CrossMark

Measurement errors in stock markets

Hachmi Ben Ameur[1] · **Fredj Jawadi[2]** ·
Abdoulkarim Idi Cheffou[3] · **Wael Louhichi[4]**

Abstract This paper points to further measurement errors in stock markets. In particular, we show that the application of usual performance ratios to evaluate financial assets can lead to inappropriate findings and consequently wrong conclusions. To this end, we analyze standard performance ratios as well as extreme loss-based financial ratios and compare the conclusions with those provided by systemic risk measures. The application of these different measures to both conventional and Islamic stock indexes for developed and emerging countries in the context of the financial crisis yields two interesting results. First, the analysis of financial performance exhibits further measurement errors. Second, the consideration of extreme loss and systemic risk in computing performance measures increases the reliability of performance analysis.

Keywords Measurement error · Financial performance · Systemic risk · Var · CoVaR and MES

JEL Classification C2 · C5 · G10

✉ Hachmi Ben Ameur
   hechba@yahoo.fr

   Fredj Jawadi
   fredj.jawadi@univ-evry.fr

[1] INSEEC Business School, 27 rue claude Vellefaux, 75010 Paris, France

[2] UFR Sciences de Gestion et Sciences Sociales, Universit of Evry, 2, rue Facteur Cheval, 91000 Evry, France

[3] EDC Paris Business School, Courbevoie, France

[4] ESSCA Business School, Angers, France

1 Introduction

Theories and models in Exact Science as in Social Science can accurately reflect reality if their estimated value reflects the true value. However, the latter is often unobserved in practice and its accuracy varies depending on the measurement instrument used. If the measuring instrument is not exact, then this will result in a measurement error. Measurement error can be defined as the difference between the true value of a quantity and the value obtained by a measurement. Measurement error is known to include both random errors and systematic errors (or statistical errors). While the latter means that a measured value contains an offset, the former implies that a repeated measurement generates a measured value different from the previous value (Taylor 1999). A measurement error can have different sources: (i) the data collection and sampling method used, (ii) selection bias, (iii) the statistical procedure and variables used, (iv) the econometric models and tests, (v) an ill-suited survey design, etc.[1]

Obviously, measurement error is a serious problem in both Rocket Science and Economics as it risks causing biased analyses and leading to incorrect conclusions. Measurement error can have severe consequences, and it is therefore essential to assess the quality of measurements in order to reduce such errors. While in Exact Science, instruments can be improved through experimentation, but it is a difficult and expensive process in Social Science.[2] Accordingly, various statistical methods have been introduced to deal with measurement errors.[3] In Economics, economists often make use of econometric tools and proceed through trial and error until they arrive at the best method or model. This requires estimating different specifications, essaying different calibrations, using different variables and data, and applying different methods and repeatability essays to validate an instrument.[4]

Furthermore, it is worth noting that measurement uncertainty and measurement error are mostly important during periods of crisis than in periods of prosperity, as further errors would amplify losses for investors and the economic system as a whole. For instance, following the subprime crisis (2007) and the global financial crisis (2008–2009), several recent studies showed the high level of measurement errors in economic, monetary and financial data as well as in the methods and models under consideration. They suggest that the macroeconomic models, financial methods and economic policies used have failed to reflect the reality of the economic system over the last few decades. Their application resulted in significant errors that were sources of bankruptcy for Some well-known banks, major losses for a number of financial markets and investors, and a considerable risk that threatened to destroy the design of the economic system, etc.

Accordingly, a growing recent body of literature has pointed to several major measurement errors, and recommended urgent solutions to improve measurement processes. In this context, a new US institution, the Society for Economic Measurement (SEM),[5] was founded in 2013 by Carnegie Mellon University, the Center for Financial Stability and the University of Kansas to promote economic measurement research. Research led by its first President, Professor William A. Barnett (University of Kansans, USA), recommended using recent developments

---

[1] For more details, see Fuller (1987), Caroll, Ruppert and Stefanski (1995), Hausman (2001), etc.

[2] See Barnett (2015) for an excellent comparative analysis between Rocket Science and Economics Science.

[3] See Cheng et al. (2011) for a survey of the treatment of classical and non-classical measurement errors. Blackwell et al. (2015) also focused on the treatment of measurement errors and developed a unified approach to deal with missing data problems and to limit measurement errors.

[4] See Chen et al. (2011) for a recent survey on Nonlinear Models of Measurement Errors.

[5] http://sem.society.cmu.edu/home.html.

in econometrics, statistics and experimental economics to improve measurement. Indeed, Barnett et al. (2009) compared simple-sum monetary aggregates with the Divisia monetary aggregate indexes to identify measurement errors in monetary aggregates. Using a Markov switching factor approach, the authors found significant differences between these indexes which, interestingly, occur more significantly around the beginnings and the ends of economic recessions and during phases of high interest rate.

Barnett and Chauvet (2011) focused on the index number theory and showed that its use could help the Fed to identify a financial crisis and to reduce the misperception of systemic risk. Barnett (2012) also showed that obsolete monetary aggregates and data were often used, especially in economic systems with complex financial instruments, leading the Fed to make inappropriate decisions that prevented it from correctly assessing the excessive risk taken by homeowners and being able to control the systemic risk. He concluded that increased financial complexity and reduced data quality (by the application of a simple-sum monetary aggregate) resulted in a toxic mix.

In spite of the large numbers of papers that have investigated the issue of measurement errors in the monetary market, there are still a few studies on measurement errors in stock markets, even though the latter were at the centre of the recent global financial crisis, since several stock markets lost over 50 % in 2008. Almeida et al. (2010) pointed to the problem of errors-in-variables that make the standard estimators inconsistent due to an attenuation bias. Drerup et al. (2014) suggested measurement error in subjective expectations and showed that these errors may affect portfolio choice due to the implementation of heterogeneous decision-making rules.

The present paper also analyses measurement errors in stock markets. In particular, we examine the topic with a focus on financial performance to show that the performance evaluation of stocks is also subject to errors.[6] This choice is justified by the fact that performance evaluation is a key issue for investors as it provides them with a benchmark to decide on their investments and to define the level of risk they might take. A measurement error when evaluating the performance of financial assets can result in an incorrect analysis and severe losses, thereby negatively impacting on the representation of individual decision-making processes. Indeed, measurement error associated with performance evaluation is observed repeated as most performance ratios require strict hypotheses (normality, symmetry, certainty, restrictive measures for risk, estimate of risk premium, perfect expectations, etc.), while financial data do not usually allow these assumptions to be checked, which makes performance ratios obsolete and their conclusions inappropriate and incorrect. The literature says little about errors in financial performance measures as few papers have investigated the issue.[7] Indeed, Eling and Schuhmacher (2007) compared the ranking of 2763 hedge funds provided by the Sharpe ratio with that of twelve other performance ratios[8] and showed that the ranking was unchanged despite significant deviation from normality. The authors concluded that it indicated the high performance of hedge funds and affirmed that the choice of performance measure does not affect their evaluation.

At the same time, in the aftermath of the global financial crisis, hedge funds were not just considered as responsible for the underlying problem but were also the crisis propagators, and they have since been sharply criticized. Moreover, several other, even well noted, finan-

---

[6] Performance is often defined and measured through a risk-adjusted return. The most widely known performance measure is given by the ratio of Sharpe (1966).

[7] To our knowledge, this is the first paper on the topic.

[8] Sharpe ratio, Treynor ratio, Jensen alpha, Omega ratio, Sortino ratio, Kappa 3, the upside potential ratio, Calmar ratio, Sterling ratio, Burke ratio, the excess return on Value at Risk, the conditional Sharpe ratio and the modified Sharpe ratio.

cial assets and investments defaulted, resulting in major losses. Our explanation is that the financial performance evaluation methods for the different assets were measured with errors, including the subprime credits and housing assets. Indeed, at least three main arguments and criticisms can be addressed. On the one hand, computation of performance ratios requires a measure of the risk. However, as the latter is not observed in practice, it is often proxied, and an incorrect proxy for risk may lead to significant measurement errors. For example, while the Sharpe ratio refers to the standard deviation to proxy total risk, the Treynor ratio only focuses on the systematic risk that is approximated by the market beta. Thus, unsurprisingly, performance ratios can diverge and the question would be: what is the best performance ratio? On the other hand, most well-known performance ratios evaluate risk taking average and strong hypotheses (normality, symmetry, etc.) into account. Thus, they cannot include further extreme risk and are therefore less reliable if normality does not hold. Furthermore, to our knowledge, no performance takes further systemic risk into account, even though the recent global financial crisis clearly showed the high level of such risk. All in all, taking these limits into account, it seems likely that existing performance ratios provide poor information and somewhat unreliable conclusions.

Accordingly, this paper contributes to present knowledge by highlighting the significant measurement errors in stock markets when evaluating performance indexes in developed and emerging markets. To this end, we first, apply the usual performance ratios for both conventional and Islamic stock indexes and evaluate their performance under normal conditions. Thereafter, further asymmetry and extreme losses are taken into account, and we compute the performance using more robust ratios and identify contradictory conclusions with regard to those of standard ratios, suggesting further evidence of measurement errors. Finally, we focus on the systemic risk issue and show that the supposedly best-performing indexes are, at the same time, systemically the riskiest.

While checking for measurement errors, we indirectly check what would be the most appropriate performance ratio and it seems that taking further extreme risk and the systemic risk into account is helpful. Interestingly, our findings lead us to advocate the use of extreme-risk adjusted performance ratios (using the VaR and Conditional VaR as proxies) that are also robust to systemic risk (e.g., using the CoVaR and MES) in order to correctly evaluate financial assets and reduce stock market measurement errors.

This paper includes four sections. Section 2 presents the econometric methodology. The main empirical results are discussed in Sect. 3. Section 4 concludes.

2 Econometric methodology

This section presents both standard and robust performance ratios. It discusses the limitations of standard ratios and the most recent performance ratios, taking into accoun trobust measures of systemic risk.

2.1 First performance ratios' generation

The idea of first performance ratios is to evaluate the performance of risky financial assets with regard to the relationship between their expected profitability-risk ratio and the return from a risk-free asset. Among the first ratios, we cite the well-known Sharpe ratio introduced by Sharpe (1964), initially used to classify American mutual funds.The simplicity of this ratio made it very popular as it enables financial performance to be evaluated using only the

two first moments. Formally, it measures the return per unit of risk [Eq. (1)], and the higher the Sharpe ratio

(S_i) of an asset i, the better the combined performance of "risk" and return.

$$S_i = \frac{E(R_i) - r_f}{\sigma(R_i)} \tag{1}$$

where $E(R_i)$ denotes the expected return of a risky asset i, $\sigma(R_i)$ denotes the standard deviation considered as a risk proxy, r_f is the risk-free return.

However, this ratio requires normality distribution for stock returns. Accordingly, with an asymmetrical distribution and investors somewhat sensitive to such asymmetry, the Sharpe ratio is no longer reliable. In addition, when using standard deviation to proxy the risk, the ensuing risk measure includes both diversifiable (idiosyncratic) and non-diversifiable (systematic) risk. Thus, not only penalizes both the upside and the downside potential of portfolio return, but the ratio is also no longer appropriate since a rational investor is expected to select an optimal portfolio that rejects idiosyncratic risk.

As an extension, Treynor (1965) corrected this bias and substituted total risk by systematic risk. The latter is given by the market beta, noted $\beta_i = \frac{\text{cov}(R_i, R_M)}{\sigma^2(R_M)}$, that can be obtained from the market model, where $\text{cov}(R_i, R_M)$ measures the covariance between the return of an asset i (R_i) and that of the market (R_M), while $\sigma^2(R_M)$ measures the total market risk. The Treynor ratio corresponds to:[9]

$$TR = \frac{E(R_i) - r_f}{\beta_i} \tag{2}$$

As for Sharpe ratio, the higher the TR of an index i, the better its performance. However, while the TR is preferred to the S_i, its main limitation is that it supposes the stability of the market beta which cannot be reliable (Ben Ameur et al. 2015).

Accordingly, Jensen introduced the Jensen's alpha measure, which provides a performance measure more guided by its theoretical performance that is based on the CAPM (Capital Asset Pricing Model). Formally, Jensen' alpha of an asset i, noted α_i, corresponds to:

$$\alpha_i = E(R_i) - \lfloor r_f + \beta_i(E(R_M) - r_f) \rfloor \tag{3}$$

where $E(R_M)$ denotes the expected market return.

In the same context, Modigliani and Modigliani (1997) proposed a risk-adjusted measure of portfolio performance, noted M^2, which presents the advantage of adjusting the portfolio risk to the market risk. They derived it from the Sharpe ratio and expressed the performance in units of percent return. This is a measure of the risk-adjusted returns of some investment portfolios, meaning that it captures the portfolio returns, adjusted for the portfolio risk relative to that of the market. Formally, it corresponds to:

$$M^2 = \left(\frac{\sigma(R_M)}{\sigma(R_i)}\right)(E(R_i) - r_f) + r_f$$
$$= S_i \sigma(R_M) + r_f \tag{4}$$

As for previous performance measures, the higher the M^2, the stronger the portfolio performance.

---

[9] Treynor ratio requires also implicitly normality distribution for stock returns as it is also based on first two moments. The main difference with Sharpe ratio is that it involves the systematic risk instead of the intrinsic risk.

The last two measures of the first generation performance ratios were introduced by Graham and Harvey (1994, 1996) and are noted GH_1 and GH_2 respectively. These measures help to overcome two drawbacks associated with S as they provide measures for performance while enabling the risk volatility asset to differ. They also allow the risk-free rate to be variable and at the same time correlated with the risky asset. These measures are particularly interesting as in practice there is a significant correlation between changes in interest rates (risk-free asset) and the short-term profitability of financial markets (risky asset). Formally, the measure GH_1 is determined by the combination of the benchmark portfolio and the risk-free asset and corresponds to:

$$GH_1 = E(R_i) - \left[R_f + \left(\frac{\sigma_i}{\sigma_m} \right) (E(R_m) - R_f) \right] \tag{5}$$

GH_2 is defined as the difference between the portfolio incorporating the risky asset and the risk-free asset and portfolio returns for the same market value volatility, and corresponds to:

$$GH_2 = \left[R_f + \left(\frac{\sigma_m}{\sigma_i} \right) (E(R_i) - R_f) \right] - E(R_m) \tag{6}$$

2.2 Second performance ratio's generation

Financial markets have experienced several major variations and corrections over the last three decades [the 1987 stock crash, the Internet Bubble (2000), the subprime crisis (2007), the global financial crisis (2008–2009), etc.] that resulted in great volatility excess and extreme losses. Furthermore, an analysis of the effects of the recent global financial crisis points to the importance of high systemic risk that stimulated the propagation of the crisis from the US to several developed and emerging financial markets. However, previous performance ratios were not adapted to take extreme losses and therefore the tail distribution risk into account, which led to several errors and incorrect notations for risky financial assets (i.e. subprime assets). In order to adjust performance ratios to extreme risk and better take the linkage increase between international financial markets into account, several extensions have been proposed that we classify in a second generation of performance ratios.

Dowd (2000) proposes an adjusted to extreme risk ratio similar to the Sharpe ratio but that proxies the risk by the Value-at-Risk (VaR) rather than total risk [Eq. (8)]. Let's recall that the VaR describes the possible loss of an investment, which is not exceeded for a given probability and a given horizon.

Under the normal distribution hypothesis, the Value-at-Risk of an asset i corresponds to:

$$VaR_i(\alpha) = - \left(R_i^d + z_\alpha \sigma_i \right) \tag{7}$$

where z_α is the α-quantile of the standard normal distribution, R_i^d is the mean return, σ_i is the volatility parameter,

Regarding the estimation of the VaR,[10] the historical method is preferred in practice as it does not require any assumption of the distribution, which is determined from the historical

---

[10] In the literature, the parametric and non-parametric approaches can be used to estimate the Value-at-Risk model. We identified three main methods: variance–covariance approach, historical simulation and Monte Carlo simulation. The first is a parametric method based on the normality assumption of the distribution of the market parameters and index. The second, non-parametric method is the easiest approach as only historical data are used to determine the VaR for the market and the index. The third is also a non-parametric method that requires two steps. In the first step, the volatilities and correlation parameters are calibrated using the historical data. In the second step, simulation of the stochastic processes is used to establish the return distribution, and the VaR can then be determined from this distribution.

data. However, as the normality distribution hypothesis is often rejected for financial data, Cornish–Ficher developed the Modified Value-at-Risk that has the advantage of taking further asymmetry (skewness excess) and leptokurtic character (Kurtosis excess) into account.

$$RVaR_i(\alpha) = \frac{E(R_i) - R_f}{VaR(R_i)} \tag{8}$$

While the VaR enables the probability of extreme risk to be taken into consideration, it has been subject to some criticism as it is neither convex nor sub-additive. To overcome these limitations, Martin et al. (2003) proposed another extension, the STARR ratio, whose performance measure is based on the Conditional Value-at-Risk (CVaR). This risk measure is more informative than the VaR as it evaluates the mean of large losses beyond VaR-level. The CVaR also satisfies certain plausible axioms of Artzner et al. (1999) and is considered as a coherent measure of risk in the literature (Pflug 2000). Formally, the CVaR is defined as:

$$CVaR_i(\alpha) = E\left[R_i \left(R_i \leq VaR_i(\alpha)\right)\right] \tag{9}$$

where R_i denotes the return of a risky asset i.

Hence, the corresponding performance ratio becomes:

$$RCVaR_i(\alpha) = \frac{E(R_i) - R_f}{CVaR(R_i)} \tag{10}$$

Biglova et al. (2004) proposed a performance measure called the Rachev ratio, based on the ratio between average earnings and mean losses. The Rachev ratio is computed using the same idea as a CVaR risk measure, with the particularity that it focuses on the right-tail reward potential. Formally, the Rachev ratio corresponds to:

$$R - Ratio\,(\alpha, \beta) = \frac{CVaR_{1-\alpha}\left(r_{risk\text{-}free(benchmark)} - r_{portfolio(asset)}\right)}{CVaR_{1-\alpha}\left(r_{portfolio(asset)} - r_{risk\text{-}free(benchmark)}\right)} \tag{11}$$

where $r_{portfolio(asset)}$ the return of the portfolio or the index and $r_{risk\text{-}free(benchmark)}$ the risk free return or the market return.

The higher the Rachev ratio, the better the performance. However, this ratio focuses on return distribution extremes and ignores most of the returns in the middle.

In the same context, the Darolles et al. (2009) L-performance measure is also recommended as it is robust to outliers and can accommodate thick-tailed distributions. Formally, it is based on the notion of trimmed L-moments, used in the analysis of catastrophic events such as extreme floods or low flows, and corresponds to:

$$L_{p,\alpha} = VaR_{R_i,0.5} \times \left(VaR_{-R_i,\alpha} - VaR_{R_i,\alpha}\right)^{-1} \tag{12}$$

where R_i denotes the return of a risky asset i and α the quantile.

Finally, to overcome the limitations of previous ratios, the Omega ratio [Eq. (13)], proposed by Keating and Shadwick (2002), provides a measure that incorporates all the moments of the return distribution including skewness and kurtosis. As financial assets do not show a normal distribution, the performance analysis based on their mean and variance may induce measurement and evaluation errors. Accordingly, the use of performance ratios that take the probability of extreme losses into account would help to improve the analysis of financial performance and the selection of optimal investments.

$$\Omega = \frac{\int\limits_{r}^{b} (1 - F(x))\, \mathrm{d}x}{\int\limits_{a}^{r} F(x)\, \mathrm{d}x} \tag{13}$$

Overall, the second generation of performance ratios contributes by taking further extreme losses into consideration. However, the matter of systemic risk is still not included, something that is more than required as, in the aftermath of the global financial crisis (2008–2009), the importance of systemic risk was noted to be responsible for major losses even for financial assets, investments and institutions with supposedly superior financial performance (compared to the previous performance ratios). Accordingly, besides the evaluation of the financial performance with these ratios, the consideration and the evaluation of the systemic risk in evaluating the financial performance is required to complete the analysis of financial performance and to minimize measurement and evaluation errors.

2.3 The measure of systemic risk contribution

With regard to systemic risk, this sub-section discusses two different measures that help to improve the evaluation of financial performance while taking the further contribution of systemic risk into account: (i) the CoVaR measure, and (ii) the marginal expected shortfall approach.

2.3.1 The CoVaR measure

Adrian and Brunnermeier (2011) proposed a measure for systemic risk using the conditional Value-at-Risk, called the CoVaR. If we consider two assets i and j, the CoVaR is defined as the VaR of one asset at some probability quantile that is conditional on another asset that is at its VaR level for the same quantile. In other words, while the VaR describes the possible loss of an investment, it does not exceed a given probability of $1 - \alpha$ $P(R_i \leq VaR_i(\alpha)) = \alpha$, where R_i denotes the asset return i. The VaR of asset j, conditional on the event $(R_i = VaR_i(\alpha))$, is denoted by, $CoVaR_{j(i}(\alpha) = VaR_i(\alpha)$, where α is the quantile. Accordingly, we write:

$$P\left(R_i \leq CoVaR_{j(i)}(\alpha) \,|\, R_i = VaR_i(\alpha)\right) = \alpha \tag{14}$$

Thus, the contribution of the asset i to the risk of the asset j corresponds to:

$$\Delta CoVaR_{j(i)}(\alpha) = CoVaR_{j(R=VaR_i)}(\alpha) - CoVaR_{j(R=median_i)}(\alpha) \tag{15}$$

where $\text{CoVaR}_{j(R=median_i)}(\alpha)$ corresponds to the CoVaR when the market is not in stress.

By using the $CoVaR$ and the $\Delta CoVaR$ measures with regard to the Treynor ratio, we take into account the further dynamic impact of the world market risk on each index. Accordingly, our contribution consists of proposing the use of a dynamic measure to evaluate financial performance. Indeed, measuring systemic risk with the CoVaR captures the direct and indirect spillover effect based on the covariation between the distribution tails of specific indices and the benchmark index, while the use of the $\Delta CoVaR$ enables us to incorporate new information at each time t provided by the new sample data sets. In practice, the estimation of the $CoVaR$ measure is determined by the quantile regressor method proposed by Koenker and Bassett (1978), which has the advantage of not requiring any hypothesis regarding the returns distribution. Furthermore, the quantile regression method analyzes the impact on the distribution tails rather than on the average, as is the case with the Least Square method. It

also allows us to resolve the problem of aberrant values of widely dispersed errors as the quantiles are less sensitive to further outliers.

As for the $\Delta CoVaR$, we use the sliding window technique, which enables us to capture the contribution of extreme market movements on the extreme risk of each index. This method allows us to focus exclusively on the distribution tails of each index while reproducing further spillover effect between the world market risk (asset j) and a given asset i. Accordingly, we proceed in different steps. First, we estimate the sensitivity of each specific index (i) relative to the benchmark index (j). Second, we estimate the contribution of each specific index to the systemic risk of the global market index.

2.4 The marginal expected shortfall approach

The expected shortfall (ES) risk measure, introduced by Artzner et al.(1999), refers to the VaR, but is more appropriate than the VaR as it is not affected by the problem of subadditivity. The idea consists of measuring the tail dependence in the return of market indices and considering the tail dependence estimates as a measure of systemic risk. Acharya et al. (2010) define the marginal expected shortfall (MES) as the marginal contribution of firm J to the expected shortfall of the financial system. In this paper, we adapt the MES Approach from the risk study of financial firms to the study of international systemic risk. This enables us to analyze the further contagion effect related to extreme price variation and to identify the markets with the highest contribution of systemic risk across time on other markets.

Formally, the MES for firm J is the expected value of the stock return R_J conditional on the market portfolio return R_i being at or below the sample p-percent quantile. Let C be a constant that corresponds to the level of market "tail risk" that we can choose, we can write:

$$MES(R_J, p) = E\left(R_J \mid R_i < VaR\left(R_i, p\right) = C\right) \tag{16}$$

The assets that contribute most significantly to the market decline are those with the highest levels of MES. These assets are the most likely candidates to be systemically risky.

In practice, we use the methodology proposed by Brownlees and Engle (2010) to estimate MES from market index i and the market index j. This approach consists of a multi-stage estimation. In the first step, GARCH models are used to generate conditional volatilities and standardized residuals. In the second step, the ADCC specification, which allows us to determine the conditional asymmetric correlation and then the estimates of the tail expectation, is determined by a nonparametric kernel estimation approach proposed by Scaillet (2005). The main advantage of this approach is that the estimator does not require any hypothesis regarding the distribution of the index returns. Furthermore, Brownless and Engle (2010) showed that this approach yields smooth estimates of the MES, implying that the dynamic MES estimates are not over sensitive to small changes in the systemic loss threshold level.

Formally, let $r_{i,t}$ and $r_{j,t}$ be respectively the return of two market indices, with:

$$r_{i,t} = \sigma_{i,t}\varepsilon_{i,t}$$
$$r_{j,t} = \sigma_{j,t}\varepsilon_{j,t} \tag{17}$$

$$r_{i,t} = \sigma_{i,t}\rho_{i,t}\varepsilon_{j,t} + \sigma_{i,t}\sqrt{1 - \rho_{i,t}^2}\,\zeta_{i,t}$$
$$With : \left(\varepsilon_{j,t}, \zeta_{i,t}\right) \rightarrow F \tag{18}$$

where $\sigma_{i,t}$, $\sigma_{j,t}$ and $\rho_{i,t}$ are respectively the volatilities of the market index j and i and the conditional correlation between the two market indices. $\varepsilon_{j,t}$, $\zeta_{i,t}$ the disturbances are assumed to be independently and identically distributed. The joint distribution F allows for the random variables to be uncorrelated but not independent.

3 Empirical analysis

3.1 The data and preliminary analysis

The data includes conventional and Islamic stock indexes from developed and emerging markets, including the Dow Jones Developed Market (DJDM), the Dow Jones Emerging Market (DJEM), the Dow Jones Islamic Developed Market (DJIDM), the Dow Jones Islamic Emerging Market (DJIEM) and the Dow Jones Global Market (DJGM) as the benchmark index. While using the DJ indexes for both conventional and Islamic indexes, we did minimize the risk of heterogeneity in the data under consideration. Furthermore, this sample allows us to conduct a dual comparison for financial performance: Developed versus Emerging markets and Conventional versus Islamic markets. As for the risk-free interest rate (r_f), we use the 90-Day Treasury Bill as a proxy. Data is weekly and covers the period April 2004–March 2014, which allows to evaluate financial performance in the context of the financial crisis as well as during calm periods (before July 2007). Interestingly, our sample covers several crises as well as stylized facts linked to the Gulf War, the subprime crisis and the recent crisis, suggesting further evidence of time-variation in stock markets and investors' behavior. Stock indexes are closing prices, and all data are obtained from the Bloomberg database.

To get a preliminary idea of the performance of the different stock markets under consideration with regard to the benchmark index (the DJGM), we plot each index with the DJG index in Fig. 1 and note several results. First, all of the indexes seem to have undergone a major correction during the subprime and financial crises, while they appear to have performed better over the last few years. Second, the pegging of stock indexes to the DJGM is more marked for developed indexes than for emerging prices. Third, the spread between the DJGM and the Islamic indexes seems significantly reduced. Fourth, the strong time-variation of these indexes suggests that their performance also evolves over time and that they share a common trend by the end of the period, which may be associated with further systemic risk effect.

Next, we applied three unit root tests (ADF, PP and KPSS) and illustrated that all of the series are I(1). We then focused on stock return series, computed as a first logarithm difference of stock prices.[11] From Table 1, we note that all stock returns are close to zero on average, which possibly reflects the correction induced by the financial crisis. The DJEM shows the highest mean returns and standard deviation, which is in line with previous studies on emerging markets. The third and fourth moment statistics point to significant asymmetry and leptokurtic excess, while the strong rejection of normality confirms that the distribution queues are higher than those of normal distribution. Furthermore, the negative value for skewness may suggest further nonlinearity and non-normality in the data, which makes the use of some financial performance ratios inappropriate

---

[11] The results of the unit root tests are not reported but are available upon request.

Fig. 1 The dynamics of Dow
Jones indexes
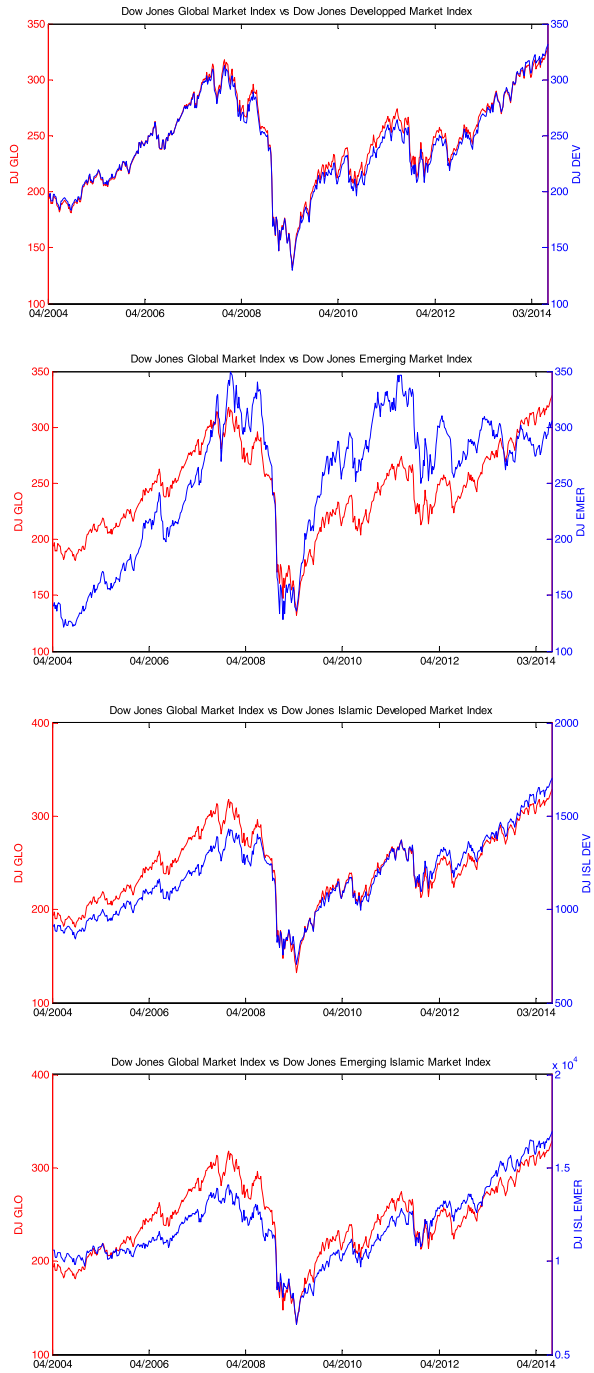
Table 1 Descriptive statistics and normality test

	DJ Global	DJ Dev	DJ Emer	DJ Isla Emer	DJ Isla Dev
Min return	−0.0960	−0.0960	−0.0964	−0.0870	−0.0927
Max return	0.0500	0.0498	0.0751	0.0465	0.0455
Mean return	0.0004	0.0004	0.0006	0.0004	0.0005
STD	0.0069	0.0114	0.0146	0.0104	0.0111
Variance return	0.0001	0.0001	0.0002	0.0001	0.0001
Skewnes	−1.4584	−1.4831	−1.0311	−1.0206	−1.3690
Kurtosis	11.3630	11.6546	7.9084	10.7083	11.1564
Jarque–Bera test	2.9286e+003	3.0777e+003	1.4202e+003	2.5270e+003	2.8077e+003

Table 2 First generation performance measures for (a) the whole period, (b) sub-period 1 and (c) sub-period 2

	DJDM	DJEM	DJIEM	DJIDM
(a)				
Sharpe	0.0093	0.0204	0.0078	0.0175
Treynor	0.0001	0.0002	9.9290.10?5	0.00002
Jensen Alpha	3.98e–06	1.81e–04	−3.7e–06	9.60e–05
M^2	3.58e–06	1.31e–04	−1.39e–05	9.75e–05
Graham Harvey 1	3.54e–06	1.67e–04	−1.26e–05	9.41e–05
Graham Harvey 2	3.56e–06	1.31e–04	−1.39e–05	9.75e–05
(b)				
Sharpe	0.0042	0.1263	−0.013	0.0821
Treynor	0.0002	0.00098	−0.0001	0.0002
Jensen Alpha	−5.3e–05	8.27e–04	−3.7e–04	−1.16e–04
M^2	−5.44e–05	5.08e–04	−4.34e–04	−1.22e–04
Graham Harvey 1	−5.34e–05	7.56e–04	−4.14e–04	−1.26e–04
Graham Harvey 2	−5.43e–05	5.07e–04	−4.34e–04	−1.22e–04
(c)				
Sharpe	0.0013	−0.0113	0.0138	0.014
Treynor	1.798 e–05	−0.00016	0.0002	0.0002
Jensen Alpha	3.28 e–06	−1.68e–04	−1.77e–04	1.92e–04
M^2	3.32e–06	−1.34e–04	−1.99e–04	2.01e–04
Graham Harvey 1	3.28e–06	−1.66e–04	−1.78e–04	1.92e–04
Graham Harvey 2	3.32e–06	−1.34e–04	−1.99e–04	2.01e–04

3.2 Measurement errors in financial performance analysis

We computed the different financial performance ratios and analyzed their results. To better apprehend the evolution of the financial performance of stock indexes, we break down the sample under consideration into two subsamples: (i) sample 1: before the subprime crisis (April 2004–July 2007), (ii) sample 2: after the subprime crisis (August 2007–March 2014)

and we recomputed these ratios. We reported the main results of the computation of the first generation performance ratios in Table 2a, b, c.

3.2.1 Financial performance analysis with first generation ratios

Our analysis over the whole period points to higher financial performance for the DJEM, while the DJIEM seems to show the lowest performance (Table 2a). After breaking down the sample into two subsamples to take the crisis effect into account, we show that before the crisis, the DJIEM is still the best performing index of all the first generation performance ratios, while the DJIDM supplants the other indexes after the crisis.

This conclusion is in line with several previous studies (Jawadi et al. 2014; Arouri et al. 2013; etc.) which show that the crisis had a knock-on effect on Islamic finance that became a form of hedging and refuge for investors. However, these results should be considered with caution as the data properties make the application of the first generation performance ratios inappropriate and its conclusions incorrect. Indeed, the data under consideration do not follow a normal and symmetrical distribution, which makes the reference to the Sharpe ratio incorrect. Furthermore, Trenor and Jesen's conclusion may not operate as they assume stability of the market beta, while Fig. 1 highlights further evidence of time-variation. Finally, it is important to note that all of these ratios adjust return excess to risk (total or systematic) that is adjusted in the mean, but they do not take further extreme losses in the market into account, as observed in 2008. Accordingly, the application of these ratios may result in incorrect conclusions and measurement errors. To remedy this inconsistency, we also applied the second generation ratios for the full period and the two sub-periods.

3.2.2 Financial performance analysis with second generation ratios

We compute the second generation ratios and reported them in Table 3a, b, c. Accordingly, we obtain several interesting results. First, over the whole period, the DJEM seems to be the best performer according to the RVaR, the RCVaR and the OMEGA ratios, while the DJIEM is significantly better for the Rachev ratio. However, the DJDM performs better than the other indexes according to the L-performance ratio. Second, when we focus on the two subsamples, we note that the DJEM is preferred for the RVaR, the RCVaR, the OMEGA and the L-performance ratios, while the Rachev ratio gives preference to the DJIDM. With regard to the second subsample, the divergence between ratios is more explicit (with the same ratio too, but for different levels of probability). Indeed, the DJDM is preferred only by the L-performance ratio, while the RCVaR preferred the DJEM. As for the VaR (1 %) and the Rachev ratio, the DJIEM supplants the other markets, while the RVaR (5 %) and the OMEGA ratios are higher for the DJIDM. Overall, these findings point to two interesting conclusions. On the one hand, it seems important and necessary to include further extreme losses and price variations when measuring financial performance as this enables us to take the asymmetry and high-tail distribution inherent to the data into consideration. On the other hand, the analysis based on the second generation ratios improves the previous analysis when using more robust financial performance ratios. For example, the Rachev ratio takes the specific properties of these indices into account, such as the fact that asset return distributions are fat-tailed and skewed, and also by taking coherent risk measure into account and focusing on some quantile of the distribution. The RVaR

Table 3 Second generation performance measures for (a) the whole period, (b) subsample 1 and (c) subsample 2

	DJDM	DJEM	DJIEM	DJIDM
(a)				
RVaR 5 %	0.006	0.013	0.0045	0.0111
RVaR 1 %	0.0028	0.006	0.0031	0.0056
RCVaR 5 % (V0 = 100)	3.606e–05	0.023	1.991e–05	6.941e–05
RCVaR 1 % (V0 = 100)	2.050e–05	0.044	3.3407e−05	3.940e−05
Rachev	0.7146	0.755	0.901	0.737
L-performance	0.0464	0.045	0.025	0.044
OMEGA excess return	0.9914	1.027	0.987	1.015
(b)				
RVaR 5 %	0.0254	0.0766	–0.0075	0.0171
RVaR 1 %	0.0142	0.0878	–0.0047	0.0106
RCVaR 5 % (V0 = 100)	0.000178	0.0166	–5.85e–05	0.000134
RCVaR 1 % (V0 = 100)	0.00013	0.0836	–4.5831e–05	0.000102
Rachev	0.7022	0.6733	0.7217	0.7517
L-performance	0.0609	0.1032	0.0636	0.068
OMEGA excess return	1.0669	1.3277	0.9817	1.0396
(c)				
RVaR 5 %	0.00084	−0.0075	0.0068	0.0096
RVaR 1 %	0.00044	−0.0084	0.0059	0.0050
RCVaR 5 % (V0 = 100)	5.115e–06	0.0245	5.968e–05	5.4426e–05
RCVaR 1 % (V0 = 100)	2.927e–06	0.0534	3.313e–05	3.0109e–05
Rachev	0.6869	0.7496	0.8328	0.6977
L-performance	0.0822	0.0094	0.027	0.0284
OMEGA excess return	0.9587	0.9255	0.9936	0.9953

and RCVaR for (1 and 5 %) focus on extreme values of the historical distribution. Over-all, this yields results that differ from previous analyses, suggesting further evidence of measurement and evaluation errors. However, the results of the second generation analyses are not unanimously conclusive as they vary per sample and according to the ratio under consideration.

3.2.3 The systemic risk issue

Furthermore, while the ratios in Table 3 are more robust and adjusted to risk, none of them can take the systemic risk issue into account. The latter is even more important in that it can be transmitted rapidly from one market to another (as in 2008), and can consequently cripple the other markets, even if they show high performance. Accordingly, it is essential to take the contribution of systemic risk into account when evaluating financial performance, as recommended in the Basel III agreement. Before moving on to examine the systemic risk contribution, we illustrate the level of market linkages during calm and crisis periods while computing the unconditional correlation matrix that we reported in Table 4a, b, c.

	DJGM	DJDM	DJEM	DJIEM	DJIDM
Table 4 Correlation matrix for (a) the whole period, (b) subsample 1 and (c) subsample 2					
(a)					
DJGM	1				
DJDM	0.9957	1			
DJEM	0.8923	0.8663	1		
DJIEM	0.9063	0.9216	0.7234	1	
DJIDM	0.9802	0.9829	0.8597	0.9132	1
(b)					
DJGM	1				
DJDM	0.9986	1			
DJEM	0.8672	0.8447	1		
DJIEM	0.872	0.8805	0.6477	1	
DJIDM	0.9712	0.9723	0.8281	0.8633	1
(c)					
DJGM	1				
DJDM	0.9953	1			
DJEM	0.8987	0.8722	1		
DJIEM	0.9117	0.9279	0.7381	1	
DJIDM	0.9818	0.9849	0.8665	0.9209	1

Accordingly, we note high unconditional bilateral correlation between all of the stock markets as well with the global benchmark, indicating strong linkages between these markets, and suggesting further evidence of systemic risk between these areas. We also note that most correlations (even with Islamic stock markets) increase after the global financial crisis. However, once again, computation of this matrix also requires normality as it is based on the Pearson correlation test. The rejection of normality and leptokurtic excess in our data makes it less reliable, however.

In order to fill the gap, we check for systemic risk by investigating the sensitivity of our indexes to the global markets while carrying out a quantile regression estimation. This framework has a triple advantage. First, it provides a robust measure of the sensitivity of our indexes to the global market while investigating the quantile properties, enabling the high level tail of the distribution of our data to be taken into account. This provides a more robust beta measure from the quantile regression than that employed by the Treynor index. Second, this robust investigation indirectly enables us to evaluate the level of systemic risk between the markets under consideration and the global markets. Third, the quantile regression estimates are used to generate estimates of the $\Delta CoVaR$ measure. We report the main results of the quantile regression in Table 5.

From Table 5, we observe that all parameters are statistically significant. The DIEM is more sensitive to global market variations than the other markets, while the beta is higher than 1 for the two quantile values $\alpha = 1\%$ and $\alpha = 5\%$. This suggests that the DJEM may amplify the global market risk, and it also suggests strong evidence of systemic risk transmission between this market and the benchmark market. We can also note that the standard emerging and developed index for the two quantiles $\alpha = 1\%$ and $\alpha = 5\%$ are more sensitive to the global market than the Islamic indexes. For example, the DJIEM presents a lower value of beta. Overall, this analysis confirms the hypothesis of systemic risk and points

Table 5 Beta estimation with quantile regression method

|		Quantile regression beta	
		b0	b1
y = DJ Emer x = DJ GLO α = 1 %	Parameters	−0.0175	1.4138
	t stat	−13.1839	4.9308
y = DJ Emer x = DJ GLO α = 5 %	Parameters	−0.0102	1.2437
	t stat	−19.1270	20.9152
y = DJ Isla Emer x = DJ GLO α = 1 %	Parameters	−0.0121	0.7803
	t stat	−10.5613	2.3205
y = DJ Isla Emer x = DJ GLO α = 5 %	Parameters	−0.0066	0.8877
	t stat	−17.8851	28.5055
y = DJ Dev x = DJ GLO α = 1 %	Parameters	−0.0019	0.9908
	t stat	−11.532	43.8592
y = DJ Dev x = DJ GLO α = 5 %	Parameters	−0.0012	0.9874
	t stat	−12.9692	61.3165
y = DJ Isla Dev x = DJ GLO α = 1 %	Parameters	−0.0061	0.9411
	t stat	−10.4149	12.8017
y = DJ Isla Dev x = DJ GLO α = 5 %	Parameters	−0.0034	0.9775
	t stat	−15.6965	37.0758

to a risk that financial performance ratio calculations might fail to take this systemic risk into account.

3.2.4 The measure of the systemic risk effect

Hereafter, in order to measure the contribution of systemic risk and identify which market appears to show greater systemic risk, we apply two different non-parametric approaches: (i) the $\Delta CoVaR$ approach and (ii) the marginal expected shortfall. Both require dynamic analyses, but while the former only provides a threshold for losses, the latter enables us to measure the average losses for a given probability. In practice, for both approaches, our estimation covers two levels: $\alpha = 1\%$ and $\alpha = 5\%$. We report the main results of the $\Delta CoVaR$ in Fig. 2a,b, while Fig. 3a, b report those of the marginal expected shortfall.

In practice, we first apply the dynamic $\Delta CoVaR$ for two quantiles $\alpha = 1\%$ and $\alpha = 5\%$ using a sliding window approach for the period from 04/2004 to 03/2014 such as the estimation interval at each time t is $[t - K, t]$ with K the time delay equal to two years as in previous studies. This choice gives us a long enough period for a distribution with sufficient information on the distribution tails. Accordingly, we note that both DJEM and DJIEM indexes show significant sensitivity toward the DJGM index, with the DJEM displaying higher dependency. The latter, considered by the first generation financial performance ratio as the best performing ratio, is the most systemically risky index too. In other words, this index is the source of the greatest systemic risk, involving major losses. This finding illustrates a form of measurement and evaluation error when the notation of this financial asset is provided by standard performance ratios. It is also important

a

b
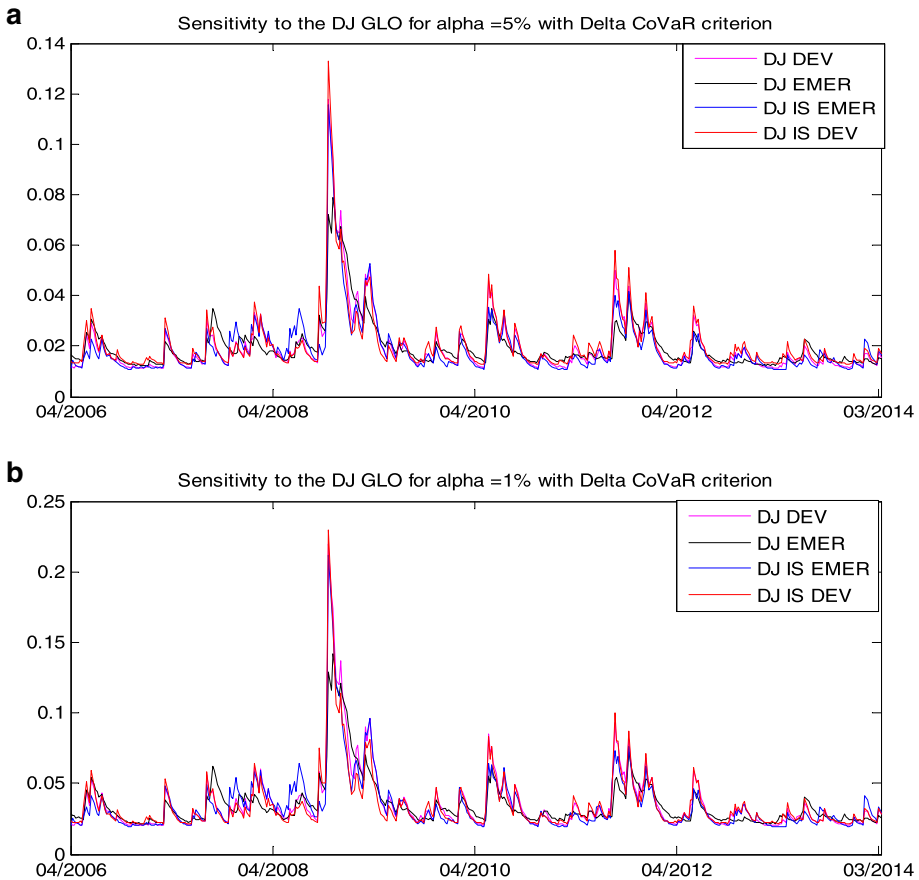
Fig. 2 **a** Delta CoVaR for $\alpha = 5$ %. **b** Delta CoVaR for $\alpha = 1$ %

to note that the systemic risk apparent from Fig. 2a, b reached its maximum level in 2008, which implies that measurement errors are higher in crisis periods than in calm periods.

However, the $\Delta CoVaR$ method, even based on a dynamic approach and quantile regression, only computes the loss threshold. In order to compute expected losses in the mean, we use the marginal expected shortfall (MES) approach and reporte the main results on Fig. 3a, b. From these figures, we also note that the DJEM has the highest level of systemic risk as it is associated with the highest losses. It is also the most closely linked to the DJGM index. This means that, even when considering the evaluation of losses in mean with the MES approach, our findings identify the DJEM index as the most important systemic index. As for the $\Delta CoVaR$ method, the findings from the MSE approach point to further measurement and evaluation stock market errors when considering usual performance ratios. These ratios overnote subprime credits, which, because of the rapid transmission of the systemic risk, led to a major global financial crisis in 2008.

a

b
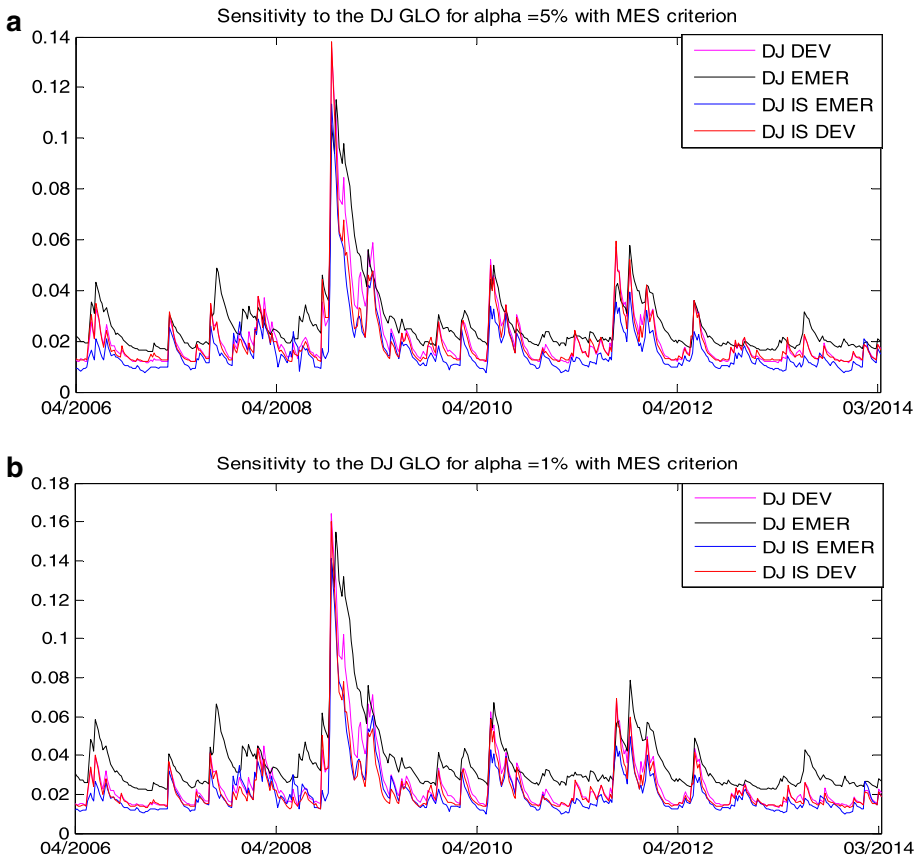
Fig. 3   **a** MES for $\alpha = 5$ %. **b** MES for $\alpha = 1$ %

4 Conclusion

This paper focuses on a hot topic linked to measurement errors. Tackling it from a financial performance perspective, it shows that the use of first generation performance ratios can lead to incorrect conclusions. Indeed, these ratios are conditioned by strong assumptions associated with normality and symmetry, etc. The frequent violation of these hypotheses makes these ratios inappropriate. Accordingly, the recent financial literature identified a second generation of financial performance ratios, which have the advantage of not requiring such hypotheses and of being risk adjusted while taking further extreme losses into account. While these robust ratios improve the performance analysis and provide a more precise result, they are not suitable for taking further evidence of systemic risk into consideration. However, the systemic risk is considerable, especially during crisis periods. From an empirical point of view, the evaluation of systemic risk contribution with two non-parametric approaches over the period 2004–2014 shows that the DIEM index, which is considered as the best performing financial asset, appears to be the major source of systemic risk. In addition, DJIEM, despite being considered as a hedge area, seems to amplify systemic risk. These findings highlight the importance of measurement errors in stock markets due to the application of such financial performance ratios. These errors not only induce major losses and even bankruptcy for

several banks, financial institutions and markets, but also lead to a global financial crisis and a major economic downturn for several developed and emerging economies. A further future extension of this research would be to develop a new financial performance ratio that incorporates the systemic risk dimension.

References

Acharya, V. V., Pedersen, L. H., Philippon, T., Richardson, M. P. (2010). Measuring systemic risk. SSRN: http://ssrn.com/abstract=1573171.

Adrian, T., & Brunnermeier, M. (2011) . CoVaR, Working Paper, National Bureau of Statistics.

Almeida, H., Campello, M., & Galvao, A. F. (2010). Measurement errors in investment equations. *The Review of Financial Studies*, *23*(9), 3279–3328.

Arztner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *3*, 203–228.

Barnett, W. A. (2012). *Getting it wrong: how faulty monetary statistics undermine the fed, the financial system, and the economy*. Cambridge: MIT Press.

Barnett, W. A. (2015). Collaboration with or without coauthorship: Rocket science versus economic science. In M. Szenberg & L. Ramrattan (Eds.), *Intellectual collaborative experiences*. Cambridge: Cambridge University Press.

Barnett, W. A., & Chauvet, M. (2011). How better monetary statistics could have signaled the financial crisis. *Journal of Econometrics*, *161*(1), 6–23.

Barnett, W. A., Chauvet, M., & Tierney, H. L. R. (2009). Measurement error in monetary aggregates: a markov switching factor approach. *Macroeconomic Dynamics*, *13*(S2), 381–412.

Ben Ameur, H., Idi Cheffou, K., Jawadi, F., & Louhichi, W. (2015). Modeling beta changes with three regime threshold market model, *Working Paper presented at the 9th CFE conference*, London, December 12–14, 2015.

Biglova, A., Ortobelli, S., Rachev, S. T., & Stoyanov, S. (2004). Different approaches to risk estimation in portfolio theory. *Journal of Portfolio Management*, *31*, 103–112.

Blackwell, M., Honaker, J., & King, G. (2015). A unified approach to measurement error and missing data: Overview. Sociological Methods & Research.

Brownlees, C. T., & Engle, R. F. (2010). Volatility, correlation and tails for systemic risk measurement. SSRN: http://ssrn.com/abstract=1611229.

Carroll, R., Ruppert, J., & Stefanski, L. A. (1995). *Measurement error in nonlinear models: A modern perspective*. London: Chapman and Hall.

Chang, C. C., Chen, S. S., Chou, R. K., & Hsin, C. W. (2011). Intraday return spillovers and its variations across trading sessions. *Review of Quantitative Finance and Accounting*, *36*, 355–390.

Chen, X., Hong, H., & Nekipelov, D. (2011). Nonlinear models of measurement errors. *Journal of Economic Literature*, *49*(4), 901–937.

Darolles, S., Gouriéroux, C., & Jasiak, J. (2009). L-performance with an application to hedge funds. *Journal of Empirical Finance*, *16*(4), 671–685.

Dowd, K. (2000). Adjusting for risk: An improved sharpe ratio. *International review of Economics and Finance*, *9*(3), 209–222.

Drerup, T., Enke, B., & Von Gaudecker, H. M. (2014). Measurement error in subjective expectations and the empirical content of economic models. Netspar Discussion Paper No. 10/2014-043.

Eling, M., & Schuhmacher, F. (2007). Does the choice of performance measure influence the evaluation of hedge fund? *Journal of Banking and Finance*, *31*(9), 2632–2647.

Fuller, W. (1987). *Measurement Error Models*. New York: Wiley.

Graham, J., & Harvey, C. (1994). Market timing ability and volatility implied in investment newsletters' asset allocation recommendations. Unpublished working paper, National Bureau of Economic Research, Cambridge, MA.

Graham, J. R., & Harvey, C. R. (1996). Market timing ability and volatility implied in investment newsletters' asset allocation recommendations. *Journal of Financial Economics*, *42*, 397–422.

Hausman, J. (2001). Mismeasured variables in econometric analysis: problems from the right and problems from the left. *The Journal of Economic Perspectives*, *15*, 57–67.

Keating, C., & Shadwick, W. (2002). A universal performance measure. *Journal of Performance Measurement*, *6*(3), 59–84.

Martin, D., Rachev, S., & Siboulet, F. (2003). Phi-alpha optimal portfolios and extreme risk management, Wilmott Magazine of Finance, November/2003, pp. 70–83.

Modigliani, F., & Modigliani, L. (1997). Risk-adjusted performance. *Journal of Portfolio Management*, *23*(2), 45–54.

Pflug, G. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In S. Uryasev (Ed.), *Probabilistic constrained optimization: methodology and applications*. Berlin: Kluwer Academic Publishers.

Scaillet, O. (2005). Nonparametric estimation of conditional expected shortfall. *Insurance and Risk Management Journal*, *74*, 639–660.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, *39*(1), 119–138.

Taylor, J. R. (1999). *An introduction to error analysis: The study of uncertainties in physical measurements*. Sausalito: University Science Books.

Treynor, J. (1965). How to rate management of investment funds. *Harvard Business Review*, *43*(1), 63–75.