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.

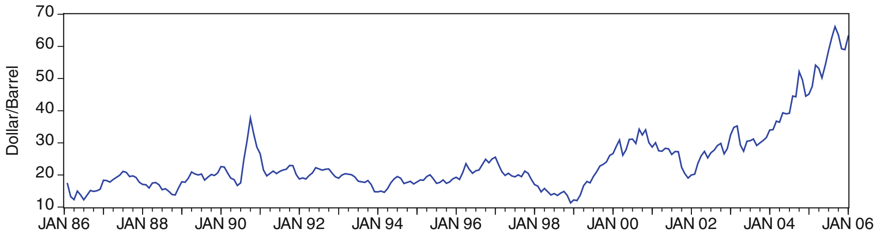

Fig. 1 WTI spot price for the period January 1986–December 2005 (monthly data)

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:

$$S_t = \alpha_0 + \alpha_1 S_{t-1} + \sum_{j=0}^{5} \psi_j D01_{jt} + \lambda S99_t + \sum_{i=0}^{k} \beta_i RIS_{t-i} + \sum_{i=0}^{k} \left(\gamma_i LIS_{t-i} + \delta_i LIS_{t-i}^2\right) + \sum_{i=0}^{k} \left(\phi_i HIS_{t-i} + \varphi_i HIS_{t-i}^2\right) + \varepsilon_t \tag{5}$$

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 $S99$ and a set of dummy variables capturing the effects of 11 September 2001 ($D01$):

$$s_t = \alpha + \beta RIS_t + \delta ppi_t + \varphi od_t + \lambda S99_t + \gamma D01_t + \varepsilon_t \tag{18}$$

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

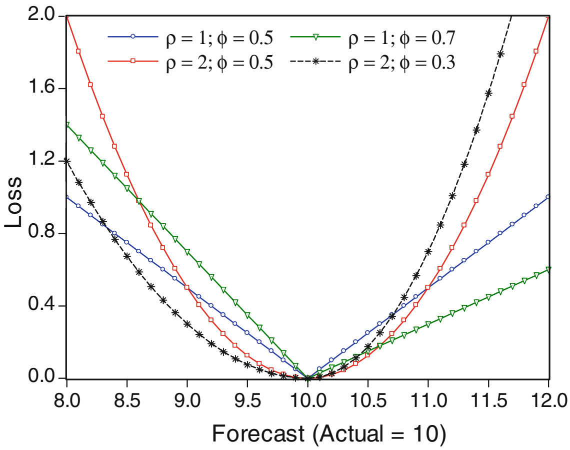

Fig. 2 Generalized loss function.
Notes The generalized loss function refers to Elliott et al. (2005); Forecasts are shown on the horizontal axis; The actual value is equal to 10; Over-prediction, $u < 0$, (under-prediction, $u > 0$) occurs to the *right* (*left*) of the actual value; The graph shows 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.

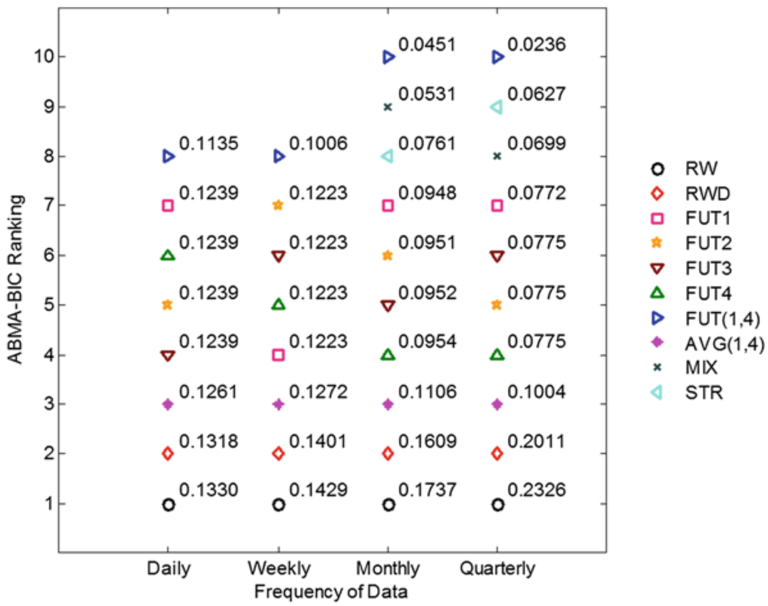

Fig. 3 Ranking of models using ABMA weights. *Notes* Models RW and RWD are described in Sect. 3.2.1 (Eqs. (9) and (10)); Models FUT1—FUT4 are described in Sect. 3.2.2 (Eqs. (13) and (14)); Models FUT(1,4) and AVG(1,4) are described in Sect. 3.2.2 (Eqs. (15) and (16)); Models MIX and STR are described 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

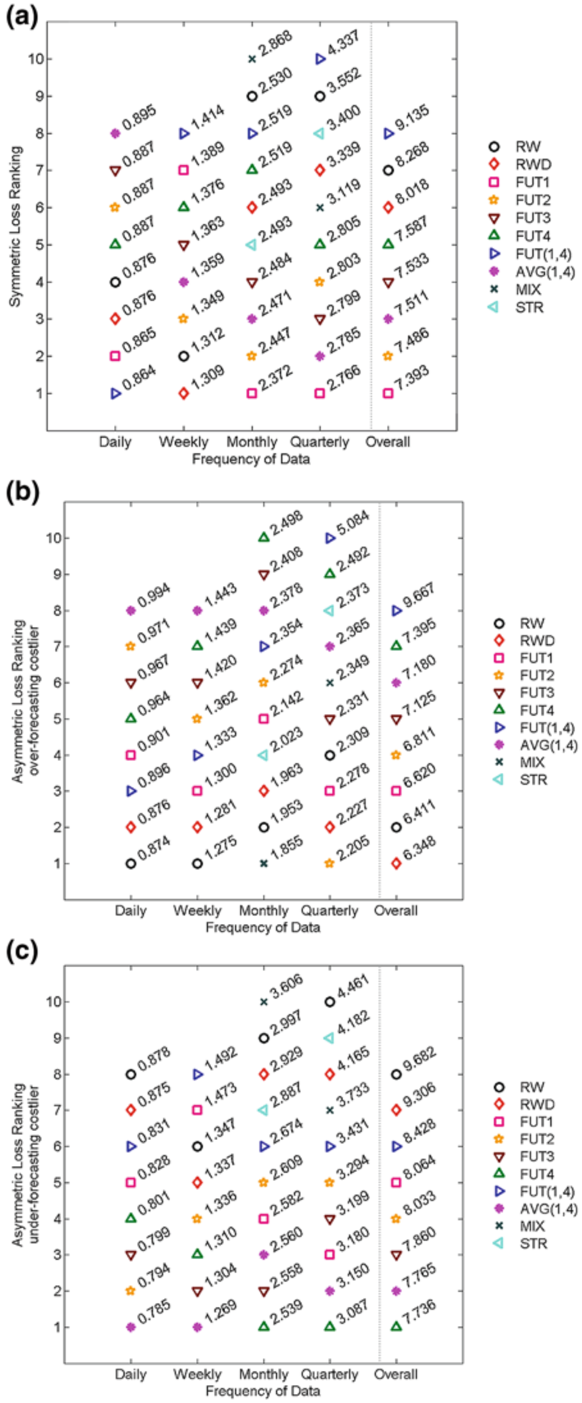

Fig. 4 Ranking of models using the generalized loss function. Notes See Notes of Fig. 3; Panel **a** reports the ranking based on MSE; Panel **b** reports the ranking based on the asymmetric loss function, under the assumption that over-forecasting is costlier; Panel **c** reports the ranking based on 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