to curb market power and ensure the solvency of the firm. Price variation was minimal and under the strict control of regulators, who determined prices largely on the basis of average costs. In contrast, a wholesale market is based on competitive bidding of supply and demand, and prices are set by market clearance. Given that electricity demand has frequent fluctuations (e.g., extreme temperatures) and that there are no inventories to buffer shocks, prices would fully absorb shocks. Price jumps and spikes in volatility are then inevitable outcomes that must be monitored. Concerns about market power were substantiated by the California crisis in 2000–2001, when market power and exploitation of market design imperfections caused an explosion in wholesale prices.

The deregulated nature of the U.S. electricity market as well as its fragmented structure with many wholesale markets, makes it an interesting case for analyzing the dynamics of prices after deregulation. The literature comparing U.S. electricity prices in different locations is scant. Hadsell et al. (2004) compare electricity volatility in five regions of the U.S. for the period 1996–2001 using a TARCH model; Park et al. (2006) use a vector autoregressive (VAR) model to analyze spot prices in different parts of the U.S. for the period 1998–2002. They find that electricity markets in the Western U.S. are separated from the Eastern markets at contemporaneous time, but this separation disappears for longer time horizons. The relationships between the markets depend on physical assets (such as transmission lines) and institutional arrangements.

Our study analyzes price dynamics of U.S. regional markets by regime-switching models (RSM). These models, introduced by Hamilton (1989), have been extensively used to model electricity prices as they accommodate well electricity price features such as asymmetric volatility, jumps, and spikes.[1] The computational burden in model estimation, which increases with the number of time series, observations, and regimes is a hindrance to their empirical application. Our estimation algorithm overcomes these limitations and allows the study of the cyclical behavior of several electricity price time series in a parsimonious way, providing new insights on the existence of common regimes and the synchronization between them. Moreover, this approach recognizes different regime-switching dynamics of electricity prices, so far not addressed in the literature. In addition, the flexible modeling of observed returns using Gaussian mixture distributions makes it more appropriate for non-Gaussian returns (see, e.g., McLachlan and Peel 2000; Dias and Wedel 2004).

To study price dynamics in different regions of the U.S., we take the Dow Jones U.S. Electricity Price Indexes. These price indexes cover several geographical regions of the United States. We conclude that prices in the same U.S. region share the same regime dynamics, i.e., prices of the East (West) coast markets behave similarly. The best model parametrization has four regimes. The extremely high volatility regime describes West Coast prices during the California electricity crisis, but prices of the East coast markets are also frequently in that regime. Regional electricity markets seem to differ in the time spent in each regime. West market prices spend more time in the low volatility regime than East coast markets. Strikingly, the time they spent in the spike regime is similar despite the episode of the California crisis. To address the question of whether prices of the East and West coasts are in the same regime at the same time, we compute synchronization measures between and within regimes. We find evidence of price synchronization in the lowest and highest volatility regimes, i.e., prices from the East and West coasts tend to be in those regimes at the same time.

The rest of the chapter is organized as follows. Section 2 gives an overview of the main changes in the U.S. electricity markets. Section 3 describes the data. Section 4 introduces the econometric methodology. Section 5 presents and discusses the empirical results. Section 6 analyzes the synchronization between the different electricity markets and, finally, Sect. 7 concludes the paper.

[1] See, e.g., Fong and See (2002), Huisman and Mahieu (2003), Bierbrauer et al. (2007), Haldrup et al. (2010), and Janczura and Weron (2010).

2 The Establishment of a Wholesale Market

The deregulation process targeted two key features of the utility sector: monopolies and natural barriers to entry. Joskow (1997) describes that the deregulation process had two main goals. First, to separate the potentially competitive functions of generation and retail from the natural monopoly functions of transmission and distribution. Second, to establish a wholesale electricity market and a retail electricity market.

Ideally, a wholesale market should have a sound free-market base such as competitive supply offers, demand bids and prices set by market-clearance. To achieve this, it urged to push for the breakdown of barriers to entry and attraction of new players into the market. In 1996 a set of measures were implemented to ease entry and enhance competition. For instance, established transmission owners had (i) to provide access to their networks at cost-based prices, (ii) to end discriminatory practices against unaffiliated generators and marketers, (iii) to expand their transmission networks if they did not have the capacity to accommodate requests for transmission service, and (iv) to provide non-discriminatory access to information required by third parties to make effective use of their networks.

These measures were reinforced by the FERC Order 2000 issued in December 1999. This contained a new set of regulations designed to facilitate the "voluntary" creation of large regional transmission organizations to solve problems created by the balkanized control of U.S. transmission networks and alleged discriminatory practices affecting independent generators and energy traders seeking to use the transmission networks of vertically integrated firms.

The particular features of the electricity operations are a hindrance to competition. Monopolies emerge as an outcome of economies of scale of the generation process and the losses from long-distance transmission. To truly compete in the distribution sector, rival firms should duplicate wire networks. However, the duplication of infrastructures is inefficient as there is a need to keep the system adequacy, i.e., the balance between inflow and outflow at all times. The failure to balance leads to the collapse of the grids which has severe economic losses.[2,3]

Market power also arises because of the inelasticity of energy demand. This naturally leads to high prices at peak times as demand rises above the production capacity of generators and further price increases result in little additional supply or reduction of demand. The prices then naturally reflect the scarcity of supply relative to demand.

---

[2] The grid needs to be constantly surveyed and cannot be under or overloaded. This implies that if wires owned by different companies were allowed to interconnect to form a single network, then the flow on one line could affect the capacity of other lines in the system to carry power creating risky unbalances.

[3] A recent case of grid collapse happened in India. India has increased the number interconnections between regional grids, approaching a single national grid. A breakdown in one part of the grid loaded other parts of the grid massively making the system collapse.

Given that electricity is not storable, inventories cannot be used to load the grid and smooth prices over time. As a result, deregulated prices are characterized by volatility that varies over time and occasionally reaches extremely high levels, commonly known as “price spikes”.

Market power and imperfect competition have well-known economic implications such as high profits for sellers at the cost of higher prices for consumers contradicting the aims of the deregulation process. Moreover, increased volatility and subsequent losses represent additional risks for market participants which for instance has led to the emergence of power derivatives markets. Market power has other detrimental effects on economic growth because high energy costs imply an increase of costs for firms and price volatility also creates uncertainty which tends to postpone investment decisions.

Finally, market power affects the reliability and credibility of wholesale markets. The California electricity crisis in 2000–2001 is a good illustration of what can go wrong in the deregulation process due to imperfections in the deregulated market. Energy traders created artificial shortages in days of peak demand to increase prices and company profits. The explosion of prices and the rolling blackouts adversely affected many businesses dependent upon a reliable supply of electricity, and inconvenienced a large number of retail consumers.[4] The California state suffered from multiple large-scale blackouts, and one of the state’s largest energy companies collapsed with harmful economic effects.[5]

3 Data

We use Dow Jones U.S. Electricity Price Indexes to analyze electricity prices in different regions of the U.S. Indexes. These prices cover different regions of the U.S. market, namely the West and East coasts. From the West region, and conditional on data availability, we use California and Oregon Border (COB), Four Corners (Utah, Colorado, New Mexico and Arizona), Mid Columbia (Washington) and Palo Verde (Arizona) prices indexes; from the East region, we use CINERGY (Ohio, Indiana) and PJM (Pennsylvania) interconnection which is the world’s largest competitive wholesale electricity market. These indexes are volume-weighted averages of wholesale electricity transactions and provide a clear spot market indication for over-the-counter trading in that region.

Our sample covers prices from 6th January 1999 through 7th July 2010, for a total of 601 price observations. Prices are weekly, from Wednesdays like Mjelde and Bessler (2009), and in U.S. dollars. Let P_{it} be the observed weekly closing price

[4] Energy traders took power plants offline for maintenance in days of peak demand. This increased power prices sometimes by 20 times its normal price.

[5] For a detailed explanation of California electricity crisis, we refer to Faruqui et al. (2001), Moulton (2005), and Woo (2001).

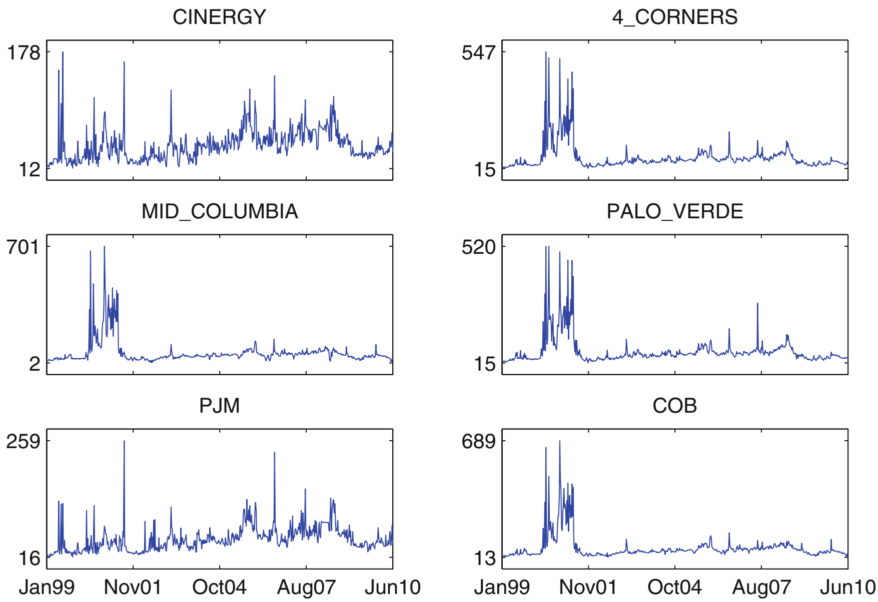


Fig. 1 Dow Jones electricity price indexes

of market i on day t, $i = 1, \ldots, n$ and $t = 0, \ldots, T$. Thus, the weekly rates of return are defined as the log-rate percentage: $y_{it} = 100 \times \log\left(P_{it}/P_{i,t-1}\right),\ t = 1, \ldots, T$.

Figure 1 depicts electricity prices for the entire period. Electricity prices show extraordinary volatility during 2000–2001, the period of the California electricity crisis. Prices in the East coast—CINERGY and PJM—also tend to show frequent price spikes.

Table 1 summarizes the descriptive statistics for the returns. The mean is positive for all series, except for CINERGY. As expected, electricity returns show high dispersion (standard deviation) and kurtosis. Interestingly, West region prices tend to show positive skewed distributions, whereas East coast series are negatively skewed. The heavy tails and skewness of the distributions turn out to reject the normality for all time series (Jarque Bera test, *p-value* < 0.001).

The stylized characteristics of these price returns—cyclical behavior, jumps, and spikes—provide ground for applying regime switchings models.

4 Methodology

The methodology applied in this work falls within the regime switching framework. Regime switching models (RSM) have been extensively applied in economics and finance research and the modeling of electricity prices is no exception (see, e.g.,

Table 1 Summary statistics

	Mean	Std. deviation	Skewness	Kurtosis	Jarque-Bera test	
					Statistics	p-value
CINERGY	−0.014	36.079	−0.319	8.038	623.84	0.000
4_CORNERS	0.096	28.441	0.367	14.303	3117.11	0.000
MID_COLUMBIA	0.070	34.939	0.194	13.226	2543.13	0.000
PALO_VERDE	0.067	28.988	0.292	14.280	3099.35	0.000
PJM	0.049	34.327	−0.460	9.383	1007.80	0.000
COB	0.064	29.718	0.270	16.089	4171.58	0.000

This table reports descriptive statistics and the Jarque-Bera test of normality for electricity prices returns. The returns are percentage log-rate returns (weekly data) and are from 06-01-1999 to 07-07-2010

Deng 1998; Ethier and Mount 1998). In a meta-analysis of several econometric approaches, Bierbrauer et al. (2007) conclude that a major strength of regime switching models over other econometric models is its flexibility in accommodating extreme observations. In particular, the model allows for consecutive spikes in a very natural way, as well as the switching of prices to the 'normal' regime after a spike.[6] In short, these models are a parsimonious representation the unique characteristics of power prices. Moreover, regimes are able to describe the price jumps caused by different levels of demand and supply (see, e.g., Andreasen and Dahlgren 2006; Bierbrauer et al. 2007; Deng 1998; Ethier and Mount 1998; Huisman and Mahieu 2003; Janczura and Weron 2010, 2012). In particular, they capture specific characteristics such as the spiky and nonlinear behavior of electricity prices (Bierbrauer et al. 2007; Mari 2006; Weron et al. 2004). Thus, the introduction of nonlinearities by the regime-switching mechanism admits temporal breaks in model dynamics.

The application of RSM has been hindered by two (related) practical issues: computational burden and the number of regimes allowed. Because of the computational burden, seminal works set up two regimes a priori (see Deng 1998 and Ethier and Mount 1998). Huisman and Mahieu (2003) are the first to propose a three-regime model, but with constraints: the initial jump regime is immediately followed by the mean-reversing regime and then moves back to the base regime. Using electricity price data from the Dutch, German, and the United Kingdom markets, they found that a regime-switching model performs better than a stochastic jump model specification for both mean-reversion and spikes. Our work departs from previous studies because we do not impose a priori the number of regimes that

[6] Other econometric approaches such as stochastic jump models have been applied in energy price modeling. Comparisons show that regime-switching models present many advantages in modeling the spiky and nonlinear behavior of electricity prices over competing techniques (Bierbrauer et al. 2007; Janczura and Weron, 2010; Mari 2006; Weron et al. 2004).

best captures the features of the electricity time series and proposes a joint analysis of distinct electricity markets.

We use the heterogeneous regime-switching model (HRSM) in Dias et al. (2008, 2009) and Ramos et al. (2011). This statistical model defines classes of regime-switching models based on the similarity of the dynamics within each class. An advantage of this approach is that we can see whether different time series share regimes (or regime dynamics). This model assumes two different types of discrete latent variables or states:

1. each time series belongs to a specific group or cluster, say $w \in \{1, \ldots, S\}$. A model with S clusters is called a HRSM-S;
2. each specific time series is modeled as a regime-switching model with K regimes and $z_{it} \in \{1, \ldots, K\}$ for all $t = 1, \ldots, T$ is the state occupied by the time series i at time t. Transitions between the K regimes over time follow a first-order Markov process.

Based on the definition of y_{it} introduced previously, let $f(\mathbf{y}_i; \varphi)$ be the density function of the electricity time series i. The HRSM-S is defined by:

$$f(\mathbf{y}_i; \varphi) = \sum_{w_i=1}^{S} \sum_{z_{i1}=1}^{K} \cdots \sum_{z_{iT}=1}^{K} f(w_i) f(z_{i1}|w_i) \prod_{t=2}^{T} f\left(z_{it}|z_{i,t-1}, w_i\right) f(\mathbf{y}_i|w_i, z_{i1}, \ldots, z_{iT}). \quad (1)$$

where: (a) $f(w_i)$ is the probability of time series i belongs to cluster w; (b) $f(z_{i1}|w_i)$ is the initial-regime probability, i.e., the probability that time series i starts the sequence in regime k conditional on belonging to the cluster w; (c) $f\left(z_{it}|z_{i,t-1}, w_i\right)$ is a latent transition probability, i.e., the probability of being in a particular regime at time t conditional on the regime at time $t-1$ and within the cluster w. Assuming a time-homogeneous transition process, $p_{jkw} = P\left(Z_{it} = k|Z_{i,t-1} = j, W_i = w\right)$ is the relevant parameter. Thus, for cluster w the transition probability matrix is

$$\mathbf{P}_w = \begin{pmatrix} p_{11w} & \cdots & p_{1Kw} \\ \vdots & \ddots & \vdots \\ p_{K1w} & \cdots & p_{KKw} \end{pmatrix},$$

with $\sum_{k=1}^{K} p_{jkw} = 1$. Thus, the HRSM-S extends the traditional RSM as it allows cluster specific regime-switching dynamics.

The last term in Eq. (1) is the observed data density conditional on the regimes, $f(\mathbf{y}_i|w_i, z_{i1}, \ldots, z_{iT})$. Assuming that the observed return at a particular time depends only on the regime at that time, i.e., conditional on the latent state z_{it}, the response y_{it} is independent of returns and regimes at other time points:

$$f(\mathbf{y}_i|w_i, z_{i1}, \ldots, z_{iT}) = \prod_{t=1}^{T} f(y_{it}|z_{it}). \quad (2)$$

The probability density of the return i at time t conditional on the regime occupied at time t, $f(y_{it}|z_{it})$, is assumed to have a normal density function. For regime k, this distribution is characterized by the parameter vector $\boldsymbol{\theta_k} = (\boldsymbol{\mu_k}, \boldsymbol{\sigma_k^2})$, i.e., the expected return or mean (μ_k) and risk or variance (σ_k^2). The right-hand side of Eq. (1) shows that we are dealing with a mixture model consisting of time-constant latent variable w_i and T realizations of the time-varying latent variable z_{it}. As in any mixture model, the observed data density $f(\mathbf{y}_i; \boldsymbol{\varphi})$ results from marginalizing over the latent variables, in this case over the $S \cdot K^T$ mixture components (see McLachlan and Peel 2000). Since $f(\mathbf{y}_i; \boldsymbol{\varphi})$ is a mixture of densities across clusters and regimes, it defines a flexible Gaussian mixture model that can accommodate deviations from normality in terms of skewness and kurtosis (see, e.g., Dias and Wedel 2004 and Pennings and Garcia 2004).

The estimation of the HRSM-S parameters is performed by the maximum likelihood (ML) method. Given the presence of missing data (clusters and regimes), the expectation-maximization (EM) algorithm (Dempster et al. 1977) is a natural choice for maximizing the log-likelihood function: $\ell(\boldsymbol{\varphi}; \mathbf{y}_i) = \sum_{i=1}^{n} \log f(\mathbf{y}_i; \boldsymbol{\varphi})$. Since the EM algorithm at the Expectation-step requires the computation and storage of $S \times K^T$ entries of $f(w_i, z_{i1}, \ldots, z_{iT}|\mathbf{y}_i)$ for each time series, computation time and computer storage increases exponentially with the number of time points. However, for regime-switching models, a special variant of the EM algorithm has been proposed that is usually referred to as the forward-backward or Baum-Welch algorithm (Baum et al. 1970) and will be used here.

A key issue in regime-switching modeling is the decision on the optimal number of regimes needed. For the HRSM-S, the selection of the number of clusters (S) and regimes (K) is based on the Bayesian information criterion (BIC) of Schwarz (1978) given by

$$BIC_{S,K} = -2\ell_{S,K}(\hat{\boldsymbol{\varphi}}; \mathbf{y}) + N_{S,K} \log n, \qquad (3)$$

where $N_{S,K}$ is the number of free parameters in the regime-switching model and n is the sample size. The combination (S, K) with the minimum BIC identifies the best model.

5 Empirical Results

This section reports the estimates of the HRSM-S applied to electricity indexes. We estimate models with the density function given by Eq. (1) for different values of S ($S = 1, \ldots, 8$) and K ($K = 1, \ldots, 8$). For each combination, we use 1000 different sets of random starting values to minimize the impact of local maxima. A solution with two latent classes ($S = 2$) and four regimes ($K = 4$) yields the lowest BIC value (log-likelihood = −16258.8; number of free parameters = 39; and BIC = 32587.6). This means that the best solution incorporates two types of regime dynamics and four regimes.

Table 2 Estimated prior and posterior probabilities and modal classes

	Latent class 1	Latent class 2	Modal
Prior probabilities	0.643	0.357	
Posterior probabilities			
CINERGY	0.000	1.000	2
4_CORNERS	1.000	0.000	1
MID_COLUMBIA	1.000	0.000	1
PALO_VERDE	1.000	0.000	1
PJM	0.000	1.000	2
COB	1.000	0.000	1

This table reports the electricity prices level probabilities and modal latent class. Prior probabilities provide the size of each latent class or cluster and posterior probabilities express the evidence that a given electricity time series belongs to a given latent class. The maximum posterior probability indicates the modal latent class

Table 2 summarizes the results for the distribution of electricity prices across latent classes. Each latent class indicates a cluster, i.e., a group of prices that shares the same regime dynamics. Electricity prices are classified into two clusters, indicating that East coast electricity prices have different dynamics from those of the West coast (i.e., CINERGY and PJM are in latent class 2, whereas other price indexes are in latent class 1). The class assignments always have probability one, i.e., there is no uncertainty about the classification of these time series.

Regimes are described in Table 3. The first set of rows shows the estimates of the probability $P(Z)$: the average proportion of returns in each regime over time. Overall, electricity prices are in regime 1 16.2 % of the time, in regime 2 7.4 % of the time, in regime 3 58.4 % of the time, and in regime 4 18.0 % of the time.

The next set of rows presents the expected returns and variance of each regime. Regimes are sorted by mean returns; regime 1 has the lowest returns and regime 4 the highest. Regimes 1–3 have negative mean returns. Regime 1 has very negative mean returns and high volatility, while regime 2 has negative mean returns and the highest volatility of all regimes; regime 3 has negative mean returns and the lowest volatility, which resembles 'the base regime'.[7] Regime 4 has positive returns and the variance is similar to that of regime 1, it is the 'up spike' or the 'up' regime. The daily standard deviation for regimes 1, 2, 3 and 4 are 22.7, 88.0, 11.6 and 24.6, respectively. The extremely high volatility of regime 2 should be noted as it shows levels not reported in previous studies.

Results in Table 4 shows why electricity prices do not share the same dynamics, or are in different clusters. The first row gives the estimated probabilities of being in

[7] We will apply terminology common to previous papers to characterize regimes: base, reverse, and spike regimes.

Table 3 Description of regimes

Regimes	1	2	3	4
P(Z)	0.162	0.074	0.584	0.180
	(0.023)	(0.018)	(0.052)	(0.028)
Return (mean)	−25.453	−0.756	−0.103	23.833
	(1.986)	(5.546)	(0.286)	(2.050)
Risk (variance)	517.497	7763.911	134.130	604.573
	(49.701)	(874.390)	(7.657)	(54.417)

This table reports the estimated marginal probabilities of regimes—$P(Z)$: is the average proportion of markets in each regime over time, means, and variances. Standard errors are reported in round brackets

a particular regime for each cluster, i.e., electricity time series have different regime probabilities across classes.

West coast prices (latent class 1) have 0.67 probability of being in regime 3 (the base regime), whereas in the East coast (latent class 2) this probability is reduced to 0.43. East coast prices spend more time in spike regimes with probabilities of regimes 1 and 4 adding up to 0.488; on the other hand, probabilities of spike regimes from the West coast add up to only 0.261. Notwithstanding, both have a similar probability of being in the crisis regime, regime 2, despite the well-known crisis in California.

In the second row, we present the transition probabilities between the regimes for each group. It means that the closer the diagonal value is to one, the higher the regime persistence. In other words, once an electricity price enters a given regime, it is likely to stay in the same regime for some period of time.

All prices show regime persistence for regimes 2 and 3, those with the highest and lowest volatility. Inversely, regimes 1 and 4 do not show persistence, i.e., the likelihood of continuing in regime 1 and 4 is very small (spike regimes). West coast prices have a 0.802 probability of jumping from regime 4 to regime 1 and East coast prices a 0.733 probability. This means that after spiking up, there is a high probability that prices will go down. It is likely that prices from regime 1 jump to regime 3, the base regime, or spike again to regime 4, highlighting a very dynamic nature.

The (mean) sojourn time is the expected time that a price takes to move out of a given regime and is measured in weeks. It is given for regime k and conditional on the latent class w by $1/(1-p_{kkw})$. Naturally, regimes with regime persistence have higher sojourn times. Regimes 2 and 3, the ones that show persistence, have sojourn times of 8 and 14 weeks for West coast, while mean times for spike regimes are around 1 week. Prices from the East coast stay in regime 3, the base, for shorter periods of time. Again, the evidence suggests that returns in the East coast are more volatile and change more often between regimes than those of the West coast.

Figures 2 and 3 show the regime-switching dynamics in electricity prices in each group through time. It depicts the posterior probability of being in each regime at period t. Electricity has a dynamic nature and its frequent switches between regimes

Table 4 Estimated regime occupancy and transition probabilities within each latent class

Regimes	1	2	3	4
Latent class 1				
$P(Z\|W)$	0.126	0.070	0.670	0.135
	(0.014)	(0.022)	(0.035)	(0.016)
Regime 1	0.138	0.067	0.370	0.425
	(0.049)	(0.022)	(0.048)	(0.051)
Regime 2	0.001	0.876	0.001	0.122
	(0.007)	(0.035)	(0.003)	(0.036)
Regime 3	0.001	0.000	0.929	0.070
	(0.003)	(0.002)	(0.010)	(0.010)
Regime 4	0.802	0.001	0.003	0.194
	(0.058)	(0.007)	(0.018)	(0.056)
Sojourn time	1.160	8.052	14.124	1.240
Latent class 2				
$P(Z\|W)$	0.226	0.082	0.430	0.262
	(0.025)	(0.029)	(0.042)	(0.028)
Regime 1	0.131	0.001	0.362	0.506
	(0.053)	(0.005)	(0.081)	(0.079)
Regime 2	0.022	0.857	0.081	0.040
	(0.090)	(0.051)	(0.045)	(0.068)
Regime 3	0.004	0.001	0.794	0.202
	(0.018)	(0.003)	(0.034)	(0.038)
Regime 4	0.733	0.043	0.004	0.220
	(0.063)	(0.016)	(0.019)	(0.059)
Sojourn time	1.151	6.974	4.843	1.282

This table reports the estimated probabilities of being in a regime conditional on the latent class: $P(Z|W)$. Rows below report transition probabilities between regimes. The last row in each panel reports the mean sojourn time, i.e., the expected time a stock market takes to move out of a given regime. It is given for regime k and conditional on the latent class w by $1/(1-p_{kkw})$. Standard errors are reported in round brackets

are notorious. The figures depict how the groups of prices have different patterns of regime switching. Light grey areas correspond to regime 3, the base regime, as revealed already by the probabilities.

West coast prices were usually in regimes 3 (light grey), 4 (white), and 2 (dark grey) during the California crisis. The dynamics of East coast prices are consistent with the information in the tables, namely the shorter durations of regimes and the frequent switching. Regime 2, the one with the highest volatility, occurs frequently

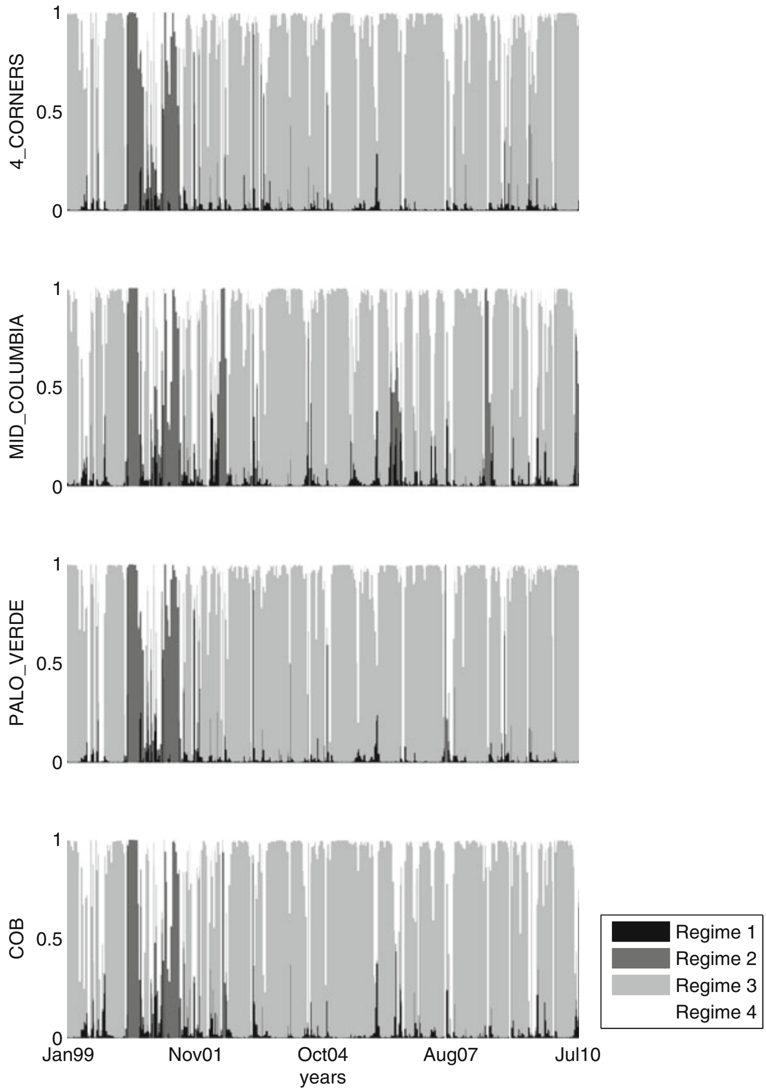


Fig. 2 Price dynamics in the U.S. West Coast. This figure shows the estimated posterior regime probability in latent class 1

in the East coast. Interestingly, the period of the California electricity crisis is clearly identified by the dark grey area in the figure. This episode, well captured by regime 2, seems to be time specific and has not occurred again in the West coast.[8]

[8] The case of California led to specific measures in order to prevent similar cases. For instance, Moulton (2005) mentions the introduction of mitigation procedures after the energy crisis in California (2000–2001).

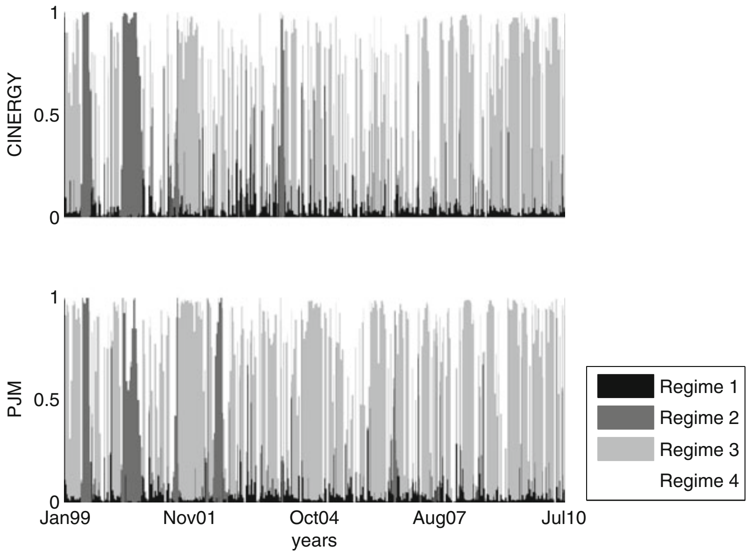


Fig. 3 Price dynamics in the U.S East Coast. This figure shows the estimated posterior regime probability in latent class 2

It is also interesting to note that in the East coast, periods of extremely high volatility occurred frequently after 2001 and prices spikes often seem to occur.

Our results support a four regime parametrization contrasting with previous works such as Huisman and Mahieu (2003), Bierbrauer et al. (2007), Janczura and Weron (2010) that used three regimes; however, their study did not use U.S. data which included the particular episode of California Crisis with extremely high volatility due to price manipulation.

6 Electricity Synchronization

In this section we look at the synchronization of the regimes. To measure synchronization and co-movement in the electricity price series, we compute the association between prices using the posterior probability of being in regime k. In other words, synchronization is measured by the likelihood that prices share regime k at the same period t.

Let $\hat{\alpha}_{it}$ be the estimated probability that electricity price i at time t will be in regime k. To obtain a number in the full range of real numbers, this probability is expressed using the logit transformation:

Table 5 Synchronization between electricity price regimes

		Time series					
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: regime 1							
CINERGY	(1)	1					
4_CORNERS	(2)	−0.031	1				
MID_COLUMBIA	(3)	−0.031	0.594	1			
PALO_VERDE	(4)	−0.047	0.916	0.554	1		
PJM	(5)	0.675	−0.115	−0.043	−0.105	1	
COB	(6)	−0.063	0.753	0.841	0.716	−0.093	1
Panel B: regime 2							
CINERGY	(1)	1					
4_CORNERS	(2)	0.513	1				
MID_COLUMBIA	(3)	0.390	0.707	1			
PALO_VERDE	(4)	0.498	0.942	0.669	1		
PJM	(5)	0.789	0.383	0.432	0.384	1	
COB	(6)	0.513	0.890	0.873	0.860	0.470	1
Panel C: regime 3							
CINERGY	(1)	1					
4_CORNERS	(2)	0.355	1				
MID_COLUMBIA	(3)	0.313	0.660	1			
PALO_VERDE	(4)	0.362	0.952	0.638	1		
PJM	(5)	0.677	0.326	0.331	0.328	1	
COB	(6)	0.365	0.836	0.867	0.833	0.384	1
Panel D: regime 4							
CINERGY	(1)	1					
4_CORNERS	(2)	0.028	1				
MID_COLUMBIA	(3)	−0.011	0.596	1			
PALO_VERDE	(4)	0.021	0.917	0.555	1		
PJM	(5)	0.663	−0.044	−0.025	−0.056	1	
COB	(6)	−0.019	0.749	0.847	0.708	−0.049	1

This table reports the correlation between electricity prices based on the logit of the posterior probability of being in regimes 1, 2, 3 and 4, see Eq. (4)

$$\log it_{itk} = \log\left(\frac{\hat{\alpha}_{itk}}{1 - \hat{\alpha}_{itk}}\right). \tag{4}$$

Synchronization is quantified using the product-moment correlation between the logits for two time series. Our logit-based measure does not suffer from distortion caused by outliers because it filters out extreme observations of prices.

Table 5 shows the correlation between price time series. Panel A shows the probability of two electricity price series being in regime 1 at the same time, and the other panels for the other regimes.

For prices in the same cluster, or geographical area, it is likely that correlation within the cluster is high since they share the same regime dynamic. However, if electricity price indexes are in different clusters, it is interesting to see whether there is synchronization so as to gain some insights about common drivers.

We find a clear distinction of synchronization of regimes. For regimes with regime persistence, there is synchronization between groups (Regimes 2 and 3 present synchronization within and between groups) but we do not find evidence of synchronization for the other two regimes. To put it simply, when prices of the West coast are in the base (or the highest volatility) regime, it is likely that prices of the East coast are also in the base (or the highest volatility) regime. Conversely, it is not likely that prices of the East and West coasts will be found in regimes 1 and 4 at the same time. The correlation is high between returns within classes, but close to zero between the different geographical areas.

7 Conclusion

The 1980s saw the implementation of a wave of deregulatory reforms in the U.S. electricity sector. Wholesale electricity markets were transformed from a highly regulated government controlled system into deregulated local markets. The increase in competition of wholesale markets changed price dynamics and increased price volatility, exposing consumers and producers to significantly greater risks.

We draw on the literature that has proposed multi-regime frameworks to characterize electricity prices. We depart from previous work because we do not impose a fixed number of regimes a priori. Our findings suggest that a four-regime parametrization offers a better characterization of the price dynamics: a base regime, an extremely high volatility regime, a spike up regime, and a reverse regime. Our results show that electricity prices from West and East coasts have different regime dynamics with the latter prices switching more often between regimes. Additionally, our methodology suggests that electricity prices are better parameterized by four regimes: the base regime with low volatility; a spike up and a reverse regime both with high volatility and short duration; and a fourth one with extremely high volatility. The extremely high volatility regime describes West coast prices during the California electricity crisis, but East coast prices are also frequently in that regime. We find evidence of price synchronization in the lowest and highest volatility regimes, i.e., prices from the East and West coasts tend to be in those regimes at the same time.

In summary, in this chapter we describe and compare the price dynamics of electricity prices in the wholesale electricity markets of U.S. East and West coasts. The characterization of joint price dynamics is of great importance to financial market participants and may be useful in making optimal risk management decisions.

References

Andreasen J, Dahlgren M (2006) At the flick of a switch. Energy Risk 71–75

Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in statistical analysis of probabilistic functions of Markov chains. Ann Math Stat 41(1):164–171

Bierbrauer M, Menn C, Rachev ST, Truck S (2007) Spot and derivative pricing in the EEX power market. J Bank Finance 31(11):3462–3485

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via EM algorithm. J R Stat Soc Ser B-Methodol 39(1):1–38

Deng S (1998) Stochastic models of energy commodity prices and their applications: mean reversion with jumps and spikes. PSERC Working Paper 98–28

Dias JG, Vermunt JK, Ramos S (2008) Heterogeneous hidden Markov models. In: Brito P (ed) Proceedings in computational statistics COMPSTAT 2008. Physica/Springer, Heidelberg, pp 373–380

Dias JG, Vermunt JK, Ramos S (2009) Mixture hidden Markov models in finance research. In: Fink A, Lausen B, Seidel W, Ultsch A (eds) Advances in data analysis, data handling and business intelligence. Springer, Berlin, pp 451–459

Dias JG, Wedel M (2004) An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. Stat Comput 14(4):323–332

Ethier R, Mount T (1998) Estimating the volatility of spot prices in restructured electricity markets and the implication for option values. Working Paper, Department of Applied Economics and Management, Cornell University. Ithaca, NY

Faruqui A, Chao H, Niemeyer V, Platt J, Stahlkopf K (2001) Analyzing California's power crisis. Energy J 22(4):29–52

Fong WM, See KH (2002) A Markov switching model of the conditional volatility of crude oil futures prices. Energy Econ 24(1):71–95

Hadsell L, Marathe A, Shawky HA (2004) Estimating the volatility of wholesale electricity spot prices in the US. Energy J 25(4):23–40

Haldrup N, Nielsen F, Nielsen M (2010) A vector autoregressive model for electricity prices subject to long memory and regime switching. Energy Econ 32:1044–1058

Hamilton JD (1989) A new approach to the economic-analysis of nonstationary time-series and the business-cycle. Econometrica 57(2):357–384

Huisman R, Mahieu R (2003) Regime jumps in electricity prices. Energy Econ 25(5):425–434

Janczura J, Weron R (2010) An empirical comparison of alternate regime-switching models for electricity spot prices. Energy Econ 32(5):1059–1073

Janczura J, Weron R (2012) Efficient estimation of Markov regime-switching models: An application to electricity wholesale market prices. AStA Adv Stat Anal 96:385–407

Joskow PL (1997) Restructuring, competition and regulatory reform in the U.S. electricity sector. J Econ Perspect 11(3):119–138

Knittel C, Roberts MR (2005) An empirical examination of restructured electricity prices. Energy Econ 27(5):791–817

Mari C (2006) Regime-switching characterization of electricity prices dynamics. Phys A Stat Theor Phys 371(2):552–564

McLachlan G, Peel D (2000) Finite mixture models. Wiley, New York

Mjelde JW, Bessler DA (2009) Market integration among electricity markets and their major fuel source markets. Energy Econ 31(3):482–491

Moulton JS (2005) California electricity futures: the NYMEX experience. Energy Econ 27 (1):181–194

Park H, Mjelde JW, Bessler DA (2006) Price dynamics among U.S. electricity spot markets. Energy Econ 28(1):81–101

Pennings JME, Garcia P (2004) Hedging behavior in small and medium-sized enterprises: the role of unobserved heterogeneity. J Bank Finance 28(5):951–978

Ramos S, Vermunt J, Dias J (2011) When markets fall down: are emerging markets all equal? Int J Finance Econ 16:324–338

Schwarz G (1978) Estimating dimension of a model. Ann Stat 6(2):461–464

Weron R, Bierbrauer M, Trück S (2004) Modeling electricity prices: jump diffusion and regime switching. Phys A 336:39–48

Woo C (2001) What went wrong in California's electricity market? Energy 26(8):747–458

Pricing Futures and Options in Electricity Markets

Fred Espen Benth and Maren Diane Schmeck

Abstract In this paper we derive power futures prices from a two-factor spot model being a generalization of the classical Schwartz–Smith commodity dynamics. We include non-Gaussian effects by introducing Lévy processes as the stochastic drivers, and estimate the model to data observed at the European Electricity Exchange in Germany. The spot and futures price models are fitted jointly, including the market price of risk parameterized from an Esscher transform. We apply this model to price call and put options on power futures. It is argued theoretically that the pricing measure for options may be different to the pricing measure of futures from spot in power markets due to the non-storability of the electricity spot. Empirical evidence pointing to this fact is found from option prices observed at the European Electricity Exchange.

Keywords Energy markets · Pricing measures · Jump processes · Spot price · Futures and forwards · Options

1 Introduction

In the last two decades markets for power have been liberalized in Europe and other places world-wide. Nowadays, we find well-functioning markets for purchase of electricity in many countries on the European continent, in the Nordic countries and

F.E. Benth (✉)
Centre of Mathematics for Applications, University of Oslo, 1053 Blindern, 0316 Oslo, Norway
e-mail: fredb@math.uio.no
URL: http://folk.uio.no/fredb/

M.D. Schmeck
Department of Mathematics, University of Cologne, Weyertal 86-90, 50931 Cologne, Germany
e-mail: mschmeck@math.uni-koeln.de
URL: http://www.mi.uni-koeln.de/mschmeck/

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_10

in the UK. Furthermore, there exists markets in North America, Australia and some places in Asia. Typically, these markets separate between a day-ahead spot market for electricity, and financial contracts for future delivery of power. In some, more developed markets, one also trades in derivatives like plain vanilla call and put options on the futures contracts. This takes place in for example the Nordic market NordPool and the German market European Electricity Exchange (EEX).

In this paper we focus the attention on pricing spot, futures and options jointly in the power market. Our aim is to argue for a separation of the modelling of the risk premium charged in the futures market and the risk neutral measure used for options pricing. The classical approach to futures pricing is to specify a stochastic dynamics of the spot price, and define the futures price as the conditional risk-adjusted expected average spot price over the delivery period. The risk-adjustment is modelled by a specification of pricing probabilities, which changes the characteristics of the spot dynamics (see Benth et al. 2008 for a discussion and application of this approach to energy markets). Usually, as this approach yields a risk neutral (or martingale) dynamics of the futures price, one would price options using the same probability. We argue here that there is no violation of no-arbitrage pricing to have another pricing measure for options, as long as this is an equivalent martingale measure for the futures price dynamics. The economic argument in favour of this is the non-storability of the electricity spot price.

Based on a small data set of option prices at the EEX, we also argue empirically for this possibility. Fitting a two-factor model for the spot price dynamics to EEX data, we price futures and calibrate the risk premium using a parametric class of pricing probabilities stemming from the Esscher transform (see Benth et al. 2008). Although the access to option data at the EEX is poor due to a rather illiquid market, we find evidence for a different risk neutral pricing measure than the one used to derive futures prices from the spot dynamics. We benchmark our results to the Black-76 prices derived from historical volatility.

Our two-factor spot model is a generalization of the Schwartz–Smith dynamics (see Schwartz and Smith 2000), consisting of a long-term non-stationary factor and a short-term stationary factor. The Schwartz–Smith model has been applied to electricity markets by Lucia and Schwartz (2002), who analysed spot and futures data at the NordPool market. As the Schwartz–Smith model is Gaussian, it fails to account for the large spikes in the market. We extend the model to include Lévy process driven noises, which also accounts for the high variability in EEX prices in non-spike periods. Our proposed model is a simplification of the dynamics proposed and analysed in Benth et al. (2011) and Barndorff-Nielsen et al. (2013). The fitting of the spot and futures dynamics goes by filtering the non-stationary factor by using futures prices of contracts far from delivery.

The presentation of our results are separated into several sections. In the next section we present the rationale behind pricing of futures in power markets. Furthermore, we discuss the pricing of options, and why one may use a different probability for this purpose. Section 3 first defines the two-factor spot model, and presents theoretical futures prices based on this dynamics. The joint spot and futures price model is estimated to EEX data in Sect. 4, while Sect. 5 analyses empirically the option pricing performance of our futures price model. This section argues in favour of a different pricing measure for options. Finally, in Sect. 6, we conclude and outline some future research directions.

2 The Relation Between Spot, Futures and Options in Power Markets

Typically, the liberalized power markets are divided into a day-ahead spot market, a financial market for futures (and/or forwards[1]) contracts on power, and a market for plain vanilla call and put options on the futures. The futures contracts deliver the underlying power over an agreed period of time, and the delivery is settled financially, i.e., the money-equivalent of the spot is delivered. These contracts are denominated in a "currency" per MWh and work essentially as a *swap* contract where one exchanges a floating spot price against a fixed over the contracted period.

For example, in the German EEX market the swaps have delivery periods being months, quarters or years. The swap price is naturally denominated in Euro per MWh, and the contract is accounted against the hourly power spot price. One distinguishes between base and peak load contracts, where the peak load take into account only the power spot prices in the peak hours, defined as the working days from 8 in the morning to 8 in the evening. The base load contracts are settled against the spot price of all hours in the delivery period.

The power spot prices are determined in an auction-based system, where the traders hand in prices and volumes for production or consumption for given hours the next day. Based on these bids, the exchange creates demand and supply curves for each hour the following day, and at 2 p.m. the EEX publishes these spot prices for the 24 h next day. We emphasise that the trade in the power spot market is physical, and one therefore needs to have facilities for either producing or consuming (retailing) electricity. Unlike most other assets that can be traded, one cannot form a portfolio and use the spot for investment or speculation purposes. By the very nature of electricity, it is not possible to store. There are some exceptions, since one may in fact use water reservoirs, say, as storage of power in terms of potential energy. However, this is only possible for a limited segment of the market, namely the hydro power producers.

The options traded in the market are written on specific financial swap contracts. At the EEX power options are written on the Phelix Base futures with monthly, quarterly and yearly delivery periods. The EEX offers only European style call and put options, where the exercise takes place four trading days prior to the beginning of the delivery period of the underlying futures.

[1] Some markets have both forwards and futures traded. We shall not make a distinction between these two asset classes here, but stick to the notion of futures.

Let us discuss at a more technical level the relationship between spot prices, swaps, and options. For illustration, consider first a market where the spot is a liquidly tradeable asset, like for example an exchange-traded stock. We denote $S(t)$ as the spot price at time $t \geq 0$, and consider a futures contract which delivers the spot at a maturity time T. The futures price at time $t \leq T$ is denoted by $f(t,T)$, and from standard no-arbitrage arguments based on the cash and carry strategy (see e.g. Duffie 1992), it can be determined as

$$f(t,T) = S(t)\mathrm{e}^{r(T-t)} \,. \tag{1}$$

Here, $r > 0$ is the deterministic risk-free interest rate, where we have supposed that interest rates are continuously compounded. As is known from classical financial theory, (1) can be established without any model assumptions on the spot price.

Assume that we are given a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\tilde{T}]}, P)$. We interpret $\tilde{T} < \infty$ as the time horizon of the market, including the maturities of all options and futures relevant in our analysis. If $S(t)$ is a semi-martingale process, then there exists (at least one) equivalent martingale measure Q such that

$$f(t,T) = \mathbb{E}_Q[S(T)\,|\,\mathcal{F}_t] \,. \tag{2}$$

We refer to Shiryaev (1999) for the rigorous argumentation with conditions leading to this representation of $f(t,T)$. In a complete market, i.e., a market where all derivatives on S can be replicated, the probability measure Q is uniquely defined. In the case of an incomplete market, one may have many such measures Q. The question is to determine *one* relevant for pricing of derivatives. But, once such a measure is pinned down, we can price futures and next use the same probability for pricing options. Thus, for example the price of a European option with payoff $g(f(\tau,T))$ at exercise time $\tau \leq T$ becomes

$$C(t) = \mathrm{e}^{-r(\tau-t)}\mathbb{E}_Q[g(f(\tau,T))\,|\,\mathcal{F}_t] \,,$$

for $0 \leq t \leq \tau$. Note that we use the same Q for both the futures and the option, as is the customary when pricing several derivatives based on an asset in an incomplete market situation. Note, however, that we may use different equivalent martingale measures for pricing different derivatives, as long as there exists at least *one* measure Q that is an equivalent martingale measure for *all* products.

To see this, suppose that we have two derivatives on the spot with payoffs given by the random variables X and Y, respectively. Let the prices at time zero be $C_X = \mathbb{E}_{Q_X}[X]$ and $C_Y = \mathbb{E}_{Q_Y}[Y]$, where we for the moment assume that the interest rate is zero to simplify the argument. The probabilities Q_X and Q_Y are equivalent martingale measures. If there exists an equivalent martingale measure Q, such that the price processes S, C_X and C_Y, are all Q-martingales, then the market is arbitrage-free. However, as long as Q is equivalent to P, it has to be equivalent to Q_X and Q_Y

as well. Furthermore, by the no-arbitrage theory we must have that $C_X = \mathbb{E}_Q[X]$ and $C_Y = \mathbb{E}_Q[Y]$. This implies that

$$\mathbb{E}_{Q_X}\left[X\frac{dQ}{dQ_X}\right] = \mathbb{E}_{Q_X}[X],$$

and

$$\mathbb{E}_{Q_Y}\left[Y\frac{dQ}{dQ_Y}\right] = \mathbb{E}_{Q_Y}[Y].$$

These two equalities put strong conditions on the range of possible probabilities Q_X, Q_Y and Q.

In the case of power markets, the situation is completely different since the probability measure used to price futures can theoretically be completely detached from the measure pricing options on futures. As we have already argued, the power spot price cannot be traded in the normal financial sense, and it works as a *reference index* for the settlement of futures contracts. With this view at hand, the *pricing measure* Q used to derive the futures price on the spot does not need to be an *equivalent martingale* measure, but is required only to be an *equivalent* measure. However, the futures is a tradeable asset and its price dynamics must be a Q-martingale in order for the market to be free of arbitrage opportunities. Pricing using conditional expectation as in (2) ensures this by definition.

In a specification of the market, one would typically model the spot price evolution using some stochastic process $S(t)$, and choose a parametric class of equivalent probability measures Q. Based on a selected probability Q from this class, the standard approach to price electricity futures is to define it as

$$F(t, T_1, T_2) = \mathbb{E}_Q\left[\frac{1}{T_2 - T_1}\int_{T_1}^{T_2} S(t)\,dt \,|\, \mathcal{F}_t\right]. \tag{3}$$

Here, we consider a contract delivering electricity over the time interval $[T_1, T_2]$, and the contract is entered at time $t \leq T_1$, with settlement at the end of the delivery period T_2. Note that the price is denoted in MWh, and therefore is normalized by the length of the delivery period. This gives a theoretical swap price dynamics which we next calibrate to the observed prices by fitting the parameters of the probabilities Q. This will pin down a probability $\widehat{Q}$ under which we model the risk-neutral futures price dynamics. Note that the risk-neutral dynamics of F is a $\widehat{Q}$-martingale. Since by construction $\widehat{Q}$ is equivalent to P, we can also (in principle) derive the market dynamics of the futures. Note that in general $\widehat{Q}$ is not a probability for which the spot price dynamics becomes a martingale after discounting.

In reality, the above procedure in specifying a probability $\widehat{Q}$ for pricing futures is an approach to find a parametric representation of the price process $F(t, T_1, T_2)$, where we calibrate to represent *the risk premium* in the market, i.e., to explain the difference between the observed futures prices and the predicted average spot price. The latter is calculated by relation (3) using $Q = P$. Apriori there are two extreme choices one can make on Q. First, ignoring the existence of a risk premium, one could select $Q = P$. Alternatively, assuming the electricity spot is tradeable, one could force Q to be a *martingale* measure. Note that depending on the model for S, one could have many possible martingale measures, so the latter choice is not necessarily unique. Both alternatives are theoretically viable, but hardly reasonable from the characteristics of electricity markets.

Our next problem is to price call and put options written on the futures. Following the standard no-arbitrage pricing framework discussed above, a first thought would be to use $\widehat{Q}$ and compute the option price using this probability. To be more specific, let us suppose that we have a call option with exercise time $\tau \leq T_1$ written on a swap with dynamics $F(t, T_1, T_2)$ given in (3) for the pricing measure $\widehat{Q}$. The price of this call at time $t \leq \tau$ is

$$C(t) = \mathrm{e}^{-r(\tau - t)} \mathbb{E}_{\widehat{Q}}[\max(F(\tau, T_1, T_2) - K, 0) \,|\, \mathcal{F}_t].$$

However, in general, there will exist several equivalent measures Q for which $t \mapsto F(t, T_1, T_2)$ is a Q-martingale. In fact, since typically the power spot price dynamics involves jump processes, the futures price will follow a jump dynamics as well. Under certain conditions, such models admit the existence of a continuum of equivalent martingale measures Q. In this case we pin down a pricing measure $\tilde{Q}$ by selecting it from a parametric class of equivalent martingale measures Q for $F(t, T_1, T_2)$. One could derive this probability by calibrating to observed option prices in the market, or to appeal (partial) hedging arguments (see Cont and Tankov 2004 for a discussion of hedging and pricing in incomplete markets).

Note that finding $\tilde{Q}$ for option pricing follows in principle the same scheme as choosing $\widehat{Q}$ for the futures prices. The fundamental difference is that $\widehat{Q}$ does not need to be a martingale measure for the spot price, whereas $\tilde{Q}$ has to be a martingale measure for the futures price. Both probability measures are equivalent to P. In the next sections we shall estimate a particular two-factor model to spot price data collected from the EEX, and apply this to futures pricing based on a class of probabilities defined by Esscher transformation. Using option price data, we shall argue that the spot-futures probability $\widehat{Q}$ is not the right probability for pricing options on the futures, pointing towards $\tilde{Q} \neq \widehat{Q}$.

Our analysis is not restricted to power markets only. In the weather markets, like the temperature market at the Chicago Mercantile Exchange (CME), futures on temperature indices measured in various cities world-wide are traded. In addition, plain vanilla call and put options on these futures are traded. The underlying “spot” price here is the temperature in a given city, for example Chicago itself. Given a stochastic model for the temperature $S(t)$, one can derive the resulting futures price written on an index of the temperature. Typically, one chooses to price using a conditional expectation analogous to (3), where a pricing measure is selected. Obviously, temperature itself is not a tradeable commodity, and we can use the same argumentation as above to defend choosing the pricing probabilities which are not necessarily martingale measures for the temperature dynamics. On the other hand, the futures contracts are tradeable financial assets, and to price the options with these as underlying, we need to use a probability measure Q which turns the futures price into a Q-martingale. As in the case of power, the futures pricing measure $\widehat{Q}$ does not need to be the same as the option pricing measure $\tilde{Q}$. We note in passing that CME also organize a market for precipitation derivatives based on snow and rainfall indices in some cities in the US. Further, there has been trials to create an organized market for wind futures and options at the now closed US Futures Exchange. Here our discussion makes sense as well.

3 The Spot Price Dynamics and Implied Futures Prices

We consider a simple arithmetic two-factor spot price dynamics in the spirit of Lucia and Schwartz (2002). The occurrence of negative spikes at the EEX, and, even more, the observation that these spikes may even lead to negative prices, indicate that an arithmetic model may be suitable. To this end, suppose that $S(t)$ follows the dynamics

$$S(t) = \Lambda(t) + X(t) + Y(t)\,. \tag{4}$$

Here, $\Lambda : [0, \tilde{T}] \mapsto \mathbb{R}$ is a measurable deterministic function, modelling the mean seasonal variation in spot prices. Usually, this function consists of a linear trend and a periodic function (a linear combination of sines and cosines, with different frequencies), and as such is a smooth function. The *base component* $X(t)$ in the spot price dynamics is assumed to be non-stationary and defined to be a Lévy process, i.e.,

$$dX(t) = dL_1(t)\,. \tag{5}$$

In Lucia and Schwartz (2002), it is assumed that $L_1(t) = \gamma t + \sigma B(t)$ with γ and σ being constants and $B(t)$ a Brownian motion. The volatility σ is naturally assumed to be positive. One may think of the base component as stochastic variations from market activity as well as long term effects like inflation in fuel prices and limited resources, as well as entry of new sources of energy (like renewables). As it will turn out from our empirical analysis of EEX spot price data, a drifted Brownian motion is unsuitable for modelling the true dynamics of the non-stationary term, and a Lévy process is much more appropriate.

Typically in power markets spot prices may exhibit random shocks due to imbalances in supply and demand. These shocks are seen as spikes in the price path, imposed from an unexpected increase in demand due to colder weather, say, or shut down of a major power plant yielding a drop in supply. The prices will in both these cases exhibit a major price jump upward, which is followed typically by a strong decline since demand will be significantly reduced by higher prices, or expensive power production plants are ramped up (like coal-fired plants in Denmark in the NordPool area). In the EEX market one observes many negative spikes, which is caused by wind power mainly. By political legislation, wind power and other renewable energy sources have priority into the electricity grid, and hence an unexpected increase in wind power production (due to more wind where the farms are…) may create bigger than expected supply (since it takes time to ramp down or adjust other power plants fueled by gas and coal or producing nuclear energy). In fact, one observes negative prices in the EEX market due to over-supply, where some producers choose to pay for power consumption rather than shut down their production.

From this discussion, we see that there is ample evidence for a mean-reverting short-time factor of the form

$$dY(t) = -\eta Y(t)\,dt + dL_2(t)\,. \tag{6}$$

Here, the constant $\eta > 0$ is expected to be rather big, since spikes created by the Lévy process $L_2(t)$ are reverting fastly back to normal price levels. We suppose that $L_2(t)$ may have both positive and negative jumps, i.e., $L_2(1)$ is distributed on $\mathbb{R}$.

Notice that in Lucia and Schwartz (2002), both an arithmetic and geometric two-factor model were analysed theoretically and empirically on NordPool data. In their approach, the second factor Y was also assumed to be driven by a Brownian motion. We believe that a jump factor for the noise is more appropriate in order to explain the sudden spikes in prices, exhibiting a jump like behaviour in the price path. Also, most empirical studies of power spot prices point strongly towards non-Gaussianity in prices, and hence the need to use other processes than the Brownian motion to drive the dynamics (see discussion in Benth et al. 2008). We remark that Lucia and Schwartz (2002) let the short and long term factors correlate through their driving noise.

We denote $L = (L_1, L_2)$, and assume that L is a bivariate Lévy process with cumulant (log-characteristic function) defined by

$$\psi(\mathbf{x}) = \mathrm{i}\mu'\mathbf{x} - \frac{1}{2}\mathbf{x}'C\mathbf{x} + \int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}\mathbf{x}'\mathbf{z}} - 1 - \mathrm{i}\mathbf{x}'\mathbf{z}1(|\mathbf{z}| \leq 1)\,\ell(d\mathbf{z})\,, \tag{7}$$

with $\mathbf{x} = (x, y)' \in \mathbb{R}^2$, $\mu \in \mathbb{R}^2$, C a symmetric non-negative definite 2×2 matrix and $\ell(d\mathbf{z})$ a Lévy measure on $\mathbb{R}^2 \backslash \{\mathbf{0}\}$. Here $\mathbf{x}'$ denotes the transpose of the vector, and $\mathrm{i} = \sqrt{-1}$ is the imaginary unit. In the case of independence between L_1 and L_2, we can express the cumulant as a sum

$$\psi(x, y) = \psi_1(x) + \psi_2(y)$$

where ψ_i, $i = 1, 2$ are cumulants of the univariate Lévy processes L_1 and L_2. Our general model allows for a dependency between L_1 and L_2, although we shall assume independence in the empirical study on EEX data below.

In Benth et al. (2011) they use a more general model. The stationary short time variations are modelled as a continuous-time autoregressive moving average (CARMA) process, where the driving process L_2 is an α-stable Lévy process. As it includes mean reversion, a CARMA model is comparable to the standard approach of commodity spot price modelling, i.e., to describe the spot as a sum of several Ornstein–Uhlenbeck processes with different speeds of mean reversion and stochastic drivers (see Benth et al. 2008). In Benth et al. (2011), a CARMA(2,1) dynamics is proposed and fitted empirically to EEX spot price data. Such a dynamics is similar to a two-factor model, with each factor being an Ornstein–Uhlenbeck process. Although we find strong indications of a two-factor dynamics in our empirical study, we simplify the considerations here to a one-factor model as a first order approximation of the short-term factor. This makes the fitting of data significantly easier, and is in line with the more classical two-factor model of Lucia and Schwartz (2002). Moreover, it turns out that we can do well with a much more regular Lévy process than the α-stable to model the random fluctuations.

Our first concern is to introduce a parametric class of equivalent probabilities Q which is appropriate for pricing swaps. For $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, define the equivalent probability Q_θ, where the density process of Q_θ with respect to P is

$$\frac{dQ_\theta}{dP}|_{\mathcal{F}_t} = \exp\{\theta L(t) - \psi(-\mathrm{i}\theta)t\}. \tag{8}$$

In order for this to be well-defined, we must of course assume exponential integrability conditions on $L(1)$. Hence, suppose from now on that there exists a constant $c > 0$ such that

$$\int_{\mathbb{R}^2} \mathrm{e}^{\mathbf{x}'\mathbf{z}} \ell(d\mathbf{z}) < \infty, \tag{9}$$

for all $|\mathbf{x}| \leq c$. This ensures finite exponential moments for $L(1)$ up to order c.

The probability Q_θ parameterized by θ is known as the Esscher transform of L (see Benth et al. 2008). The probability Q_θ is equivalent to P by definition of the Radon-Nikodym densities. We emphasize, however, that we do not demand Q_θ to be a *martingale* measure, in the sense that the power spot dynamics becomes a Q_θ-martingale (the reader should note that this is technically impossible anyway with the Esscher transform on an Ornstein-Uhlenbeck process, see Benth and Sgarra (2012)). The reason is the non-storability of the spot which makes it non-tradeable, i.e., one cannot create portfolios with spot investments in electricity. Once purchased, it must be consumed. The parameter θ is restricted to the subspace of $\mathbb{R}^2$ defined by $|\theta| \leq c$.

In the next Lemma we characterize the process L under Q_θ:

Lemma 3.1 *The process L is a Lévy process with respect to Q_s with cumulant function*

$$\psi_{Q_\theta}(\mathbf{x}) = \psi(\mathbf{x} - \mathrm{i}\theta) - \psi(-\mathrm{i}\theta).$$

Hence, the drift is

$$\mu + \theta' C + \int_{|\mathbf{z}|<1} (\mathrm{e}^{\theta \mathbf{z}} - 1)\mathbf{z}\ell(d\mathbf{z})$$

and the Lévy measure

$$\ell_{Q_\theta}(d\mathbf{z}) = \mathrm{e}^{\theta \mathbf{z}}\ell(d\mathbf{z}),$$

while the covariance matrix C remains the same.

Proof Using Bayes' Theorem along with the density process of Q_θ and the independent increment property of the Lévy process, yield that the conditional log-characteristic function of $L(t)$ given $\mathcal{F}_s$ for $t \geq s \geq 0$ is

$$\ln \mathbb{E}_{Q_s}\left[\mathrm{e}^{\mathrm{i}\mathbf{x}' L(t)} \,|\, \mathcal{F}_s\right] = (\psi(\mathbf{x} - \mathrm{i}\theta) - \psi(-\mathrm{i}\theta))(t-s).$$

Hence, L is a Lévy process under Q_θ as well. By a direct computation, we find the drift and the Lévy measure as claimed. □

Note that if we have a (bivariate) drifted Brownian motion as Lévy process, i.e., $\ell(d\mathbf{z}) = 0$, then the Esscher transform is simply a Girsanov transform of the Brownian motion with a constant parameter θ. For Lévy processes with jumps, the Lévy measure is exponentially tilted by the Esscher transform. We may interpret this as a rescaling of the size and intensity of jumps.

We remark that the expected value of $L(1)$ under Q_θ is given by

$$\mathbb{E}_\theta[L(1)] = -\mathrm{i}\nabla\psi(-\mathrm{i}\theta),$$

where ∇ is the gradient and $\mathbb{E}_\theta[\cdot]$ is the expectation operator with respect to the probability Q_θ. Thus, the Lévy process $\tilde{L}(t) = L(t) + \mathrm{i}\nabla\psi(-\mathrm{i}\theta)t$ becomes a martingale under Q_θ as it has expectation zero. This means in particular that under Q_θ, the dynamics of X and Y are, respectively,

$$dX(t) = -\mathrm{i}\psi_x(-\mathrm{i}\theta)\,dt + d\tilde{L}_1(t) \tag{10}$$

and

$$dY(t) = \left\{-\mathrm{i}\psi_y(-\mathrm{i}\theta) - \eta Y(t)\right\} dt + d\tilde{L}_2(t)\,. \tag{11}$$

Here, we have used the notation ψ_x and ψ_y as the partial derivatives of ψ with respect to the two variables x and y, respectively. The solution $Y(s)$ at time $s \geq t$, conditioned on $Y(t)$, of this Ornstein–Uhlenbeck process is

$$Y(s) = Y(t)\mathrm{e}^{-\eta(s-t)} + \frac{-\mathrm{i}\psi_y(-\mathrm{i}\theta)}{\eta}(1 - e^{-\eta(s-t)}) + \int_t^s \mathrm{e}^{-\eta(u-t)}\,\tilde{L}_2(du)\,. \tag{12}$$

Next, we consider pricing of swaps in this market. Let us start with analysing the implied swap price dynamics for the arithmetic model. The following result holds:

Proposition 3.2 *The swap price $F(t, T_1, T_2)$ is given by*

$$\begin{aligned} F(t, T_1, T_2) =& \bar{\Lambda}(T_1, T_2) + X(t) + Y(t)\bar{\eta}(t, T_1, T_2) \\ & - \frac{1}{2}\mathrm{i}\psi_x(-\mathrm{i}\theta)(T_2 - T_1) - \mathrm{i}\psi_x(-\mathrm{i}\theta)(T_1 - t) + \frac{-\mathrm{i}\psi_y(-\mathrm{i}\theta)}{\eta}(1 - \bar{\eta}(t, T_1, T_2))\,, \end{aligned}$$

where

$$\bar{\eta}(t, T_1, T_2) = \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} \mathrm{e}^{-\eta(s-t)}\,ds$$

and $\bar{\Lambda}(T_1, T_2)$ is the average value of the seasonality function $\Lambda(s)$ over the interval $[T_1, T_2]$.

Proof From the expression in (10), we find (for $s \geq t$)

$$\mathbb{E}_{Q_\theta}[X(s)|\mathcal{F}_t] = X(t) - \mathrm{i}\psi_x(-\mathrm{i}\theta(s - t)\,,$$

after appealing to the independent increment property of the Q_θ-Lévy process $\tilde{L}_1$ with zero mean, and the $\mathcal{F}_t$-measurability of $X(t)$. Similarly, from the independent increment property of the Q_θ-Lévy process $\tilde{L}_2$, having mean zero, we find from (12)

$$\mathbb{E}_{Q_\theta}[Y(s)|\mathcal{F}_t] = Y(t)\mathrm{e}^{-\eta(s-t)} - \frac{\mathrm{i}\psi_y(-\mathrm{i}\theta)}{\eta}(1 - \mathrm{e}^{-\eta(s-t)}).$$

Since

$$F(t,T_1,T_2) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \left\{ \Lambda(s) + \mathbb{E}_{Q_\theta}[X(s) + Y(s)|\mathcal{F}_t] \right\} ds$$

the result follows after using the Fubini Theorem. □

We note that $\bar{\eta}$ is the average value of the "volatility function" $\exp(-\eta(s-t))$ over the delivery period $[T_1, T_2]$, and takes the form

$$\bar{\eta}(t,T_1,T_2) = \frac{1}{\eta(T_2 - T_1)} \left(e^{-\eta(T_1 - t)} - e^{-\eta(T_2 - t)} \right), \tag{13}$$

or,

$$\bar{\eta}(t,T_1,T_2) = e^{-\eta(T_1 - t)} \frac{1}{\eta(T_2 - T_1)} \left(1 - e^{-\eta(T_2 - T_1)} \right). \tag{14}$$

In the representation (14), $T_1 - t$ is time left until start of delivery, and $T_2 - T_1$ is length of delivery. We recognize the exponential damping factor $\exp(-\eta(T_1 - t))$ as the Samuelson effect on the volatility, i.e., the volatility of the spot is increasing as time to start of delivery is decreasing. The classical Samuelson effect says that the volatility of the futures price is exponentially increasing in time to maturity to the spot volatility (see Samuelson 1965, Benth et al. 2008). We note here that $\bar{\eta}(t,T_1,T_2)$ is not converging to the "spot volatility", being one in this context, but to a value less than this. The delivery period creates this violation of the classical Samuleson effect. It is natural from a financial and empirical point of view that the volatility of the electricity futures price is not converging to that of the spot as the futures price is the average of the spot over a delivery period.

We derive the dynamics of F in the next proposition

Proposition 3.3 *The Q_θ dynamics of the swap price is*

$$dF(t,T_1,T_2) = d\tilde{L}_1(t) + \bar{\eta}(t,T_1,T_2)\, d\tilde{L}_2(t).$$

Proof Since $\bar{\eta}'(t,T_1,T_2) = \eta\bar{\eta}(t,T_1,T_2)$, the result follows after applying the Itô formula for jump processes and the Q_θ-dynamics of X and Y. □

As is apparent from the definition of $F(t,T_1,T_2)$, it is a Q_θ-martingale process for $t \leq T_1$. Thus, it defines an arbitrage-free model for the stochastic evolution of electricity futures prices.

4 An Empirical Study of EEX Spot and Futures Prices

In this section we want to estimate the parameters in the spot model, and calibrate it to futures prices where we derive the market price of risk θ. It turns out that a joint estimation of spot and futures is most efficient, where one can make use of the asymptotic behaviour of futures prices to filter out the non-stationary factor in the spot. This approach is analogous of the calibration procedure in Schwartz and Smith (2000), with a more sophisticated version of it found in Benth et al. (2011).

The following asymptotic result of the futures price with respect to time to delivery plays a crucial role in the estimation algorithm.

Proposition 4.1 *It holds that*

$$\lim_{T_1 - t \to \infty} \left\{ F(t, T_1, T_2) - \bar{\Lambda}(T_1, T_2) - \Psi(t, T_1, T_2; \theta) - X(t) \right\} = 0,$$

where

$$\Psi(t, T_1, T_2; \theta) = -\frac{1}{2} \mathrm{i}\psi_x(-\mathrm{i}\theta)(T_2 - T_1) - \mathrm{i}\psi_x(-\mathrm{i}\theta)(T_1 - t) - \frac{\mathrm{i}\psi_y(-\mathrm{i}\theta)}{\eta}.$$

Proof Recalling the explicit dynamics of $F(t, T_1, T_2)$ in Proposition 3.2, the result follows after observing that $\exp(-\eta(T_1 - t)) \to 0$ as $T_1 - t \to \infty$. □

Hence, asymptotically the futures price behaves like

$$F(t, T_1, T_2) \approx \bar{\Lambda}(T_1, T_2) + \Psi(t, T_1, T_2; \theta) + X(t), \tag{15}$$

for $T_1 - t \to \infty$. This means that in the long end of the futures market, the prices fluctuate as the non-stationary factor $X(t)$ plus some non-stochastic adjustment term $\bar{\Lambda}(T_1, T_2) + \Psi(t, T_1, T_2; \theta)$ involving the market price of risk θ. From these considerations we can derive an algorithm for estimating the model. It goes as follows.

For a fixed delivery period $[T_1, T_2]$,

(1) Fit a seasonal function $\Lambda(t)$ to the spot prices $S(t)$.
(2) Fit the autocorrelation function of $Y(t)$ to the deseasonalized spot prices to have an apriori estimate of η. Use this η to find a threshold $\widehat{T}$ for which "$T_1 - t = \infty$", i.e., how big should $T_1 - t$ be for the asymptotic behaviour of F in (15) to be acceptable.
(3) Subtract $\overline{\Lambda}(T_1, T_2)$ from the observed futures prices to "deseasonalize" them. Call this time series $\tilde{F}(t, T_1, T_2)$.
(4) Observe that we have for $T_1 - t \geq \hat{T}$

$$\tilde{F}(t,T_1,T_2) \approx c(\theta,T_1,T_2) - \mathrm{i}\psi_x(-\mathrm{i}\theta)(T_1-t) + X(t),$$

where

$$c(\theta,T_1,T_2) = -\frac{1}{2}\mathrm{i}\psi_x(-\mathrm{i}\theta)(T_2-T_1) - \frac{\mathrm{i}\psi_y(-\mathrm{i}\theta)}{\eta}.$$

Hence, for all observed futures prices $F(t,T_1,T_2)$ for which $T_1 - t \geq \hat{T}$, estimate the "constants" $c(\theta,T_1,T_2)$ and $-\mathrm{i}\psi_x(-\mathrm{i}\theta)$ by linear regression of $\tilde{F}$ with respect to $T_1 - t$.

(5) Using the estimated regression coefficients $\hat{c}$ and $\hat{a}$, we filter out $X(t)$ from the observations,

$$\tilde{F}(t,T_1,T_2) - \hat{c} - \hat{a}(T_1-t)$$

for all $T_1 - t \geq \hat{T}$.

(6) Subtract the filtered data series $X(t)$ from the deseasonalized spot prices. This results in a time series which is modelled by $Y(t)$. Re-estimate η based on linear regression of $Y(t)$ against $Y(t-1)$.

(7) Fit a Lévy process L to the residuals of the Y process and the time series of the X process obtained above. From the fitted Lévy process L, we obtain the cumulant ψ.

(8) For the given cumulant ψ, find the estimated market price of risk θ by solving the system of equations

$$\hat{a} = -\mathrm{i}\psi_x(-\mathrm{i}\theta),$$
$$\hat{c} = -\frac{1}{2}\mathrm{i}\psi_x(-\mathrm{i}\theta)(T_2-T_1) + \frac{-\mathrm{i}\psi_y(-\mathrm{i}\theta)}{\eta}.$$

This calibration algorithm provides us with a full specification of both the spot and the futures price model, including the estimation of the market prices of risk $\theta = (\theta_1, \theta_2)$. We next apply it to spot and futures price data collected from the European Energy Exchange (EEX).

We have available daily Phelix base load spot prices from 02.01.2006 to 19.10.2008, constituting altogether 1,022 daily observations. We remark that we include weekend prices as we are going to apply base load futures prices in our estimation routine. These futures are settled on the spot prices including the weekends. To the spot price data, we fit the seasonality function taken from Barndorff-Nielsen et al. (2013),

$$\Lambda(t) = \xi_0 + \xi_1\cos(\frac{\tau_1 + 2\pi t}{365}) + \xi_2\cos(\frac{\tau_2 + 2\pi t}{7}) + \xi_3 t + \xi_4\mathbf{1}_{\mathrm{Sat}}(t) + \xi_5\mathbf{1}_{\mathrm{Sun}}(t).$$

This function takes annual and weekly seasonality into account along with a trend. As prices on weekends are in general lower than during the rest of the week due to a different demand situation, we introduce additionally a weekend-correction to capture these effects. Here $\mathbf{1}_{\mathrm{Sat}}(t)$ and $\mathbf{1}_{\mathrm{Sun}}(t)$ are equal 1, if the weekday corresponding to t is a Saturday and Sunday, respectively.

A non-linear least squares estimation on the spot data yields the parameters reported in Table 1. Figure 1 (left) displays the spot price data and its estimated seasonality function. The estimated seasonality follows the general movements of the spot, on a weekly pattern as well as a yearly one.

Next we continue the calibration algorithm with filtering the non-stationary factor X from the futures data with long time to delivery. For this purpose we use base load futures contracts with 1 month delivery period from the EEX, for which we have available price data for the same dates as the spot (weekends and holidays are excluded, as there is no trade in futures).

We first need to determine the threshold $\hat{T}$ for which the futures prices are asymptotically given by (15). This depends, obviously, on the value of η, the speed of mean reversion in the factor process Y. We can estimate this parameter from the autocorrelation function of Y which is known to be exponentially decaying at the rate η (see Benth et al. 2008). However, at this point in the estimation procedure we have not yet filtered the time series of Y from the spot data, so the empirical autocorrelation function is unknown to us. Therefore, we do a rough estimation of η by looking at the empirical autocorrelation of the deseasonalized spot, which is modelled by $X(t) + Y(t)$. We observe a decaying autocorrelation structure, and fit an exponentially decaying function to the first five lags obtaining the pre-estimate $\hat{\eta} = 0.1781$. We derive $\hat{T} = 16$ as the threshold when $Y(t)\bar{\eta}(t, T_1, T_2) \approx 1$ using $Y(t)$ being three times the standard deviation of spot price data. Note that we expect the presence of X to make the beta smaller than the "true" one. A larger value for η would lead to a smaller threshold. Hence, our decision to apply $\hat{T} = 16$ is a conservative choice.

We construct a time series of futures prices with "infinite" time to delivery from the base load contracts as follows: if the time to delivery is more than 16 days, we choose the futures which has the first coming month as delivery period. Otherwise, we switch to the contract with delivery in the following month. That is, we use the price series of front-month contracts as long as these are farther than 16 days to delivery, and switch to the next month when the front-month contracts have less than 16 days to delivery. Like this we make sure that for each date we have a futures price with time to delivery of more than 16 days. These prices will not, at least approximately, have any influence from the stationary component Y. As the futures are not traded on weekends and holidays, we use as a substitute for missing

Table 1 Estimated parameters of the seasonal function

ξ_0	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	τ_1	τ_2
738.733	4.360	−11.716	0.020	1.000	1.000	−13637.760	40.401

values at the weekend the price of the preceeding Friday. On holidays, we use the price of the last trading day before the holidays.

To deseasonalize the constructed futures price series we subtract the average seasonality of the delivery period. We have fitted the seasonality function to data until October 2008, such that we take October 2008 as the last delivery period and let our futures price series end at 14.09.2008. A linear regression of this time series delivers the estimates $\hat{a} = 0.030$ and $\hat{c} = 3.406$. We filter the non-stationary time series $X(t)$ from the futures prices corresponding to step (5) in the algorithm, and afterwards retrieve the stationary time series $Y(t)$ from the spot prices as in step (6). The plot on the right in Fig. 1 shows the filtered factor $X(t)$ along with the deseasonalized spot prices. It seems to reflect a long-term stochastic trend in the price data.

Next we estimate the mean reversion parameter η. The autocorrelation function of the time series $Y(t)$ is plotted in Fig. 2. Re-estimating η over the first five lags results in $\hat{\eta} = 0.359$. The initial decrease of the autocorrelation function seems to be captured well by using an exponential function. However, it decays too rapidly for larger lags. Including more lags to fit the autocorrelation function (i.e., η) results in a poor fit in the first lags. To get a better fit over all lags, one could use two (or more) exponential components. This would mean that we model the factor Y by two or more Ornstein–Uhlenbeck processes, or by a higher-order CARMA model. Benth et al. (2011) indicate that one should indeed use a higher-order CARMA

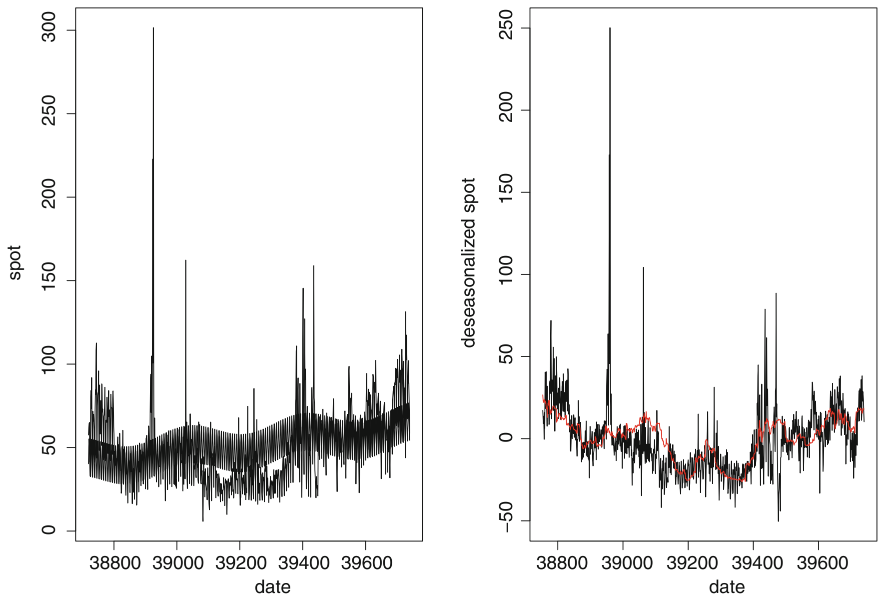


Fig. 1 *Left* empirical spot price data together with the estimated seasonality function. *Right* deseasonalized spot price data with the filtered data series $X(t)$

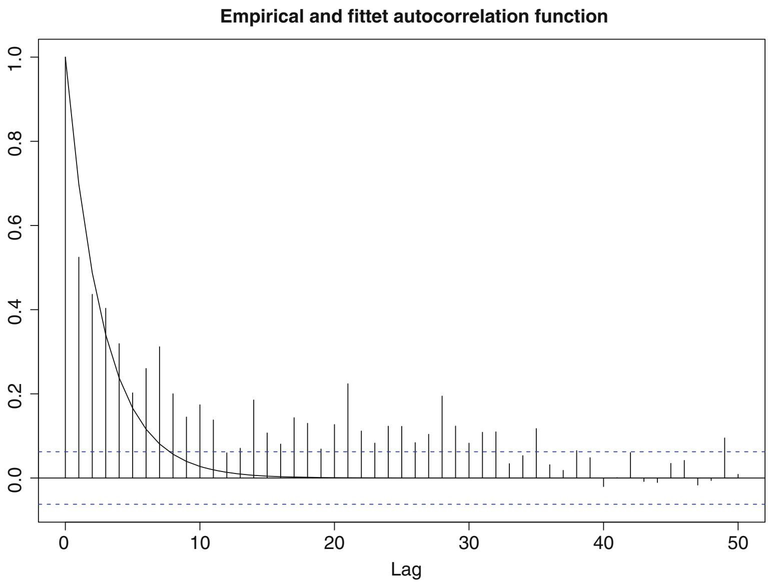


Fig. 2 Autocorrelation function of $Y(t)$

model. However, such models are much more complex to estimate, and we apply the one-factor assumption on Y here as a first approximation of the dynamics.

The next step is to fit a bivariate Lévy process $L = (L_1, L_2)$ to the time series $X(t)$ and $Y(t)$. For simplicity, we assume that L_1 and L_2 are independent, meaning that there is no dependency between the short-term and long-term price fluctuations. In the Schwartz–Smith model (see Schwartz and Smith (2000), or Lucia and Schwartz (2002) for the case of electricity) L is assumed to be a bivariate Brownian motion. However, the Gaussian assumption on the increments $\Delta X(t)$ is not realistic, and we propose to fit the dynamics of X with a normal inverse Gaussian (NIG) Lévy process, i.e., a Lévy process with NIG distributed marginals. The NIG distribution seems to be a good choice for modelling the residuals of $Y(t)$ as well.

The NIG distribution is a four parameter family of distributions successfully applied to model the log returns of financial data. For its applications to finance and a detailed probabilistic analysis of the NIG family, we refer the interested reader to Barndorff-Nielsen (1998). Assuming $L_1(t)$ to be a NIG Lévy process, its cumulant (i.e., the logarithm of the characteristic function) function at time 1 is given by

$$\Psi(x) = \delta\{\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + ix)^2}\} + \mu ix, \tag{16}$$

for the four parameters μ, β, $\delta > 0$ and $\alpha > 0$. The skewness of the NIG distribution is described by β, where $\beta > 0$ means a positively skewed distribution, and $\beta < 0$ negatively skewed. For a symmetric NIG distribution, i.e., when $\beta = 0$, μ is the mean. Otherwise, μ is the location parameter. δ is the scale and α the tail heaviness parameter. Note that the NIG distribution has semi-heavy tails, with the normal distribution as a limiting case. We easily find the expectation from (16) as

$$\kappa_1 = \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}} + \mu.$$

The estimated parameters of $L_1(1)$ based on maximum-likelihood are given in Table 2. We remark in passing that the NIG distribution has been applied in studies of energy prices in Benth and Šaltytė-Benth (2004) and Börger et al. (2009).

We fit another NIG Lévy process L_2 to the residuals of Y. The estimates are reported in Table 2. The estimated densities of $L_1(1)$ and $L_2(1)$ are displayed together with the empirical ones in Fig. 3. The fit seems to be good, and we find the NIG distribution as a satisfactory choice for modelling L_1 and L_2. Recall that we assumed independence of L_1 and L_2. Empirically, the correlation between the data series for L_1 and L_2 is given by -0.16. A more realistic model should take this into account, which requires an analysis of the dependency structure. We relegate this to

Table 2 Estimated NIG parameters of L_1 and L_2

	α	β	δ	μ
L_1	0.0946	−0.0099	0.3136	0.02421
L_2	0.0402	0.0071	14.3407	−2.9488

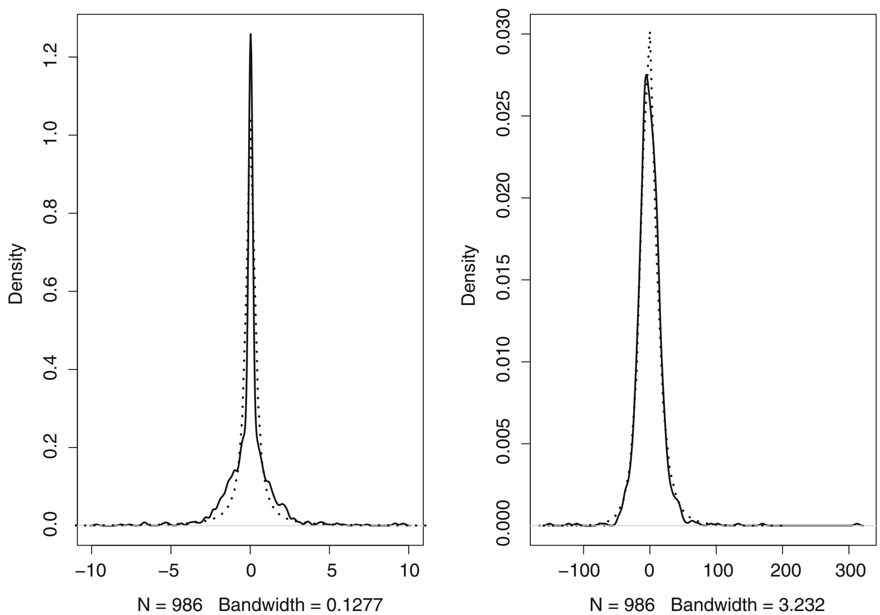


Fig. 3 Empirical density of L_1 (*left*) and L_2 (*right*) as well as the fitted NIG density (*dashed line*)

future studies. From the estimates in Table 2 we observe that the NIG distributions for L_1 and L_2 are close to symmetric.

Following step (8), the market price of risk $\theta = (\theta_1, \theta_2)$ is given by

$$\theta_1 = \frac{\alpha_1 \frac{\hat{a}-\mu}{\delta}}{\sqrt{(\frac{\hat{a}-\mu_1}{\delta_1})^2 + 1}} - \beta_1 \tag{17}$$

$$\theta_2 = \frac{\alpha_2 K}{\sqrt{K^2 + 1}} - \beta_2, \tag{18}$$

where

$$K = \frac{\beta_2}{\delta_2}\left(\hat{c} - \frac{1}{2}\hat{a}(T_2 - T_1) - \frac{\mu_2}{\beta_2}\right).$$

Here, the subscript in the parameters α, β, δ and μ refer back to L_1 and L_2. Using the estimates for the NIG distributions, we can derive the values of θ_1 and θ_2. These are reported in Table 3 along with the expected values of L_1 and L_2 with respect to the probabilities P and the fitted Q_θ. We note that the market price of risk is positive, and that the expected value of L_1 and L_2 are moved from being negative under P to positive under Q_θ. The fitted market price of risk is shifting the distribution of L_1 and L_2 towards the right, roughly meaning that we get more positive jumps and less negative. Furthermore, quite nicely the NIG distribution is preserved under a constant Esscher transform. Hence, L is a bivariate NIG Lévy process both under P and Q_θ, where only the skewness parameter is different under the two measures.

Let us comment on the risk premium implied by our estimated model. The risk premium is defined as the difference between the futures price and the predicted average spot price over the delivery period. In mathematical terms,

$$R_P(t, T_1, T_2) = F(t, T_1, T_2) - \mathbb{E}\left[\frac{1}{T_2 - T_1}\int_{T_1}^{T_2} S(t)\,dt \,|\, \mathcal{F}_t\right]. \tag{19}$$

From Proposition 3.2 we find

Table 3 The market price of risk derived from the fitted NIG parameters together with the expectation of L_1 and L_2 under P and Q_θ

i	θ_i	$\mathbb{E}_P[L_i(1)]$	$\mathbb{E}_\theta[L_i(1)]$
1.	0.0115	−0.0087	0.0296
2.	0.0010	−0.3583	0.0211

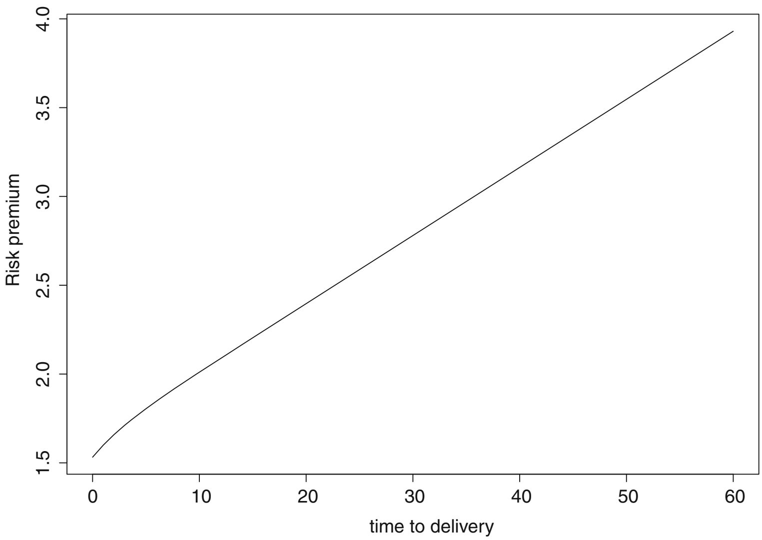


Fig. 4 Theoretical risk premium for the estimated model parameters

$$\begin{aligned} R_P(t,T_1,T_2) =& \frac{1}{2}(\mathbb{E}_\theta[L_1(1)] - \mathbb{E}[L_1(1)])(T_2-T_1) \\ &+ (\mathbb{E}_\theta[L_1(1)] - \mathbb{E}[L_1(1)])(T_1-t) \\ &+ (\mathbb{E}_\theta[L_2(1)] - \mathbb{E}[L_2(1)])\frac{1}{\eta}(1-\bar{\eta}(t,T_1,T_2)). \end{aligned}$$

The non-stationary factor gives a linear contribution in time to delivery $T_1 - t$, while the stationary factor gives an exponential shape and converges fastly to a constant when $T_1 - t \to \infty$. A plot of the risk premium for the estimated model parameters is shown in Fig. 4. As a result of the positive market price of risk, the risk premium also becomes positive. This tells us that the consumers in the market are willing to pay a premium for locking in electricity prices in the futures market. Note that we use data from the relative short end of the market, using the front-month (or second month) contracts.

5 Pricing of Options on Futures

At EEX, the market for options is rather illiquid, however, there exists traded contracts. In 2008, 12 options on baseload futures with delivery period 1 month were traded, 11 of them in the period we consider. Out of these 11, four are call options, and seven puts. We use these for further analysis and discussion.

In Tables 4 and 5 we list the calls and puts with their main characteristics. We have decided to label the contracts by Ci, $i = 1,2,3,4$ for the calls and Pi,

Table 4 Traded call options in 2008 with delivery period 1 month

Contract	Trading day	Delivery period	Strike	Futures price	Settlement price
C1	06.02.2008	Mar 2008	57	56.81	1.900
C2	28.01.2008	Mar 2008	57	57.00	2.270
C3	15.01.2008	Feb 2008	75	70.50	1.065
C4	09.01.2008	Feb 2008	74	68.50	0.928

Table 5 Traded put options in 2008 with delivery period 1 month

Contract	Trading day	Delivery period	Strike	Futures price	Settlement price
P1	08.07.2008	Aug 2008	74	74.77	3.233
P2	08.07.2008	Aug 2008	75	74.77	3.835
P3	03.07.2008	Aug 2008	73	78.00	1.989
P4	08.04.2008	May 2008	55	55.35	1.522
P5	04.03.2008	Apr 2008	58	58.70	1.911
P6	28.02.2008	Apr 2008	58	61.75	0.955
P7	08.01.2008	Feb 2008	65	69.00	1.179

$i = 1, \ldots, 7$ for the puts. Recall that the exercise time τ of the options is four trading days before the delivery period of the underlying futures starts. The historical data available from the EEX provides settlement prices for traded option contracts. For all derivatives traded, a settlement price is established on all exchange trading days. In the case that a settlement price cannot be determined on basis of the order book situation, a so-called Chief Trader Procedure applies, where all trading participants can take part with a representative. The EEX Market Supervision makes a standardised form available for all those trading participant volunteering to specify a market price for the respective derivatives. The settlement price is then determined as the average of the expectations of the market participants. We note that options on peakload futures are not traded at all at EEX, explaining why we use baseload spot data in our empirical analysis above.

We first look at the "classical" approach to price options on futures in commodity markets, namely pricing using the Black-76 formula (see Black 1976). For the convenience of the reader, we state the Black-76 formula in a Proposition.

Proposition 5.1 *Suppose the risk-neutral futures price dynamics is a geometric Brownian motion*

$$\frac{dF(t,T_1,T_2)}{F(t,T_1,T_2)} = \sigma\, dB(t),$$

for a constant $\sigma > 0$. Then, the price at time $t \leq \tau$ of a call option with strike K and exercise time $t \leq \tau \leq T_1$, is given by $C_{B76}(t, F(t,T_1,T_2))$ with

$$C_{B76}(t,x) = e^{-r(\tau-t)}\left[x\Phi(d_1(x)) - K\Phi(d_2(x))\right],$$

for Φ *being the cumulative standard normal distribution function, and*

$$\begin{aligned} d_1(x) &= \frac{\ln\left(\frac{x}{K}\right) + \frac{1}{2}\sigma^2(\tau - t)}{\sigma\sqrt{\tau - t}}, \\ d_2(x) &= d_1 - \sigma\sqrt{\tau - t}. \end{aligned}$$

In the Black-76 formula, one boldly assumes the futures price dynamics to be a geometric Brownian motion, a dynamics which is far from the one we have estimated to the electricity futures prices at the EEX. The volatility σ is also constant, an assumption that is not likely to be true. Based on the historically estimated volatility of the futures contracts in question, we can price the call options. The Black-76 prices are reported in Table 6 along with the actual settlement prices as quoted on the EEX. Appealing to the put-call parity, we report the put prices in Table 7. In both tables, we have also reported the historical volatility σ used in the Black-76 formula, as well as the implied volatility so that Black-76 matches the settlement price. We estimate the historical volatility of the logreturns of the underlying futures from the last month of daily price data. Furthermore, we choose $r = 5\%$ which is about the average yearly Euro LIBOR rate in 2008. We find that the price of all options are substantially underestimated by Black-76. Due to the low

Table 6 Black-76 pricing of the call options

Contract	Settlement price	Black-76	Mispricing (%)	Hist. vol.	Impl. vol.
C1	1.900	0.464	−76	0.1046	0.3770
C2	2.270	0.725	−68	0.1100	0.3560
C3	1.065	0.000	−100	0.0788	0.5030
C4	0.928	0.000	−100	0.0821	0.4450

Table 7 Black-76 pricing of the put options

Contract	Settlement price	Black-76	Mispricing (%)	Hist. vol.	Impl. vol
P1	3.233	0.693	−79	0.1491	0.521
P2	3.835	1.158	−70	0.1491	0.532
P3	1.989	0.055	−97	0.1496	0.509
P4	1.522	0.177	−88	0.0679	0.357
P5	1.911	0.295	−85	0.1014	0.394
P6	0.955	0.001	−100	0.0797	0.366
P7	1.179	0.000	−100	0.0842	0.437

volatility, those options that are far out of the money have a Black-76 price being essentially 0 (P6 and P7, and C3 and C4). The implied volatility becomes very high compared to the historical volatility. Indeed, the historical volatility is in the modest range of 8–11 % for the underlying futures of the call, whereas the implied volatilities are estimated to be from 35 to 50 %. The mispricing is rather dramatic, as the percentages ranging above 70 % tells. One might be tempted to speculate that the market is adding a huge risk premium for effects like illiquidity of the options and non-normality of the futures price dynamics. The issuer runs a big risk selling call options, since it is difficult to turn around the position in the option market. However, the underlying future is reasonably liquid, so delta hedging is possible. This removes some of the liquidity risk for the issuer.

One can in theory create synthetic investment strategies mimicking to a large extent the payoff of a call or put option. This could be used in order to exploit potential arbitrages in the option market. However, if the futures dynamics is not a geometric Brownian motion, there will be a large residual error in such strategies, which theoretically can be made perfect by delta hedging in the Black-76 framework. The empirical study of spot and futures pricing in the previous Section strongly points towards non-Gaussian models, hence ruling out this possibility.

In any case, the conclusion so far is that Black-76 in its simplest form is inadequate for pricing of options in the EEX market. As our proposed futures price dynamics is far more sophisticated than a simple geometric Brownian motion, we now move on to analyse the implied option prices from this model with the hope that it can improve the situation.

The call option price is then given by

$$C(t) = e^{-r(\tau-t)}\mathbb{E}_Q[F(\tau, T_1, T_2) - K \,|\, \mathcal{F}_t]. \tag{20}$$

The pricing probability Q is an equivalent martingale measure for $F(t, T_1, T_2)$, and we let this be given by Q_θ. The Q_θ-dynamics of $F(t, T_1, T_2)$ is given by Proposition 3 and Q_θ is determined through the market price of risk (17) from the spot-futures analysis above. We evaluate the expectation through a Monte-Carlo simulation. To simulate the Lévy processes L_1 and L_2 under Q_θ, we use that NIG-Lévy processes are stable with respect to an Esscher change of measure. In fact, it can be seen (see Benth et al. 2008) that if, for $i = 1, 2$, $L_i(1)$ is NIG distributed under P with parameters $\alpha_i, \beta_i, \delta_i$ and μ_i, then the $L_i(1)$ is again NIG distributed under Q_θ with the same parameters except the skewness, which becomes $\beta_i + \theta_i$.

Based on a simulation of 1,000,000 paths we compute the option prices based on the average payoff. To simulate the NIG distribution, we applied the algorithm implemented in the R-package fBasics, which is based on the normal variance-mean mixture of the NIG distribution.

The resulting numbers are reported in Tables 8 and 9. We have also included the mispricing and computed the implied volatility of the simulated price using the Black-76. From the tables, we see that the picture is more mixed, with both over and

Table 8 Simulated prices of the call options

Contract	Settlement price	Simulated price	Mispricing (%)	Impl. vol.
C1	1.900	2.748	45	0.4820
C2	2.270	3.525	56	0.4882
C3	1.065	0.821	−23	0.4255
C4	0.928	1.006	9	0.4037

Table 9 Simulated prices of the traded put options

Contract	Settlement price	Simulated price	Mispricing (%)	Impl. vol.
P1	3.233	2.476	−23	0.4256
P2	3.835	2.964	−23	0.4239
P3	1.989	1.438	−28	0.4260
P4	1.522	2.397	57	0.5358
P5	1.911	2.659	39	0.5368
P6	0.955	1.889	98	0.5384
P7	1.179	1.376	17	0.4032

underpricing of the calls and puts. Moreover, at the first glance, the mispricing seems to be less severe than in the case of Black-76, although admittedly still very big.

Our spot and futures price model includes non-Gaussian noise as both factors in the spot are driven by an NIG Lévy process. Note that the futures price is depending on the non-stationary factor directly, whereas the short-term factor is dampened and negligible for contracts far from delivery. From our estimation procedure, the non-stationary long-term factor is estimated from the futures prices, so if the market would price according to our futures price dynamics with the given pricing measure Q_θ, at least options with long time until exercise should be priced reasonably accurate. Looking at C1 and C2, these have the longest time to exercise in our sample of call options. However, the simulated option prices from our model for these two contracts are approximately 50 % higher than the quoted prices. This means that our model is pricing in *too much* risk. From the spot dynamics we estimated positive market prices of risk which pushes the skewness of the NIG distribution to more positive jumps. The more positive market price of risk, the higher values of the options. Thus, it seems like the futures model inherits far too much risk premium from the spot when it comes to option pricing. We reach the conclusion that the option market is not including the same risk perception as the one inherited from the spot in the futures market. This is a clear sign that a completely different pricing measure Q is used in the option market than in the futures pricing. Note that C1 and C2 are both (approximately) at-the-money, so a big portion of the distribution of the futures is taken into account in the pricing.

The contracts C3 and C4 are far out-of-the-money and slightly closer to exercise time than C1 and C2. Noteworthy is that the mispricing of these are significantly less than for C1 and C2, being respectively −23 and 9 %. If we have based our calculations of the call prices on the wrong risk premium, it will be more influential far from exercise than close since we span out more of the risk the longer into the future we simulate the futures price. Close to exercise, the misspecification of the tails under the chosen Q will be relatively much smaller than when we move futures. Maybe more importantly is that a smaller portion of the price distribution of F is taken into account for these two out-of-the-money options than C1 and C2, and hence a wrongly chosen Q matters less. This discussion conforms with the observations above for C1 and C2.

Note that contract P6 is farthest from exercise among the put options, as well as being out-of-the-money. This contract has the highest mispricing by our model. All the other put contracts in our sample have shorter time left to exercise. P1, P2, P4 and P5 are all approximately at-the-money put options with almost the same time left until exercise. The mispricing of these are significantly less than for P6. In fact, for P1 and P2 our model gives a price $-23\,\%$ less than the settlement prices. P4 and P5 are contracts on futures with delivery in the spring months May and April, respectively. Temperature predictions may influence the futures price expectations, as the spring may become colder or warmer than usual. We also note that it is the left-tail of the futures price distribution that counts when pricing an out-of-the-money put option. An underpricing can be the result of the distribution being moved to the right by a positive risk premium.

P3 and P7 are out-of-the-money put options where the mispricing of our model is rather modest ($-28\,\%$ and 17 %, respectively). P7 is the only put option written on a futures with delivery in the winter period, namely February. For the calls C3 and C4, which also are written on February futures contacts, we observe a relatively small pricing error. It seems like the model captures best the futures price evolution in the winter term. We also remark that the poor fit of the autocorrelation of the stationary factor Y may lead to wrong assessments of the spike influence. However, we believe that this is to some extent compensated for by the good fit of the Lévy process L_2 using a NIG distribution.

All in all, it seems like the futures price dynamics based on the pricing measure Q_θ implied by the spot-futures relationship provides a significantly better prediction of option prices than Black-76. However, the prices are far from satisfactory, and we find clear evidence that the risk-adjustments should be different than those given by Q_θ. Based on our findings, we dare to conclude that another pricing measure $\tilde{Q}$ should be used for power option pricing, a pricing measure which attributes a different loading on the distributions of the Lévy processes L_1 and L_2. In fact, based on the differences between summer and winter contracts in the pricing analysis above, one may suspect that such a measure change should incorporate seasonalities as well. Furthermore, it may also account for the state of the futures price, so that one can capture out-of and in-the-money option price differences better. One can also think of pricing measures which not only changes the characteristics of the jump processes L_1 and L_2, but as well change the dynamics. For example, it is possible to define measures which change the speed of mean reversion of the Y factor. This could for example lead to a slower risk-neutral speed of mean reversion, essentially saying that a spike lasts longer in a risk neutral context than under the market probability.

As our futures price dynamics consists of two jump components, it gives rise to an incomplete market model. The selection of risk neutral probabilities for pricing options in such markets is frequently based on utility indifference pricing techniques (see Rouge and El Karoui 2000). Such a method, which is based on a risk averse, utility optimizing investor, leads to a partial hedging strategy of the option. The utility indifference method is particularly useful when pricing options in illiquid markets, where one is stuck with the option investment. Other approaches to pricing is based on deriving optimal partial hedges, where the optimality criterion may be the futures investment hedge which minimizes the variance of the hedging error (see Cont and Tankov 2004). All these various approaches lead to a pricing measure Q. It is of great interest and application to see whether such prices will explain the settlement option prices in the EEX market, and whether our conjecture $\widehat{Q} \neq \tilde{Q}$ is true.

As we have mentioned earlier, the option market at the EEX is rather illiquid, and a liquidity premium is naturally associated to the observed prices. This premium will be part of the risk premium as we have estimated. Once bought or sold an option, one might be stuck with the position taken until exercise or having to accept a significant loss by reversing it. This will impact the settlement prices as buyers and sellers know that it is difficult to get out of the position at a later time. On the other hand, the underlying futures market is reasonably liquid, so any position can (in theory, at least) be hedged to a certain degree. Considering our derived Black-76 prices, which were consistently too high, one could speculate whether the sellers had to accept a discount in prices due to illiquidity. However, for the much more realistic futures price dynamics that we considered, the picture was more mixed with both over- and underpricing. There is no doubt that a liquidity premium exists in the market, but it is hard from our analysis to conclude anything on its size and structure. Moreover, liquidity might also be an issue in the futures market, further complicating matters.

6 Conclusion

We have argued that in power markets one may use a probability measure $\widehat{Q}$ for futures pricing based on spot modelling which can be different than the equivalent martingale measure $\tilde{Q}$ used for pricing options on the futures. There is no violation of no-arbitrage pricing theory that $\widehat{Q} \neq \tilde{Q}$, and the argument hinges on the fact that electricity spot cannot be stored. Due to the non-storability, $\tilde{Q}$ can be chosen as an equivalent measure which is not necessarily turning the discounted spot dynamics into a $\tilde{Q}$-martingale. On the other hand, $\widehat{Q}$ is an equivalent measure such that the futures price becomes a $\widehat{Q}$-martingale.

We introduce a two-factor model for the spot price dynamics being a generalization of the classical commodity model of Schwartz and Smith (2000). Both the long-term and the short-term factors are driven by normal inverse Gaussian Lévy processes, a choice based on empirical arguments using data collected at the EEX. The spot model allows for analytical futures pricing, where the Esscher transform provides an parametric class of probability measures to model the risk premium. We perform a joint estimation of the spot and futures, where the crucial step is to apply long-dated futures contracts to filter out the non-stationary long-term factor of the spot.

Applying Monte Carlo simulations we priced call and put option prices for our proposed futures dynamics. We compared the simulated prices where we chose $\widehat{Q} = \tilde{Q}$ with observed option prices in the market. This led to a significant mis-pricing, and we argued that the results pointed to the fact that $\widehat{Q} \neq \tilde{Q}$. Our results were benchmarked against the Black-76 prices using the historical volatility of the underlying futures as input. The proposed spot-futures model was a clear improvement over this in predicting option prices.

We did not suggest any probability $\widehat{Q}$ for the futures price which could remedy the situation. There exists many potential approaches to produce such risk neutral probabilities taken from the theory of derivatives pricing in incomplete markets. But before setting off such a study, one should make the spot dynamics even more sophisticated to take into account some defiance like the misspecification of the autocorrelation structure of the stationary factor. CARMA processes could be a choice here, or even more general Lévy semistationary processes. However, this will require more advanced estimation procedures to fit to data. On the other hand, such improvements will make the conclusions on option pricing and choice of risk neutral measures less prone to model error. A further issue is to open for more flexible pricing measures for the futures price, taking into account random changes in the risk premium and impacts from fuels and weather.

Illiquidity of the power option markets is a clear issue which can question our analysis. Power options are relatively little traded, but we believe that in the future these markets will emerge as important one for hedging and speculation of power. The results in our paper will hopefully provide an important guideline in the challenges when pricing spot, futures and options simultaneously.

Acknowledgments Fred Espen Benth acknowledges the financial support from the project "Energy markets: modelling, optimization and simulation" (EMMOS), funded by the Norwegian Research Council under grant evita 205328/v30.

References

Barndorff-Nielsen OE (1998) Processes of normal inverse Gaussian type. Finance Stoch 2(1):41–68

Barndorff-Nielsen OE, Benth FE, Veraat AED (2013) Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. Bernoulli 19(3):803–845

Benth FE, Šaltytė-Benth J, Koekebakker S (2008) Stochastic modelling of electricity and related markets. World Scientific, Singapore

Benth FE, Klüppelberg C, Müller G, Vos L (2011) Futures pricing in electricity markets based on stable CARMA spot models (Preprint)

Benth FE, Šaltytė-Benth J (2004) The normal inverse Gaussian distribution and spot price modelling in energy markets. Int J Theor Appl Finance 7(2):177–192

Benth FE, Sgarra C (2012) The risk premium and the Esscher transform in power markets. Stoch Anal Appl 30:20–43

Black F (1976) The pricing of commodity contracts. J Financial Econ 3:169–179

Börger R, Cartea A, Kiesel R, Schindlmayer G (2009) A multivariate commodity analysis and applications to risk management. J Futures Mark 29(3):197–217

Cont R, Tankov P (2004) Financial modelling with jump processes. Chapman & Hall/CRC, Boca Raton

Duffie D (1992) Dynamic asset pricing theory. Princeton University Press, Princeton

Lucia JJ, Schwartz ES (2002) Electricity prices and power derivatives: evidence from the Nordic power exchange. Rev Deriv Res 5(1):5–50

Rouge R, El Karoui N (2000) Pricing via utility maximization and entropy. Math Finance 10:259–276

Samuelson P (1965) Proof that properly anticipated prices fluctuate randomly. Ind Manag Rev 6:41–44

Shiryaev AN (1999) Essentials of stochastic finance. World Scientific, Singapore

Schwartz E, Smith JE (2000) Short-term variations and long-term dynamics in commodity prices. Manag Sci 46(7):893–911

Switching from Feed-in Tariffs to a Tradable Green Certificate Market

Aitor Ciarreta, Maria Paz Espinosa and Cristina Pizarro-Irizar

Abstract Feed-in tariffs have been the key support system for electricity from renewable sources in Spain and other European countries. However, given the growing criticism of this incentive scheme mainly due to its financial burden, the Spanish government has recently cancelled subsidies for any new electricity from renewable sources (RD-l 1/2012 2012). Since tariffs do not benefit from market signals, subsidies to some technologies may be either too high or too low to attain the regulator's objectives. Existing literature on tradable green certificates suggests that a switch to a green certificates setup could be an efficient solution when substantial investments in renewable energy are already in place and technologies are at a mature stage. This chapter analyzes the implementation of a green certificates scheme as an instrument to foster renewables. We solve a sequential game where we analyze the interaction between the electricity pool and the tradable green certificates market. We focus on the retailer regulation design that would give lead to a decreasing green certificates demand and simulate the effect of such regulation on the price of certificates.

Keywords Tradable green certificates · Energy policy · Electricity market · Renewable energy

---

We would like to thank the Ministry for Science and Innovation (ECO2012-35820), the Basque Government (DEUI, IT-IT783-13) and UPV/EHU (UFI 11/46 BETS) for financial support. Cristina Pizarro-Irizar also thanks the financial support from a predoctorate grant from the Basque Government (BFI-2011-301).

A. Ciarreta · M.P. Espinosa · C. Pizarro-Irizar (✉)
Department Foundations of Economic Analysis II, University of the Basque Country, UPV/EHU, Lehendakari Aguirre 83, 48015 Bilbao, Spain
e-mail: mariacristina.pizarro@ehu.es

A. Ciarreta
e-mail: aitor.ciarreta@ehu.es

M.P. Espinosa
e-mail: mariapaz.espinosa@ehu.es

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_11

1 Introduction

New or emerging electricity from renewable sources (RES-E) is not profitable in a free market, due to relatively high production costs, and support instruments are hence introduced to help the penetration of renewable technologies (Menanteau et al. 2003). This promotion seeks to improve market efficiency, internalize external costs, accelerate investments in research and provide temporary incentives for early market development as such new technologies approach commercial readiness (Sims et al. 2003). Additionally, RES-E support systems increase the amount of RES-E produced through the merit order effect, since electricity production from conventional fossil-fuel sources (marginal plants) is then substituted by RES-E and the wholesale price of electricity drops.[1] However, the net effect on the consumer price level will depend on the way in which the RES-E support system is financed (Rathman 2007). The burden of RES-E deployment usually rests on final consumers and the choice of the promotion instrument and how it is implemented is crucial (Haas et al. 2011).

In this sense, RES-E has been promoted in some countries through feed-in tariff (FIT) schemes, or its variant feed-in premium (FIP). Under this system RES-E producers may sell their entire output at a guaranteed price that is set above the wholesale market clearing price. This higher price allows the generators of some renewable sources of energy to cover the higher costs of this type of energy and stay in the market.

A different way of promoting green sources of energy is the creation of tradable green certificate (TGC) markets. The regulator may create a demand for the renewable energy through the distributors' obligation to meet a specified share of green energy. Green certificates, which are also referred to as renewable energy certificates (REC), tradable REC (T-REC), tradable renewable certificates or credits (TRC), renewable portfolio standards (RPS), green tickets or green tags, rely on market mechanisms for resource allocation. These markets aim at the promotion of green energy sources through the separation of electricity as a commodity (traded in the wholesale market) from the ecological attributes of electricity (avoiding CO_2 emissions, etc.), which are traded as a different product on the green certificate market. Indeed, both markets are separated but there are strong interactions between the determination of the price of the certificate, the price of the electricity and the role of regulation (Jensen and Skytte 2002).

Comparing both systems in Europe, the FIT approach (price-based mechanism) is generally more popular than the TGC approach (quantity-based mechanism), as it guarantees the price and removes the risk from investors in renewable generation; whereas the TGC scheme may involve higher uncertainties, due to market outcomes, and investors consequently require higher payments (Neuhoff 2005).

[1] For an analysis of the effect of renewable electricity production in the Spanish electricity market see Ciarreta et al. (2012a, b) and for an analysis of the effect of regulation in the electricity prices in Spain see Ciarreta and Espinosa (2012).

This conclusion may partly rest on the European experience, where FIT regimes in Germany or Spain outperformed the TGC scheme in other countries, such as the United Kingdom (Buttler and Neuhoff 2008). However, it is argued that FIT may serve mainly to shift the risk to other agents (i.e. consumers) but does not reduce it to society as a whole. Moreover, the problem with FIT is the need to set the tariff at an appropriate level, risking that it may be too high, creating excessive rents for some generators, or too low, restricting investment (Green and Vasilakos 2011). In fact, as has been seen, regulatory uncertainty is one of the main problems of the FIT system.

Although a TGC system provides less market certainty than price-based mechanisms, price fluctuations and market dynamics can be partly influenced by the design of the regime (Gan et al. 2005). Another source of evidence in favor of TGC is effectiveness in the achievement of the goal to secure a certain share of renewables in electricity consumption (Bye 2003). It is expected that competition between producers and increasing supply of green certificates will lead to a decline in the price of electricity from renewable sources, so in this respect, the green certificate system is considered as a cost effective way to meet the renewable energy target (Schaeffer et al. 1999). One more argument in favor of a TGC market is the issue of equity, i.e. the fairness of the distribution of costs and benefit between different actor groups (Bergek and Jacobsson 2010). The market decides the level of support given to renewable electricity production, so apart from guaranteeing the production of a certain quantity of RES-E, green certificates are added to the revenue that the producer can get for the electricity itself. Additionally, the introduction of market forces on the 'non-electricity' attributes of energy is supposed to bring about greater efficiency. The transition to market based solutions leads to effective competition between different forms of power from renewable energy sources, since producers must try to benefit from technical progress due to the pressures of bidding processes in the certificates market (Menanteau et al. 2003).

At present there is no general agreement on the appropriateness of the different schemes. Existing literature supports that the type of allowance given to each renewable technology must be adapted according to their stages of maturity (Christiansen 2001; Meyer 2003; Jacobsson and Lauber 2006). Technological maturity is closely related to the cost per MWh of each technology. In this sense, three main categories may be distinguished according to their merit order (Jensen 2003):

- Cost-competitive technologies. These technologies are not eligible for policy support, since their production cost is similar to (or even lower than) conventional sources. This category includes large hydro.
- Moderately non-competitive technologies. These technologies are to be complemented with a relatively modest support system. The production cost of RES-E included in this category is higher than the cost of electricity generated by some conventional sources. Such technologies may include small hydropower, some biomass-based technologies and onshore wind power.

- Non-competitive technologies. Those technologies that are still far from being market-ready, but have the potential to join the first category in the longer term, should be supported by incentive schemes. These technologies would not survive without incentives, since the investment in R&D needed to make them competitive would never take place. Expensive technologies and technologies in the technical development phase, such as solar-photovoltaic or offshore wind power technologies, are included in this category.

When designing appropriate renewable energy support frameworks, one of the main criticisms of TGC markets concerns the competition between renewable technologies at different stages of development. On the one hand, if the certificate price corresponds to the most expensive renewable technology included in the system, all technologies with lower costs would receive an extra profit (Verbruggen 2004) and the promotion of the total renewable portfolio would be more expensive than necessary. On the other hand, if the certificate price corresponds to the moderately non-competitive technologies, one possible solution is to reserve the green market for the most mature renewable technologies (Meyer 2003); and so, photovoltaics being at an early stage could benefit from a FIT approach, while wind or biomass would be ready for competition in a TGC market (Midttun and Gautesen 2007). Moreover, instead of FIT, additional investment subsidies for solar power could be available, improving the economic incentives for investments in solar electricity.

In this chapter, we therefore suggest that, since many renewable technologies are nowadays at a quite mature stage (moderately non-competitive technologies, e.g. wind, biomass or small hydro), green certificates could work properly as a promotion instrument in a country such as Spain. Additionally, the European Commission considers the model of green certificates as the preferable candidate for a European common support scheme for renewable energy (Ringel 2006). The Commission also claims that the existing promotion framework should be improved in order to reach the target of a 20 % share of renewables in the EU's total energy consumption by 2020 at the lowest possible cost (European Commission 2011). Moreover, some authors claim that allowing for EU-wide trade in green certificates can ensure a cost-effective distribution of renewable energy production (it cuts the overall cost of achieving the EU's renewable target by almost 70 %), but differentiated renewable targets across countries reduce the cost-effectiveness of the TGC system as national targets prevent a cost-effective distribution of energy (Aune et al. 2012). We therefore examine, both theoretically and empirically, using actual data of the Spanish electricity market, the feasibility of a TGC market in Spain and conclude that it may help to reduce the financial burden derived from the FIT-FIP system.

Regarding the instruments for the promotion of RES-E, price-driven instruments (i.e. FIT) are designed to support the costs of electricity production, whereas quantity-driven instruments (i.e. TGC) fix a capacity target to be met. Generally, those instruments are implemented separately, but there are three cases in which the combination of them could be relevant for the promotion of RES-E (Schaeffer et al. 1999):

- During the transition phase from one instrument to another.
- When a country with a certain promotion scheme decides to allow producers to take part in a trading system with other countries, where another promotion scheme is implemented.
- In order to compensate for the disadvantages of one instrument, a permanent combination of instruments could be useful.

In the first case, a gradual introduction of the TGC market should be applied and transition schemes for existing plants should be established to ensure that the new investors have stable economic conditions until the market is fully functional. Regarding the second case, international trade with certificates can exploit comparative resource advantages and lower costs, as long as common rules for the trade of certificates and the period of validity are implemented. In the latter case, combining two different instruments may help to offset some of the disadvantages of each instrument. For instance, the market-based nature of TGC helps to reduce the regulatory uncertainties of the FIT scheme, but at the same time, non-competitive technologies might not be ready for a TGC system, so they would need a FIT to survive.

The chapter is structured as follows. Section 2 develops a theoretical model to analyze the interaction between the electricity pool and the TGC market. Simulations and results for the Spanish electricity market are presented in Sect. 3. Finally, Sect. 4 summarizes the main conclusions of this work.

2 The Model

In this section, we set up a model for the certificates market, the electricity market and the interaction between them. Electricity considered as a commodity is a homogenous good, independently of the energy source, and it is sold as such in a liberalized physical electricity market. The eco-services provided by some sources of energy are sold separately on the green certificates market. The ecological impact of different renewable energy sources may be different, along with the cost associated to the electricity system management. However, we assume here for simplicity that the ecological services of the renewable sources of energy are also a homogeneous good, ignoring for example differences between hydro and wind sources of energy.

We present a two-stage model under autarky (we consider a one country closed-economy without international trading) and we assume that both markets work under perfect competition. The game takes place in two stages:

- Stage 1: Electricity generators take supply decisions at the pool, retailers take demand decisions and the wholesale market clears. Generators are awarded certificates depending on their green production.
- Stage 2: Generators decide how many green certificates to sell. Retailers buy certificates to fulfill their obligations. The market for certificates clears.

Here it is assumed that the TGC are issued at the end of stage 1 to be sold at stage 2. We solve the game using backwards induction, i.e. we solve first the market for certificates (stage 2). We assume that each generator produces both renewable and non-renewable electricity and that each producer is endowed a certain amount of TGC depending on the clean electricity delivered to the network. We assume a one-to-one link between the number of green certificates endowed and the number of MWh produced by renewable technologies (i.e. 1 MWh = 1 TGC). Those TGC are assumed to have a regulator's defined life of one period[2] so banking is not allowed in our model and unused certificates are withdrawn from the market when the period expires.

With regard to the technology mix of RES-E, some authors are in favor of a technology neutral design in order to promote competition between the certificate-eligible technologies, so that the market decides which technologies are preferable to achieve the target, which encourages a cost-efficient deployment of renewable energy sources (Nilsson and Sundqvist 2007). On the contrary, other authors (Schmalensee 2011) suggest that technology-specific multipliers could be used to penalize some intermittent technologies, such as wind, for the costs they impose on the electric power system or even to reward some technologies because of the perceived external effect of induced learning-by-doing if their production is increased, such as biomass. For the sake of simplicity, we consider the technology neutral design and we treat all renewables as a whole in our analysis.

All things considered, there are three main actors in our model: the regulator, retailers (demand side), and generators (supply side); and two interacting markets. We do not consider any uncertainty for the sake of simplicity. Since retailers and generators behave competitively in both markets, they are unable to affect the price of electricity.

The notation used in the model is compiled in Table 1.

2.1 The Tradable Green Certificates Market

2.1.1 Regulation in the TGC Market

The market for TGC should be regulated due to information asymmetries: the energy attribute being sold in this market is not observable for the end-use consumer. The regulator therefore needs to certify the resources used in the energy production process and to assign the property rights of a TGC for each MWh produced by a generator (for example, issuing a certificate with a serial number). These certificates can then be marketed and their sale and use should be closely monitored.

[2] Certificates may have a longer life, and there may be certificates in the market with different lifespans and different trading prices. We ignore this issue for the moment.

Table 1 Notation of the model

Generating sector	
c_b	Intercept of the marginal cost function of black electricity
c_g	Intercept of the marginal cost function of green electricity
h	Parameter of the cost function of each generator $(h > 0)$
H	Parameter of the aggregate cost function
q_b	Quantity of black electricity (non-renewable) sold by one generator
q_g	Quantity of green electricity (renewable) sold by one generator
q_G	Total quantity of electricity (non-renewable + renewable) sold by one generator $(q_G = q_b + q_g)$
Q_b	Aggregate supply of black electricity
Q_g	Aggregate supply of green electricity
Q_G	Aggregate supply of electricity $(Q_G = Q_b + Q_g)$
x_G	Amount of TGC sold by one generator
X_G	Aggregate supply of TGC
Retailing sector	
a	Parameter of the demand function for electricity $(a > 0)$
b	Parameter of the demand function for electricity $(b \geq 0)$
q_R	Total quantity of electricity bought by one retailer
Q_R	Aggregate demand for electricity
x_R	Amount of TGC bought by one retailer
X_R	Aggregate demand for certificates
Market prices	
p_c	Price of the certificates at the TGC market
p_e	Price of electricity at the pool
Policy variables (regulated)	
α	Quota of green electricity imposed by the policy maker $(0 \leq \alpha \leq 1)$
d	Parameter of the penalty function of one retailer $\left(-\frac{1}{2} < d \leq 0\right)$
D	Parameter of the aggregate penalty function
f	Parameter of the penalty function of one retailer $(f > 0)$
F	Parameter of the aggregate penalty function
s	Price to the end-users of electricity
x	Retailers' obligation to purchase TGC $(x = \alpha q_R)$

Certificates are generally issued by government decision. This obligation could be transferred to the supply side (e.g. Italy) or to the demand side (e.g. retailers in the UK or end-users in Sweden). Our model considers that the obligation to buy TGC is set on retailers (calculated on the basis of the desired share of renewable consumption), in order to avoid the free-riding problem due to the public-good nature of the ecological benefit of green electricity (Menanteau et al. 2003). Relying renewable electricity demand on consumer choice has also been proposed as an alternative to obligatory schemes, but this option seems to have little impact on the deployment of renewable energy technologies (EWEA 2004), since most consumers prefer renewable energy but are not willing to purchase it at higher prices (Rader and Norgaard 1996). Moreover, we would expect the demand coming from end-use consumers to be so low that the equilibrium price would not reflect the social value of the ecological benefit of green energy. Thus, a mandatory quota of TGC for retailers may solve this market failure.

Clear consistent government policy is thus needed to set a stable green certificate system (Schaeffer et al. 1999). In order to protect both TGC producers and consumers, minimum and maximum prices could be established. Minimum values are secured when the government itself also acts as a buyer of green certificates (e.g. the Walloon region in Belgium), whereas maximum values are set through a penalty system for non-compliance. The role of policy makers in our TGC model lies in the establishing of (i) the amount of certificates that each green producer receives in relation to the proportion of green electricity produced (here one-to-one relationship), (ii) the retailers' obligation to purchase a minimum number of TGC (quota α) and (iii) the payment penalty if retailers do not meet their obligation.

2.1.2 The Role of Retailers in the TGC Market

Two parties are involved on the demand side: the end-users of electricity and the retailers. Retailers get their margins from buying wholesale and selling to end-users. We model demand for TGC as reflecting the retailer's obligation to pay for the environmental attributes of energy, which are related to the way it has been produced. Regulation may establish the obligation for retailers to meet some renewable energy requirements, and these obligations determine the demand on the TGC market. Each retailer must buy a fraction of total consumption.

In our analysis, electricity retailers have an incentive to buy certificates from the producers, because penalties are set if they are not able to meet their obligation. Retailers must pay a non-compliance fine depending on the number of certificates not bought. Our analysis models the penalty as a linear-quadratic loss function that leads to a decreasing demand for TGC. Therefore, the demand function has a price-cap, since no retailer would demand green certificates at a higher price than the penalty incurred for non compliance. Retailers not complying with the target would pay depending on the number of certificates not acquired. Retailers buying more than the target would neither pay for it nor receive any reward for the extra certificates acquired. The penalty function for a retailer is then given by:

$$P(x_R) = \begin{cases} f\left[\frac{1}{2}(x - x_R)^2 + d(x - x_R)\right] & \text{if } x_R < x \\ 0 & \text{if } x_R \geq x, \end{cases}$$

where $f > 0$ and $-\frac{1}{2} < d \leq 0$ are the parameters of the penalty function, x is the retailers' obligation to purchase TGC and x_R is the amount of TGC bought by the retailer.

By allowing retailers to choose the amount of TGC they want to buy in the market, we give an active role to demand. The elasticity of the demand for certificates will depend on the obligation x and the parameters of the penalty function, f and d. If we set $f = 0$, there would be no penalty in case the obligation to purchase certificates is not met. Retailers decide x_R in the second stage and the number of TGC traded is determined endogenously. Thus, the optimization problem for retailers is defined as follows:

$$\max_{x_R} \pi_R = q_R(s - p_e) - x_R p_c - f\left[\frac{1}{2}(x - x_R)^2 + d(x - x_R)\right],$$

where q_R is the total amount of electricity bought by one retailer, s is the price to the end-users of electricity, p_e is the price of electricity at the pool and p_c is the certificate price at the TGC market.

Since the retailer's obligation to acquire TGC depends on the demand for electricity in the previous period and the government target, the relation $x = \alpha q_R$ holds, where α is the quota of renewable electricity imposed by policy makers and $0 \le \alpha \le 1$. Hence, the optimization problem is equivalent to the following one:

$$\max_{x_R} \pi_R = q_R(s - p_e) - x_R p_c - f\left[\frac{1}{2}(\alpha q_R - x_R)^2 + d(\alpha q_R - x_R)\right]$$

Notice that the price s that end-consumers pay is perceived by the retailer as given. Likewise, the demand for electricity q_R and the selling price p_e are given at this stage, since when the TGC market opens, the energy production decisions have already been made and the energy market has cleared.

The first order condition reads:

$$\frac{\partial \pi_R}{\partial x_R} = -p_c - f[(\alpha q_R - x_R)(-1) - d] = 0$$

It follows that a retailer's demand for certificates is:

$$x_R = \alpha q_R + d - \frac{p_c}{f},$$

where $0 \le \alpha \le 1$ and $f > 0$.

The certificate system therefore is steered by the two parameters of the penalty function f and d, but also influenced by the regulated obligation α.

Aggregate demand for TGC is the total demand for certificates in the retailing sector:

$$X_R = \begin{cases} \alpha Q_R + D - \frac{p_c}{F} & \text{if } p_c < F(\alpha Q_R + D) \\ 0 & \text{if } p_c \ge F(\alpha Q_R + D), \end{cases} \tag{1}$$

where $X_R = \sum x_R$ is the aggregate demand for certificates, $Q_R = \sum q_R$ is the aggregate electricity consumption by end-users, $D = \sum d$ and $\frac{1}{F} = \sum \frac{1}{f}$.

We therefore may conclude that retailers' demand for TGC depends on the total amount of energy sold to the final consumer, the price of the certificates, the TGC percentage requirement α and the parameters of the penalty function D and F. The number of TGC that a retailer is willing to buy depends negatively on the certificate price. Zero demand occurs when $p_c \geq F(\alpha Q_R + D)$, while there is a positive demand for certificates as long as $p_c < F(\alpha Q_R + D)$ holds. Since the price of the certificates cannot be negative, the condition $F(\alpha Q_R + D) > 0$ must always hold.

2.1.3 The Role of Generators in the TGC Market

Since generators hold the property rights on the energy they produce and the renewable attribute of energy, TGC supply is determined by the optimal generators' decisions concerning the selling of green certificates. Each generator can produce both renewable and non-renewable electricity and it is endowed a certain amount of TGC depending on the clean electricity delivered to the network. Hence, total supply of certificates is constrained to the production of green electricity. We assume a one-to-one link between the number of green certificates endowed and the number of MWh produced by renewable technologies.

Regarding costs, we assume additively separable cost functions with respect to the quantities of conventional and renewable energy sources. We also assume linearly increasing marginal costs. Total costs for black and green generation are respectively:

$$C_b(q_b) = c_b q_b + \frac{1}{2} h q_b^2,$$
$$C_g(q_g) = c_g q_g + \frac{1}{2} h q_g^2,$$

with $c_g > c_b$, since the technologies subject to green certificates are classified as moderately non-competitive (see Sect. 1). By assuming the same parameter h $(h > 0)$ for both cost functions, we ensure that the marginal cost function of renewable electricity is always higher than the marginal cost function of fossil electricity (the marginal cost curves do not cross).[3]

We assume that there is perfect competition in the certificates market, so firms are not able to modify the market price by means of changing its own certificates production or demand.

[3] Other authors have modelled these cost functions with two different parameters to allow for differences in the level of marginal costs of black and green electricity (Ciarreta et al. 2011).

The maximization problem that each generator solves reads as follows:

$$\max_{x_G} \pi_G = p_e(q_b + q_g) + x_G p_c - (c_b q_b + \frac{1}{2} h q_b^2) - (c_g q_g + \frac{1}{2} h q_g^2)$$

subject to:

$$x_G \leq q_g,$$

where q_g and q_b are the quantity of green and black electricity sold by one generator, x_G is the amount of TGC sold by the generator and h, c_g and c_b are parameters of the cost function.

Perfect competition in the certificates market ensures that $x_G = q_g$ and $X_G = Q_g$, so the aggregate supply of certificates under perfect competition is the electricity produced by green sources.

2.1.4 Market Balance for Green Certificates

In order to determine the equilibrium certificate price we use the condition of market balance for tradable green certificates. The total number of certificates has to be equal to the demand for certificates: $X_G = X_R$. From (1), the market balance equation for the TGC market is therefore given by:

$$Q_g = \alpha Q_R + D - \frac{p_c}{F}$$

Hence, the price of certificates may be written as:

$$p_c^* = F[D + \alpha Q_R - Q_g]$$

The higher the deviation $\alpha Q_R - Q_g$, the higher the price. With no deviation, $\alpha Q_R - Q_g = 0$, the certificate price is $p_c = FD$. The price increases with the deviation from the objective and therefore provides the incentives for investment in green energy sources.

Finally, since $Q_R = Q_g + Q_b$, the TGC price and quantity in equilibrium can be expressed as:

$$p_c^* = F[\alpha Q_b + D + (\alpha - 1)Q_g] \tag{2}$$

$$X^*(p_c^*) = Q_g \tag{3}$$

This means that the certificate price would be zero if the quantity of green electricity (Q_g) were higher than the target $(\alpha(Q_b + Q_g))$. Remember that by the time the certificate market meets, the electricity market has already cleared, so the volume of green energy produced will be known.

2.2 The Electricity Market

2.2.1 The Generators' Behavior in the Electricity Market

We assume that each generator has renewable (green) and non-renewable (black) energy production (q_g and q_b, respectively) and that both types of production plants are necessary to satisfy the demand for energy. We also consider that there are no capacity constraints and that production costs are higher for renewable energy production. The generator decides about its level of electricity supply in stage 1. Hence, the optimization problem to be solved by each generator is:

$$\max_{q_b, q_g} \pi_G(q_b, q_g) = p_e(q_b + q_g) + p_c q_g - (c_b q_b + \frac{1}{2} h q_b^2) - (c_g q_g + \frac{1}{2} h q_g^2)$$

The first order conditions are:

$$\frac{\partial \pi_G}{\partial q_b} = p_e - c_b - h q_b = 0$$
$$\frac{\partial \pi_G}{\partial q_g} = p_e + p_c - c_g - h q_g = 0$$

Generators are assumed perfectly competitive[4] in both markets and they produce green electricity so that marginal revenue $(p_e + p_c)$ equals marginal cost $(c_g + h q_g)$.

The supply functions of black and green energy are respectively:

$$q_b = \frac{p_e - c_b}{h}$$
$$q_g = \frac{p_e + p_c - c_g}{h}$$

Under a TGC system the payment received by green producers for each certificate should cover the extra costs involved in producing green electricity in comparison with fossil fuel-based electricity. Thus, the certificate price corresponds to the difference between the marginal cost of renewables $(c_g + h q_g)$ and the market price for electricity (p_e). As long as firms are able to cover their costs they will be willing to stay in operation.

The aggregate supply is:

$$Q_b = \frac{p_e - c_b}{H}$$
$$Q_g = \frac{p_e + p_c - c_g}{H}$$
$$Q_G = Q_b + Q_g = \frac{2p_e + p_c - (c_b + c_g)}{H},$$

where $\frac{1}{H} = \sum \frac{1}{h}$.

[4] For an analysis of market power in electricity markets see Ciarreta and Espinosa (2010a, b).

2.2.2 The Retailers' Behavior in the Electricity Market

We assume a linear demand function for electricity with parameters a $(a > 0)$ and b $(b \geq 0)$:

$$Q_R = a - bp_e$$

We assume that b is not large and that the condition $b(H - \alpha F) + 2 > 0$ holds.

2.2.3 Market Balance for Electricity

In equilibrium, total supply of electricity $(Q_G = Q_b + Q_g)$ has to be equal to the net demand for electricity (Q_R):

$$Q_G = Q_b + Q_g = Q_R$$

And, thus, we get the following electricity price in terms of the expected certificate price:

$$p_e^*(p_c) = \frac{aH + c_b + c_g - p_c}{2 + bH} \tag{4}$$

This result shows that there is a negative relationship between the electricity price and the certificate price: the higher the expected certificate price, the lower the electricity price.

Similarly, the quantity of electricity in equilibrium is as follows:

$$Q_b^*(p_e^*) = \frac{p_e^* - c_b}{H}$$

$$Q_g^*(p_e^*) = \frac{p_e^* + p_c - c_g}{H}$$

$$Q^*(p_e^*) = Q_b^*(p_c) + Q_g^*(p_c) = \frac{2p_e^* + p_c - (c_b + c_g)}{H}$$

Inserting the price (4) in the quantity functions yield:

$$Q_b^*(p_c) = \frac{aH - p_c + c_g - c_b(1 + bH)}{H(2 + bH)} \tag{5}$$

$$Q_g^*(p_c) = \frac{aH + p_c(1 + bH) + c_b - c_g(1 + bH)}{H(2 + bH)} \tag{6}$$

$$Q^*(p_c) = Q_b^*(p_c) + Q_g^*(p_c) = \frac{2a + p_c b - b(c_b + c_g)}{2 + bH} \tag{7}$$

Equations (5)–(7) show that the price of certificates increases the production of green electricity and decreases the production of black electricity. But this effect seems to be stronger in the green production and, hence, the total production of electricity is positively affected by the price of certificates, showing a greater influence as parameter b of the demand function rises. Regarding costs, the supply of non-renewable electricity is positively affected by the cost parameter of renewable energy, whereas the supply of green electricity is increased by the cost parameter of black production. Both supplies are negatively affected by their own generation costs.

2.3 Equilibrium in the Electricity Market and the Green Certificates Market

As stated before, the game is played sequentially (electricity market clears first and TGC market second), so we proceed using backward induction. We thus start from the demand for certificates Eq. (1), determined in stage 2, and we substitute the expression for the equilibrium quantity of electricity (7). Since $Q_R = Q_G = Q_b + Q_g$, the demand for certificates can be expressed in terms of the certificate price as:

$$X_R(p_c) = \frac{F\alpha[2a - b(c_b + c_g)] + DF(2 + bH) + (bF\alpha - 2 - bH)p_c}{F(2 + bH)} \tag{8}$$

In order to have a negative relationship between the number of TGC and the certificate price, we need $b(H - \alpha F) + 2 > 0$, which holds by assumption. Note that for values of b close to zero, this inequality always holds.

Additionally, from the equilibrium of the TGC market (3) we get that $X_R^*(p_c^*) = Q_g^*(p_c*)$, so using (8) and (6) we get the final expression for the certificate price in equilibrium.

$$p_c^* = \frac{aFH(2\alpha - 1) - c_b F(1 + bH\alpha) + c_g F[1 + bH(1 - \alpha)] + DFH(2 + bH)}{bFH(1 - \alpha) + H(2 + bH) + F} \tag{9}$$

Equation (9) shows that a decrease in the costs of renewable electricity decreases the certificate price. Therefore, any efficiency improvement in the production of green electricity would have the effect of decreasing the price of the certificates, even if the regulator were not aware of the efficiency gain. This is an advantage of TGC versus FIT.

3 Quantitative Analysis of the Effect of TGC in Spain

We simulate the implementation of a TGC incentive scheme in this section. In Spain, renewable producers are under a pure FIT or a FIP scheme, which is a fixed premium on top of the market price (RD 661/2007 2007). Under the FIT system, renewable generators sell their electricity under a guaranteed fixed tariff, whereas under the FIP option they take part in the daily market and get the price of the pool plus a guaranteed premium. In case of FIP, a cap and floor system has been introduced in order to protect renewable generators when the market price is too low and prevent excessive gains when the price is high enough. The floor is the lowest level of premium plus the electricity price, whereas the cap is the maximum electricity price to where a premium is still paid (Interactive EurObserv'ER Database 2012). The FIT approach isolates renewable generators from market prices and risks and consumers carry the price risk. In contrast, the FIP option let RES-E producers face some market risk. Generators are then exposed to market price signals and the premium is adjusted to keep both generators' risks and revenues within a particular range (Klessmann et al. 2008).

As Table 2 shows, only a minor part of the solar electricity is sold under the FIP scheme in Spain, whereas the majority of wind uses this promotion option rather than FIT. The high financial support level given to solar generators under the FIT scheme induced solar producers to choose the FIT system rather than FIP. However, due to the great burden of the current tariffs in the deficit of regulated activities,[5] FIT-FIP to new capacity have been removed, affecting all technologies. It is commonly argued that solar energy may not be competitive enough to take part successfully in a TGC system, but the market participation shares also show that there are other technologies that may be able to compete in a TGC market, such as wind power.

We simulate a switch to a TGC system in this section. The correct definition of the regulation parameters could lead to an efficient TGC system, achieving a certificate price that could send the correct signals for investment in renewable energy sources.

We calibrate our model according to the data of the year 2010 (see Table 3). For the sake of simplicity, as final price we take the sum of the price of the pool (p_e) and the premium (p_{FIP}), even though there are other components in the final price of the Spanish electricity system.[6] The price p_{FIP} is computed as the equivalent premium divided by the total renewable electricity sold in the pool and under tariff (provided by the CNE).

We compute the costs of renewable and non-renewable electricity by applying the equilibrium condition marginal revenue equals marginal cost. The relationship

---

[5] See Espinosa (2013a, b) and (Espinosa and Pizarro-Irizar 2012).

[6] Electricity costs include the daily market (pool, bilateral contracts and intraday market) (OMIE 2007) and other costs such as restrictions, capacity payments, transport and distribution costs, diversification and security of supply and other access costs (Mejía 2010).

Table 2 Share of the energy sold under FIP in the pool over the total energy sold (FIT + FIP) by the Special Regime (%)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Cogeneration	4.90	37.56	42.02	54.06	77.44	60.58	62.52	66.82	15.23	17.55
Solar	0.00	0.00	0.00	0.00	93.46	0.48	0.59	1.03	9.72	0.00
Wind	0.00	0.00	2.51	63.80	37.98	93.77	92.66	92.99	90.52	77.37
Small hydraulic	0.00	0.00	4.36	24.46	55.62	53.52	67.10	65.74	50.74	49.51
Biomass	0.00	0.00	2.05	44.82	47.57	68.21	68.59	62.83	31.74	28.67
Waste	0.00	0.00	11.10	34.75	24.42	60.11	87.20	95.42	80.31	82.21
Waste treatment	0.00	0.00	0.00	15.38	24.42	14.75	0.00	0.00	0.00	0.00

Source Own elaboration using data from the National Energy Commission (CNE)

Table 3 Electricity market in Spain

$Q = Q_b + Q_g$	193,345
Q_b	99,243
Q_g	94,101
Q_g/Q	0.49
p_e	38.01
p_{FIP}	78.13
$p_e + p_{FIP}$	116.14

Actual data 2010. Energy in GWh and prices in €/MWh
Source Ciarreta and Espinosa (2012), premium computed from CNE (2011)

$p_e = c_b + HQ_b$ provides the intercept of the marginal cost function of black electricity is $c_b = 28.09$. From the expression $p_e + p_{FIP} = c_g + HQ_g$ we get the parameter for green electricity $c_g = 106.73$. The value for the parameter $H = 10^{-7}$ is chosen in order to scale the cost functions for green and black electricity.

We assume an inelastic demand for electricity for the year 2010, taking $a = 193,345$ and $b = 0$. We fix $D = -0.01$ and $F = 50$ and see how the certificate price would have been for different values of the obligation and cost of green electricity. The values for the parameters have been selected so that the certificate price is high enough to promote investment.[7]

If the regulator knows the cost function for each generator, the TGC system is equivalent to the FIT-FIP in the sense that F can be set at a value that replicates the outcome of the FIT-FIP system (see columns (1) and (2) in Table 4). The advantage is that the TGC market reacts to efficiency gains in the production of green energy even if the regulator does not observe these gains or react to them (see column (3) in Table 4). If the decrease in c_g, from 106.73 to 82,[8] is unobserved to the regulator,

[7] However, changes in the value of the parameter D do not affect the price substantially.

[8] According to Sallé et al. (2012), average costs of wind power are 82 €/MWh.

Table 4 Simulations of the TGC market in Spain

		$\alpha = 0.49$		$\alpha = 0.60$	
	FIP	TGC	TGC efficiency gain	TGC	TGC efficiency gain
Parameters					
α	–	0.49	0.49	0.60	0.60
c_b	28.09	28.09	28.09	28.09	28.09
c_g	106.73	106.73	82	106.73	82
H	–	10^{-7}	10^{-7}	10^{-7}	10^{-7}
F	–	50	50	50	50
D	–	–0.01	−0.01	−0.01	−0.01
a	–	193,345	193,345	193,345	193,345
b	–	0	0	0	0
Energy traded (GWh)					
Q_b	99,243	99,243	99,243	77,338	77,338
Q_g	94,101	94,101	94,101	116,007	116,007
Prices (€/(MWh)					
p_e	38.01	38.01	38.01	35.82	35.82
p_{FIP}	78.13	–	–	–	–
p_c	–	78.13	53.40	82.51	57.78
$p_e + p_{FIP}$	116.14	–	–	–	–
$p_e + p_c$		116.14	91.41	118.33	93.60
$p_e + p_c - p_{FIP}$	–	0	−24.73	2.19	−22.54
Preferred technology	–	–	TGC	FIP	TGC

Source Own elaboration with market data for 2010

the FIT-FIP would not reflect this efficiency gain in the cost of promoting renewable energy. The TGC market however would translate the lower cost into a lower price of the certificate and the cost to the system of producing a green MWh (from 116.14 to 91.41 €/MWh).

Additionally, if the regulator increases the obligation for renewable electricity, say from 0.49 to 0.60, the TGC market would determine the price of certificates necessary to implement the new share of renewables (see column (4) in Table 4). With such an ambitious renewable target, considering actual costs for green production and the values chosen for the penalty function, the FIT-FIP system would be more cost-effective (the certificate price would be 82.51 €/MWh, higher than the price for the FIT-FIP system). This proves the importance of the correct setting of the regulation parameters and the renewable target. However, as the renewable cost falls, the TGC system once again shows this efficiency gain and the certifcate price drops without changing the penalty function (see column (5) in Table 4).

Finally, our analysis assumes that agents are price takers and behave competitively. In particular, the supply of certificates must be competitive. For this reason, the proposal of a TGC market would not be appropriate if the number of producers of a given technology is not large enough. Thus, a FIT-FIP could be more adequate for the initial stages.

4 Conclusions

We have analyzed the interaction between the TGC market and the electricity market when both markets work under perfect competition, even though a high concentration in generation and a low demand elasticity may indicate the presence of market power (Green and Newbery 1992; Cardell et al. 1997; Fabra and Toro 2005). The analysis of these markets with price-maker agents is left for future research.

We have modelled a decreasing demand for TGC and we have shown, both theoretically and empirically, that a decrease in the costs of renewable electricity may decrease the certificate price. Therefore, any efficiency improvement in the production of green electricity would have the immediate effect of decreasing the price of the certificates, even if the regulator were not aware of the efficiency gain. This transmission of market signals makes the TGC more efficient when compared to the FIT system.

However, the certificate price may be too low for the non-competitive technologies. In order to avoid this problem, we propose the combination of TGC and FIT, even if there is a wide variety of promotion schemes. The combination of a certificate system and a feed-in scheme could be used for a more efficient promotion of RES-E. It is known that the FIT scheme helps to maintain investor confidence, so technologies not being competitive under a TGC system (e.g. solar energy), because of their high cost and their need of R&D investments, could adopt a FIT regulation; whereas competitive renewable technologies (e.g. wind power) would work under a purely TGC setup. With this approach, the less mature technologies will be protected and gradually integrated into the certificate market. Moreover, the TGC scheme is more cost-effective, because the certificate price would be set by a low-cost technology rather than a high-cost one, avoiding the windfall profits for low-cost technologies that could offset the potential efficiency gains of a TGC system when the price is set by a high-cost technology. Thus, instead of separating FIT and TGC, an interesting possibility for solar energy could be the implementation of a complementary FIT added to the TGC price. This regulation could help to finance solar technologies, since they would receive a higher price than the certificate (FIT + TGC), and at the same time it would help to reduce the actual burden, because one component of the incentive would be market-based (TGC) and would adjust easily to the efficiency gains of the technologies. This system that combines both renewable promotion schemes tries to take advantage of both TGC and FIT.

References

Aune FR, Dalen HM, Hagem C (2012) Implementing the EU renewable target through green certificate markets. Energy Econ 34:992–1000

Bergek A, Jacobsson S (2010) Are tradable green certificates a cost-efficient policy driving technical change or a rent-generating machine? Lessons from Sweden 2003–2008. Energy Policy 38(3):1255–1271

Buttler L, Neuhoff K (2008) Comparison of feed-in tariff, quota and auction mechanisms to support wind power development. Renew Energy 33:1854–1867

Bye T (2003) On the price and volume effects from green certificates in the energy market. Discussion papers 351. Statistics Norway, Oslo

Cardell JB, Hitt CC, Hogan WW (1997) Market power and strategic interaction in electricity networks. Resour Energy Econ 19:109–137

Christiansen AC (2001) Technological change and the role of public policy: an analytical framework for dynamic efficiency assessments. Report of the Fridtjof Nansen Institute 74 pp

Ciarreta A, Espinosa MP (2010a) Market power in the Spanish electricity auction. J Regul Econ 37 (1):42–69

Ciarreta A, Espinosa MP (2010b) Supply function competition in the Spanish wholesale electricity market. Energy J 31(4):137–157

Ciarreta A, Georgantzis N, Gutiérrez-Hita C (2011) On the promotion of renewable sources in spot electricity markets. The role of feed-in-tariff systems. In: 4th annual conference on competition and regulation in network industries, Brussels

Ciarreta A, Espinosa MP, Pizarro-Irizar C (2012a) The effect of renewable energy in the Spanish electricity market. Lect Notes Inform Technol 9:431–436

Ciarreta A, Espinosa MP, Pizarro-Irizar C (2012b) Efecto de la energía renovable en el mercado diario de electricidad. Escenario 2020. Cuadernos Económicos de ICE 83:101–116

Ciarreta A, Espinosa MP (2012) The impact of regulation on pricing behavior in the Spanish electricity market (2002–2005). Energy Econ 34(6):2039–2045

CNE (2011) Información Estadística sobre las Ventas de Energía del Régimen Especial

European Wind Energy Association. Wind Energy (2004) The Facts, an analysis of wind energy in the EU-25, funded by EU Commission

Espinosa MP, Pizarro-Irizar C (2012) Políticas para la reducción del déficit tarifario. Papeles de Economía Espaola 134:117–126

Espinosa MP (2013a) Understanding the electricity tariff deficit and its challenges. Span Econ Fin Outlook 2(2):47–55

Espinosa MP (2013b) An austerity-driven energy reform. Span Econ Fin Outlook 2(5):51–61

European Commission (2011) Communication from the commission to the European parliament and the council. Renewable energy: progressing towards the 2020 target. COM(2011) 31 final

Fabra N, Toro J (2005) Price wars and collusion in the Spanish electricity market. Int J Ind Organ 23:155–181

Gan L, Eskeland GS, Kolshus HH (2005) Green electricity market development: lessons from Europe and the US. Energy Policy 35:144–155

Green RJ, Newbery DM (1992) Competition in the British electricity spot market. J Polit Econ 100 (5):929–953

Green R, Vasilakos N (2011) The economics of offshore wind. Energy Policy 39(2):496–502

Haas R, Panzer C, Resch G, Ragwitz M, Reece G, Held A (2011) A historical review of promotion strategies for electricity from renewable energy sources in EU countries. Renew Sustain Energy Rev 15(2):1003–1034

Interactive EurObserv'ER Database (2012) http://www.eurobserv-er.org

Jacobsson S, Lauber V (2006) The politics and policy of energy system transformation-explaining the German diffusion of renewable energy technology. Energy Policy 34(3):256–276

Jensen JC (2003) Policy support for renewable energy in the European Union. Energy Research Centre of the Netherlands, ECN Publication, ECN-C-03-113

Jensen S, Skytte K (2002) Interaction between the power and green certificate markets. Energy Policy 30:425–435
Klessmann C, Nabe C, Burges K (2008) Pros and cons of exposing renewables to electricity market risks—a comparison of the market integration approaches in Germany, Spain, and the UK. Energy Policy 36:3646–3661
Mejía P (2010) Los retos del mercado. I Foro permanente de la energía. Desafíos energéticos del futuro
Menanteau P, Finon D, Lamy M-L (2003) Prices versus quantities: choosing policies for promoting the development of renewable energy. Energy Policy 31(8):799–812
Meyer NI (2003) European schemes for promoting renewables in liberalised markets. Energy Policy 31(7):665–676
Midttun A, Gautesen K (2007) Feed in or certificates, competition or complementarity? Combining a static efficiency and a dynamic innovation perspective on the greening of the energy industry. Energy Policy 35:1419–1422
Nilsson M, Sundqvist T (2007) Using the market at a cost: How the introduction of green certificates in Sweden led to market inefficiencies. Util Policy 15:49–59
Neuhoff K (2005) Large-scale deployment of renewables for electricity generation. Oxf Rev Econ Policy 21(1):88–110
OMIE (2007) Daily and intraday electricity market activity rules, version 01/07/2007
Rader N, Norgaard R (1996) Efficiency and sustainability in restructured electricity markets: the renewables portfolio standard. Electr J 9(6):37–49
Rathman M (2007) Do support systems for RES-E reduce EU-ETS-driven electricity prices? Energy Policy 35(1):342–349
Real Decreto 661/2007, May 25 de mayo. Regulation of renewables. Boletín Oficial del Estado 126, 22846-22886
Real Decreto-ley (2012). Elimination of incentives to new renewable capacity. Boletín Oficial del Estado 24:8068–8072
Ringel M (2006) Fostering the use of renewable energies in the European Union: the race between feed-in tariffs and green certificates. Renew Energy 31:1–17
Sallé C, López JM, Muñoz M, Martín I, Sáez de Miera G, Laverón F (2012) El déficit de tarifa y la importancia de la ortodoxia en la regulación del sector eléctrico. Cuadernos de Energía 35:1–32
Schaeffer GJ, Boots MG, Martens JW, Voogt MH (1999) Tradable green certificates—a new market-based incentive scheme for renewable energy: introduction and analysis. Energy Research Centre of the Netherlands. ECN-I-99–004
Schmalensee R (2011) Evaluating policies to increase the generation of electricity from renewable energy. MIT Center for Energy and Environmental Policy Research. CEEPR WP 2011-008
Sims REH, Rogner H-H, Gregory K (2003) Carbon emission and mitigation cost comparisons between fossil fuel, nuclear and renewable energy resources for electricity generation. Energy Policy 31(13):1315–1326
Verbruggen A (2004) Tradable green certificates in Flanders (Belgium). Energy Policy 32(2): 165–176

Unobserved Heterogeneous Effects in the Cost Efficiency Analysis of Electricity Distribution Systems

Per J. Agrell, Mehdi Farsi, Massimo Filippini and Martin Koller

Abstract The purpose of this study is to analyze the potential effects of unobserved heterogeneity on the cost efficiency measurement of electricity distribution systems within the framework of incentive regulation schemes such as price- or revenue cap. In particular, we decompose the benchmarking process into two steps: In the first step, we attempt to identify classes of distribution system operators functioning in similar environments and with comparable network and structural characteristics. For this purpose, we apply a latent class model. In the second step, best practice is obtained within each class, based on deterministic and stochastic frontier models. The results show that the decomposition of the benchmarking process into two steps and the consideration of technology classes can reduce the unobserved heterogeneity within classes, hence, reducing the unexplained variation that could be mis-specified as inefficiency.

Keywords CPI-X regulation · Efficiency measurement · Unobserved heterogeneity · Latent class model

The views expressed in this paper are those of the authors and do not necessarily reflect the opinion of the institutions with which they are affiliated. The responsibility for all errors lies solely with the authors.

P.J. Agrell
Louvain School of Management, Université Catholique de Louvain,
Voie Du Roman Pays 34, 1348 Louvain-La-Neuve, Belgium

M. Farsi
Faculty of Economics and Business, University of Neuchâtel,
Pierre-à-Mazel 7, 2000 Neuchâtel, Switzerland

M. Filippini (✉) · M. Koller
Department of Management, Technology and Economics, ETH Zurich,
Zürichbergstrasse 18, 8032 Zürich, Switzerland
e-mail: mfilippini@ethz.ch

M. Filippini
Department of Economics, University of Lugano, Via Giuseppe Buffi 13,
6900 Lugano, Switzerland

S. Ramos and H. Veiga (eds.), *The Interrelationship Between Financial and Energy Markets*, Lecture Notes in Energy 54, DOI 10.1007/978-3-642-55382-0_12

1 Introduction

In the last two decades the electricity distribution sector in Europe has witnessed a wave of regulatory reforms aimed mainly at improving the economic efficiency. Thereby, information on several efficiency concepts in production theory, including scale and scope efficiency as well as cost efficiency has become very important. The concept of cost efficiency is a measure of the regulated electricity distribution company's ability to minimize costs, given specific demand and market conditions. Cost inefficiency, also called 'X-inefficiency', occurs when the company fails to produce with full efficiency at the cost frontier, defined by best-practice companies.

Regulatory authorities increasingly use empirical cost norms, such as parametric or non-parametric benchmarking methods, in various incentive regulation schemes (Haney and Pollitt 2009). One of the most widely used regulatory regimes in electricity networks is price- or revenue-cap regulation (often denoted CPI-X regulation, cf. Littlechild (1983)). This method determines a maximum price or revenue index in real terms, less a productivity improvement parameter, referred to as the 'X-factor'.[1] The X-factors include a general productivity improvement requirement (usually called the 'general X-factor') and potentially an individual efficiency improvement parameter (frequently denoted the 'Xi-factor' or the individual X-factor). Whereas the purpose of the general X-factor is to share the productivity gains in the sector between the consumers and the companies, the individual term is intended to eliminate incumbent efficiency differences between companies. The exact translation of an estimated static cost inefficiency to an annual real productivity target (Xi) depends on the allowed period to catch up inefficiency, the type of inefficiency detected (capital and/or operating costs) and the type of by-pass mechanism (Z) used for certain costs that may be proportional to the inefficiency (e.g. network losses). Notwithstanding, the mechanism allows the regulator to set differentiated price or revenue caps based on the individual company's empirically estimated productive efficiency performance.[2] An alternative to the CPI-X regulation, addressing the arbitrariness of the adjustment parameters and the risk induced by the lag, is the yardstick regulation paradigm (cf. Shleifer 1985). In this model, the reimbursement of the regulated firm is linked to a dynamic norm, excluding the cost report of the specific company in its calculation. Although Shleifer presented the model for a stylized cost function, the use of frontier analysis tools enables the application of yardstick methods also to multi-output production

---

[1] In addition to inflation, the changes beyond companies' control may include changes in input factor prices and exogenous changes in demand and network characteristics, generally referred to as 'Z-factors'.

[2] The level of productive efficiency or cost efficiency of a firm is composed by the levels of technical and allocative efficiency. For a discussion of these concepts see Kumbhakar and Lovell (2003).

and service provision. Several regulators in Europe, thereof Germany and Norway, use DEA for dynamic yardstick regimes in electricity distribution regulation.[3]

However, the increasing use of efficiency analysis has raised serious concerns among regulators and companies regarding the reliability of efficiency estimates.[4] In fact, empirical evidence suggests that the estimates are sensitive to the adopted efficiency measurement approach.[5] This implies that the choice of the approach may have important effects on the financial situation of the companies as well as on the industrial structure of the regulated sector.

One important dimension affecting the reliability of efficiency estimates is the presence of *unobserved factors.* The regulated companies operate in different regions with various environmental and network characteristics that are only partially observed. This heterogeneity in the service area is an important factor to consider in a benchmarking analysis. Recall that the purpose of the benchmarking method is to create a cost norm for efficient, structurally comparable companies under similar operating conditions. Some methods of estimating efficiency take account of such unobserved factors, but in different ways. Generally, in deterministic models such as the non-parametric linear programming approach, the unobserved factors that influence the level of production costs are not considered in the analysis. The explicit assumption in these approaches is that all relevant cost differences are captured by observed variables. The few efficiency analysis models addressing part of the unobserved heterogeneity factors are parametric and based on panel data. The seminal paper for the development of models for unobserved heterogeneous factors is Greene (2005). The main idea is to introduce an individual effect in an econometric model capturing the unobserved heterogeneous factors that remain constant over time. The main problem hereby is that the individual effects can capture also part of the inefficiency that remains constant over time. In addition, the complexity of the models developed by Greene (2005) and the entailed assumptions remain important obstacles in applying panel data models in regulatory practice. Given that the unobserved factors are considered differently in various models, the resulting estimates can vary across methods. The magnitude of variation depends on the importance of the unobserved factors, which might change from one case to another.

To address this problem, we propose an alternative strategy for improve efficiency measurement methodology in the presence of unobserved heterogeneity. In our strategy, we decompose the benchmarking process into two steps: In the first step, we attempt to identify classes of companies that operate in similar environments and with comparable network and structural characteristics. For this purpose, we apply a latent class model. In the second step, the best practice is obtained

[3] The theory for dynamic applications of DEA in yardstick and a comparison with a conventional CPI-X approach are found in Agrell et al. (2005a).

[4] Shuttleworth (2005) provides a critical overview of the problems coming along with the use of benchmarking in the regulation of electricity networks.

[5] See e.g. Jamasb and Pollitt (2003), Estache et al. (2004), Farsi and Filippini (2004) or Farsi et al. (2006).

within each class, based on deterministic and stochastic frontier models. Provided that the identified classes contain reasonably comparable cases and assuming a reasonable explanatory power for the variables included in the model specification, any deterministic or stochastic approach can be used to estimate efficiency.

The outline of this chapter is as follows: Sect. 1 reviews some of the most commonly used approaches to efficiency measurement. Section 2 addresses the cost model specifications and estimation methods. Section 3 introduces the data and Sect. 4 provides the estimation results for both steps and measures of cost efficiency for different frontier models in the second step. We draw our conclusions in Sect. 5.

2 Review on Approaches to Efficiency Measurement

This section briefly reviews some of the most commonly used frontier approaches to cost efficiency measurement, based on more extensive reviews in Kumbhakar and Lovell (2003), Murillo-Zamorano (2004), Coelli et al. (2005), Cornwell and Smith (2008), Greene and William (2008), Kumbhakar and Lovell (2003), and Farsi and Filippini (2009).[6] The focus here is mainly on cost efficiency and on cost functions, the argumentation is analogously valid for production functions and productive efficiency (under a set of regularity conditions, cf. Shepard (1953) and Nerlove (1963)). The frontier approach assumes that full cost efficiency is defined by those companies that are identified as the best-practice peers. All other companies are assumed to operate above the cost frontier, hence to have non-zero inefficiency.

Economic literature has developed two different frontier paradigms to empirically measure cost efficiency.[7] The first is based on a non-parametric deterministic and the second on an econometric approach, sometimes also referred to the parametric approach.

Non-parametric approaches, such as the Data Envelopment Analysis (DEA), proposed by Farrell (1957) and Charnes et al. (1978), use linear programming to construct a company's efficiency frontier, which is considered as a deterministic function of the observed variables. These methods are non-parametric in the sense that they do not impose any specific functional form or distribution assumption, i.e. it is assumed that the data are free of noise. Thanks to their relative simplicity and availability, such methods, in particular DEA, are quite popular among both researchers and regulators in energy distribution networks. The DEA models can be input- or output-oriented and one of the a priori assumptions concerns the returns to scale. The models can be specified as constant returns to scale (CRS), variable returns

[6] The latter review includes also sections on the traditional production theory and on scale and scope economies.

[7] A third paradigm, the Bayesian approach is only little-known in applied science. Readers interested in Bayesian stochastic frontier models (sometimes also assigned to non-parametric models) are referred to van den Broeck et al. (1994).

to scale (VRS), non-increasing returns to scale (NIRS), non-decreasing returns to scale (NDRS), free disposal hull (FDH) and free replicability hull (FRH), where the latter two merely impose disposability and additivity, but not convexity of the production space. A basic DEA formulation calculating the minimal cost under VRS for company i in a sample of N companies with k inputs and m outputs would be

$$\begin{aligned} &\min_{\lambda, x_i} w_i' x_i \\ &s.\,t.:\ -y_i + Y\lambda \geq 0;\ x_i - X\lambda \geq 0;\ N'\lambda = 1;\ \lambda \geq 0 \end{aligned} \tag{1}$$

where w_i and x_i are $k \times 1$ vectors representing input prices and quantities for company i; y_i is an $m \times 1$ vector representing the given output bundle; X and Y are input and output matrices namely, a $k \times N$ and an $m \times N$ matrix consisting of the input and output bundles for all companies in the sample; N is an $N \times 1$ vector of ones; and λ is an $N \times 1$ vector of non-negative constants to be estimated. The VRS property is satisfied through the convexity constraint ($N'\lambda = 1$) that ensures that only similar-sized companies are benchmarked against each other. The linear programming algorithm finds a piece-wise linear isoquant in the input-space, which corresponds to the minimum costs of producing the given output at any given point. Cost efficiency (CE) finally is measured by the minimum feasible input bundle for each company relative to its actual input bundle, i.e. $CE_i = w_i' x_i^* / w_i' x_i^0$.

In contrast to non-parametric methods, most of the econometric approaches include estimating an empirical cost function, where the observed variables should include a vector of outputs (**q**) and a vector of input prices (**p**). The remaining unobserved part, the residual, is completely (in deterministic models) or partially (in stochastic models) assigned to inefficiency.

The first econometric frontier models that appeared in the literature were deterministic and estimated by OLS. Usually, their cost function is expressed in logarithms as

$$\ln C_{it} = f(\boldsymbol{q}_{it}, \boldsymbol{p}_{it}; \boldsymbol{\beta}) + \alpha + \varepsilon_{it} \tag{2}$$

where C_{it} is total cost incurred by the unit i at time t, $f(.)$ is a parametric cost function, $\boldsymbol{q}_{it}$ and $\boldsymbol{p}_{it}$ are vectors of outputs and input prices, respectively, $\boldsymbol{\beta}$ is the vector of parameters and α the intercept to estimate, and ε_{it} is the residual. As the error term in deterministic models only reflects the inefficiency, it is assumed to be non-negative. Therefore, Winsten (1957) suggested shifting the estimated intercept down by the minimal residual value. This model is called Corrected OLS (COLS). The cost efficiency of unit i in the COLS model is thus given by $\exp(-u_{it})$ with $u_{it} = \varepsilon_{it} - \min(\varepsilon_{it}) \geq 0$. Afriat (1972) proposed a slightly different model, usually referred to as Modified OLS (MOLS), where the OLS intercept is shifted by the expected value of the inefficiency term that is, $E(u_{it})$. The cost efficiency of unit i at time t in the MOLS model is thus given by $\exp(-u_{it})$ with $u_{it} = \varepsilon_{it} + E(u_{it})$. The efficiency term u_{it} is not necessarily positive (some units are below the cost frontier). Truncation at zero assigns the respective units with full efficiency.

Deterministic models are similar to DEA and other linear programming models in that the best practice (the cost frontier) is a fixed function that does not vary across observations or units. As main drawback, these models attribute the residual entirely to inefficiency, i.e. they do not account for other sources of stochastic variation such as measurement errors.[8,9] Nevertheless, deterministic models are still widely used in applied economic literature and in regulation (see e.g. Haney and Pollitt (2009)).

To overcome the drawbacks of deterministic models, Aigner et al. (1977) and Meeusen and van den Broeck (1977) proposed a stochastic frontier model (SFA), which divides the residual ε_{it} into two parts: u_{it} is reflecting inefficiency, and v_{it} is capturing the random noise. The basic cost function of the stochastic frontier model can be written as

$$\ln C_{it} = f(\boldsymbol{q}_{it}, \boldsymbol{p}_{it}; \boldsymbol{\beta}) + \alpha + u_{it} + v_{it} \tag{3}$$

With certain distribution assumptions on u_{it} and v_{it}, this model can be estimated using the Maximum Likelihood (ML) estimation method. Typically, it is assumed that the inefficiency term u_{it} has a one-sided non-negative distribution that is, a normal distribution truncated at zero: $u_{it} \sim |N(0,\sigma_u^2)|$,[10] and the random noise term v_{it} is normally distributed: $v_{it} \sim N(0,\sigma_v^2)$. Additionally, u_{it} and v_{it} are considered as being independently distributed from each other. As in the models above, one would expect the most efficient unit to take $u_{it} = 0$, and the efficiency value to be calculated as $\exp(-u_{it})$. Unfortunately, $E(u_{it})$ cannot be calculated for an individual unit. Jondrow et al. (1982) proposed therefore a different estimator to measure efficiency. This estimator is based on the conditional expectation function of the residual, $(E[u_{it}|\varepsilon_{it}])$, and is known as the JLMS estimator referring to the authors.[11] This is a highly non-linear function that only slightly increases the inefficiency for units close to the frontier leaving no unit with full efficiency. The other estimator proposed by these authors is based on the conditional mode $(M[u_{it}|\varepsilon_{it}])$ that normally assigns full efficiency to several units. It has been used much less in the empirical literature than the JLMS estimator.

---

8 Semi-parametric frontier models such as quantile regression (Koenker and Bassett 1978) sometimes count as deterministic models. Unlike least squares methods, quantile regression techniques do not approximate the conditional mean of the response variable, but either its median or quantiles and offer therefore a systematic strategy for examining the entire distribution of the population. Readers interested in applied quantile regression models for efficiency measurement are referred to Behr (2010) and to Knox et al. (2007).

9 In real regulatory application, regulators use specific outlier detection and elimination methods to reduce the impact of, and incentives for, errors in the reference set, see Agrell and Niknazar (2014).

10 Other extensions of the SFA model have considered exponential, gamma, or truncated normal distributions for the inefficiency term.

11 Jondrow et al. in (1982).

The models described so far can be applied either to cross-sectional or panel data. However, the panel structure in the data is ignored, as these models require pooling all observations and treating them as being independent from each other. Temporal variations can be captured using time trends or time-interactions. Moreover, these models are not suited to account for unmeasured, i.e. unobserved heterogeneity. This is due to the fact that with pooled data, each observation is considered as a single, discrete unit. With only one observation per unit, it is not possible to disentangle efficiency and time-invariant, unit-specific heterogeneity. Therefore, the presence of unobserved heterogeneity influences the estimation results of the regressors in case of correlation, or the residuals (referred to heterogeneity bias, Chamberlain (1982)). The structure of panel data offers the opportunity to apply models that account for the individual effect that should capture the unobserved heterogeneity and hence free from the heterogeneity bias. The time dimension in panel data sets allows us to observe the same unit repeatedly over a certain time span. This enables us to extract time-invariant factors such as unit-specific characteristics that do not necessarily accrue to the unit's inefficiency, but do affect the costs across different networks. Especially structural inefficiencies (inefficiency that is constant over time) and inefficiencies following a certain time path can be better identified using panel data. Most of the developments of the panel data models go back to the stochastic frontier models of Aigner et al. (1977) and Meeusen and van den Broeck (1977) expressed in Eq. (3).

An early application to panel data of this stochastic frontier model was the Random Effects (RE) model by Pitt and Lee (1981) which was estimated by ML and assumed that the inefficiency u_{it} is fixed through time, but still half-normally distributed: $u_i \sim |N(0,\sigma_u^2)|$. Important variations of this model were presented by Schmidt and Sickles (1984) who relaxed the distribution assumption, and by Battese and Coelli (1988) who assumed a truncated normal distribution. Schmidt and Sickles (1984) also proposed a Fixed Effects (FE) model to avoid the possible heterogeneity bias in case of correlation of u_{it} with the explanatory variables. One of the drawbacks of models with time-invariant efficiency is that time-varying components of heterogeneity are entirely interpreted as random noise. Therefore, Cornwell et al. (1990), Kumbhakar (1990) and Battese and Coelli (1992) suggested the first stochastic models allowing the cost efficiency to vary over time. However, the first two models developed were vulnerable to multicollinearity and the third was characterized by a deterministic functional form of the inefficiency term over time.

The main restriction of all of the models presented above is that unobserved factors are assumed to be random over time. This implies that time-invariant factors such as physical network and environmental characteristics are not considered as heterogeneity. The family of 'true' panel data models (Kumbhakar 1991) and Polachek and Yoon (1996) as precursor models of Greene (2005) extend the original stochastic frontier model as it is formulated in Eq. (3) by adding a unit-specific time-invariant factor accounting for the individual effect.[12] Hence, apart from the

[12] The term 'true' refers to the FE and RE models fully described in Greene (2005).

random noise component, these models include two stochastic terms for unobserved heterogeneity, one for time-varying and one for time-invariant individual effects. This model can be written as

$$\ln C_{it} = f(\boldsymbol{q}_{it}, \boldsymbol{p}_{it}; \boldsymbol{\beta}) + \alpha_i + u_{it} + v_{it} \tag{4}$$

where α_i is the time-invariant unit-specific factor and the model is estimated by Maximum Simulated Likelihood (MSL). In a RE framework, α_i is an *iid* random component and must not be correlated with the observed variables. In a FE framework, α_i is a constant parameter for every unit.[13] As in all ML models, the inefficiency component can be measured by the JLMS estimator of Jondrow et al. (1982). Assuming that physical network and environmental characteristics do not vary considerably over time and that the inefficiency is time-varying, these models help to separate unobserved time-invariant effects from efficiency estimates. However, if inefficiency is persistent over time, these models underestimate the inefficiency systematically, e.g. if managers take wrong decisions in every period or make the same mistakes again and again, the corresponding consequences in terms of inefficiency are detected as time-invariant unit-specific heterogeneity and not as inefficiency. As noted in Greene (2008), the 'truth' doubtless lies somewhere between the two strong assumptions.

The idea of observed parameter variability was early applied to a precise indication of heterogeneity of the production environment by Kalirajan and Obowona (1994) in the stochastic frontier context. A similar random parameter (RP) model was also formulated by Greene (2005), which is a generalization of the True Effects models in that not only the constant but also the parameters of the observed variables are unit-specific indicating the effect of different environments or technologies. This model is estimated by MSL. As noted by Greene (2008), the estimation of the MSL of this model can be numerically cumbersome.

Another approach to accommodate heterogeneity among units into the model is followed by latent class (LC) models. Originally introduced by Lazarsfeld and Henry (1968), LC identifies distinct class membership among subjects regarding their cost structure and estimates a separate cost function for each of these classes simultaneously.[14] LC models can be regarded as the discrete counterparts of RP models. With a sufficient large number of classes, LC approximates a fully parameterized RP model. The LC model can be written as:

$$\ln C_{it} = f\left(\boldsymbol{q}_{it}, \boldsymbol{p}_{it}; \boldsymbol{\beta}_j\right) + \alpha_j + u_{it}\big|_j + v_{it}\big|_j \tag{5}$$

[13] An alternative version of the True FE model uses dummy variables for every unit. However, this specification may be affected by the 'incidental parameter problem', especially in short panel data sets.

[14] Latent class analysis has been applied in different fields of science and industry sectors, e.g. in the banking sector (Orea and Khumbhakar 2004) or more recently in the electricity distribution (Cullmann 2010).

The subscript i denotes the unit, and u_{it} and v_{it} are defined as above. α_j is the constant and $\boldsymbol{\beta}_j$ is a vector of discrete random parameters identified in $j = 1, 2, \ldots, J$ classes, assuming that each observation in the sample follows a specific technology. These technologies differ from each other in the values of model parameters $\{\alpha_i, \boldsymbol{\beta}_i, \sigma_i\}$. This vector includes also a set of prior probabilities that determines the fraction of each latent class in the sample. It is defined as a discrete random vector with the following distribution:

$$\{\alpha_i, \boldsymbol{\beta}_i, \sigma_i\} = \left\{\alpha_j, \boldsymbol{\beta}_j, \sigma_j\right\} \text{ with probability } P_j, \text{ where} : j = 1, 2, \ldots, J, \text{ and} \sum_{j=1}^{J} P_j = 1 \quad (6)$$

The subscript j denotes the latent class with J being the number of classes. The choice of J is usually based on diagnostic criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC).[15] These criteria indicate the optimal number of classes from an informational perspective, but cannot be used for statistical inference. After the estimation of the LC model, posterior probabilities $\widehat{P}_j$ can be calculated for each observation from Bayes rule.

The choice of the econometric models presented so far is usually not straightforward. For instance, Farsi and Filippini (2009) have found several studies that report discrepancies in efficiency estimates between different models and approaches.[16] Such discrepancies are partly due to methodological sensitivity in the estimation of individual efficiency scores and partly due to different consideration of unobserved heterogeneity factors, which are particularly relevant in network industries such as electricity distribution. Panel data models can be used to control for the firm- or network specific unobserved heterogeneity. The use of panel data models is especially interesting as data for several years have become available to an increasing number of regulators in many countries. The complexity of such models remains however an important obstacle in applying panel data models in regulation. The effort in disentangling inefficiency variations from unobserved factors such as statistical noise due to error and omitted variables is a crucial element of all frontier models, in both cross sectional and panel data. The statistical modeling challenge has a parallel in practice: benchmarking can only be effective to the extent that for any specific company with given characteristics, there exists a set of comparable companies upon which a 'best practice' can be constructed.

Therefore, as previously discussed, we propose an alternative strategy in this paper to consider unobserved heterogeneity factors in that we decompose the

---

15 However, compared to the BIC, the AIC corrects the likelihood function only by the sample size and not by the number of parameters to estimate. This is a clear disadvantage with increasing number of classes.

16 See e.g. Jensen (2000), Jamasb and Pollitt (2003), Street (2003), Estache et al. (2004), Farsi and Filippini (2004). The results show substantial variations in estimated efficiency scores and, for some of them, in efficiency rankings across different approaches (econometric and non-parametric) and among model specifications.

benchmarking analysis into two steps: In the first step, we attempt to identify classes of companies that operate in similar environments and with comparable network and structural characteristics. For that step, we use a latent class model. In the second step, best practice is obtained within each class applying different benchmarking methods. Provided that the identified classes include reasonably comparable cases and assuming a reasonable explanatory power for the variables included in the model specification, any deterministic or stochastic approach can provide accurate values of efficiency. Therefore, we use the DEA, MOLS and SFA methods for the second step. In the next section, we will apply this approach using a sample of Norwegian electricity distribution companies.

3 Cost Model Specification and Estimation Methods

We specify a cost model that explains total costs of the Norwegian electricity distribution system operators (DSO) with two input and one output variable, one environmental factor and one network characteristics. We write this model as follows:

$$TC = f(P_L, P_C, Q, D, S) \tag{7}$$

where the dependent variable TC represents the total costs of the DSO. P_L and P_C are the input prices of labor and capital, respectively. Q is the delivered electricity, D the network density and S, finally, the share of high voltage network. For a complete description of the data and variables, see Sect. 4.

For the identification of the comparable technology classes in the first step, we apply a Latent Class (LC) approach (cf. Lazarsfeld and Henry (1968), see Sect. 2)[17] to estimate the cost model in Eq. (7).[18] Using a Cobb-Douglas functional form and imposing the linear homogeneity restriction, the LC model in Eq. (5) can be adapted to:

$$\ln\frac{TC_{it}}{P_{Cit}} = \alpha_{0j} + \beta_{Pj}\ln\frac{P_{Lit}}{P_{Cit}} + \beta_{Qj}\ln Q_{it} + \beta_{Dj}\ln D_{it} + \beta_{LSj}S_{it} + \varepsilon_{it|j} \tag{8}$$

where subscript i denotes the electricity distribution company $i = 1, 2, \ldots, I$, subscript t the years 1998–2002, and $\varepsilon_{it} \sim N(0,\sigma_i)$ the error term. The subscript j denotes the latent class with J being the number of classes.

After the identification of comparable technology classes, we estimate the cost efficiency in the second step separately for each class. As the heterogeneity within

[17] Different models could be considered to identify technology classes. LC is a statistical method that has been used in literature to identify classes (see Orea and Khumbhakar (2004) or Greene (2005)).

[18] All estimations have been conducted by Nlogit software version 4.0.

classes is expected to be low due to comparable technologies, any deterministic or stochastic approach can be considered. For general overviews on approaches to efficiency measurement, see e.g. Murillo-Zamorano (2004) or Greene (2008), or, for an empirical application, Farsi and Filippini (2009). With respect to current regulatory practice (see Haney and Pollitt (2009) for an overview over 40 countries), we apply the three following, most prevalent methods: The Data Envelopment Analysis (DEA, proposed by Farrell (1957) and Charnes et al. (1978)), the Modified OLS (MOLS, proposed by Afriat (1972)) and the Stochastic Frontier Analysis (SFA, proposed by Aigner et al. (1977)).

DEA is a non-parametric method to calculate cost efficiency as a deterministic function of the observed variables, i.e. it is assumed that the data are free of stochastic variation due to measurement errors or noise. The cost model given in Eq. (7) can be readily used for the efficiency measurement with the DEA method. Assuming variable returns to scale (VRS), the Eq. (1) reduces to the following minimization problem:

$$\begin{aligned} &\min_{\lambda} TC_{it} \\ &s.t.: \; -Y_{it} + Y\lambda \geq 0; \; TC_{it} - TC\lambda \geq 0; \; N'\lambda = 1; \; \lambda \geq 0 \end{aligned} \tag{10}$$

where Y_{it} represents the vector of the output bundle including output Q_{it} and output characteristics D_{it} and S_{it}, as both characteristics take resources. However, in the DEA model, D is defined as the inverse of the network density, since a higher network density implies lower costs. N and λ are vectors of ones and non-negative constants, respectively. Cost efficiency (CE) is measured as the minimum feasible costs for each company relative to its actual costs, i.e. $CE_{it} = TC^{*}/TC_{it}$.

MOLS and SFA are parametric methods that use regression techniques to construct the efficiency frontier. Both require the specification of a functional form of the cost function as well as assumptions about the error term(s). Similar to Eq. (8) in the first step, we estimate cost model in Eq. (7) using a Cobb-Douglas functional form and impose the linear homogeneity restriction. The MOLS and SFA models in Eqs. (2) and (3) can be adapted to:

$$\ln\frac{TC_{it}}{P_{Cit}} = \alpha_0 + \beta_P \ln\frac{P_{Lit}}{P_{Cit}} + \beta_Q \ln Q_{it} + \beta_D \ln D_{it} + \beta_{LS} S_{it} + \varepsilon_{it} \tag{11}$$

The MOLS approach is based on the OLS estimation. The residuals ε_{it} are corrected using a constant shift, which is the expected value of the inefficiency term, $E(u_{it})$. The cost efficiency in the MOLS is thus deterministic and given by $CE_{it} = \exp(-u_{it})$ with $u_{it} = \varepsilon_{it} + E(u_{it})$. u_{it} is not necessarily positive, as some units may lie below the cost frontier. Truncation at zero assigns the respective units with full efficiency.

The SFA approach is based on the Maximum Likelihood estimation. The residuals ε_{it} are composed of the inefficiency term u_{it} and the random noise term v_{it}.

In this study, it is assumed that u_{it} follows one-sided non-negative distribution, i.e. a normal distribution truncated at zero: $u_{it} \sim |N(0,\sigma_u^2)|$, and that v_{it} is normally distributed: $v_{it} \sim N(0,\sigma_v^2)$. Additionally, u_{it} and v_{it} are considered as being independently distributed from each other. The cost efficiency in the SFA is thus stochastic and given by $CE_{it} = \exp(-u_{it})$.

In order to compare the results from this two-step approach with that of a conventional analysis, we estimate the three models (DEA, MOLS, SFA) also in one step, i.e. without consideration of classes, but for the whole sample. The resulting tables are given in the Appendix.

4 Data

The data we use for this study consist of a balanced panel of 555 observations from 111 companies that have operated in the Norwegian power distribution sector from 1998 to 2002.[19] The available information includes total costs, labor costs, full time equivalents, total transformer capacity, distributed electricity, number of customers, line length for each low and high voltage, and year dummies. Table 1 provides a descriptive summary of the balanced panel data set for the variables included in the models.

From this data, we calculated the variables included in the models as follows: The dependent variable (*TC*) is the total network costs excluding the cost of purchased electricity. It is measured in millions Norwegian Kroner (NOK) and is in real terms; hence it is adjusted for inflation. *TC* includes all DSO's network costs consisting of both operating and capital expenditures. The explanatory variables involve two input price variables, one output variable and one environmental and one network characteristic, hence the DSO's are here considered to be single-product firms. The input price variables include a price for labor (P_L) and a price for capital (P_C). We derived P_L by dividing labor costs by the number of full-time equivalents. P_C is an approximation to the real capital price, calculated as a residual price by dividing non-labor costs by the installed transformer capacity. The output is given by the delivered electricity (Q), measured in gigawatt hours (GWh). The environmental variable is the network density (D), represented by total number of customers divided by total network length in kilometers. The network characteristic (S) is modeled by the share of high voltage network length and total network length.

[19] In order to get a balanced panel data set, we extracted this data from the data that has been used in several scientific studies (Agrell et al. (2005a, b)) as well as in a research project financed by the Norwegian Water Resources and Energy Directorate partly reported in Agrell and Bogetoft (2009) and a research project financed by Swiss Federal Institute of Energy reported in Filippini et al. (2011).

Table 1 Descriptive summary, 1998–2002

Variable description	Variable mean		Standard deviation	Min	p25	Median	p75	Max
Total cost (10^6 NOK)	TC	36.5	73.4	1.28	11.7	18.1	35.8	862
Labor cost (10^6 NOK)		10.2	16.8	0.47	3.70	5.58	10.2	161
Fulltime equivalents (FTE)		30.0	48.9	2.00	10.0	16.0	31.0	419
Price labor (10^3 NOK/FTE)	P_L	349	52.9	120	312	344	383	586
Transformer capacity (MVA)		243	698	4.08	38.8	75.7	201	7,944
Price capital (10^3 NOK/MVA)	P_C	176	83.6	31.6	112	160	220	528
Distributed electricicity (GWh)	Q	316	942	6.86	66.4	127	267	11,200
Number of customers		11,445	32,622	288	2,812	5,002	10,176	373,290
Line length (km)		1,351	1,871	56.5	467	762	1,609	13,583
Density (customers/km)	D	7.00	3.25	1.32	5.16	6.23	7.66	29.0
Line length, high voltage (km)		492	703	10.0	159	267	557	4,995
Line length, low voltage (km)		859	1,209	0.00	292	493	930	10,090
Share HV network	S	0.37	0.11	0.09	0.30	0.36	0.43	1.00
Year dummies	dyear	0.20	0.40	0.00	0.00	0.00	0.00	1.00

T = 5 (panel of years 1998–2002), i = 111, N = 555

5 Results

In the first step, we first determine the optimal number of classes J of the LC model. Using the model specification in Eq. (8), we applied LC models to the data in Table 1 with two to six classes.[20] The specification diagnostics obtained by this analysis show that $J = 4$ is the optimal number of classes for the BIC and $J = 6$ for the AIC. In cases with $J > 4$, we observed some implausible values for the regression coefficients, e.g. statistically insignificant values for the output. Considering the appealing statistical features of the BIC, we adopted this criterion and selected four classes.

The estimation results of this LC model estimated in the first step are summarized in Table 2. These results show four distinctive technology classes with significant coefficients in most of the cases. Differences in the coefficients indicate that there are variations in marginal costs and technological characteristics across these classes. We see throughout all classes that total cost increases with higher input prices and higher outputs and in three classes with an increasing share of high voltage networks. As expected, operation with density reduces costs. Differences in coefficients indicate that there are variations in marginal costs and technical characteristics across classes. Prior class probabilities indicate also different class sizes.

Table 3 provides a descriptive summary of the observed variables for each class as identified by the estimated posterior class probabilities. These probabilities show that the operators can be distinguished with high probabilities. The fact that even in the worst cases, minimum probabilities are greater than 0.5 suggests that operators can be classified without ambiguity. The resulting classes have at least 100 observations, which is large enough for reasonable degrees of freedom for the second step estimations. The values of the observed variables in each class indicate that we can distinguish in an approximate manner certain features that characterize each class. Class 1 faces low input prices and a high customer density, whereas Class 2 has high input prices and medium customer density. Classes 3 and 4 face intermediate values for most of the variables except for a relatively low customer density in Class 3.

The estimation results for the MOLS and the SFA estimated in the second step for each class separately are summarized in Table 4. Other than the first step, the estimations are based on cost model specification in Eq. (11) and on subsamples of the data in Table 1, given by the four classes of the first step. In general, the coefficients are of the same magnitude as in the LC model in the first step. The coefficients of the MOLS and the SFA differ slightly because of different assumptions on the error term. The signal-to-noise ratio λ is significant for three classes indicating skewness and existence of inefficiency. The insignificant value of λ in class two means that standard errors of the inefficiency terms are low compared to that of the noise terms, which will results in low inefficiency values for this class.

[20] Using several specifications, we also tried models with seven or more classes. Due to non-convergence we could not estimate any models with more than six classes.

Table 2 Estimation results latent class model with four classes, first step

First step variable	Class 1 Coefficient (SE)	Class 2 Coefficient (SE)	Class 3 Coefficient (SE)	Class 4 Coefficient (SE)
Input price ratio (P)	0.8581 *** (0.034)	0.4591 *** (0.020)	0.8675 *** (0.028)	0.6869 *** (0.014)
Distributed electricity (Q)	0.9083 *** (0.016)	0.7752 *** (0.008)	1.0600 *** (0.016)	0.9871 *** (0.008)
Density (D)	−0.4228 *** (0.052)	−0.3301 *** (0.026)	−0.9295 *** (0.067)	−0.0537 *** (0.020)
Share HV network (S)	0.5928 *** (0.153)	0.0875 *** (0.107)	2.7845 *** (0.197)	0.7600 *** (0.056)
Constant	5.0126 *** (0.016)	4.6728 *** (0.007)	4.6591 *** (0.012)	4.7313 *** (0.006)
Sigma (σ^2)	0.1939 *** (0.009)	0.0739 *** (0.005)	0.0939 *** (0.009)	0.0804 *** (0.004)
Prior class probability	0.2157 *** (0.040)	0.2833 *** (0.046)	0.1752 *** (0.039)	0.3258 *** (0.048)

***, **, *: significant at 1 %, 5 % and 10 %, respectively; standard errors given in brackets. T = 5 (panel of years 1998-2002), i = 111, n = 555

Table 3 Identifying the classes

Variable description	Variable	Sample	Class 1	Class 2	Class 3	Class 4
Posterior class probability			0.99 (0.01)	0.91 (0.12)	0.91 (0.16)	0.98 (0.04)
Total cost (10^6 NOK)	TC	36.5 (73.4)	65.2 (136)	14.5 (7.09)	45.9 (67.0)	33.5 (36.7)
Price labor (10^3 NOK/FTE)	P_L	349 (52.9)	338 (53.5)	355 (53.9)	352 (55.5)	348 (49.2)
Price capital (10^3 NOK/MVA)	P_C	176 (83.6)	155 (81.0)	213 (92.4)	166 (76.2)	161 (67.3)
Distributed electrcicity (GWh)	Q	316 (942)	675 (1,936)	109 (90.6)	332 (454)	272 (345)
Density (customers/km)	D	7.00 (3.25)	7.84 (5.15)	6.92 (2.91)	6.28 (1.49)	6.92 (2.49)
Share HV network	S	0.37 (0.11)	0.36 (0.10)	0.37 (0.13)	0.38 (0.07)	0.38 (0.09)
Standard errors given in brackets	*N*	*555*	*115*	*170*	*100*	*170*

Standard errors given in brackets. T = 5 (panel of years 1998-2002), i = 111, n = 555

Table 4 Estimation results MOLS and SFA for each class, second step

	Second step variable	Class 1 Coefficient (SE)	Class 2 Coefficient (SE)	Class 3 Coefficient (SE)	Class 4 Coefficient (SE)
MOLS	Input price ratio (P)	0.8588 *** (0.048)	0.4623 *** (0.018)	0.8644 *** (0.023)	0.6891 *** (0.018)
	Distributed electricity (Q)	0.9080 *** (0.020)	0.7719 *** (0.010)	1.0581 *** (0.010)	0.9856 *** (0.008)
	Density (D)	−0.4236 *** (0.056)	−0.3252 *** (0.017)	−0.9255 *** (0.055)	−0.0464 ** (0.019)
	Share HV network (S)	0.6076 *** (0.202)	0.0802 *** (0.050)	2.7761 *** (0.164)	0.7582 *** (0.072)
	Constant	5.0162 *** (0.019)	4.6692 *** (0.007)	4.6642 *** (0.010)	4.7317 *** (0.007)
SFA	Input price ratio (P)	0.8961 *** (0.050)	0.4624 *** (0.017)	0.8669 *** (0.022)	0.6880 *** (0.018)
	Distributed electricity (Q)	0.9027 *** (0.020)	0.7719 *** (0.010)	1.0587 *** (0.010)	0.9867 *** (0.008)
	Density (D)	−0.4140 *** (0.055)	−0.3253 *** (0.017)	−0.9329 *** (0.053)	−0.0469 ** (0.019)
	Share HV network (S)	0.8017 *** (0.188)	0.0802 *** (0.049)	2.7953 *** (0.161)	0.7538 *** (0.071)
	Constant	4.8158 *** (0.020)	4.6472 *** (0.015)	4.6079 *** (0.016)	4.6947 *** (0.011)
	Sigma: $\sigma^2 = \sigma_n^2 + \sigma_v^2$	0.2786 *** (0.002)	0.0764 *** (0.000)	0.1082 *** (0.001)	0.0870 *** (0.000)
	Lambda: $\lambda = \sigma_u/\sigma_v$	2.1775 *** (0.395)	0.3866 *** (0.265)	0.8413 *** (0.275)	0.6256 *** (0.214)

***, **, *: significant at 1 %, 5 % and 10 %, respectively; standard errors given in brackets. T = 5 (panel of years 1998–2002), i = 111, N = 555

Table 5 Efficiency scores, second step

	Class 1			Class 2			Class 3			Class 4		
	DEA	MOLS	SFA	DEA	MOLS	SFA	DEA	MOLS	SFA	DEA	MOLS	SFA
Mean	0.568	0.776	0.829	0.695	0.906	0.978	0.807	0.882	0.946	0.771	0.899	0.964
SDev	0.207	0.131	0.093	0.168	0.062	0.006	0.149	0.076	0.018	0.125	0.065	0.010
Min	0.267	0.362	0.531	0.374	0.732	0.916	0.481	0.695	0.883	0.559	0.740	0.933
p25	0.429	0.702	0.793	0.559	0.860	0.976	0.680	0.836	0.939	0.682	0.861	0.960
Median	0.496	0.773	0.843	0.652	0.902	0.978	0.799	0.881	0.950	0.743	0.895	0.965
p75	0.688	0.854	0.890	0.832	0.954	0.981	0.958	0.938	0.959	0.860	0.947	0.971
Max	1	1	0.962	1	1	0.986	1	1	0.976	1	1	0.983

T = 5 (1998-2002), i = 111, N = 555

The results of the efficiency analysis for the four classes and three models each are summarized in Table 5. The average efficiency value ranges from 0.56 for DEA in Class 1 to 0.98 for SFA in Class 2. In general, the average efficiency values are lowest in Class 1 for all three models and highest in Class 2 for MOLS and SFA. The highest average efficiency value for DEA is in Class 3 with 0.81. The standard deviations are highest in Class 1 for all three models and lowest in Class 2 for SFA, indicated already by the insignificant lambda in Table 4. Throughout all three models, SFA produces higher efficiency values than DEA and MOLS. This is expected since the model considers statistical noise. Another typical feature is that whereas DEA and MOLS assign full efficiency to several observations, SFA does not classify any operator as fully efficient. The minimum values are low in Class 1 for all three models. In particular, DEA attributes considerably lower minimum efficiency estimates for all classes than the other models.

In general, the efficiency values are higher and more realistic than the corresponding scores of a conventional analysis performed in one step (given in Table 7 in the Appendix). The decomposition of the benchmarking process into two steps and the consideration of technology classes has reduced unobserved heterogeneity within classes and, hence, reduced the unexplained variance previously claimed as inefficiency. Therefore, conventional cross-sectional or pooled models might underestimate cost efficiency.

6 Summary and Conclusions

Regulatory authorities increasingly use benchmarking practices to identify a company's individual efficiency in various incentive regulation schemes such as price- or revenue cap. The identification of cost efficiency in electricity distribution is a challenging task, as the companies operate in different regions with various environmental and network characteristics that are only partially observed. Therefore, the purpose of this study was to analyze cost efficiency in electricity distribution under consideration of these unobserved heterogeneity factors.

In order to disentangle cost efficiency variations from unobserved factors, we proposed an alternative strategy that decomposes the benchmarking process into two steps: The first step is to identify classes of comparable companies in order to reduce unobserved heterogeneity within classes and the second to obtain the best practice for each class.

The analysis in the first step has revealed four distinct latent classes. These classes can be characterized in an approximate manner by different observed variables, mainly by input prices and customer density. The analysis in the second step applying DEA, MOLS and SFA frontier methods has shown that average efficiency values vary considerably among methods and classes. In general, DEA has produced lowest and SFA highest values. Companies in Class 1 are on average considerably less efficient than companies in the other classes, and the variation in efficiency scores in Class 1 is highest. This class involves clearly the largest and most heterogeneous companies concerning output.

Most importantly, the efficiency values are generally higher and more plausible than the corresponding scores of a conventional single-step analysis. The decomposition of the benchmarking process into two steps and the consideration of technology classes has reduced unobserved heterogeneity within classes and, hence, reduced the unexplained variance previously claimed as inefficiency. Therefore, conventional cross-sectional or pooled models might underestimate the real cost efficiency values. This in turn could lead to too incommensurate regulatory measures in account of the affected companies, especially if price or revenue cap regulation as incentive regulation scheme is in force.

7 Appendix

See Tables 6 and 7.

Table 6 Estimation results for MOLS and COLS, conventional analysis

Variable	MOLS Coefficient (SE)	SFA Coefficient (SE)
Input price ratio (P)	0.6827 *** (0.024)	0.6637 *** (0.024)
Distributed electricity (Q)	0.9328 *** (0.011)	0.9327 *** (0.011)
Density (D)	−0.3423 *** (0.028)	−0.3226 *** (0.024)
Share HV network (S)	0.7258 *** (0.091)	0.7679 *** (0.080)
Constant	4.8223 *** (0.009)	4.5888 *** (0.011)
Sigma: $\sigma^2 = \sigma_u^2 + \sigma_v^2$		0.3175 *** (0.001)
Lambda: $\lambda = \sigma_u/\sigma_v$		2.2973 *** (0.229)
***, **, *: significant at 1 %, 5 % and 10 %, respectively N = 555		

Table 7 Efficiency scores, conventional analysis

	DEA	MOLS	SFA
Mean	0.554	0.762	0.802
SDev	0.151	0.148	0.104
Min	0.246	0.319	0.498
p25	0.453	0.645	0.727
Median	0.527	0.769	0.827
p75	0.625	0.871	0.887
Max	1	1	0.960
T = 5 (1998–2002), i = 111, N = 555			

References

Afriat SN (1972) Efficiency estimation of production functions. Int Econ Rev 13(3):568–598

Agrell PJ, Bogetoft P (2009) Harmonizing the nordic regulation of electricity distribution. In: Bjørndal E, Bjørndal M, Pardalos PM, Rönnqvist M (eds) Energy natural resources and environmental economics. Springer, 293–316

Agrell PJ, Bogetoft P, Tind J (2005a) DEA and dynamic yardstick competition in scandinavian electricity distribution. J Prod Anal 23(2):173–201

Agrell PJ, Bogetoft P, Tind J (2005b) Efficiency evaluation with convex pairs. Adv Model Optim 7:211–237

Agrell PJ, Niknazar P (2014) Structural and behaviorial robustness in applied best-practice regulation. Soc-Econ Plan Sci 48(1):89–103

Aigner D, Knox Lovell CA, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econometr 6(1):21–37

Battese GE, Coelli TJ (1992) Frontier production functions, technical efficiency, and panel data: With application to paddy farmers in India. J Prod Anal 3:153–169

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. J Econometr 38:387–399

Behr A (2010) Quantile regression for robust bank efficiency score estimation. Eur J Oper Res 200:568–581

Chamberlain G (1982) Multivariate regression models for panel data. J Econom 18:5–46

Charnes A, Cooper WW, Rhodes EL (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Coelli TJ, Prasada Rao DS, O'Donnell C, Battese GE (2005) An introduction to efficiency and productivity anlaysis. Springer, New York

Cornwell CM, Smith PC (2008) Stochastic frontier analysis and efficiency estimation. In: Sevestre P, Mátyás L (eds) The econometrics of panel data. Springer, New York, pp 697–726

Cornwell C, Schmidt P, Sickles RC (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. J Econom 46:185–200

Cullmann A (2010) Benchmarking and firm heterogeneity: a latent class analysis for German electricity distribution companies. Empirical Economics 1–23

Estache A, Rossi M, Ruzzier C (2004) The case for international coordination of electricity regulation: evidence from the measurement of efficiency in South America. J Regul Econ 25 (3):271–295

Farrell MJ (1957) The measurement of productive efficiency. J Roy Stat Soc Ser A(Part 3):253–290

Farsi M, Filippini M (2004) Regulation and measuring cost-efficiency with panel data models: Application to electricity distribution utilities. Rev Ind Organ 25(1):1–19

Farsi M, Filippini M (2009) Efficiency measurement in the electricity and gas distribution sectors. In: Hunt LC, Evans J (eds) International handbook on the economics of energy. Edward Elgar, pp 598–623

Farsi M, Filippini M, Greene WH (2006) Application of panel data models in benchmarking analysis of the electricity distribution sector. Ann Pub Coop Econ 77(3):271–290

Filippini M, Agrell PJ, Farsi M, Koller M (2011) Messung der produktiven Effizienz in EVU—Eine Anwendung zur Anreizregulierung in der Schweiz. Studie im Auftrag des Bundesamtes für Energie (BFE). Centre for Energy Policy and Economics (CEPE), ETH Zurich

Greene WH (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. J Econom 126(2):269–303

Greene WH (2008) The econometric approach to efficiency analysis. In: Fried HO, Knox, Lovell CA, Schmidt SS (eds) The measurement of productive efficiency and productivity growth. Oxford University Press, Oxford, pp 92–250

Haney AB, Pollitt MG (2009) Efficiency analysis of energy networks: an international survey of regulators. Energy Policy 37:5814–5830

Jamasb TJ, Pollitt MG (2003) International benchmarking and regulation: an application to European electricity distribution utilities. Energy Policy 31(15):1609–1622

Jensen U (2000) Is it efficient to analyse efficiency rankings? Empirical Economics 25(2):189–208

Jondrow J, Knox Lovell CA, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. J Econom 19(2–3):233–238

Kalirajan KP, Obowona MB (1994) Frontier production function: the stochastic coefficients approach. Oxford Bull Econ Stat 56(1):87–96

Knox KJ, Blankmeyer EC, Stutzman JR (2007) Technical efficiency in Texas nursing facilities: a stochastic production frontier approach. J Econ Finan 31(1):75–86

Koenker R, Bassett G Jr (1978) Regression quantiles. Econometrica 46(1):33–50

Kumbhakar SC (1990) Production frontiers, panel data, and time-varying technical efficiency. J Econom 46:201–211

Kumbhakar SC (1991) Estimation of technical inefficiency in panel data models with firm- and time-specific effects. Econ Lett 36:43–48

Kumbhakar SC, Knox Lovell CA (2003) Stochastic frontier analysis. Cambridge University Press, Cambridge

Lazarsfeld PF, Henry NW (1968) Latent structure analysis. Houghton Mifflin, Boston

Littlechild SC (1983) The regulation of British telecommunications' profitability. Department of Trade and Industry, London

Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev 18(2):435–444

Murillo-Zamorano Luis R (2004) Economic efficiency and frontier techniques. J Econ Surv 18 (1):46

Nerlove M (1963) Returns to scale in electricity supply. In: Christ C (ed) Measurement in economics: studies in mathematical economics and econometrics in memory of Yehuda Grunfeld, vol 21. Stanford University Press, Stanford

Orea L, Khumbhakar SC (2004) Efficiency measurement using stochastic frontier latent class model. Empirical Economics 29(1):169–183

Pitt MM, Lee L-F (1981) The measurement and sources of technical inefficiency in the Indonesian weaving industry. J Dev Econ 9(1):43–64

Polachek SW, Yoon BJ (1996) Panel estimates of a two-tiered earnings frontier. J Appl Econ 11:169–178

Schmidt P, Sickles RC (1984) Production frontiers and panel data. J Bus Econ Stat 2(4):367–374

Shepard RW (1953) Cost and production functions. Princeton University Press, Princeton, New Jersey

Shleifer A (1985) A theory of yardstick competition. Rand J Econ 16(3):319–327

Shuttleworth G (2005) Benchmarking of electricity networks: practical problems with its use for regulation. Util Policy 13(4):310–317

Street A (2003) How much confidence should we place in efficiency estimates? Health Econ 12 (11):895–907

van den Broeck J, Koop G, Osiewalski J, Steel MFJ (1994) Stochastic frontier models : a Bayesian perspective. J Econom 61(2):273–303

Winsten CB (1957) Discussion on Mr. Farrells paper. J Roy Stat Soc Ser A (General) 120(3):282–284