PRIFYSGOL BANGOR / BANGOR UNIVERSITY

PRIFYSGOL BANGOR UNIVERSITY

Performance ranking (dis)similarities in commodity markets

Zhang, Hanxiong; Auer, Benjamin R.; Vortelinos, Dimitrios I.

Global Finance Journal

DOI:
10.1016/j.gfj.2017.09.001

Published: 11/01/2018

Peer reviewed version

Cyswllt i'r cyhoeddiad / Link to publication

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Zhang, H., Auer, B. R., & Vortelinos, D. I. (2018). Performance ranking (dis)similarities in commodity markets. *Global Finance Journal*, *35*, 115-137. https://doi.org/10.1016/j.gfj.2017.09.001

Hawliau Cyffredinol / General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

02. May. 2024

Performance ranking similarities in commodity markets: A re-examination of recent evidence

Hanxiong Zhang[*], Benjamin R. Auer[†], Dimitrios I. Vortelinos[‡]

May 30, 2016

In this article, we revisit recent evidence indicating that the choice of performance measure appears to be irrelevant for the ranking of investment alternatives in the commodity market. Extending the previous literature by (i) using data of higher frequency, (ii) investigating ranking similarities in different market phases and (iii) considering spot market investments in addition to futures market investments, we provide the following insights into the performance rankings produced by 13 popular performance measures for 24 commodities. First, we find that, in the spot market, ranking differences are somewhat larger than in the futures market, especially when it comes to identifying the best investment opportunities. Second, when using daily instead of monthly data, performance measures modeling reward by average returns still produce similar performance rankings. However, when a higher data frequency is used for metrics measuring reward by higher partial moments, performance rankings are crucially different from the other group of measures. Finally, we find that the degree of ranking (dis)similarity appears to be time-varying. The latter two results can be explained by the two rationales that performance measures are characterized by different degrees of estimation error and that the (generalized) location-and-scale condition – under which all of our performance measures produce identical rankings of risky alternatives – may not hold to a similarly strong extent in different market phases.

Keywords: Performance ranking, commodity investments, data frequency, market phase dependency, location-and-scale condition.

JEL classification: C10, D81, G11, G29.

[*] Lincoln International Business School, University of Lincoln, UK. Contact at hzhang@lincoln.ac.uk, 0044-1522-835639 (tel.).

[†]Corresponding author. University of Leipzig, Department of Finance & CESifo Group Munich, Germany. Contact at auer@wifa.uni-leipzig.de, 0049-341-9733672 (tel.).

[‡]Lincoln International Business School, University of Lincoln, UK. Contact at dvortelinos@lincoln.ac.uk, 0044-1522-835634 (tel.).

1. Introduction

With their seminal studies *Eling and Schuhmacher* (2007) and *Eling* (2008) started an ongoing debate on whether the choice of performance measure matters in the evaluation of asset performance. For a wide variety of investment fund datasets, they document high rank correlations between the Sharpe ratio and several alternative reward-to-risk ratios based on drawdowns, partial moments and the Value-at-Risk. Because this result suggests that investors could prefer simpler performance measures to more complex ones, the findings of these studies are of high practical importance and have quickly stimulated further research.

Zakamouline (2011) reinvestigates the findings of *Eling and Schuhmacher* (2007) for hedge funds by taking a more detailed look at the rankings produced by different performance measures instead of focussing just on rank correlations. By calculating the maximum upgrade, maximum downgrade, mean absolute change, and standard deviation of the change in the rankings, she argues that a high rank correlation coefficient does not necessarily imply almost identical rank orders because there are funds that show substantial changes in ranking if the performance measure is changed from the Sharpe ratio to an alternative measure. *Ornelas et al.* (2012) reinvestigate the findings of *Eling* (2008) for mutual funds and suggest that performance measures do not yield similar rankings if their reward measures are different (e.g., when the mean excess return is replaced by a higher partial moment).[1] *Eling et al.* (2011) argue that, when looking at the Sharpe ratio and several performance measures based on partial moments (i.e., Sortino-Satchell, Farinelli-Tibiletti and Rachev ratios), the choice of performance measure in hedge fund evaluations is irrelevant only when these alternative measures are tailored to a moderate investment style. When they are used to describe aggressive investment styles, rank correlations with the Sharpe ratio decrease significantly. Finally, *Auer and Schuhmacher* (2013) use a selection of more advanced rank

[1] They also show that the rankings of the less frequently used 'manipulation proof performance measure' and the 'appraisal ratio' differ considerably from Sharpe ratio rankings.

correlation measures and find that adequately defined drawdown-based performance measures yield hedge fund rankings that are not too different from those of the Sharpe ratio when investors are primarily interested in picking the best investments and when a sufficiently large return sample is used to calculate performance measure estimates. They also highlight that the rankings are not strictly identical when small return samples are analyzed.

While most studies in this field have concentrated on the rankings of investment funds, the recent contribution of *Auer* (2015a) focuses on commodity markets where the Sharpe ratio has become the dominant measure for evaluating and comparing different commodity trading strategies (see *Erb and Harvey,* 2006; *Gorton and Rouwenhorst,* 2006; *Miffre and Rallis*, 2007; *Szakmary et al.* 2010; *Fuertes et al.* 2010, 2015; *Bianchi et al.,* 2015a, 2015b) and the question of whether one type of investment is superior to another may be answered differently when other performance measures are employed. Using a sample of 24 highly liquid commodity futures, *Auer* (2015a) shows that the Sharpe ratio and its 12 most popular alternatives yield almost identical rankings of investment alternatives in the commodity market. He also shows that his empirical findings are robust in several dimensions (e.g., the futures dataset, the use of equal-length sub-samples and the performance measure parameterization).

Given that the findings of *Eling and Schuhmacher* (2007) and *Eling* (2008) for investment funds have been challenged by several follow-up studies, the goal of our article is to analyze the robustness of *Auer's* (2015a) results in the commodity market. Specifically, we provide three important extensions of his study.

First, in contrast to the majority of studies in the field, we use daily data instead of monthly data.[2] This is because it is well-known that the accuracy of risk measure estimates greatly

[2] *Ornelas et al.* (2010) is the only study using both monthly and daily data to analyse performance ranking similarities. The rest of relevant literature uses monthly data only.

improves with the sample frequency (see *Burghardt and Walls*, 2011; *Frazzini and Pedersen*, 2014). As different kinds of risk measures are characterized by distinct degrees of estimation error in small samples (see *Schuhmacher and Auer*, 2014) and thus react differently to changes in the sample frequency, a change from monthly to daily data may have crucial impact on empirically observable commodity ranking similarities.

Second, we analyze a potential market phase dependency of ranking similarities by dividing our sample into subsamples classified by prevalent market conditions. Previous research has shown that the moments of asset returns (see *Jondeau and Rockinger*, 2003) and correlations between asset returns (see *Ang and Bekaert*, 2002) are not only time-varying but are strongly driven by the general market direction (see *Amira et al.*, 2011). Thus, we might expect that rank correlations between return-based performance measures also vary over time such that we may observe alternating periods of stronger and weaker ranking similarities.

A last contribution of our study is related to the commodity dataset used in the performance evaluation. *Auer* (2015a) focuses on commodity futures returns because they can be easily realized by investors by either trading the futures contracts at almost negligible transaction costs or by investing in exchange traded commodities (ETCs) tracking the futures. We are interested in whether switching to another kind of commodity investment crucially influences the finding of similar performance rankings. Specifically, in addition to futures data, we have a look at the commodity spot market. Here, one could realize spot market returns by physical investment. However, in practice, investors do not choose such a course of action for most commodities (except for precious metals) because they are unable to store the purchased quantities (e.g., natural gas). Instead, investors hoping to directly capture spot market movements invest in certificates and ETCs which are collateralized by holdings in the physical commodity and thus deliver spot returns (minus costs of storage) to the investors.[3]

The remainder of our article is organized as follows. Section 2 reports the main features of our dataset. Section 3 briefly describes the selection of performance measures (and their specifications) used in our study. Section 4 reports our empirical results and verbally summarizes the outcomes of several robustness checks. Section 5 concludes by discussing our empirical results in the context of recent theoretical literature on ranking similarities and by pointing out directions for future research.

2. Data

2.1 Data source and return calculation

This study employs data from the constituents (subindices) of the Standard and Poor's Goldman Sachs Commodity Index (S&P GSCI) for the period from January 7, 2002 to March 31, 2016. This index covers 24 commodities from a wide variety of sectors: six energy products (Brent crude oil, WTI crude oil, gas oil, heating oil, natural gas, unleaded gasoline), two precious metals (gold, silver), five industrial metals (aluminium, copper, lead, nickel, zinc), eight agricultural products (cocoa, coffee, corn, cotton, soybeans, sugar, Chicago wheat, Kansas wheat) and three livestock products (feeder cattle, lean hogs, live cattle). For each commodity, we obtain a total return futures index and a spot index. The futures index measures the returns accrued from investing in liquid fully-collateralized nearby futures, whereas the spot index reflects the performance of a physical commodity investment. The main advantage of the futures index is that it is completely comparable to returns from a regular investment in the S&P 500 (with dividend

[3] Such products are typically not available for perishable commodities (e.g., agricultural products and livestock). Thus, our investment rankings based on spot market data are partially hypothetical. Nonetheless, even the hypothetical rankings for perishable commodities provide important information for comparing the general price developments in different commodity market segments.

reinvestment) or a government bond.[4] This is why these indices have become popular benchmarks for evaluating investment strategies (e.g., momentum trading rules) in commodity futures markets (see *Erb and Harvey,* 2006; *Bianchi et al.,* 2015a) and important tools for identifying the best investment opportunity in a set of given commodity alternatives (see *Auer,* 2015a).

The daily data for the futures and spot indices is obtained from Thomson Reuters Datastream. Following the standard convention (see *Miffre and Rallis,* 2007), we calculate daily log returns as $r_t = (ln\ I_t - ln\ I_{t-1}) \cdot 100$, where I_t is the index value for a given commodity on day t and ln denotes for natural logarithm. Excess returns are computed by subtracting the daily US Treasury bill rate from Ibboston Associates.[5]

2.2 Characteristics of the full sample

Table 1 presents the minimum, maximum, mean, standard deviation, skewness, kurtosis, and the results of the *Jarque and Bera (1987)* test for normality for the commodity excess returns in our sample period.

[Insert Table 1 about here]

In the futures market, the highest daily losses and gains can be observed for the energy, precious metals and industrial metals sectors (which is in line with *Doran and Ronn,* 2008; *Auer*, 2014; *Sévi,* 2015). For example, silver and nickel (natural gas and WTI crude oil) exhibit the most significant losses (gains) of -19.49% and -18.26% (18.77% and 13.34%), respectively. In contrast, the livestock sector is more stable with respect to such drastic outliers. This is partially related to the fact

[4] For details on the index construction (e.g., the procedure used to roll over from one futures contract to the next), see www.goldmansachs.com/what-we-do/securities/products-and-business-groups/products/gcsi.

[5] This series is available in the data library of Kenneth French, i.e., mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

that this sector has been less subject to speculative attacks than the others (see *Auer*, 2015a). Turning to the mean excess returns, we find the highest (lowest) value for copper (natural gas), i.e., 0.04% (-0.14%). With 2.99% (0.92%), standard deviations take their highest (lowest) values for natural gas (feeder cattle). With the exceptions of gas oil, natural gas, coffee, corn and wheat, all excess return distributions are negatively skewed suggesting that there is a higher chance of realizing high negative excess returns than large positive ones. All commodities are characterized by kurtosis values larger than 3 indicating heavier tails and/or stronger peaks than a normal distribution. Given these properties, the null hypothesis of normally distributed excess returns is rejected for all commodities.[6] A brief look at the results for the spot returns yields largely similar results.

2.3 Subsample properties

In order to analyze whether ranking similarities are linked to (i) the data frequency used for performance evaluation or depend on (ii) the market phase in which the performance is measured, we divide our sample into 7 subsamples.

Auer subsample: January 7, 2002 – September 30, 2013. This sample period has been used by *Auer* (2015a). As he has used monthly data and found that the choice of performance measure is largely irrelevant, this subsample will help in analyzing whether this result also holds when using data of higher frequency.

S1 - Argentina crisis subsample: January 7, 2002 – November 30, 2002. In this period, international markets were affected by the Argentine crisis in December 2001, when the government of Argentina declared itself unable to pay its debts (see *Cho et al.,* 2015).

---

[6] These strong deviations from normality are usually the reason why many researchers discard the Sharpe ratio and resort to alternative performance measures (see *Eling and Schuhmacher*, 2007).

S2 - Growth subsample: December 1, 2002 – August 1, 2008. This is a period of economic growth with low inflation, significant international trade and large financial flows in the emerging and developing world (see NBER business cycles, http://www.nber.org/cycles).

S3 - Lehman Brothers crisis subsample: September 1, 2008 – December 7, 2010. This period covers the expansion of the FED and ECB balance sheet because of liquidity issues that seized financial markets following the collapse of Lehman Brothers (see *Cukierman,* 2013).

S4 - EU crisis subsample: December 8, 2010 – April 4, 2011. This period starts with the beginning of the EU debt crisis and ends with the most influential period of the EU debt crisis (see *Cho et al.,* 2015).

S5 - Greek crisis subsample: April 5, 2011 – March 31, 2012. This period covers the peak of the Greek sovereign crisis where the ECB's rate of balance sheet expansion was accelerated by 70.88% per annum (see *Cukierman,* 2013).

S6 - Post crisis subsample: April 1, 2012 – March 31, 2016. This period is characterized by a relaxation of the Greek sovereign crisis and lasts until the end of our sample.

Figure 1 (2) summarizes the descriptive statistics for the commodity futures (spot) excess returns in our subsamples, where, in each subfigure, we concentrate on one descriptive measure (mean, standard deviation, skewness or kurtosis). To understand our specific form of visualization, take the standard deviation as an example. In the corresponding subfigure, we plot the mean of the 24 commodity standard deviations for each subsample and also report their minimum and maximum as a band about the mean value. This way we can visualize the evolution of commodity return standard deviations over our subsamples and also illustrate dispersion across the commodities.

[Insert Figures 1 and 2 about here]

A look at the futures returns in Figure 1 shows that the characteristics of the Auer subsample do not crucially differ from our full sample. As far as our subsamples S1 to S6 are concerned, we see that the EU crisis subsample S4 was the best period for commodity investments in terms of average mean excess returns. It was also accompanied by relatively low average commodity market volatility. In contrast, the Lehman subsample S3 is characterized by rather low (and on average negative) excess returns and significantly higher average volatility. As far as skewness and kurtosis are concerned, we can observe deviations from normality in all of our subsamples. We also see that skewness and kurtosis vary over time, which is in line with time-variation in higher moments documented for other asset classes (see *Jondeau and Rockinger*, 2003; *León et al.*, 2005).

Apart from somewhat lower dispersion in the means, turning to the descriptive statistics of the spot subsamples in Figure 2 reveals results largely similar to Figure 1. Thus, we may expect that the performance evaluation based on our different performance measures yields similar results for futures and spot markets.

3. Methodology

In our study, we focus on a selection of performance measures most popular in research and practice (see *Eling and Schuhmacher,* 2007; *Auer,* 2015a). These 13 reward-to-risk measures (defined in Table 2) mainly share the same reward measure (the mean excess return) in the numerator but differ with respect to the type of risk measure in the denominator. This key difference allows building four main groups of performance measures: 'classic', 'based on drawdowns', 'based on partial moments' and 'based on the Value at Risk'.

[Insert Table 2 about here]

The Sharpe ratio is one of the most popular performance measures in the investment industry (see *Eling and Schuhmacher,* 2007). For a long time, researchers mistakenly believed that the measure has a decision theoretic foundation only under normally distributed returns (see *Auer*, 2015c). This erroneous belief and the technical defects of the measure (i.e., its vulnerability to option based manipulation strategies or distortions introduced by very high or very low returns; see *Goetzmann et al., 2007*; *Schuster and Auer, 2012*; *Auer, 2013*) made researchers and practitioners look for suitable alternatives.

Given the drawback that the standard deviation used as a risk measure in the Sharpe ratio also considers positive deviations from the mean to be risk, straightforward modifications of the Sharpe employ risk measures focusing on worst-case events. A first class of alternative performance measures following this spirit uses drawdowns to quantify risk. We use five measures of this class, namely, the Calmar ratio, Sterling ratio, Burke ratio, Pain ratio and Martin ratio (as defined in *Schuhmacher and Eling,* 2011). While the Calmar ratio quantifies risk by the maximum drawdown, the Sterling and Burke ratios use the mean and the square root of the sum of squares (which puts a stronger emphasis on large losses) of the K largest continuous drawdowns, respectively. We follow the literature standard and set K=5 (see *Auer and Schuhmacher,* 2013; *Auer,* 2015a). Finally, the Pain ratio and Martin ratio quantify risk by calculating the mean of the percentage drops from the previous peak and the square root of the mean of the squared percentage drops from the previous peak, respectively. This allows the duration of drawdowns to be taken into account.

Another class of alternative performance measures uses partial moments to quantify risk (and reward). In contrast to the standard deviation, lower partial moments focus only on negative deviations from a minimal acceptable excess return (which is zero in our definition following *Schuhmacher and Eling, 2012*). Again, we select the four most popular metrics of this kind, namely, the Omega ratio, Sortino ratio, Kappa 3 ratio and Upside potential ratio. The former three measures use the (normalized) lower partial moments of orders one, two and three, respectively, where a higher order models more risk-averse investors (see *Eling et al.,* 2011). While these measures use the mean excess return to quantify reward, the Upside potential ratio modifies the Sortino ratio by replacing the mean excess return with the higher partial moment of order one (focusing on positive deviations of the minimal acceptable return).

Finally, our last group of performance measures covers three ratios based on the Value at Risk (VaR). The VaR used in the excess return on VaR quantifies the possible percentage loss of an investment, which is not exceeded with a given probability $1 - \alpha$ in a certain period. The conditional VaR used in the conditional Sharpe ratio represents the expected percentage loss under the condition that the VaR is exceeded. We estimate both risk measures using historical simulation because it can account for non-normally distributed returns (see *Auer,* 2015b) and because it is the most popular method in the industry (see *Pérignon and Smith*, 2010).[7] Finally, the modified Sharpe ratio makes use of the modified VaR which is an extension of the standard VaR formula for normal distributions that accounts for skewness and kurtosis in the data (see *Eling,* 2008). For all of our VaR-based measures, we set $\alpha = 5\%$, which is a typical value in this context (see *Gilli and Këllezi,* 2006).

Given that negative mean excess returns can negatively influence asset rankings,

---

[7] In a survey of the VaR disclosures of 60 US, Canadian and large international banks over the period from 1996 to 2005, *Pérignon and Smith* (2010) document that 73% of the banks used the historical simulation method.

we apply the *Israelsen* (2005) correction. That is, for each performance measure with reward measure θ and risk measure φ, we do not use θ/φ when ranking commodities but use the measure $\theta/\varphi^{\theta/abs(\theta)}$. For positive excess returns, this formula is identical to the original performance measure θ/φ. In the case of an negative excess return, we get the expression $\theta\varphi$ yielding a correct ranking.

4. Empirical analysis

4.1 Commodity rankings and rank correlations

In our empirical investigation, we take the perspective of a commodity investor who has access to the dataset presented in Section 2 and is interested in identifying the best commodity investments by evaluating a commodity ranking based on historical performance.

For our full sample, Table 3 presents the rankings generated by each of our performance measures as well as a mean ranking across all measures. Subdivided into futures- and spot-based investments, the commodities are ranked from best (rank 1) to worst (rank 24) according to the performance measure realizations.

[Insert Table 3 about here]

We observe the following. First, while rankings in futures markets are widely similar across performance measures which measure reward by means of average returns, the rankings derived from the Upside potential ratio which measures reward by means of higher partial moments are crucially different. For example, while gold is the best investment based on the average return measures (which is in line with *Caporin et al.*, 2015; *Kristjanpoller and Minutolo*, 2015; *O'Connor et al.*, 2015), it only reaches rank 10 based on the Upside potential ratio. Second, in spot markets, we observe a similar outstanding role of the Upside potential ratio. We also see that the differences in rankings produced by the other performance measures are larger in spot data than in futures data. Finally, the mean ranks (across all performance measures) in futures and spot markets are quite different. For example, while gold, soybeans and copper (natural gas, lean hogs and Chicago wheat) are the three best (worst) investments based on futures data, gold, copper and coffee (aluminium, natural gas and nickel) are the best (worst) investments based on spot data. This indicates that the choice of database (futures vs. spot) can have a crucial impact on performance evaluation and observed ranking differences between alternative performance measures.

A similar picture emerges in the results for our subsamples (see Tables A1 to A7 of the appendix). In addition, they provide the following insights. First, when comparing our results based on daily data in the Auer subsample to the original results of *Auer* (2015a) based on monthly data, we find that using a different data frequency influences the ranking outcome. For example, while the Upside potential ratio produces rankings similar to the other measures in monthly data, it does not do so based on daily data. Second, our results show that the relative performance of commodities is time-varying. That is, for example, while, in futures data, the Sharpe ratio ranks gold fifth in our growth subsample, it is ranked 20th in the EU crisis subsample. A closer inspection of the subsample results also reveals that the general market direction (boom or bust) is insufficient to explain time-varying commodity performance. This is partially because commodity prices are no longer simply determined by the demand and supply but are more and more influenced by the continuing financialization of the commodity market (see *Tang and Xiong*, 2012). Finally, not only the relative performance but also the difference in the rankings of different performance measures appears to be time-varying. We focus on this point in the remainder of our analysis by introducing some compact measures of ranking similarity which can be easily compared across subsamples.

We start by computing Kendall's τ and Spearman's ρ rank correlation coefficients which are typical measures for such a purpose (see *Auer and Schuhmacher*, 2013). The main difference between both measures is that, in the calculation of Spearman's ρ, large differences in rankings have higher weights than small differences, whereas Kendalls's τ does not consider the severity of differences but concentrates on whether or not there are differences at all. Large values of both measures indicate strong ranking similarities and a value of one reflects equality of the two rankings entering the rank correlation coefficient.

Because the Sharpe ratio is the simplest of our performance measures it is typically used as the benchmark. This means that rank correlations are calculated between this measure and potential alternatives (see *Zakamouline*, 2011; *Ornelas et al.*, 2012). Table 4 reports these rank correlations for our full sample, where we concentrate on the results based on futures data because, currently, ETCs based on futures are available in a wider variety than ETCs capturing spot market prices. Thus our futures-based results have higher practical relevance (see *Garner*, 2012).[8]

[Insert Table 4 about here]

Table 4 suggests that the ranks delivered by the Sharpe ratio are highly correlated with the ranks of all alternative measures using mean returns to measure reward, which is consistent with the findings of *Auer and Schuhmacher* (2013) and *Auer* (2015a). Correlations vary from 0.9904 (Pain ratio) to 0.998 (Omega ratio, Sortino ratio, Kappa 3 ratio, Excess Return on VaR) and from 0.9348 (Pain ratio) to 0.9855 (Omega ratio, Sortino ratio, Kappa 3 ratio, Excess Return on VaR) according to Spearman's ρ and Kendall's τ, respectively. However,

[8] Detailed results for the spot market data are available from the authors upon request.

the rank correlations between the Sharpe ratio and the Upside potential ratio are significantly lower. Here, we have ρ = 0.2552 and τ = 0.1884 which is in line with the results of *Zakamouline* (2011) and *Ornelas et al.* (2012) for monthly data.

Figure 3 which follows the design of Figures 1 and 2 allows judging the evolution of rank correlations over time. For each of our subsamples, we plot the minimum, maximum and average of all rank correlations with the Sharpe ratio. Because of the distinct role of the Upside potential ratio, we create separate plots including and excluding the correlation values for this performance measure.

[Insert Figure 3 about here]

We can observe that the mean of the rank correlations is crucially affected by the results for the Upside potential ratio. After excluding this measure, all mean rank correlations are close to one (based on both correlation measures). Furthermore, we can observe time-varying rank correlations suggesting that the degree of ranking-similarity varies over time. While the variation of the rank correlations for the measures with mean return reward measures vary only little over time, changes in the correlations for the Upside potential ratio are more significant between periods.

To address the criticism of *Zakamouline* (2011) and *Ornelas et al.* (2012) that high rank correlations do not necessarily imply almost identical rankings, Table 4 presents the descriptive statistics for the differences in ranks (minimum differences, maximum differences, mean absolute differences and standard deviation of absolute differences) between the 12 alternative performance measures and the Sharpe ratio.

Supporting our rank correlation analysis, rankings do not drastically change when an alternative performance measure with mean return reward measure is applied instead of the Sharpe ratio. In the most extreme case, the Calmar ratio, one commodity moves down 2 (minimum of -2) places and another one moves up 2 (maximum of 2) places. However, when we look at the Upside potential ratio these changes are more extreme because, here, on average a commodity moves 6.42 places. This high difference is also reflected by a high standard deviation of absolute differences taking a value of 5.61.

Figure 4 which summarizes the mean absolute differences and the standard deviation of mean absolute differences for our subperiods paints a similar picture. As in Figure 3 we find that the Upside potential ratio contributes significantly to the mean of ranking differences across performance measures and that the magnitude of its deviations from the Sharpe ratio is different depending on the subsample. Interestingly, all performance measures show their strongest deviations from the Sharpe ratio in the most recent subsample.

[Insert Figure 4 about here]

4.2 Focus on best-performing commodities

Because investors are typically interested in identifying the best investments several studies have suggested that a focus on these investments may produce additional insights into ranking similarities (see *Zakamouline*, 2011; *Auer and Schuhmacher*, 2013).[9] Following this suggestion, Table 5 presents the ranking difference statistics for the five commodities with the highest Sharpe ratio in the full sample. That is, we identify the commodities with the highest Sharpe ratios, rank these commodities from best (rank 1) to worst (rank 5) based on the Sharpe ratio and the alternative performance measures and then use these ranks to

[9] Of course, if investors would like to implement momentum strategies which require short-selling the commodities with the poorest past performance (as, for example, in *Miffre and Ralllis*, 2007; *Szakmary et al.,* 2010), the worst investments also become relevant. Therefore, we have extended our analysis to the ranking differences among the worst five commodities. We obtain results similar to the top five commodities suggesting that ranking similarities are stronger for 'extreme performers' than for 'average' performers.

calculate the ranking differences.[10]

[Insert Table 5 about here]

Interestingly, we find that a focus on the best investments drastically reduces the differences in rankings for all alternative performance measures including the upside potential ratio. In the most extreme cases, ranks are changed by only one position. In the majority of cases, the ranks are not changed at all leading to mean absolute differences and standard deviations of absolute differences below 1. A similar picture emerges when repeating this analysis for our subsamples. Figure 5 shows that even though ranking differences do not seriously influence decision making, there are periods (e.g., the EU crisis subsample S4) where the rankings of the alternative measures differ more from the Sharpe ratio rankings than in others.

[Insert Figure 5 about here]

4.3 Some final robustness checks

While the robustness of our results with respect to different datasets and subsamples has already been part of our main analysis, this section covers some additional aspects of parameter choice in the calculation of the performance measures. We follow *Eling and Schuhmacher* (2007) and *Eling* (2008) by varying the significance level α in the VaR-based measures between 1% and 10% in steps of one and the number of drawdowns K in the Sterling and Burke ratios between 1 and 10 in steps of one. However, we find that such changes do not influence our overall results on ranking (dis)similarity between these measures and the Sharpe ratio.

5. Conclusion

In this article, we reinvestigate the recent evidence of *Auer* (2015a) suggesting that

[10] In contrast, *Auer and Schuhmacher* (2013) calculate the differences using the original ranks of the alternative performance measures in the ranking of all investment alternatives.

the choice of performance measure does not crucially affect the rankings of investment alternatives in commodity markets. We analyze the robustness of these findings with respect to (i) data of higher frequency, (ii) subsamples reflecting different market phases and (iii) data selection, i.e., futures versus spot market returns.

Our results show that switching from monthly to daily data leads to increases in ranking differences among our 13 performance measures. Especially the Upside potential ratio, which captures investment reward by means of higher partial moments, generates crucially different rankings than performance metrics using the mean excess returns to measure reward. This holds in both futures and spot market data. In the latter case, ranking differences are larger. Furthermore, we can observe that ranking differences appear to be time-varying. That is, there are periods of larger and smaller differences. For the Upside potential ratio, changes in ranking differences between different periods are more significant than for the other measures.

Even though these results challenge prior studies arguing that in typical empirical applications the choice of performance measure is irrelevant, our findings are still in line with the recently emerged strand of literature developing the theoretical conditions under which different performance measures produce identical rankings of risky alternatives. *Schuhmacher and Eling* (2011, 2012) show that if investment returns fulfill the location and scale (LS) condition of *Sinn* (1983) and *Meyer* (1987), the Sharpe ratio, adequately defined drawdown-based performance measure and certain performance measures based on partial moments, the VaR and other risk quantities yield identical rankings. Given that the LS condition cannot be satisfied in an environment with cross-sectionally different levels of skewness and kurtosis, *Schuhmacher and Auer* (2014) show that these performance measures also yield identical rankings when the generalized LS condition of *Meyer and Rasche* (1992) holds, which allows for cross-sectional differences in skewness and kurtosis.

To relate our results to this literature, we have to consider that, in empirical studies, we are working with small samples while the theoretical literature refers to population properties. Thus, even if the generalized LS condition holds and we have identical population rankings, the rankings in small samples may still be different because of estimation error which is different for each performance measure (see *Schuhmacher and Auer*, 2014). Thus, our detected ranking differences do not challenge the theoretical literature on ranking similarities. To challenge this literature, we would have to show that the generalized LS condition does not hold or at least has weaker empirical support in specific subsamples. This would deliver a perfect explanation for the time-varying nature of ranking similarities that we have detected. Unfortunately, the statistical techniques currently available do not allow testing the generalized LS condition such that more work on adequate statistical methodology is required to answer this question (see *Auer*, 2015c).

While this aspect offers plenty of scope for future theoretical research, our findings also suggest directions for additional empirical research. For example, recent work analyzing commodity market investments has strongly focused on the performance of futures-based momentum, reversal and term structure strategies (see *Fuertes et al.* 2010, 2015; *Bianchi et al.,* 2015a, 2015b). Here, the typical approach is to evaluate different strategies in different subsamples by means of their Sharpe ratios. In light of our results, the identification of the best strategy can yield different outcomes when another performance measure is used instead of the Sharpe ratio. Thus, from a practical perspective, it may be highly interesting to expand our study of individual commodities to a ranking of advanced commodity trading strategies (involving more than one commodity) and their different specifications.

References

Amira, K., Taamouti, A., Tsafack, G. (2011) 'What drives international equity correlations? Volatility or market direction?', *Journal of International Money and Finance*, 30(6), 1234-1263.

Ang, A., Bekaert, G. (2002) 'International asset allocation with regime shifts', *Review of Financial Studies*, 15(4), 1137-1187.

Auer, B. R. (2013) 'The low return distortion of the Sharpe ratio', *Financial Markets and Portfolio Management*, 27(3), 299-306.

Auer, B. R. (2014) 'Could diamonds become an investor's best friend?', *Review of Managerial Science*, 8(3), 351-383.

Auer, B. R. (2015a) 'Does the choice of performance measure influence the evaluation of commodity investments?', *International Review of Financial Analysis*, 38, 142-150.

Auer, B. R. (2015b) 'Extreme value theory, asset ranking and threshold choice: A practical note on VaR estimation', *Journal of Risk*, 18(1), 27-44.

Auer, B. R. (2015c) 'On the role of skewness, kurtosis, and the location and scale condition in a Sharpe ratio performance evaluation setting', *International Journal of Theoretical and Applied Finance*, 18(6), 1550037:1-13.

Auer, B. R., Schuhmacher, F. (2013) 'Robust evidence on the similarity of Sharpe ratio and drawdown-based hedge fund performance rankings', *Journal of International Financial Markets, Institutions and Money*, 24, 153-165.

Bianchi, R. J., Drew, M. E., Fan, J. H. (2015a) 'Combining momentum with reversal in commodity futures', *Journal of Banking and Finance*, 59, 423-444.

Bianchi, R. J., Drew, M. E., Fan, J. H. (2015b) Microscopic momentum in commodity futures', *Griffith University Working Paper No. 2015-10*.

Burghardt, G., Walls, B. (2011) *Managed Futures for Institutional Investors: Analysis and Portfolio Construction*. New Jersey: Wiley.

Caporin, M., Ranaldo, A., Velo, G. (2015) 'Precious metals under the microscope: A high-frequency analysis'. *Quantitative Finance*, 15(5), 743-759.

Cho, S., Hyde, S., Nguyen, N. (2015) 'Time-varying regional and global integration and contagion: Evidence from style portfolios', *International Review of Financial Analysis*, 42, 109-131.

Cukierman, A. (2013) 'Monetary policy and institutions before, during, and after the global financial crisis', *Journal of Financial Stability*, 9(3), 373-384.

Doran, J. S., Ronn, E. I. (2008) 'Computing the market price of volatility risk in the energy commodity markets', *Journal of Banking and Finance*, 32(12), 2541-2552.

Eling, M. (2008) 'Does the measure matter in the mutual fund industry?', *Financial Analysts*

Journal, 64(3), 54-66.

Eling, M., Farinelli, S., Rossello, D., Tibiletti, L. (2011) 'One-size or tailor-made performance ratios for ranking hedge funds?', *Journal of Derivatives and Hedge Funds*, 16(4), 267-277.

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

Erb, C. B., Harvey, C. R. (2006) 'The strategic and tactical value of commodity futures', *Financial Analysts Journal*, 62(2), 69-97

Frazzini, A., Pedersen, L. H. (2014) 'Betting against beta', *Journal of Financial Economics*, 111(1), 1-25.

Fuertes, A.-M., Miffre, J., Fernandez-Perez, A. (2015) 'Commodity strategies based on momentum, term structure, and idiosyncratic volatility', *Journal of Futures Markets*, 35(3), 274-297.

Fuertes, A.-M., Miffre, J., Rallis, G. (2010) 'Tactical allocation in commodity futures markets: Combining momentum and term structure signals', *Journal of Banking and Finance*, 34(10), 2530-2548.

Garner, C. (2012) *A Trader's First Book on Commodities: An Introduction to the World's Fastest Growing Market*. 2 edn. New Jersey: Pearson Education.

Gilli, M., Këllezi, E. (2006). 'An application of extreme value theory for measuring financial risk', *Computational Economics*, 27, 207-228.

Goetzmann, W., Ingersoll, J., Spiegel, M., Welch, I. (2007) 'Portfolio performance manipulation and manipulation-proof performance measures', *Review of Financial Studies*, 20(5), 1503-1546.

Gorton, G., Rouwenhorst, K. G. (2006) 'Facts and fantasies about commodity futures', *Financial Analysts Journal*, 62(2), 47-68.

Israelsen, C. (2005) 'A refinement of the Sharpe ratio and information ratio', *Journal of Asset Management*, 5(6), 423-427.

Jarque, C. M., Bera, A. K. (1987) 'A test for normality of observations and regression residuals', *International Statistical Review / Revue Internationale de Statistique*, 55(2), 163-172.

Jondeau, E., Rockinger, M. (2003) 'Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements', *Journal of Economic Dynamics and Control*, 27(10), 1699-1737.

Kristjanpoller, W., Minutolo, M. C. (2015) 'Gold price volatility: A forecasting approach using the Artificial Neural Network-GARCH model' *Expert Systems with Applications*, 42(20), 7245-7251.

León, Á., Rubio, G., Serna, G. (2005) 'Autoregressive conditional volatility, skewness and kurtosis', *Quarterly Review of Economics and Finance*, 45(4-5), 599-618.

Meyer, J. (1987) 'Two-moment decision models and expected utility maximization', *American Economic Review*, 77(3), 421-430.

Meyer, J., Rasche, R. H. (1992) 'Sufficient conditions for expected utility to imply mean-standard deviation rankings: Empirical evidence concerning the location and scale condition', *Economic Journal*, 102(410), 91-106.

Miffre, J., Rallis, G. (2007) 'Momentum strategies in commodity futures markets', *Journal of Banking and Finance*, 31(6), 1863-1886.

O'Connor, F., Lucey, B., Batten, J., Baur, D. (2015) 'The financial economics of gold - A survey', *International Review of Financial Analysis*, 41, 186-205.

Ornelas, J. R. H., Silva Júnior, A. F., Fernandes, J. L. B. (2012) 'Yes, the choice of performance measure does matter for ranking of US mutual funds', *International Journal of Finance and Economics*, 17(1), 61-72.

Pérignon, C., Smith, D. R. (2010) 'The level and quality of Value-at-Risk disclosure by commercial banks', *Journal of Banking and Finance*, 34(2), 362-377.

Schuhmacher, F., Auer, B. R. (2014) 'Sufficient conditions under which SSD- and MR-efficient sets are identical', *European Journal of Operational Research*, 239(3), 756-763.

Schuhmacher, F., Eling, M. (2011) 'Sufficient conditions for expected utility to imply drawdown-based performance rankings', *Journal of Banking and Finance*, 35(9), 2311-2318.

Schuhmacher, F., Eling, M. (2012) 'A decision-theoretic foundation for reward-to-risk performance measures', *Journal of Banking and Finance*, 36(7), 2077-2082.

Schuster, M., Auer, B. R. (2012) 'A note on empirical Sharpe ratio dynamics', *Economics Letters*, 116(1), 124-128.

Sévi, B. (2015) 'Explaining the convenience yield in the WTI crude oil market using realized volatility and jumps', *Economic Modelling*, 44, 243-251.

Sinn, H. W. (1983) *Economic Decisions Under Uncertainty*. Heidelberg: Physica-Verlag.

Szakmary, A. C., Shen, Q., Sharma, S. C. (2010) 'Trend-following trading strategies in commodity futures: A re-examination', *Journal of Banking and Finance*, 34(2), 409-426.

Tang, K., Xiong, W. (2012) 'Index investment and the financialization of commodities', *Financial Analysts Journal*, 68(5), 54-74.

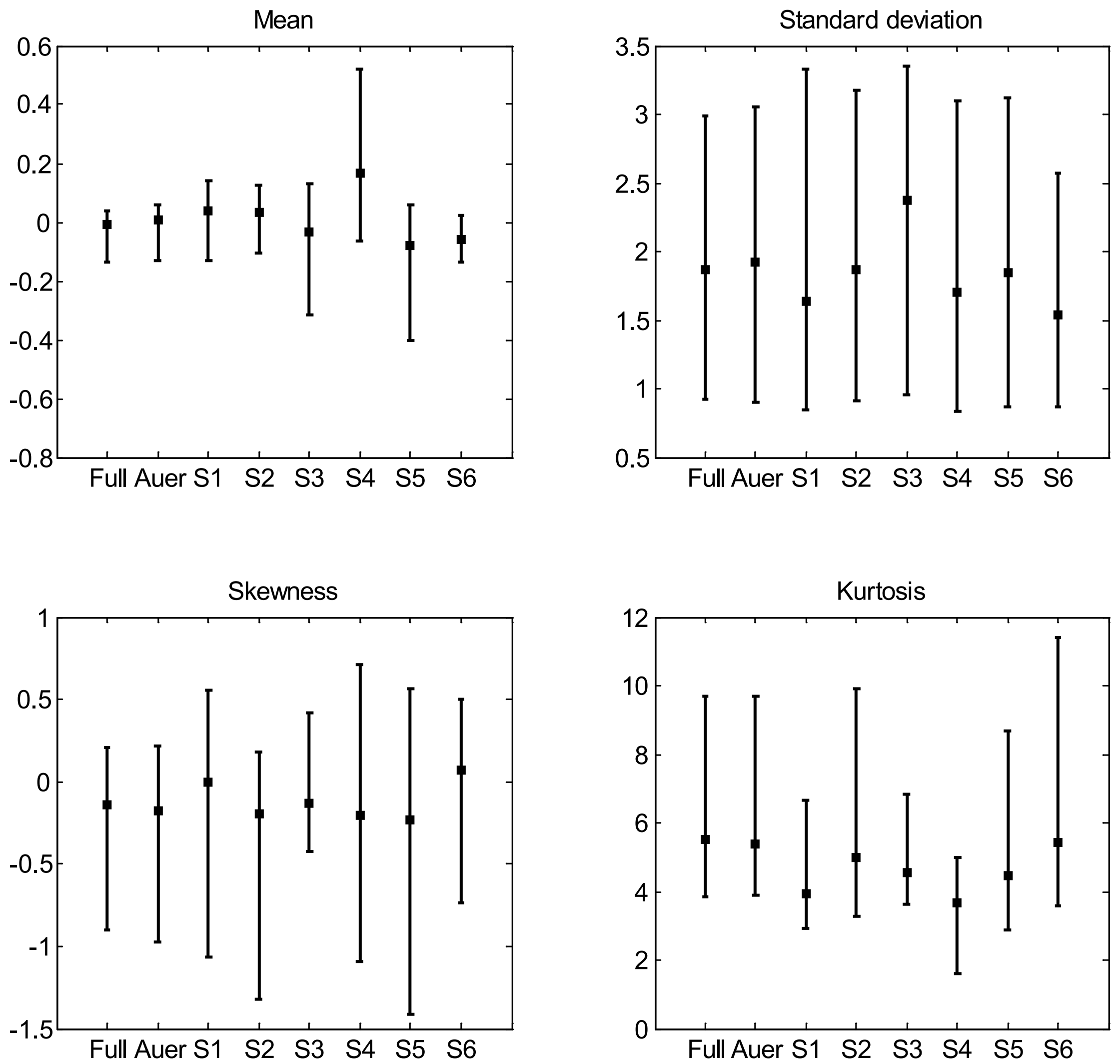
Zakamouline, V. (2011) 'The performance measure you choose influences the evaluation of hedge funds', *Journal of Performance Measurment* 15(3), 48-64.

Table 1: Descriptive statistics (full sample, futures and spot)

	Futures							Spot						
	Min	Max	Mean	SD	Skew	Kurt	JB	Min	Max	Mean	SD	Skew	Kurt	JB
Energy														
Crude oil (Brent)	-10.479	12.881	0.002	2.080	-0.112	6.021	13.690	-10.479	12.879	0.013	2.075	-0.069	6.123	14.580
Crude oil (WTI)	-13.065	13.341	-0.025	2.251	-0.141	5.753	11.430	-13.065	13.341	0.011	2.243	-0.108	5.872	12.380
Gas oil	-9.651	10.732	0.009	1.899	0.044	5.105	6.628	-9.657	10.732	0.015	1.889	0.062	5.113	6.684
Heating oil	-9.680	10.067	-0.002	2.084	-0.001	4.820	4.945	-9.678	10.068	0.015	2.082	-0.012	4.848	5.097
Natural gas	-14.641	18.768	-0.137	2.989	0.205	4.878	5.512	-14.645	18.760	-0.009	3.033	0.247	5.002	6.346
Unleaded gasoline	-11.369	12.972	0.017	2.305	-0.200	5.176	7.304	-11.181	12.971	0.018	2.320	-0.154	5.206	7.406
Precious metals														
Gold	-9.810	8.584	0.034	1.196	-0.391	7.818	35.560	-9.811	8.583	0.037	1.198	-0.391	7.799	35.280
Silver	-19.489	12.469	0.025	2.122	-0.902	9.705	71.940	-19.489	12.470	0.028	2.121	-0.901	9.700	71.850
Industrial metals														
Aluminium	-8.253	5.927	-0.018	1.409	-0.283	5.235	7.934	-8.272	5.926	-0.003	1.409	-0.283	5.236	7.944
Copper	-10.397	11.900	0.040	1.801	-0.147	7.004	24.060	-10.382	11.898	0.027	1.802	-0.146	7.017	24.210
Lead	-13.112	12.835	0.031	2.157	-0.216	6.209	15.650	-13.033	12.832	0.028	2.157	-0.216	6.200	15.560
Nickel	-18.256	13.158	0.011	2.402	-0.159	6.310	16.500	-18.224	13.154	0.004	2.402	-0.157	6.302	16.420
Zinc	-11.133	9.853	0.004	1.989	-0.149	5.563	9.935	-11.133	9.926	0.016	1.989	-0.158	5.576	10.050
Agriculture														
Cocoa	-10.014	9.098	0.010	1.843	-0.327	5.884	13.050	-10.006	8.984	0.016	1.844	-0.335	5.880	13.050
Coffee	-11.255	12.080	-0.025	2.045	0.156	5.008	6.160	-11.258	12.050	0.022	2.047	0.138	4.966	5.886
Corn	-8.128	8.670	-0.028	1.801	0.067	5.008	6.046	-8.124	8.663	0.009	1.817	0.047	4.925	5.546
Cotton	-7.123	6.939	-0.022	1.738	-0.083	4.231	2.301	-7.130	6.940	0.007	1.763	-0.081	4.192	2.161
Soybeans	-7.341	6.431	0.035	1.563	-0.220	5.058	6.608	-7.342	6.427	0.016	1.592	-0.246	4.960	6.096
Sugar	-12.369	8.553	-0.006	2.053	-0.254	5.058	6.708	-12.368	8.556	0.014	2.074	-0.248	5.048	6.623
Wheat (Chicago)	-9.973	8.793	-0.041	2.012	0.078	4.726	4.485	-9.972	8.790	0.007	2.016	0.062	4.710	4.387
Wheat (Kansas)	-8.991	8.097	-0.017	1.831	0.065	4.617	3.927	-8.994	8.096	0.009	1.830	0.067	4.625	3.968
Livestock														
Feeder cattle	-5.997	4.255	0.004	0.923	-0.235	4.406	3.282	-6.004	4.251	0.012	0.934	-0.250	4.427	3.412
Lean hogs	-6.409	5.721	-0.054	1.490	-0.048	3.860	1.118	-6.620	7.331	0.006	1.624	0.058	3.958	1.388
Live cattle	-6.359	3.694	-0.006	0.938	-0.185	4.724	4.641	-6.363	3.682	0.011	0.964	-0.200	4.554	3.843

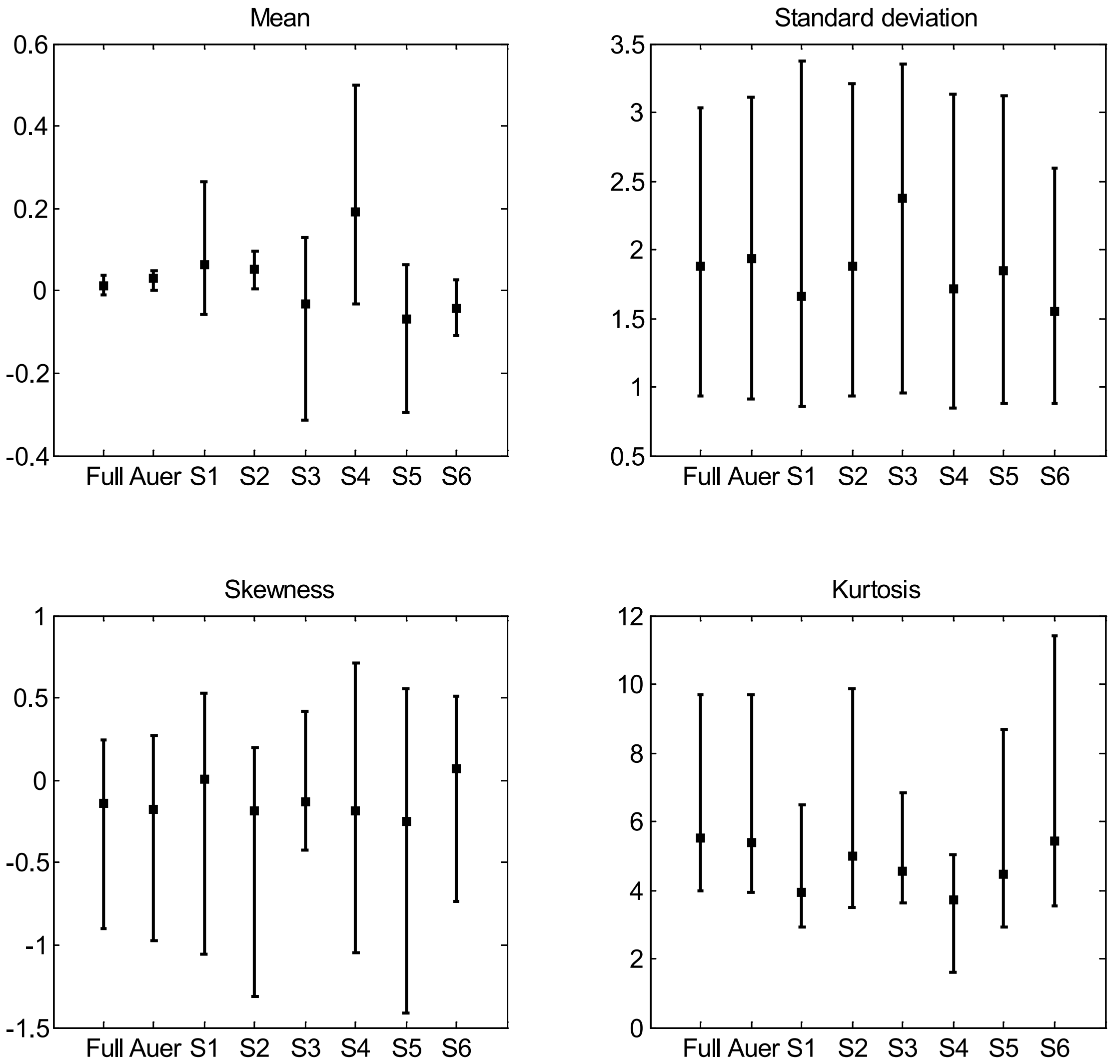
For the period from January 7, 2002 to March 31, 2016, this table reports the minimum, maximum, mean, standard deviation, skewness, kurtosis and Jarque-Berra (JB) test statistic (which, for better visualization has been divided by 100) for the daily excess returns of futures and spot commodity subindices of the S&P GSCI. The returns are given in percent. All JB test statistics are significant at the 1% level.

Figure 1: Descriptive statistics (subsamples, futures)



For our 24 commodity futures indices, this figure illustrates the cross-sectional averages (bold dots) of the sample means, standard deviations, skewness values and kurtosis values within each of our sample specifications. Furthermore, the highest and lowest realizations of these metrics are represented by a band around each average. Sample abbreviations are used as follows: Full: January 7, 2002 – March 31, 2016, Auer: January 7, 2002 – September 30, 2013, S1: January 7, 2002 – November 30, 2002, S2: December 1, 2002 – August 1, 2008, S3: September 1, 2008 – December 7, 2010, S4: December 8, 2010 – April 4, 2011, S5: April 5, 2011 – March 31, 2012, S6: April 1, 2012 – March31, 2016.

Figure 2: Descriptive statistics (subsamples, spot)



For our 24 commodity spot indices, this figure illustrates the cross-sectional averages (bold dots) of the sample means, standard deviations, skewness values and kurtosis values within each of our subsample specifications. Furthermore, the highest and lowest realizations of these metrics are represented by a band around each average. Sample abbreviations are used as follows: Full: January 7, 2002 – March 31, 2016, Auer: January 7, 2002 – September 30, 2013, S1: January 7, 2002 – November 30, 2002, S2: December 1, 2002 – August 1, 2008, S3: September 1, 2008 – December 7, 2010, S4: December 8, 2010 – April 4, 2011, S5: April 5, 2011 – March 31, 2012, S6: April 1, 2012 – March31, 2016.

Table 2: Performance measures

No.	Performance measure	Reward measure	Risk measure
Classic			
(1)	Sharpe ratio	μ	σ
Based on drawdowns			
(2)	Calmar ratio	μ	MDD
(3)	Sterling ratio	μ	$K^{-1}\Sigma_{k=1}^{K} CDD_k$
(4)	Burke ratio	μ	$[\Sigma_{k=1}^{K} CDD_k^2]^{1/2}$
(5)	Pain ratio	μ	$T^{-1}\Sigma_{t=1}^{T} DDP_t$
(6)	Martin ratio	μ	$[T^{-1}\Sigma_{t=1}^{T} DDP_t^2]^{1/2}$
Based on partial moments			
(7)	Omega ratio	μ	LPM_1
(8)	Sortino ratio	μ	$LPM_2^{1/2}$
(9)	Kappa 3 ratio	μ	$LPM_3^{1/3}$
(10)	Upside potential ratio	HPM_1	$LPM_2^{1/2}$
Based on the Value at Risk			
(11)	Excess return on Value at Risk	μ	VaR_α
(12)	Conditional Sharpe ratio	μ	$CVaR_\alpha$
(13)	Modified Sharpe ratio	μ	$MVaR_\alpha$

This table summarizes the reward-to-risk ratios applied in our study. $\mu = T^{-1}(R_1 + \cdots + R_T)$ and $\sigma = [T^{-1}\{(R_1 - \mu)^2 + \cdots + (R_T - \mu)^2\}]^{1/2}$ are the mean and the standard deviation of the excess returns $R_t = r_t - r_{ft}$, $t = 1, \ldots, T$ of a given commodity, where r_t is the daily log return and r_{ft} is the corresponding risk-free rate. MDD denotes the maximum drawdown (the largest negative cumulative excess return), CDD_k a continuous drawdown (the k-th largest negative cumulative excess return that is not interrupted by a positive excess return) and DDP_t the drawdown from a previous peak (a negative cumulative excess return from the previous peak). K is the number of continuous drawdowns incorporated in the calculation. The signs of the drawdowns are dropped to generate positive risk measures. $HPM_m = T^{-1}\sum_{t=1}^{T} max\,(R_t, 0)^m$ and $LPM_m = T^{-1}\sum_{t=1}^{T} max\,(-R_t, 0)^m$ are higher and lower partial moments of order m. VaR_α is the (historical simulation) Value at Risk, i.e. the α-quantile of the excess return distribution. The (historical simulation) conditional VaR is estimated as $CVaR_\alpha = B^{-1}\Sigma_{R_t \le -VaR_\alpha} - R_t$, where B is the number of excess returns fulfilling the summation condition. The modified VaR is estimated as $MVaR_\alpha = -[\mu + \sigma\{z_\alpha + (z_\alpha^2 - 1)\gamma/6 + (z_\alpha^3 - 3z_\alpha)\kappa/24 - (2z_\alpha^3 - 5z_\alpha)\gamma^2/36\}]$, where z_α is the α-quantile of the standard normal distribution and γ and κ denote skewness and excess kurtosis of the excess return distribution, respectively.

Table 3: Commodity rankings (full sample, futures and spot)

	Futures														Spot													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy																												
Crude Oil (Brent)	12	12	12	12	12	12	12	12	12	18	12	12	12	12.5	15	17	13	13	13	14	15	15	15	20	15	15	15	15.0
Crude Oil (WTI)	21	21	21	21	21	21	21	21	21	21	21	21	21	21.0	17	18	17	18	16	16	17	17	18	16	18	18	17	17.2
Gas Oil	9	10	10	10	9	9	9	8	8	4	9	8	8	8.5	11	15	15	15	12	12	11	11	9	10	11	9	9	11.5
Heating Oil	13	14	13	13	13	13	13	13	13	3	13	13	13	12.3	13	13	12	14	11	11	13	13	13	7	13	12	13	12.2
Natural Gas	24	24	24	24	24	24	24	24	24	23	24	24	24	23.9	24	24	24	24	24	24	24	24	24	5	24	24	24	22.5
Unleaded Gasoline	6	7	6	6	6	6	6	6	6	7	6	6	6	6.2	12	10	14	12	8	9	12	12	12	13	12	13	12	11.6
Precious metals																												
Gold	1	1	1	1	1	1	1	1	1	10	1	1	1	1.7	1	1	1	1	1	1	1	1	1	14	1	1	1	2.0
Silver	5	5	5	5	4	5	5	5	5	24	5	5	5	6.4	3	6	5	5	5	5	3	4	5	24	3	5	4	5.9
Industrial metals																												
Aluminium	16	16	16	16	16	16	16	16	16	20	16	16	16	16.3	23	23	23	23	23	23	23	23	23	18	23	23	23	22.6
Copper	3	3	3	3	2	3	2	3	3	13	2	3	2	3.5	2	5	3	3	2	2	2	2	2	22	2	2	2	3.9
Lead	4	4	4	4	5	4	4	4	4	14	4	4	4	4.8	4	9	4	4	7	6	4	3	4	21	4	4	3	5.9
Nickel	8	9	7	7	10	10	8	9	9	9	8	9	9	8.6	22	22	22	22	22	22	22	22	22	17	22	22	22	21.6
Zinc	11	11	11	11	11	11	11	11	11	16	11	11	11	11.4	10	12	8	7	15	15	10	10	10	19	10	10	10	11.2
Agriculture																												
Cocoa	7	6	8	8	7	7	7	7	7	19	7	7	7	8.0	9	3	10	9	9	8	9	9	11	23	9	11	11	10.1
Coffee	20	19	19	19	18	18	20	20	19	8	20	19	19	18.3	7	8	2	2	6	7	7	7	6	3	6	7	6	5.7
Corn	19	18	20	20	19	19	19	19	20	15	19	20	20	19.0	16	14	19	19	17	17	16	16	16	8	17	16	16	15.9
Cotton	18	20	18	18	20	20	18	18	18	17	18	18	18	18.4	19	21	18	17	21	21	19	19	19	12	19	19	19	18.7
Soybeans	2	2	2	2	3	2	3	2	2	1	3	2	3	2.2	8	7	9	8	10	10	8	8	8	15	8	8	8	8.8
Sugar	15	15	15	15	15	15	15	15	15	11	15	15	15	14.7	14	11	11	11	14	13	14	14	14	11	14	14	14	13.0
Wheat (Chicago)	23	22	22	22	22	22	22	22	23	6	22	22	22	20.9	20	19	20	20	20	20	20	20	20	2	20	20	20	18.5
Wheat (Kansas)	17	17	17	17	17	17	17	17	17	2	17	17	17	15.8	18	16	16	16	19	19	18	18	17	1	16	17	18	16.1
Livestock																												
Feeder Cattle	10	8	9	9	8	8	10	10	10	5	10	10	10	9.0	5	4	6	6	3	3	5	5	3	9	5	3	5	4.8
Lean Hogs	22	23	23	23	23	23	23	23	22	22	23	23	23	22.8	21	20	21	21	18	18	21	21	21	4	21	21	21	19.2
Live Cattle	14	13	14	14	14	14	14	14	14	12	14	14	14	13.8	6	2	7	10	4	4	6	6	7	6	7	6	7	6.0

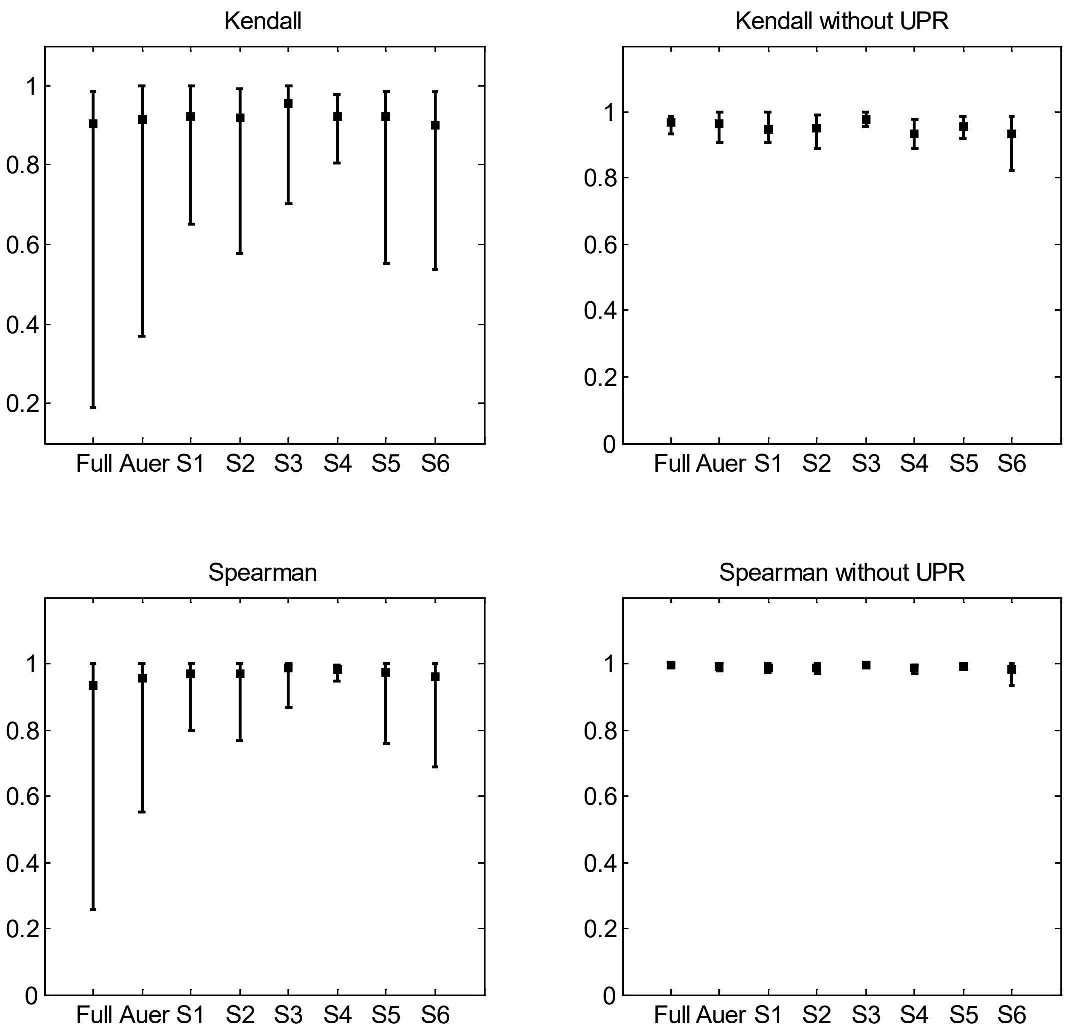
For the period from January 7, 2002 to March 31, 2016 and our different performance measures (as numbered in Table 2), this table reports the rankings of our 24 futures-based and our 24 spot-based commodity investments, where 1 (24) resembles the best (worst) investment performance. For each investment subgroup, the last column reports the mean ranking across all performance measures.

Table 4: Rank correlations and ranking differences (full sample, futures)

	Rank correlations		Ranking differences			
	τ	ρ	Min	Max	MAD	SDAD
(2)	0.9493	0.9922	-2	2	0.5833	0.6539
(3)	0.9710	0.9965	-1	1	0.3333	0.4815
(4)	0.9710	0.9965	-1	1	0.3333	0.4815
(5)	0.9348	0.9904	-2	2	0.5833	0.7755
(6)	0.9493	0.9922	-2	2	0.4167	0.7755
(7)	0.9855	0.9983	-1	1	0.1667	0.3807
(8)	0.9855	0.9983	-1	1	0.1667	0.3807
(9)	0.9855	0.9983	-1	1	0.1667	0.3807
(10)	0.1884	0.2552	-17	19	6.4167	5.6099
(11)	0.9855	0.9983	-1	1	0.1667	0.3807
(12)	0.9783	0.9974	-1	1	0.2500	0.4423
(13)	0.9710	0.9965	-1	1	0.3333	0.4815

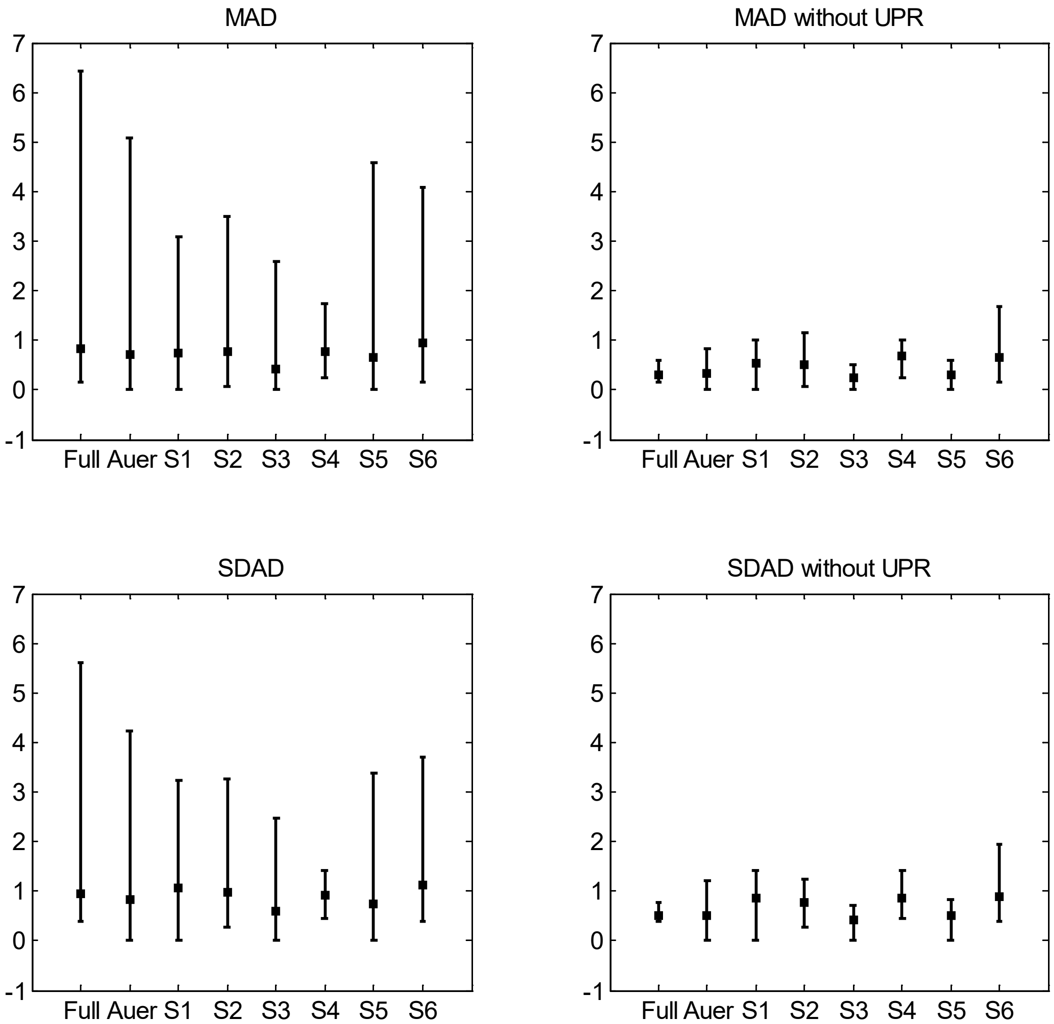
Using the futures-based rankings of Table 3, this table presents the Kendall's(τ) and Spearman's(ρ) rank correlations between the Sharpe ratio and our 12 alternative performance measures. All rank correlation coefficients are significant at the 1% level with the exception of the Upside potential ratio which is insignificant. Furthermore, it provides descriptive statistics of the ranking differences between our 12 alternative performance measure and the Sharpe ratio. These include the minima (Min) and maxima (Max) of the differences as well as the mean absolute difference (MAD) and the standard deviation of absolute differences (SDAD).

Figure 3: Rank correlations (subsamples, futures)



Similar to the visualization in Figures 1 and 2, for each of our subsamples, the left side of this figure presents the means, minima and maxima of Kendal's and Spearman's rank correlation coefficients between our 12 alternative performance measures and the Sharpe ratio. The right side shows the results when the correlation values for the Upside potential ratio (UPR) are excluded.

Figure 4: Ranking differences (subsamples, futures)



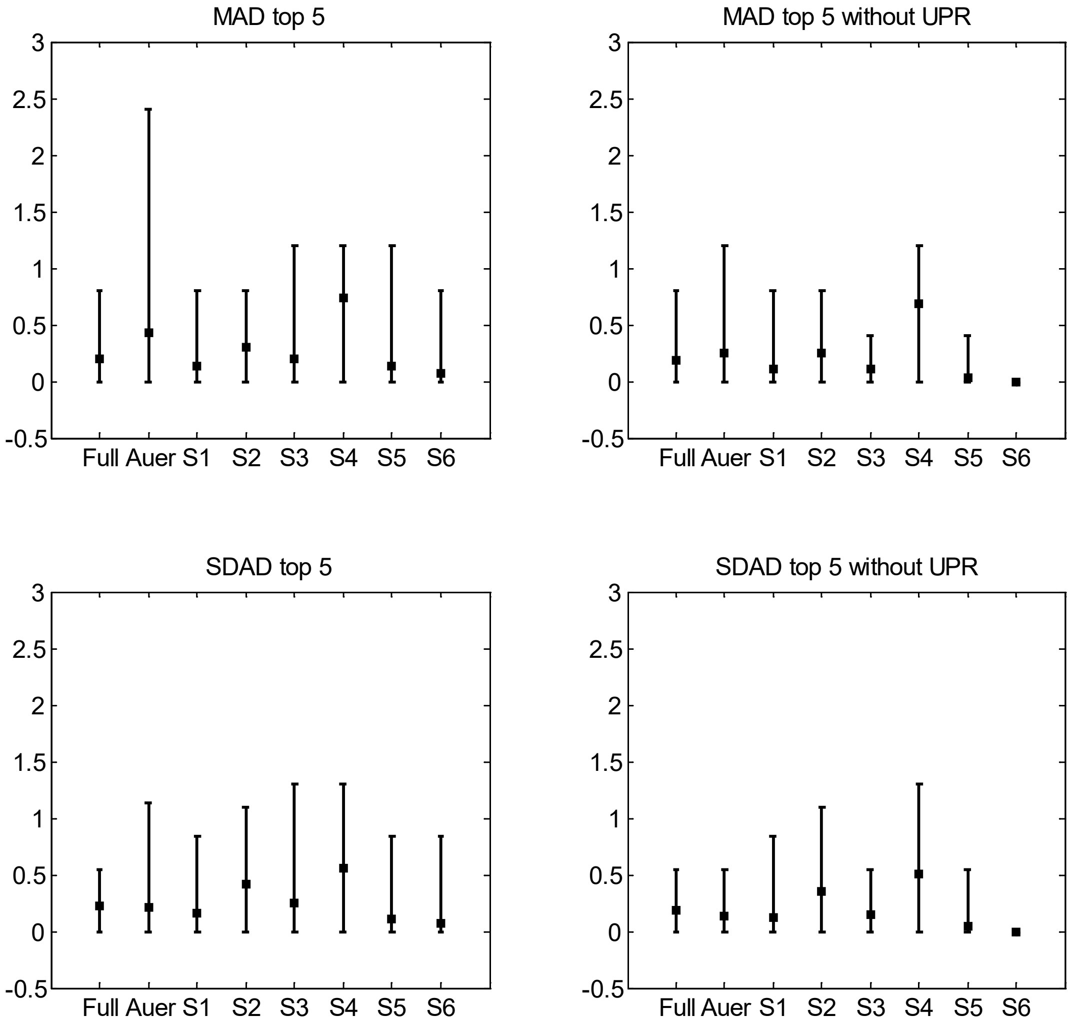
Similar to the visualization in Figures 1 and 2, for each of our subsamples, the left side of this figure presents the means, minima and maxima of mean absolute differences (MAD) and standard deviations of absolute differences (SDAD) between the rankings of our 12 alternative performance measures and the Sharpe ratio. The right side shows the results, when the MAD and SDAD values for the Upside Potential ratio (UPR) are excluded.

Table 5: Ranking differences for top five investments (full sample, futures)

	Rank correlations		Ranking differences			
	τ	ρ	Min	Max	MAD	SDAD
(2)	1.0	1.0	0	0	0.00	0.00
(3)	1.0	1.0	0	0	0.00	0.00
(4)	1.0	1.0	0	0	0.00	0.00
(5)	0.6	0.8	-1	1	0.80	0.45
(6)	1.0	1.0	0	0	0.00	0.00
(7)	0.8	0.9	-1	1	0.40	0.55
(8)	1.0	1.0	0	0	0.00	0.00
(9)	1.0	1.0	0	0	0.00	0.00
(10)	0.8	0.9	-1	1	0.40	0.55
(11)	0.8	0.9	-1	1	0.40	0.55
(12)	1.0	1.0	0	0	0.00	0.00
(13)	0.8	0.9	-1	1	0.40	0.55

Similar to Table 4, this table reports ranking difference statistics but concentrates on the five commodities with the highest Sharpe ratios. That is, we identify the commodities with the highest Sharpe ratios, use the alternative performance measures and the Sharpe ratio to rank these assets from 1 to 5 and then calculate the ranking differences.

Figure 5: Ranking differences for top five investments (subsamples, futures)



This figure is similar to Figure 4. While figure 4 includes all commodities, this figure focuses on the differences for the top five commodities with the highest Sharpe ratios.

Table A1: Commodity rankings (Auer subsample)

	Futures														Spot													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy																												
Crude Oil (Brent)	5	8	4	4	8	8	5	5	5	6	5	5	5	5.6	4	9	2	2	4	4	4	4	4	12	4	5	4	4.8
Crude Oil (WTI)	11	13	11	11	12	13	11	11	11	14	11	11	11	11.6	6	10	6	6	7	7	7	6	6	14	6	6	6	7.2
Gas Oil	4	5	6	6	7	7	4	4	4	1	4	4	4	4.6	2	5	5	4	5	6	2	2	2	3	3	2	2	3.3
Heating Oil	9	9	10	10	9	9	9	9	9	2	9	9	9	8.6	5	7	4	5	6	5	5	5	3	1	5	3	5	4.5
Natural Gas	24	24	24	24	24	24	24	24	24	22	24	24	24	23.8	23	23	23	23	23	23	23	23	23	5	23	23	23	21.6
Unleaded Gasoline	7	6	7	7	5	5	8	7	6	5	7	6	7	6.4	10	11	11	11	8	8	11	10	9	13	10	10	9	10.1
Precious metals																												
Gold	1	1	2	2	1	1	1	1	1	13	1	1	1	2.1	1	1	1	1	1	1	1	1	1	18	1	1	1	2.3
Silver	8	4	8	8	4	4	7	8	8	24	8	8	8	8.2	7	2	8	8	3	3	6	7	10	24	7	7	7	7.6
Industrial metals																												
Aluminium	18	18	18	18	18	18	18	18	18	21	18	18	18	18.2	24	24	24	24	24	24	24	24	24	22	24	24	24	23.8
Copper	2	3	3	3	2	2	2	2	2	8	2	2	2	2.7	3	6	3	3	2	2	3	3	5	19	2	4	3	4.5
Lead	6	7	5	5	6	6	6	6	7	12	6	7	6	6.5	11	12	7	7	11	11	9	11	11	20	11	11	10	10.9
Nickel	10	10	9	9	10	10	10	10	10	7	10	10	10	9.6	20	21	19	19	22	22	20	20	20	17	20	20	20	20.0
Zinc	14	14	14	14	14	14	14	14	14	19	14	14	14	14.4	17	20	18	18	20	21	17	18	19	21	18	19	18	18.8
Agriculture																												
Cocoa	12	11	13	13	11	11	12	12	12	20	12	12	12	12.5	21	15	21	21	19	19	21	21	21	23	21	21	21	20.4
Coffee	21	21	21	21	20	20	21	21	21	17	21	21	21	20.5	15	16	12	12	13	13	15	15	15	7	14	14	14	13.5
Corn	20	19	20	20	19	19	20	20	20	15	20	20	20	19.4	19	17	20	20	16	16	18	19	18	9	19	18	19	17.5
Cotton	19	20	19	19	21	21	19	19	19	16	19	19	19	19.2	14	18	14	13	18	18	14	14	14	11	15	15	15	14.8
Soybeans	3	2	1	1	3	3	3	3	3	3	3	3	3	2.6	8	4	9	9	12	12	8	8	7	15	8	9	8	9.0
Sugar	13	12	12	12	13	12	13	13	13	11	13	13	13	12.5	16	13	17	16	14	14	16	16	17	16	16	16	16	15.6
Wheat (Chicago)	22	22	22	22	22	22	22	22	22	10	22	22	22	21.1	18	19	15	15	17	17	19	17	16	4	17	17	17	16.0
Wheat (Kansas)	16	16	16	16	16	16	16	16	16	4	16	16	16	15.1	13	14	10	10	15	15	13	13	13	2	13	13	13	12.1
Livestock																												
Feeder Cattle	15	15	15	15	15	15	15	15	15	9	15	15	15	14.5	9	8	13	14	9	9	10	9	8	8	9	8	11	9.6
Lean Hogs	23	23	23	23	23	23	23	23	23	23	23	23	23	23.0	22	22	22	22	21	20	22	22	22	6	22	22	22	20.5
Live Cattle	17	17	17	17	17	17	17	17	17	18	17	17	17	17.1	12	3	16	17	10	10	12	12	12	10	12	12	12	11.5

For the period from January 7th, 2002 to September 30th, 2013 and our different performance measures (as numbered in Table 2), this table reports the rankings of our 24 futures-based and our 24 spot-based commodity investments, where 1 (24) resembles the best (worst) investment performance. For each investment subgroup, the last column reports the mean ranking across all performance measures.

Table A2: Commodity rankings (S1 - Argentina crisis subsample)

	Futures														Spot													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy																												
Crude Oil (Brent)	6	6	6	6	7	6	6	5	5	11	6	6	5	6.2	15	12	12	12	13	12	15	15	14	16	16	14	14	13.8
Crude Oil (WTI)	7	5	5	4	5	4	7	6	6	10	7	5	6	5.9	10	11	10	10	9	9	10	10	10	13	14	10	10	10.5
Gas Oil	11	9	8	8	9	9	11	11	10	8	11	10	10	9.6	9	9	9	8	6	7	9	9	8	7	9	8	9	8.2
Heating Oil	10	7	9	9	8	7	8	10	11	14	10	11	11	9.6	7	7	5	5	4	4	7	8	9	14	8	9	8	7.3
Natural Gas	8	8	7	7	10	10	9	8	8	5	8	7	8	7.9	3	3	1	1	3	3	3	3	3	3	3	3	3	2.7
Unleaded Gasoline	13	12	13	13	12	12	13	13	13	15	13	13	13	12.9	17	15	15	15	15	15	17	17	17	15	17	17	17	16.1
Precious metals																												
Gold	3	3	3	3	4	3	3	3	3	9	3	3	3	3.5	6	6	8	9	7	6	6	6	7	12	5	7	7	7.1
Silver	21	20	21	21	20	20	21	21	21	24	21	21	21	21.0	21	21	21	21	21	21	21	21	22	24	21	22	21	21.4
Industrial metals																												
Aluminium	19	19	19	19	19	19	19	19	19	18	19	19	19	18.9	19	19	19	19	19	19	19	19	19	18	19	19	19	18.9
Copper	15	16	16	16	16	16	16	15	15	7	15	15	15	14.8	16	16	17	17	16	16	16	16	16	8	15	16	16	15.5
Lead	23	23	23	23	23	23	23	23	23	22	23	23	23	22.9	23	24	23	23	24	24	24	24	24	23	23	24	24	23.6
Nickel	4	4	4	5	6	5	5	4	4	4	4	4	4	4.4	11	10	11	11	12	11	12	12	11	10	11	11	12	11.2
Zinc	22	22	22	22	22	22	22	22	22	21	22	22	22	21.9	20	20	20	20	20	20	20	20	20	21	20	20	20	20.1
Agriculture																												
Cocoa	5	11	11	11	3	8	4	7	9	19	5	9	7	8.4	13	14	14	14	10	13	13	14	15	22	12	15	15	14.2
Coffee	14	14	14	14	15	15	14	14	14	12	14	14	14	14	4	4	7	7	8	8	4	4	4	5	4	4	4	5.2
Corn	18	18	18	18	18	18	18	18	18	16	18	18	18	17.8	14	13	13	13	14	14	14	13	13	9	13	13	13	13.0
Cotton	16	15	15	15	14	14	15	16	16	13	16	16	16	15.2	5	5	4	4	5	5	5	5	5	6	6	5	5	5.0
Soybeans	1	1	1	1	1	1	1	1	1	1	1	1	1	1.0	2	1	3	2	2	2	2	2	2	2	2	2	2	2.0
Sugar	9	10	10	10	11	11	10	9	7	3	9	8	9	8.9	24	23	24	24	23	23	23	23	23	17	24	23	23	22.8
Wheat (Chicago)	12	13	12	12	13	13	12	12	12	6	12	12	12	11.8	8	8	6	6	11	10	8	7	6	4	7	6	6	7.2
Wheat (Kansas)	2	2	2	2	2	2	2	2	2	2	2	2	2	2.0	1	2	2	3	1	1	1	1	1	1	1	1	1	1.3
Livestock																												
Feeder Cattle	20	21	20	20	21	21	20	20	20	20	20	20	20	20.2	18	18	18	18	18	18	18	18	18	19	18	18	18	18.1
Lean Hogs	24	24	24	24	24	24	24	24	24	23	24	24	24	23.9	22	22	22	22	22	22	22	22	21	20	22	21	22	21.7
Live Cattle	17	17	17	17	17	17	17	17	17	17	17	17	17	17.0	12	17	16	16	17	17	11	11	12	11	10	12	11	13.3

For the period from January 7th, 2002 to November 30th, 2002 and our different performance measures (as numbered in Table 2), this table reports the rankings of our 24 futures-based and our 24 spot-based commodity investments, where 1 (24) resembles the best (worst) investment performance. For each investment subgroup, the last column reports the mean ranking across all performance measures.

Table A3: Commodity rankings (S2 - Growth subsample)

	Futures														Spot													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy																												
Crude Oil (Brent)	3	3	2	2	3	2	4	3	3	1	3	3	3	2.7	2	2	2	2	2	1	3	2	1	1	2	1	2	1.8
Crude Oil (WTI)	6	5	6	6	6	6	6	6	5	4	6	5	6	5.6	5	4	4	4	5	5	5	5	4	3	5	4	5	4.5
Gas Oil	2	2	3	3	4	3	2	2	2	3	2	2	2	2.5	3	1	5	5	4	4	4	3	2	4	3	2	3	3.3
Heating Oil	7	8	9	10	10	9	7	7	7	2	7	7	7	7.5	6	5	7	9	7	6	7	6	6	2	6	6	6	6.1
Natural Gas	24	24	24	24	24	24	24	24	24	22	24	24	24	23.8	20	20	20	20	20	20	20	20	20	16	20	20	20	19.7
Unleaded Gasoline	8	9	11	11	9	8	9	8	8	5	9	8	8	8.5	8	8	9	8	9	9	9	8	8	5	9	8	8	8.2
Precious metals																												
Gold	5	4	5	5	5	5	5	5	6	13	5	6	5	5.7	4	3	3	3	3	3	2	4	5	15	4	5	4	4.5
Silver	11	6	12	12	7	7	8	11	11	24	10	11	11	10.8	9	7	10	10	8	7	8	9	11	24	8	10	10	10.1
Industrial metals																												
Aluminium	12	10	10	9	11	11	12	12	12	15	12	12	12	11.5	12	10	8	7	10	10	12	12	12	19	12	12	12	11.4
Copper	1	1	1	1	1	1	1	1	1	6	1	1	1	1.4	1	6	1	1	1	2	1	1	3	9	1	3	1	2.4
Lead	4	7	4	4	2	4	3	4	4	7	4	4	4	4.2	7	9	6	6	6	8	6	7	7	11	7	7	7	7.2
Nickel	10	12	8	8	8	10	11	10	10	9	8	9	10	9.5	13	16	14	14	11	15	13	13	13	12	13	13	13	13.3
Zinc	13	14	13	13	13	14	13	13	13	16	13	13	13	13.4	14	18	13	12	15	16	14	16	16	20	16	17	16	15.6
Agriculture																												
Cocoa	16	16	16	16	16	16	15	16	16	21	16	16	16	16.3	19	17	19	19	21	19	19	19	19	23	19	19	19	19.3
Coffee	21	19	21	21	21	19	21	21	21	18	21	21	21	20.5	17	12	15	15	16	13	16	17	17	14	17	16	17	15.5
Corn	18	20	18	18	20	21	18	18	18	12	18	18	18	18.1	10	11	11	11	13	14	10	10	9	6	10	9	9	10.2
Cotton	23	23	23	23	23	23	23	23	23	23	23	23	23	23.0	23	23	21	21	23	23	23	23	23	17	23	23	23	22.2
Soybeans	9	11	7	7	12	12	10	9	9	10	11	10	9	9.7	11	15	12	13	17	17	11	11	10	13	11	11	11	12.5
Sugar	19	21	20	20	19	20	19	20	20	19	20	20	20	19.8	18	19	18	18	18	18	18	18	18	18	18	18	18	18.1
Wheat (Chicago)	20	18	19	19	18	18	20	19	19	11	19	19	19	18.3	15	13	17	17	14	12	17	14	14	7	15	14	15	14.2
Wheat (Kansas)	15	15	15	15	15	15	16	15	15	8	15	15	15	14.5	16	14	16	16	12	11	15	15	15	8	14	15	14	13.9
Livestock																												
Feeder Cattle	14	13	14	14	14	13	14	14	14	14	14	14	14	13.8	22	21	23	22	19	21	22	22	22	22	22	22	22	21.7
Lean Hogs	22	22	22	22	22	22	22	22	22	20	22	22	22	21.8	24	24	24	24	24	24	24	24	24	10	24	24	24	22.9
Live Cattle	17	17	17	17	17	17	17	17	17	17	17	17	17	17.0	21	22	22	23	22	22	21	21	21	21	21	21	21	21.5

For the period from December 1st, 2002 to August 1st, 2008 and our different performance measures (as numbered in Table 2), this table reports the rankings of our 24 futures-based and our 24 spot-based commodity investments, where 1 (24) resembles the best (worst) investment performance. For each investment subgroup, the last column reports the mean ranking across all performance measures.

Table A4: Commodity rankings (S3 - Lehman Brothers crisis subsample)

	Futures														Spot													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy																												
Crude Oil (Brent)	22	22	21	21	22	22	22	22	22	23	22	22	22	21.9	20	19	19	19	20	20	20	20	21	23	20	22	20	20.2
Crude Oil (WTI)	23	23	23	23	23	23	23	23	23	24	23	23	23	23.1	23	23	23	23	22	22	23	23	23	24	23	23	23	22.9
Gas Oil	20	21	22	22	21	21	20	20	20	17	20	20	20	20.3	22	22	22	22	23	23	22	22	22	18	21	19	22	21.5
Heating Oil	21	20	20	20	20	20	21	21	21	19	21	21	21	20.5	21	21	20	20	21	21	21	21	20	20	22	21	21	20.8
Natural Gas	24	24	24	24	24	24	24	24	24	20	24	24	24	23.7	24	24	24	24	24	24	24	24	24	7	24	24	24	22.7
Unleaded Gasoline	19	19	19	19	17	17	18	19	19	21	18	19	19	18.7	19	20	21	21	19	19	19	19	19	22	19	20	19	19.7
Precious metals																												
Gold	1	1	1	1	1	1	1	1	1	1	1	1	1	1.0	1	1	3	2	1	1	1	1	1	2	1	1	1	1.3
Silver	2	2	2	2	2	2	2	2	2	6	2	2	2	2.3	3	2	4	4	2	2	2	4	4	8	3	4	2	3.4
Industrial metals																												
Aluminium	15	15	15	15	15	15	15	15	15	15	15	15	15	15.0	18	18	18	18	18	18	18	18	18	17	18	18	18	17.9
Copper	7	9	7	7	7	8	7	7	7	9	7	7	7	7.4	7	11	9	9	10	11	7	7	7	14	8	7	7	8.8
Lead	10	10	10	10	10	10	10	10	10	8	10	10	10	9.8	8	7	8	8	7	7	8	8	8	9	7	8	8	7.8
Nickel	8	8	8	8	8	7	8	8	8	7	8	8	8	7.8	11	8	7	7	9	9	9	10	10	12	9	10	9	9.2
Zinc	9	7	9	9	9	9	9	9	9	5	9	9	9	8.5	6	6	6	6	6	6	6	6	6	6	6	6	6	6.0
Agriculture																												
Cocoa	11	11	11	11	11	11	11	11	11	13	11	11	11	11.2	12	12	12	12	12	12	12	12	12	19	12	12	12	12.5
Coffee	5	5	5	5	5	5	5	5	5	4	5	5	5	4.9	5	5	5	5	5	5	5	5	5	4	5	5	5	4.9
Corn	17	16	18	18	16	16	16	16	17	16	16	17	16	16.5	17	17	17	17	17	17	17	17	17	16	17	17	17	16.9
Cotton	3	3	3	3	4	4	3	3	3	3	4	3	3	3.2	4	3	1	1	4	3	3	3	2	3	4	3	3	2.8
Soybeans	6	6	6	6	6	6	6	6	6	14	6	6	6	6.6	16	16	16	16	16	15	15	16	16	21	15	16	16	16.2
Sugar	4	4	4	4	3	3	4	4	4	2	3	4	4	3.6	2	4	2	3	3	4	4	2	3	1	2	2	4	2.8
Wheat (Chicago)	18	17	16	16	19	19	19	18	18	11	19	18	18	17.4	15	15	15	15	15	16	16	15	15	11	16	15	15	14.9
Wheat (Kansas)	14	14	14	14	14	14	14	14	14	10	14	14	14	13.7	14	14	14	14	14	14	14	14	14	10	14	14	14	13.7
Livestock																												
Feeder Cattle	12	12	12	12	12	12	12	12	12	12	12	12	12	12.0	10	9	11	11	11	10	10	11	11	13	11	11	11	10.8
Lean Hogs	16	18	17	17	18	18	17	17	16	18	17	16	17	17.1	9	10	10	10	8	8	11	9	9	5	10	9	10	9.1
Live Cattle	13	13	13	13	13	13	13	13	13	22	13	13	13	13.7	13	13	13	13	13	13	13	13	13	15	13	13	13	13.2

For the period from September 1st, 2008 to December 7th, 2010 and our different performance measures (as numbered in Table 2), this table reports the rankings of our 24 futures-based and our 24 spot-based commodity investments, where 1 (24) resembles the best (worst) investment performance. For each investment subgroup, the last column reports the mean ranking across all performance measures.

Table A5: Commodity rankings (S4 - EU crisis subsample)

	Futures														Spot													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy																												
Crude Oil (Brent)	2	3	3	3	1	2	2	1	2	1	1	1	1	1.8	5	4	4	4	2	3	3	3	4	3	1	2	3	3.2
Crude Oil (WTI)	14	12	13	13	12	12	13	13	13	11	11	13	13	12.5	11	8	9	12	9	8	11	10	10	9	10	10	9	9.7
Gas Oil	1	1	1	1	2	1	1	2	1	4	4	4	2	1.9	2	1	1	1	1	1	2	2	2	5	6	5	2	2.4
Heating Oil	3	2	2	2	3	3	3	3	3	2	3	3	3	2.7	6	2	3	3	3	2	6	4	3	2	5	4	4	3.6
Natural Gas	24	24	24	24	24	24	24	24	24	20	24	24	24	23.7	24	24	23	23	24	24	24	24	23	20	24	23	24	23.4
Unleaded Gasoline	5	4	4	4	4	4	4	6	7	7	5	7	5	5.1	3	5	5	5	4	4	4	6	6	7	4	7	6	5.1
Precious metals																												
Gold	20	20	20	20	20	20	20	20	20	21	20	20	20	20.1	20	20	20	20	20	20	20	20	20	22	20	20	20	20.2
Silver	8	9	8	8	10	10	7	8	9	14	13	12	8	9.5	10	11	10	9	12	12	9	11	13	13	13	13	11	11.3
Industrial metals																												
Aluminium	11	8	10	9	9	9	12	11	10	9	10	9	11	9.8	13	10	11	10	10	9	13	13	11	12	12	11	13	11.4
Copper	17	16	16	16	17	17	17	17	16	17	18	16	17	16.7	17	17	17	17	18	17	17	17	17	18	18	17	17	17.2
Lead	13	13	14	14	13	13	14	14	14	13	14	14	14	13.6	14	14	14	14	14	14	14	14	14	14	14	14	14	14.0
Nickel	18	18	18	18	18	18	18	18	18	16	17	18	18	17.8	18	19	18	19	17	18	18	18	18	16	17	18	18	17.8
Zinc	19	19	19	19	19	19	19	19	19	18	19	19	19	18.9	19	18	19	18	19	19	19	19	19	17	19	19	19	18.7
Agriculture																												
Cocoa	22	22	22	22	22	22	22	22	22	23	22	22	22	22.1	22	23	24	24	22	22	22	22	22	23	22	22	22	22.5
Coffee	9	11	9	10	11	11	10	9	8	8	8	8	9	9.3	12	13	13	13	13	13	12	12	12	11	11	12	12	12.2
Corn	7	10	6	5	6	7	8	7	6	5	7	6	7	6.7	9	12	7	7	8	10	8	8	8	8	9	8	7	8.4
Cotton	4	5	7	7	8	6	6	4	4	3	6	2	6	5.2	7	7	12	11	11	11	10	7	7	4	7	3	8	8.1
Soybeans	16	17	17	17	16	16	16	16	17	19	16	17	16	16.6	16	16	16	16	16	15	16	16	16	19	16	16	16	16.2
Sugar	21	21	21	21	21	21	21	21	21	24	21	21	21	21.2	23	22	22	22	23	23	23	23	24	24	23	24	23	23.0
Wheat (Chicago)	23	23	23	23	23	23	23	23	23	22	23	23	23	22.9	21	21	21	21	21	21	21	21	21	21	21	21	21	21.0
Wheat (Kansas)	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	16	15	15	15	15	15	15	15	15.1
Livestock																												
Feeder Cattle	6	6	5	6	5	5	5	5	5	6	2	5	4	5.0	4	6	6	6	6	6	5	5	5	6	3	6	5	5.3
Lean Hogs	12	14	11	11	14	14	11	12	11	10	12	10	12	11.8	1	3	2	2	5	5	1	1	1	1	2	1	1	2.0
Live Cattle	10	7	12	12	7	8	9	10	12	12	9	11	10	9.9	8	9	8	8	7	7	7	9	9	10	8	9	10	8.4

For the period from December 8th, 2010 to April 4th, 2011 and our different performance measures (as numbered in Table 2), this table reports the rankings of our 24 futures-based and our 24 spot-based commodity investments, where 1 (24) resembles the best (worst) investment performance. For each investment subgroup, the last column reports the mean ranking across all performance measures.

Table A6: Commodity rankings (S5 - Greek crisis subsample)

	Futures														Spot													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy																												
Crude Oil (Brent)	3	3	3	3	3	3	3	3	3	12	3	3	3	3.7	5	5	5	5	5	5	5	5	5	16	5	5	5	5.8
Crude Oil (WTI)	10	11	10	11	12	11	10	12	12	16	11	12	12	11.5	9	9	9	9	9	9	9	9	9	15	9	9	9	9.5
Gas Oil	4	5	5	5	4	4	4	5	5	3	4	5	5	4.5	6	6	6	6	6	6	6	6	6	5	6	6	6	5.9
Heating Oil	8	7	8	8	8	7	8	8	8	13	8	8	8	8.2	7	7	7	7	7	7	7	7	7	12	7	7	7	7.4
Natural Gas	24	24	24	24	24	24	24	24	24	24	24	24	24	24.0	24	23	24	24	23	23	24	24	24	23	23	24	24	23.6
Unleaded Gasoline	2	1	2	2	2	2	2	2	2	5	2	2	2	2.2	3	3	3	3	3	3	3	3	3	14	3	3	3	3.8
Precious metals																												
Gold	1	2	1	1	1	1	1	1	1	6	1	1	1	1.5	1	1	1	1	2	1	1	1	1	6	1	1	1	1.5
Silver	16	16	16	16	16	16	16	16	17	22	16	17	17	16.7	17	17	19	19	17	17	17	17	20	22	17	19	18	18.2
Industrial metals																												
Aluminium	14	15	14	14	15	15	14	14	14	20	14	14	14	14.7	14	16	14	14	15	15	14	14	13	17	13	13	13	14.2
Copper	12	12	12	12	11	12	11	10	11	10	12	11	10	11.2	11	11	11	11	10	10	10	11	11	11	11	11	11	10.8
Lead	18	17	22	22	17	17	18	18	18	15	20	18	18	18.3	21	19	22	22	19	19	21	21	21	18	21	21	21	20.5
Nickel	21	19	21	21	19	19	20	21	21	18	22	21	21	20.3	22	21	21	21	22	22	22	22	22	19	22	22	22	21.5
Zinc	15	14	15	15	14	14	15	15	15	9	15	15	15	14.3	15	15	16	16	14	14	15	15	15	9	15	15	15	14.5
Agriculture																												
Cocoa	17	18	17	17	18	18	17	17	16	8	17	16	16	16.3	18	20	18	18	18	18	18	18	17	8	18	17	17	17.2
Coffee	19	22	20	20	20	20	19	19	19	21	19	19	19	19.7	20	22	20	20	21	21	20	20	19	21	20	20	20	20.3
Corn	13	13	13	13	13	13	13	13	13	19	13	13	13	13.5	13	13	13	12	13	13	13	13	14	20	14	14	14	13.8
Cotton	23	23	23	23	23	23	23	23	23	23	23	23	23	23.0	23	24	23	23	24	24	23	23	23	24	24	23	23	23.4
Soybeans	6	8	6	6	7	8	6	6	6	4	6	6	6	6.2	4	4	4	4	4	4	4	4	4	4	4	4	4	4.0
Sugar	5	4	4	4	5	5	5	4	4	1	5	4	4	4.2	12	12	12	13	12	12	12	12	12	3	12	12	12	11.4
Wheat (Chicago)	22	21	18	18	22	22	22	22	22	17	21	22	22	20.8	16	14	15	15	16	16	16	16	16	10	16	16	16	15.2
Wheat (Kansas)	20	20	19	19	21	21	21	20	20	14	18	20	20	19.5	19	18	17	17	20	20	19	19	18	13	19	18	19	18.2
Livestock																												
Feeder Cattle	7	6	7	7	6	6	7	7	7	2	7	7	7	6.4	2	2	2	2	1	2	2	2	2	1	2	2	2	1.8
Lean Hogs	11	10	11	10	10	10	12	11	10	7	10	10	11	10.2	10	10	10	10	11	11	11	10	10	7	10	10	10	10.0
Live Cattle	9	9	9	9	9	9	9	9	9	11	9	9	9	9.2	8	8	8	8	8	8	8	8	8	2	8	8	8	7.5

For the period from April 5th, 2011 to March 31st, 2012 and our different performance measures (as numbered in Table 2), this table reports the rankings of our 24 futures-based and our 24 spot-based commodity investments, where 1 (24) resembles the best (worst) investment performance. For each investment subgroup, the last column reports the mean ranking across all performance measures.

Table A7: Commodity rankings (S6 - Post crisis subsample)

Futures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy														
Crude Oil (Brent)	23	23	23	22	22	23	23	23	23	24	23	23	23	22.9
Crude Oil (WTI)	24	24	24	24	24	24	24	24	24	22	24	24	24	23.8
Gas Oil	20	22	22	23	20	22	21	20	20	23	21	20	20	21.1
Heating Oil	21	21	21	21	18	19	20	21	21	20	20	21	21	20.4
Natural Gas	22	20	20	20	19	18	22	22	22	9	22	22	22	20.0
Unleaded Gasoline	16	18	19	19	12	12	16	16	17	18	18	17	16	16.5
Precious metals														
Gold	6	8	7	7	9	9	6	7	8	19	6	7	7	8.2
Silver	18	16	17	17	21	20	15	18	19	21	17	19	18	18.2
Industrial metals														
Aluminium	11	11	11	11	11	11	11	10	10	6	10	10	10	10.2
Copper	10	10	10	10	10	10	10	11	11	17	11	11	11	10.9
Lead	7	6	6	6	6	6	7	6	6	2	7	6	6	5.9
Nickel	17	17	15	15	13	15	18	17	16	16	16	16	17	16.0
Zinc	5	5	5	5	5	5	5	5	5	4	5	5	5	4.9
Agriculture														
Cocoa	1	1	1	1	1	1	1	1	1	1	1	1	1	1.0
Coffee	19	15	18	18	17	17	19	19	18	11	19	18	19	17.5
Corn	12	12	13	13	14	13	12	12	12	13	12	12	12	12.5
Cotton	9	7	9	9	8	7	9	9	9	15	9	9	9	9.1
Soybeans	2	2	2	2	2	2	2	2	2	3	2	2	2	2.1
Sugar	15	19	16	16	23	21	17	15	15	14	15	15	15	16.6
Wheat (Chicago)	13	13	12	12	16	14	13	13	13	7	13	13	13	12.7
Wheat (Kansas)	14	14	14	14	15	16	14	14	14	8	14	14	14	13.8
Livestock														
Feeder Cattle	4	4	4	4	4	4	4	4	4	12	4	4	4	4.6
Lean Hogs	8	9	8	8	7	8	8	8	7	10	8	8	8	8.1
Live Cattle	3	3	3	3	3	3	3	3	3	5	3	3	3	3.2

Spot

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Mean
Energy														
Crude Oil (Brent)	24	24	23	23	24	24	24	24	24	24	24	24	24	23.8
Crude Oil (WTI)	23	22	21	21	22	23	23	23	23	20	23	23	23	22.3
Gas Oil	21	23	24	24	21	22	20	20	20	23	20	20	20	21.4
Heating Oil	22	21	20	20	18	21	22	22	22	21	22	22	21	21.1
Natural Gas	6	6	5	5	5	5	6	6	6	2	6	6	6	5.4
Unleaded Gasoline	20	20	22	22	19	19	21	21	21	18	21	21	22	20.5
Precious metals														
Gold	8	9	8	8	9	9	8	8	9	19	8	9	8	9.2
Silver	19	18	19	19	23	20	18	19	19	22	19	19	19	19.5
Industrial metals														
Aluminium	9	8	9	9	8	8	9	9	8	4	9	8	9	8.2
Copper	14	15	12	12	14	14	14	14	14	17	15	14	14	14.1
Lead	7	5	7	7	7	6	7	7	7	3	7	7	7	6.5
Nickel	18	19	17	18	17	17	19	18	18	15	18	18	18	17.7
Zinc	4	4	4	4	4	4	4	4	4	5	4	4	4	4.1
Agriculture														
Cocoa	1	1	1	1	1	1	1	1	1	1	1	1	1	1.0
Coffee	16	10	13	13	10	10	16	16	16	9	16	16	16	13.6
Corn	17	17	18	17	20	18	17	17	17	14	17	17	17	17.2
Cotton	13	11	15	15	11	11	13	13	13	16	13	13	13	13.1
Soybeans	11	14	16	16	15	15	11	12	12	13	10	12	12	13.0
Sugar	15	16	14	14	16	16	15	15	15	12	14	15	15	14.8
Wheat (Chicago)	10	12	10	10	12	12	10	10	10	8	11	10	10	10.4
Wheat (Kansas)	12	13	11	11	13	13	12	11	11	7	12	11	11	11.4
Livestock														
Feeder Cattle	3	3	3	3	3	3	3	3	3	11	3	3	3	3.6
Lean Hogs	5	7	6	6	6	7	5	5	5	10	5	5	5	5.9
Live Cattle	2	2	2	2	2	2	2	2	2	6	2	2	2	2.3

For the period from April 1st, 2012 to March 31st, 2016 and our different performance measures (as numbered in Table 2), this table reports the rankings of our 24 futures-based and our 24 spot-based commodity investments, where 1 (24) resembles the best (worst) investment performance. For each investment subgroup, the last column reports the mean ranking across all performance measures.