AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Factor Structure in Commodity Futures Return and Volatility

Peter Christoffersen, Asger Lunde and Kasper V. Olesen

CREATES Research Paper 2014-31

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Factor Structure in Commodity Futures Return and Volatility*

Peter Christoffersen[a] **Asger Lunde**[b]
Kasper V. Olesen[b]

[a]*Rotman School of Management, University of Toronto, CBS & CREATES*

[b]*Aarhus University, Department of Economics and Business & CREATES*

September 8, 2014

Abstract

Using data on more than 750 million futures trades during 2004-2013, we analyze eight stylized facts of commodity price and volatility dynamics in the post financialization period. We pay particular attention to the factor structure in returns and volatility and to commodity market integration with the equity market. We find evidence of a factor structure in daily commodity futures returns. However, the factor structure in daily commodity futures volatility is even stronger than in returns. When computing model-free realized commodity betas with the stock market we find that they were high during 2008-2010 but have since returned to the pre-crisis level close to zero. The common factor in commodity volatility is nevertheless clearly related to stock market volatility. We conclude that, while commodity markets appear to again be segmented from the equity market when only returns are considered, commodity volatility indicates a nontrivial degree of market integration.

Keywords: Factor structure, financial volatility, beta, high-frequency data, commodities, financialization

JEL Classifications: G13, Q02

*All authors acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Christoffersen was also supported by Bank of Canada, GRI, and the Social Sciences, and Humanities Research Council of Canada. Lunde and Olesen thank the AU Ideas Pilot Center, Stochastic and Econometric Analysis of Commodity Markets, for financial support. Address correspondence to Peter Christoffersen, Rotman School of Management, University of Toronto, 105 St. George St., M5S 3E6 Toronto, Canada; E-mail address: peter.christoffersen@rotman.utoronto.com.

1 Introduction

The deregulation of commodity markets in the early 2000s, and the subsequent large inflows of investment capital to commodity futures and related securities, has sparked a heated debate in the popular press as well as in academia. A large part of the discussion of this so-called "financialization" of commodity markets has focused on bubbles in commodity prices and increases in commodity price volatility supposedly caused by futures market trading. The relationship between commodity futures and other asset markets, and in particular risk sharing across markets, has received substantial attention in academia. Cheng & Xiong (2013) provide an excellent overview of existing work on the financialization of commodity futures markets.

We use more than 750 million commodity futures transactions to construct daily realized volatility for fifteen commodities during the 2004-2013 period. To our knowledge, we are the first to construct realized covariances for each commodity with the stock market which we use to develop time-varying but model-free stock market betas as well as systematic risk contributions for each commodity. We use these measures along with daily futures returns to address the following research questions:

First, what are the stylized facts of commodity futures returns and volatility post financialization? Second, is there a factor structure in commodity futures returns? Third, is the stock market driving the common component of commodity futures returns? Fourth, do the stock market betas of commodity futures returns vary significantly over time? Fifth, does the ratio of commodity futures return volatility explained by the stock market beta change over time? Sixth, is there a factor structure in commodity futures volatility? Finally, is stock market volatility driving the common component of commodity futures volatility? The stylized facts that we uncover are:

Fact #1: *Daily realized commodity futures volatility has extremely high persistence.*

Fact #2: *The logarithm of realized commodity futures volatility is close to normally distributed.*

Fact #3: *There is some evidence of a factor structure in daily commodity futures returns excluding meats.*

Fact #4: *The factor structure in daily commodity futures volatility is much stronger than the factor structure in returns.*

Fact #5: *There is little evidence of a time-trend in the degree of integration across commodity futures markets during the 2004-2013 period.*

Fact #6: *The strong common factor in commodity volatility is largely driven by stock market volatility.*

Fact #7: *Commodity betas with the stock market were high during 2008-2010 but have since returned to a level close to zero.*

Fact #8: *Commodity futures returns standardized by expected realized volatility are closer to normally distributed than the returns themselves but still display leptokurtosis.*

Most authors, including Baker (2014), Baker & Routledge (2012), and Hamilton & Wu (2014), date the financialization to take effect sometime in the 2004-2005 period. We follow these papers and begin our analysis in January 2004. Our analysis is based on approximately 3/4 billion trades in commodity futures contracts observed from January 2004 through December 2013. We choose the three most heavily traded commodities in the energy, metals, grains, softs, and meats categories for a total of 15 futures contracts. Our work builds heavily on recent advances in model-free volatility and covariance measurement using high-frequency data. See for example Andersen, Bollerslev, Christoffersen & Diebold (2013), Barndorff-Nielsen & Shephard (2007), and Hansen & Lunde (2011) for recent surveys.

Bakshi, Gao & Rossi (2013) develop a three-factor commodity pricing model, using the average commodity return, a basis spread factor, and a momentum factor. They find that the model captures both the cross-sectional and time-series variation in commodity returns. Szymanowska, De Roon, Nijman & Van Den Goorbergh (2014) also find that the basis spread factor explains the cross-section of commodity returns. Daskalaki, Kostakis & Skiadopoulos (2014) also explore common factors in the cross-section of commodity futures returns. They test various asset pricing models motivated either by the traditional empirical equity market studies, or by available commodity pricing theories. They find that none of the models they investigate are successful and conclude that commodity markets are heterogeneous and segmented from the equity market. From the apparent conflict in these recent studies we conclude that the factor structure of commodity returns needs to be studied further.

The evidence on the integration of commodity and equity markets is also mixed. For early evidence on segmentation, see Bessembinder (1992), Bessembinder & Chan (1992), and Gorton & Rouwenhorst (2006). For recent evidence on integration, see Tang & Xiong (2012), Henderson, Pearson & Wang (2013), Basak & Pavlova (2014), and Singleton (2014). We contribute to this discussion by constructing daily model-free realized market betas from high-frequency returns on commodity futures and market index futures.

Inspired by Longin & Solnik (2001) and Ang & Chen (2002) who study equity markets, we analyze the threshold correlations between commodity futures returns and stock market returns. We find strong evidence of tail-dependence between commodity and equity market returns suggesting important nonlinear and non-normal dependencies between commodity and equity markets.

In recent work focusing on the U.S. equity market, Chen & Petkova (2012), Duarte, Kamara, Siegel & Sun (2014), and Herskovic, Kelly, Lustig & Van Nieuwerburgh (2014) find strong evidence of factor structure in idiosyncratic volatility. Herskovic et al. (2014) use daily data to compute annual firm-specific realized volatilities on a large number of stocks. They then look for and find strong commonality in the volatility of individual equities. The commonality in equity volatility is strong even after the market factor is removed. This work is important because it has the potential to explain the so-called risk anomaly documented in Ang, Hodrick, Xing & Zhang (2006) and Ang, Hodrick, Xing & Zhang (2009) showing that stocks with low volatility this quarter have high returns next month and vice versa. Inspired by this equity market literature, we investigate the factor structure of individual commodity volatility. We find strong evidence of a common factor in daily commodity futures volatility which we compute from intraday returns.

Our study complements Tang & Xiong (2012) who focus on the connection between the large inflow of commodity index investments and the large increase in commodity price co-movements. They find that prices of non-energy commodity futures in the

United States have become increasingly correlated with oil prices after 2004. Also, they find that the correlation between commodity returns and the MSCI Emerging Markets Index has been rising in recent years. Substantial increases in return correlation between commodities and stocks were also found in Büyüksahin, Haigh & Robe (2010) using the Dynamic Conditional Correlation (DCC) model of Engle (2002). In recent work, Boons, Roon & Szymanowska (2013) find that stocks that have high betas with a commodity index earned a relatively low average return in the pre-financialization period and a relatively high return in the post-financialization period. We investigate the dynamic relationship between commodities and the stock market using model-free realized betas. Our time-varying betas can be viewed as dynamic measures of integration with the stock market. Our work is therefore related to the research on time-varying integration of emerging markets, see, for example, Carrieri, Errunza & Hogan (2007) and Bekaert, Harvey, Lundblad & Siegel (2011).

The remainder of the paper is structured as follows. In Section 2 we describe the data and the methodology for computing returns and realized volatilities. Section 3 investigates the multivariate properties of commodity futures returns and volatility. Section 4 analyzes the relationship between commodities and the stock market, and Section 5 investigates the distributional properties of commodity returns and shocks. Section 6 concludes. Supplementary figures and tables are placed in the appendix.

2 Constructing Returns and Realized Volatility

We collect data from two main sources. High-frequency data on commodity futures are obtained from Tick Data Inc. and transaction data on the Spyder ETF are obtained from the NYSE Trade and Quote (TAQ) database.[1] Spyder is used as a proxy for the stock market index. The commodity futures dataset includes intraday transactions and quotes for all traded maturities on more than 60 commodities traded across the world.

[1] Standard & Poor's Depository Receipt (Spyder) is a tradable security that represents ownership in the S&P 500 index. By using Spyder, stale price effects associated with using the S&P 500 cash index at high frequencies are avoided.

2.1 Commodity Categories

We study the post-financialization period from January 1, 2004 to December 31, 2013, and focus on commodities that are traded either in Chicago or New York during the entire sample period. Gorton, Hayashi & Rouwenhorst (2012) classify commodities into five categories: metals, softs, grains, energy, and meats. We use this classification and focus on the three most actively traded commodities in each category. This leaves $K = 15$ commodities for our analysis, namely light crude, natural gas, heating oil; gold, silver, copper; soybeans, corn, wheat; sugar, coffee, cotton; live cattle, lean hogs, and feeder cattle.[2] See Table 1 for details on our fifteen commodities. The normalized closing prices of the commodities and Spyder are shown in Figure 1.

Table 1: The 15 Selected Commodities.

Classification	Commodity	Exchange	Time Zone	Transactions
Energy:	Light Crude (CL)	NYMEX/CME	New York	217,709,630
	Natural Gas (NG)	NYMEX/CME	New York	73,536,635
	Heating Oil (HO)	NYMEX/CME	New York	31,931,052
Metals:	Gold (GC)	COMEX/CME	New York	125,901,304
	Silver (SV)	COMEX/CME	New York	43,657,064
	Copper (HG)	COMEX/CME	New York	34,840,077
Grains:	Soybeans (SY)	CBOT/CME	Chicago	62,056,841
	Corn (CN)	CBOT/CME	Chicago	66,933,216
	Wheat (WC)	CBOT/CME	Chicago	34,032,067
Softs:	Sugar #1 (SB)	ICE	New York	22,666,799
	Coffee "C" (KC)	ICE	New York	10,798,589
	Cotton #2 (CT)	ICE	New York	8,798,184
Meats:	Live Cattle (LC)	CME	Chicago	10,760,576
	Lean Hogs (LH)	CME	Chicago	9,274,472
	Feeder Cattle (FC)	CME	Chicago	1,905,263

Notes: For each commodity futures contract we report the total number of transactions available during our 2004-2013 sample period.

Not surprisingly, light crude is by far the most heavily traded commodity futures with more than 200 million transactions as seen from the last column of Table 1. Each transaction consists of at least one futures contract being traded in which 1.000 barrels of crude oil is the underlying commodity. Table 1 illustrates the massive size of the crude oil futures market compared with other commodities. Gold is second, natural gas third, corn fourth, soybeans fifth, and feeder cattle is the least traded among the fifteen

[2]The remaining commodities traded in Chicago or New York are palladium, ICE energy futures, lumber, cocoa, orange juice, oats, soybean meal, and soybean oil. We also do not consider platinum which only began trading in October 2007, ethanol which began trading in January 2010, rough rice that traded from January 2010, and pork bellies that traded until July 2011.

commodities with less than 2 million registered transactions in our sample period. Figure 2 illustrates the development in the daily average transactions, volume, and dollar volume per year. Trading in all commodities has increased remarkably since 2006. The availability of large amounts of transactions data forms the basis of our analysis on commodity futures volatility.

2.2 Transaction Data Cleaning

The raw daily transactions data for the fifteen commodity futures series is cleaned using the Tick Data validation process and subsequently the algorithm in Barndorff-Nielsen, Hansen, Lunde & Shephard (2009).[3] We thus use the median price if multiple transactions have the same time stamp, and we delete entries for which the transaction price is more than five mean absolute deviations from a rolling centered median of the 25 preceding and the 25 subsequent observations. For the Spyder contract, entries with a transaction price equal to zero, entries with corrected trades, entries with an abnormal sale condition, and entries with prices that are above (below) the ask plus (minus) the bid-ask spread are deleted. For the univariate realized variance analysis, the widest feasible estimation window is used, and for the bivariate realized covariance analysis the window for each commodity is defined by its trading span overlap with Spyder as discussed further in Section 2.4 below. For most commodities, electronic trading is available 24 hours a day, only paused by short breaks. All this is illustrated in Figure 3, which shows the trading periods and breaks on December 30, 2013, for the 15 commodities.

2.3 Modeling the Futures Roll and Computing Daily Returns

For each of the fifteen commodities, a continuous price series is constructed from the nearest-to-maturity contracts in the sample period. Rollover to the subsequent

[3] *"Algorithmic data filters are employed to identify bad prints, decimal errors, transposition errors, and other data irregularities. These filters take advantage of the fact that since we are not producing data in real-time, we have the ability to look at the tick following a suspected bad tick before we decide whether or not the tick is valid. We have developed a number of filters that identify a suspect tick and hold it until the following tick confirms its validity. The filters are proprietary, and are based upon recent tick volatility, moving standard deviation windows, and time of day."* Source: www.tickdata.com.

contract occurs on days when the daily, day-session tick volume of the back-month contract exceeds the daily tick volume of the current month contract.[4] This procedure is intended to mimic the behaviour of the majority of market participants. Rollover dates are stored to allow for analysis with and without roll-returns. Roll-returns are sometimes extreme, but investors in commodity markets are exposed to them, and it is important to recognize their implications. The number of roll-returns for a commodity depends on the availability and periodicity of futures contracts and it varies in our sample period from 38 for cotton to 120 for the three energy commodities. Rollover is always done in the afternoon trading break illustrated in Figure 3.

For the univariate analysis, the first observed price for a commodity on day t is taken as the opening price, F_t^o, and the closing price, F_t^c, is the last observed price before the afternoon trading break. This is indicated for Spyder in the top line in Figure 3. The commodities are treated similarly. Daily log-returns of the closing prices, $r_t = \ln\left(F_t^c\right) - \ln\left(F_{t-1}^c\right)$, are plotted in Figure A.1 in the appendix.

Table 2.a reports various descriptive statistics for the daily log returns on our 15 commodity futures.[5] Consider for now the dynamic properties of commodity futures returns. The first-order autocorrelation, ACF(1), is significantly negative (at the 1% level) for light crude and wheat. The largest autocorrelation is 7%. The Ljung-Box test that the first 5 and 21 daily autocorrelations are jointly zero are rejected for natural gas, and for light crude, copper, soybeans, and feeder cattle when using 21 lags but not when using 5 lags. Figure A.2 in the appendix contains the autocorrelation functions for the first 60 lags which do not show any strong systematic evidence for daily return dynamics. Altogether we conclude that daily returns show little evidence of predictability based on sample autocorrelations. We therefore do not model expected return dynamics below. The distribution of commodity futures returns will be discussed in detail in Section 5.1 below.

[4] This can be done using "AutoRoll" in the TickWrite 7 software provided by Tick Data Inc. This way the data used in our analysis should be straightforward to reproduce.

[5] Table A.1.a in the appendix contains descriptive statistics for returns when roll-returns are removed.

2.4 Constructing Realized Volatilities

Trading in the late evening hours is omitted for the purpose of computing realized volatility. The effect of this is expected to be negligible as the fraction of observations discarded by this procedure is small as indicated with the numbers in grey in Figure 3. Within the trading span, a 1-minute time-grid is constructed using previous-tick interpolation for each commodity. This results in $(n+1)$ 1-minute prices where the first price in the grid is F_t^o. From these prices, n 1-minute log-returns on day t are calculated as

$$r_t^{(j)} = \ln\left(F_{t_j}\right) - \ln\left(F_{t_{j-1}}\right),$$

where $t_j - t_{j-1}$ equals one minute. We can then define $(n-4)$ 5-minute returns using

$$\tilde{r}_t^{(k)} = \sum_{k=j}^{j+4} r_t^{(j)},$$

which is a set of overlapping 5-minute returns. The 5-minute realized variance with 1-minute subsampling is then

$$RV_t^{oc} = \frac{n}{n-4} \cdot \frac{1}{5} \sum_{k=1}^{n-4} \left(\tilde{r}_t^{(k)}\right)^2,$$

where the scaling factor ensures that the realized variance is unbiased. The subsampling technique uses 5-minute returns to minimize the effect of market microstructure noise on our volatility estimate and the subsampled 5-minute returns are then averaged to increase the efficiency of the estimator. This estimator is a simplified version of the estimators advocated by Zhang et al. (2005).

Finally, variance and covariance measures are matched to close-close returns following Hansen & Lunde (2005). That is, the realized variance for the whole day based on intermittent high-frequency data is given by

$$RV_t = \hat{\omega}_1 \left(r_t^{co}\right)^2 + \hat{\omega}_2 RV_t^{oc}, \tag{1}$$

where $\hat{\omega}_1$ and $\hat{\omega}_2$ are estimated weights as given in the notes to Table A.2 in the appendix and r_t^{co} is the realized close-to-open return, i.e. $r_t^{co} = \ln(F_t^o) - \ln(F_{t-1}^c)$.[6] The $RVol_t$ measure is then computed as the square root of RV_t. The time series of $RVol_t$ for our 15 commodities are shown in Figure A.3 in the appendix.

2.5 Properties of Realized Commodity Volatility

Table 3.a reports various sample statistics for the daily realized volatilities. Table 3.a shows that $RVol_t$ has high positive skewness, as well as positive excess kurtosis, which is of course not surprising. More importantly, $RVol_t$ is extremely persistent. The first-order autocorrelation is large and significant for all 15 commodities and the Ljung-Box test is highly significant for both 5 and 21 lags.

Table 3.b reports the same sample statistics as Table 3.a but now for the natural logarithm of $RVol_t$. Note again the extremely high persistence as evidenced by the ACF(1), Q(5), and Q(21) statistics. Figure 4 plots the autocorrelation functions for the first 60 lags for $\log(RVol_t)$. The level of first-order autocorrelation varies a bit across commodities but the strong persistence is evident across our 15 commodities and we can write our first stylized fact:

Fact #1: *Daily realized commodity futures volatility has extremely high persistence.*

Comparing Table 3.b with Table 3.a we also see that the $\log(RVol_t)$ is much closer to normally distributed than is $RVol_t$ itself; skewness is close to zero and kurtosis is close to 3. Figure 5 reports the quantile-quantile plot of $\log(RVol_t)$ which visualizes the normality of $\log(RVol_t)$. The log-normal feature of $RVol_t$ is well-known in other asset markets, see for example Andersen, Bollerslev, Diebold & Labys (2001) for foreign exchange and Andersen, Bollerslev, Diebold & Ebens (2001) for equities. To our knowledge, we are the first to document that:

Fact #2: *The logarithm of realized commodity futures volatility is close to normally distributed.*

[6]Table A.2 is similar to Table 2 in Hansen & Lunde (2005), and shows the estimated values used in the computation of the realized variances for the individual commodities. The 1% largest squared overnight returns and the 0.5% largest realized covariances were omitted from the estimation.

Given the approximately normal distribution of log-realized volatility found in Table 3.b, we model the expected $\log(RVol_t)$ and use an ARMA(1,1) specification, defined by

$$\log(RVol_t) = \phi_0 + \phi_1 \log(RVol_{t-1}) + \theta_1 e_{t-1} + e_t. \tag{2}$$

We choose an ARMA(1,1) specification to capture the strong persistence in $\log(RVol_t)$ and also to capture the unavoidable measurement error in the realized variance measure defined in equation 1 above.[7] Table A.3.a in the appendix contains the ARMA-coefficient estimates. Figure 6 contains time series plots of the expected one-day ahead $\log(RVol_t)$ computed as

$$E_{t-1}[\log(RVol_t)] = \phi_0 + \phi_1 \log(RVol_{t-1}) + \theta_1 e_{t-1}.$$

For comparison, we also plot the realized stock market volatility (in grey) using the Spyder futures contract.[8] Note that the commodity $\log(RVol_t)$ - unlike the $RVol_t$ levels - are fairly well-behaved over time and in some cases shows a slightly decreasing trend over time. The concern of increased commodity market volatility often raised in the popular press does therefore not appear to be warranted.

3 Factor Structure in Commodity Returns and Volatility

In this section we investigate the multivariate properties of commodity returns and volatility. We pay particular attention to the factor structure in the cross-section of commodities.

3.1 A Common Factor in Commodity Returns?

In order to get a quick first glance at the cross-commodity return dependence, the upper triangle of the matrix in Table 4 reports the sample correlations for daily futures returns. Note the high correlations for energy and metals, the somewhat lower correlations for

[7] Using a fractionally integrated ARFIMA$(1, d, 1)$ model to capture the long memory in volatility does not change any of our conclusions nor does recursively estimating the ARMA(1,1) model.

[8] Figure A.4 in the appendix contains time series plots of the raw $\log(RVol_t)$ series.

grains and softs, and the close-to-zero correlations for meats. The average correlation with all other commodities is highest for light crude at 34% and lowest for lean hogs at 12%. The diversification benefits thus varies greatly across commodities. The average correlation across all pairs of commodity returns is 25%.

We now investigate the evidence of a common factor in our 15 commodity returns. Using the sample correlations in the upper triangle of Table 4, we compute the principal components (PCs) for the 15 return series and in Figure 7.a we plot the cumulative return for the first four PCs. The first four PCs explain 30%, 15%, 10%, and 7% respectively, for a total of 62% of the cross-sectional variation in the 15 commodity futures returns. Comparing the top-left panel in Figure 7.a with the top-left panel in Figure 1 we see that the first PC appears to capture the 2007-08 run up, and subsequent crash in oil prices.

Table 5.a reports the regressions of each commodity on the first four PCs. For each commodity we run a separate PC analysis based only on the other 14 commodities to avoid endogeneity issues in these regressions. See from the R^2 how all 15 commodities load positively on the first PC. Note also that the first four factors explain much of the variation in oil, metals and grains, much less so for softs, and virtually none for meats. The average R^2 across the 15 commodity return regressions is 23%. We conclude:

Fact #3: *There is some evidence of a factor structure in daily commodity futures returns excluding meats.*

3.2 A Common Factor in Commodity Volatility?

Our daily realized volatility measures computed from intraday returns allow us to view volatility as an observed time series albeit measured with error. We therefore now investigate the multivariate properties of our 15 commodity $\log(RV_t)$. The bottom triangle of the matrix in Table 4 contains the sample correlations for $\log(RVol_t)$. Note that the correlations for log volatility is higher than for returns for all of the 15 commodities. This is particularly the case for meats, where the average return correlation for live cattle, lean hogs, and feeder cattle is 20%, 12%, and 14% respectively, whereas their average $\log(RVol_t)$ correlations are 52%, 48%, and 49%. The average correlation across all pairs of commodity volatilities is 46% compared with 25% for returns.

The first four principal components corresponding to the ARMA(1,1) filtered $\log(RVol_t)$ are reported in Figure 7.b. The first four PCs capture 50%, 11%, 10%, and 6% respectively, for a total of 77% of the total variation. Note again that the first PC for expected $\log(RVol_t)$ in the top left panel of Figure 7.b resembles quite closely the time series of expected $\log(RVol_t)$ for light crude in the top left panel of Figure 6.

Table 5.b reports the regressions of each expected $\log(RVol_t)$ on the first four PCs. For each commodity, we again run a separate PC analysis based only on the other 14 commodities to avoid endogeneity issues in these regressions. Note that all 15 commodities load positively on the first factor. Table 5.b also reports the regression fit, R^2, of each commodity. They show that the first four PCs capture a substantial share of the variation for all commodities except perhaps coffee. The average R^2 is 54% for volatility compared with 23% for returns. Compared with the commonality in returns, the commonality in volatility is much greater. We conclude:

Fact #4: *The factor structure in daily commodity futures volatility is much stronger than the factor structure in returns.*

3.3 Time-Varying Commodity Market Integration

It is natural to ask if the principal component analysis in Table 6 is stable over time. In a study of international equity markets by Pukthuanthong & Roll (2009), they introduce a measure of time-varying market integration which is based on a time-varying principal component analysis. Following their approach we regress the return for each commodity on the first 10 PCs computed separately for each year and computed using only the other 14 commodities.[9] Using only daily returns within a given calendar year, the measure of time-varying market integration of Pukthuanthong & Roll (2009) consists of the time series of the annual adjusted R^2 from these return regressions. While Pukthuanthong & Roll (2009) only carry out the integration analysis on returns, we conduct their analysis first for returns and then for expected log-realized volatility.

Our results are reported in Figure 8. The adjusted R^2 from the return regressions are shown with a black line and the adjusted R^2 from the volatility regressions are shown

[9] Our results are robust when changing the number of PCs used.

in grey. Three important conclusions are obtained from Figure 8. First, the evidence for commodity market integration is generally stronger when based on volatility than when based on returns. This is especially true for livestock. Second, the degree of market integration varies greatly by commodity. It is strongest for oil and metals. Third, there is no obvious evidence of a time-trend in the degree of market integration during our 2004-2013 sample period. Indeed, the Pukthuanthong & Roll (2009) market integration measure for returns decreased in 2013 for all commodities except gold and silver. Note further that for several commodities the market appears to be less integrated in 2013 than in 2004 by this measure. We conclude:

Fact #5: *There is little evidence of a time-trend in the degree of integration across commodity futures markets during the 2004-2013 period.*

4 The Stock Market as Factor for Commodity Returns and Volatility

In this section we study the extent to which daily commodity returns and volatilities are integrated with equity market returns and volatilities. We use intraday trades on the Spyder futures contract to compute realized stock market volatility as described in Section 2 above. We then assess the ability of Spyder returns and realized volatility to explain commodity futures and returns in comparison with the four principal components used above. Finally, we construct realized covariance measures which enable us to construct realized betas and realized systematic risk ratios.

4.1 Spyder as an Observed Factor for Commodity Returns and Volatility

The second to last line in Table 4 shows the sample correlation between Spyder returns and the return of each commodity. The correlations range from 6% for lean hogs to 45% for copper. The average correlation with Spyder returns is 22%. The last line in Table 4 shows the correlation between Spyder volatility and the volatility of each commodity. The correlations are generally high with an average of 37%. They are highest for light crude and the three metals. It is lowest for coffee and natural gas which are outliers in this regard. The Spyder volatilities are plotted in grey along with each commodity volatility in Figure 6.

In Table 5.a we assessed the ability of the first four principal components for the commodities to explain the variation in returns and volatility over time. The principal components can be viewed as unobserved factors and the obvious next step is to ask if any observed factors can capture the variation in the PCs and thus in the commodity returns and volatility? We now want to assess the ability of Spyder to serve as an observed factor for the commodity futures market.

To this end we first regress each commodity PC on Spyder to obtain a Spyder-orthogonalized PC series from the residuals. We then regress the return for each commodity on the Spyder return as well as on the four orthogonalized principal components. Again, the PCs used for each commodity are constructed from the remaining 14 commodities only.

Table 6.a contains the results for returns. The results show that the Spyder return is significant in all 15 commodity return regressions. The coefficient ranges from around 15% for livestock and gold to around more than 50% for the other metals and oil. The average regression R^2 is 24%.

In Table 6.b we report on regressions for expected realized volatility. Again, we first orthogonalize each of the PC expected volatility factors with expected Spyder volatility. Table 6.b shows that the Spyder coefficients are significant in 13 out of 15 cases. The exceptions are natural gas and coffee. The average R^2 is 55%.

Consider finally the sample correlation between Spyder volatility and the first four commodity volatility PCs plotted in Figure 7.b. The four correlations are65%, 32%, -2%, and -20% respectively (not reported in the tables). This shows that the most important factor for commodity volatility is very highly correlated with stock market volatility, and we can write:

Fact #6: *The strong common factor in commodity volatility is largely driven by stock market volatility.*

4.2 Constructing Realized Covariance Measures

Our next task is to compute daily realized covariance measures for each commodity with Spyder. The ultimate goal is to compute stock market betas for each commodity that vary daily without imposing a particular dynamic model a priori.

For the bivariate analysis with Spyder, overlapping trading spans between commodity i and Spyder is required to avoid a bias towards zero in the realized covariances. Therefore, opening and closing prices are now redefined as the most recently observed price before the start and end of the overlapping trading span between commodity i and Spyder, respectively. This is illustrated by the grey shading in Figure 3. Spyder trades on NASDAQ from 9.30 to 16.00 in the full sample period, but for commodities the futures trading spans have changed several times. Therefore, the overlapping trade spans varies across commodities and over time. See Table A.4 in the appendix for details.

The construction of realized covariances is similar to the construction of the realized variances above. Within overlapping trade spans, a synchronised 1-minute time-grid between the commodities and Spyder is constructed. From the synchronised prices, 1-minute log-returns on day t are calculated. Then, using overlapping 5-minute returns as in Section 2.4, the realized covariance with 1-minute subsampling is for commodity i calculated as,

$$RCov_{i,t}^{oc} = \frac{n(i)}{n(i)-4} \cdot \frac{1}{5} \sum_{k=1}^{n(i)-4} \tilde{r}_{i,t}^{(k)} \tilde{r}_{SPY,t}^{(k)},$$

where we again use a rescaling to make sure the realized covariance estimate is unbiased. The subsampling technique is used to eliminate bias from market microstructure noise and non-synchronous trades within the overlapping trading span, see e.g. Barndorff-Nielsen & Shephard (2004).

Finally, realized variance and covariance measures are matched to close-close returns following Hansen & Lunde (2005). That is, the realized covariance matrix for the whole day based on intermittent high-frequency data is for commodity i and Spyder given by

$$RCov_{i,t} = \hat{\omega}_1 r_{i,t}^{co} r_{SPY,t}^{co} + \hat{\omega}_2 RCov_{i,t}^{oc}, \tag{3}$$

where $r^{co}_{i,t}$ and $r^{co}_{SPY,t}$ is the realized close-to-open return, e.g. $r^{co}_{i,t} = \ln\left(F^{o}_{i,t}\right) - \ln\left(F^{c}_{i,t-1}\right)$, for commodity i and Spyder respectively, and $\hat{\omega}_1$ and $\hat{\omega}_2$ are again estimated weights similar to the ones reported for the univariate case in Table A.2 in the appendix.

4.3 Realized Spyder Betas for Commodities

We now investigate the ability of the stock market to explain the variation in returns on the 15 commodity futures contracts. Recognizing that the relationship between each commodity and the stock market is likely changing over time, we follow Andersen, Bollerslev, Diebold & Wu (2005) and Patton & Verardo (2012) who use intraday data to compute a daily model-free realized beta for each asset defined by

$$R\beta_{i,t} = \frac{RCov_{i,t}}{RV_{SPY,t}}. \tag{4}$$

The realized covariance, $RCov_{i,t}$, is calculated on the cross-product of the intraday commodity and stock market return using the estimator in equation 3.

In order to filter the realized betas in equation 4 we estimate an ARMA(1,1) on each realized beta series. The time series of the filtered betas are plotted in Figure 9 and the ARMA coefficients are reported in Table A.3.b in the appendix. The appendix also contains a plot of the raw realized betas in Figure A.5.

We plot the beta series along with bootstrapped 75% and 90% confidence intervals constructed by resampling with replacement from the ARMA residuals. The confidence intervals are based on $10,000$ bootstrap samples.

Figure 9 shows that the realized betas were close to zero until 2008, then rose dramatically in many cases to and even beyond one. The decrease in commodity betas since 2010 are equally interesting. By the end of 2013 all the realized betas were back to zero, the level at which they began at the onset of financialization in 2004. The betas were highest for energy and metals. The remarkable rise and fall of the commodity betas are shared by all except for lean hogs and feeder cattle which stayed close to zero throughout the period. We assert:

Fact #7: *Commodity betas with the stock market were high during 2008-2010 but have since returned to a level close to zero.*

The beta of an asset does not tell us how much of the variance in the asset's return is driven by the market factor. To this end we define the Systematic Risk Ratio (SRR) for commodity i by

$$SSR_{i,t} = \frac{R\beta_{i,t}^2 \cdot RV_{SPY,t}}{RV_{i,t}}. \tag{5}$$

The SSR can thus be interpreted as the fraction of commodity i variance that is explained by market variance. Clearly, $0 \leq SSR_{i,t} \leq 1$. Our use of high-frequency data enables us to compute a systematic risk ratio for each commodity on each day. We again estimate an ARMA(1,1) on the raw SRR to filter the series. The ARMA coefficients and the raw SRR series are provided in the appendix. Figure 10 plots the time series of the filtered SRR along with the average SRR over the sample period with bootstrapped 75% and 90% confidence intervals for the ARMA residuals. Note how the SRR was close to zero for all commodities before 2008, it then rose to substantial levels - in particular for energy and metals - before returning to zero at the end of 2013.

4.4 Capturing Nonlinear Dependence with Spyder

So far we have focused on linear dependence between commodity futures and the stock market. In this section we investigate nonlinear relationships.

Figure 11 shows the threshold correlations for each daily commodity futures returns versus the daily return on the stock index ETF. Threshold correlations are computed as

$$\rho_{ij}(u) = \begin{cases} \text{Corr}\left(r_i, r_j \,|r_i < F_i^{-1}(u) \text{ and } r_j < F_j^{-1}(u)\right) & \text{if } u < 0.5 \\ \text{Corr}\left(r_i, r_j \,|r_i \geq F_i^{-1}(u) \text{ and } r_j \geq F_j^{-1}(u)\right) & \text{if } u \geq 0.5, \end{cases}$$

where u is a threshold between 0 and 1, and $F_i^{-1}(u)$ is the empirical quantile of the univariate distribution of r_i. The horizontal axis in Figure 11 denotes the percentile used to define each threshold for the correlation on the vertical axis. We only compute the threshold correlation when at least 60 observations are available, and again we perform the calculations with (black lines) and without (grey lines) rollover returns.

The dashed lines in Figure 11 denotes the threshold correlation from a bivariate normal distribution with a correlation equal to the sample correlation from the futures data. Figure 11 shows that the deviation from bivariate normality is large for oil, all metals, and for some of the agriculturals as well. While the simple linear correlations can be small, the threshold correlations are often large. It is important to keep in mind that the threshold correlations are unconditional - they are computed once from the entire sample. Figure 9 above showed that the stock market exposure of each commodity varies greatly over time. This dynamic beta may be the cause of the large threshold correlations we see in Figure 11.

We next compute threshold correlations for the expected $\log(RVol_t)$ for each commodity versus $\log(RVol_t)$ in Spyder. They are plotted in Figure 12 which shows that large positive shocks to volatility in the stock market are generally highly correlated with large positive shocks to the volatility of each commodity. This is particularly true for oil and metals but also for most of the agriculturals.

5 The Distribution of Commodity Returns and Shocks

So far, we have focused on the dependence structure across commodity futures returns and volatilities. To obtain a fully specified model, we need to make distributional assumptions as well. We therefore now investigate the distribution of futures returns and the extent to which standardizing the daily commodity futures returns by expected realized volatility will produce a close to normally distributed series of commodity futures shocks.

5.1 The Distribution of Commodity Futures Returns

As mentioned above, Table 2.a reports various sample statistics for the daily log returns on our 15 commodity futures. During our 2004-2013 sample period, the average log return was highest for silver at 4.7 bps per day and lowest for natural gas at -1.5 bps per day. Heating oil, copper, gold, and light crude appreciated strongly during the sample as well. Natural gas had by far the highest average daily volatility at more than 3.3% per day compared with feeder cattle which only had a volatility of less than 1%

per day. These findings also hold when the rollover returns are removed. We note that the cross-commodity variation in commodity return mean and volatility is large.

Table 2.a also provides information on the unconditional distribution of daily returns. Cotton contains by far the largest negative daily return at -39.3%, which is a 19 standard deviation move that occurred on February 16, 2011. This corresponds to a rollover return. Lean hogs has the largest positive daily return at 25.9% which also corresponds to a rollover return. However, even with rollovers removed, the minimum and maximum observations for all commodities are quite extreme; perhaps with the exception of live and feeder cattle. Not surprisingly, given the extreme minimum value, cotton has the largest negative skewness at -2.5. Lean hogs has the highest positive skewness at 2.0. The cross-sectional range in skewness is thus large. Kurtosis is in excess of the Gaussian value of 3 for all 15 commodities. It is not surprisingly largest at 54.2 for cotton whereas coffee has the lowest value at 3.9. Results are very different for both skewness and kurtosis for most commodities when rollover returns are removed. In this case, silver is the commodity with both the largest (negative) skewness and kurtosis. These findings are not surprising because sample estimates of skewness and kurtosis are well-known to be heavily influenced by a few large outlying observations, see Kim & White (2004).

Figure A.6 in the appendix reports QQ plots against the normal distribution which provides further evidence on the non-normality of raw returns and the important impact on the tail of the distribution from roll-returns. For now, we simply note that the unconditional distribution of commodity returns is highly leptokurtotic with substantial variation in skewness across commodities.

5.2 Commodity Futures Shocks

If we assume that e_t is i.i.d. normally distributed, which is sensible given Table 3.b, then we can use the moment-generating function of a normal variable to construct our measure of expected $RVol_t$, as

$$E_{t-1}[RVol_t] = \exp\left[E_{t-1}[\log(RVol_t)]\right] \cdot \exp\left[\frac{1}{2}\sigma_e^2\right],$$

where $\sigma_e^2 = Var[e_t]$. We can now compute the time series of unexpected commodity futures shocks, z_t, via

$$z_t = \frac{r_t - \frac{1}{T}\sum_{i=1}^{T} r_i}{E_{t-1}[RVol_t]}.$$

Note that we deliberately standardize returns by the ex-ante expected volatility and not by the ex-post realized volatility as is sometimes done in the literature. We are interested in the conditional distribution of one-day-ahead returns, that is, the distribution of z_t. We use the full-sample average as an estimate of the mean return, but using a recursive or rolling sample average has no impact on the conclusions below.

Table 2.b reports the sample statistics for z_t for each of our 15 commodities. As expected, the mean of the shock is approximately zero and the standard deviation is close to one for each commodity. More importantly, note that the non-normality strongly evident in Table 2.a is much less apparent in Table 2.b. Figure 13 shows the quantile-quantile plots of z_t. There is a drastic difference between the quantile-quantile plots of returns in Figure A.6 and the quantile-quantile plots of shocks in Figure 13, even though the data quantiles on the vertical axis still exhibits fatter tails than the standard normal quantiles on the horizontal axis particularly for the agricultural commodities. We conclude that:

Fact #8: *Commodity futures returns standardized by expected realized volatility are closer to normally distributed than the returns themselves but still display leptokurtosis.*

Finding a parametric distribution that fits the return shocks in Figure 13 is important in option valuation and risk management applications but not for our analysis and we therefore leave this task for future work. The normal-lognormal mixture model suggested in Andersen, Bollerslev, Diebold & Labys (2003) or the t-distribution used in Maheu & McCurdy (2011) should be viable approaches in this regard.

Note also that we have assumed that the volatlity of volatility, σ_e^2, is constant. Dynamics could be modelled using the approach in Maheu & McCurdy (2011) but they find that it makes little difference for the purpose of density forecasting and so we simply let σ_e^2 be constant. We instead turn our attention to our main task which is uncovering the stylized fact of the joint distribution of commodity futures returns as well as of their volatilities.

Figure A.7 in the appendix shows the threshold correlations for the return shocks, z_t. Note that they are very close to zero for all commodities except coffee. Coffee is clearly affected by rollover returns, but is also the only commodity with periodicity in volatility as it is evident from Figure 5. Comparing Figure 11 with Figure A.7 suggests that the large threshold correlations in returns are driven by time-varying volatility, so that when dynamically standardized, commodity returns show little evidence of significant threshold correlations.

6 Conclusion

We have addressed the following questions:

First, what are the stylized facts of commodity futures returns and volatility post financialization? Second, is there a factor structure in commodity futures returns? Third, is the stock market driving the common component of commodity futures returns? Fourth, do the stock market betas of commodity futures returns vary significantly over time? Fifth, does the ratio of commodity futures return volatility explained by the stock market beta change over time? Sixth, is there a factor structure in commodity futures volatility? Finally, is stock market volatility driving the common component of commodity futures volatility?

Analyzing almost a billion trades on 15 commodity futures contracts for the 2004-2013 period, we have uncovered the following:

Fact #1: *Daily realized commodity futures volatility has extremely high persistence.*

Fact #2: *The logarithm of realized commodity futures volatility is close to normally distributed.*

Fact #3: *There is some evidence of a factor structure in daily commodity futures returns excluding meats.*

Fact #4: *The factor structure in daily commodity futures volatility is much stronger than the factor structure in returns.*

Fact #5: *There is little evidence of a time-trend in the degree of integration across commodity futures markets during the 2004-2013 period.*

Fact #6: *The strong common factor in commodity volatility is largely driven by stock market volatility.*

Fact #7: *Commodity betas with the stock market were high during 2008-2010 but have since returned to a level close to zero.*

Fact #8: *Commodity futures returns standardized by expected realized volatility are closer to normally distributed than the returns themselves but still display leptokurtosis.*

An appropriate option valuation model or portfolio risk model for commodity futures needs to incorporate these features. We have deliberately taken a model-free approach in this paper. But our results suggest that the parametric models developed in Andersen et al. (2003) and Maheu & McCurdy (2011) present viable approaches.

Our results also show that the fear of increased volatility in the commodity markets as a consequence of financialization is largely overblown. Commodity volatility has not trended up nor has commodity return covariance with the stock market trended up. Commodity returns are not riskier now than a decade ago and commodities do not appear to have lost their ability to diversify equity market exposure.

Finally, our results have important implications for understanding the cross-section of commodity futures returns. In recent work focusing on the U.S. equity market, Chen & Petkova (2012), Duarte, Kamara, Siegel & Sun (2014), and Herskovic, Kelly, Lustig & Van Nieuwerburgh (2014) find very strong evidence of factor structure in idiosyncratic volatility. We find the same to be true in commodity futures volatility. Developing an asset pricing framework that can capture this feature presents an important challenge for future work.

References

Andersen, T., Bollerslev, T., Christoffersen, P. & Diebold, F. (2013), Financial risk measurement for financial risk management, *in* G. Constantinides, M. Harris & R. Stulz, eds, 'Handbook of the Economics of Finance', Vol. 2B, Elsevier, pp. 1127–1220.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**(453), 42–55. Correction published in 2003, volume 98, page 501.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Wu, J. (2005), 'A framework for exploring the macroeconomic determinants of systematic risk', *American Economic Review* **95**(2), 398–404.

Ang, A. & Chen, J. (2002), 'Asymmetric correlations of equity portfolios', *Journal of Financial Economics* **63**(3), 443–494.

Ang, A., Hodrick, R. J., Xing, Y. & Zhang, X. (2006), 'The cross-section of volatility and expected returns', *Journal of Finance* **61**(1), 259–299.

Ang, A., Hodrick, R. J., Xing, Y. & Zhang, X. (2009), 'High idiosyncratic volatility and low returns: International and further U.S. evidence', *Journal of Financial Economics* **91**(1), 1–23.

Baker, S. D. (2014), The financialization of storable commodities, Working paper, University of Virginia.

Baker, S. D. & Routledge, B. (2012), The price of oil risk, Working paper, Carnegie Mellon University.

Bakshi, G., Gao, X. & Rossi, A. (2013), Asset pricing models that explain the cross-section and time-series of commodity returns, Working paper, University of Maryland.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2009), 'Realised kernels in practice: Trades and quotes', *Econometrics Journal* **12**, C1–C32.

Barndorff-Nielsen, O. E. & Shephard, N. (2004), 'Econometric analysis of realised co-variation: High frequency based covariance, regression and correlation in financial economics', *Econometrica* **72**, 885–925.

Barndorff-Nielsen, O. E. & Shephard, N. (2007), Variation, jumps and high frequency data in financial econometrics, *in* R. Blundell, T. Persson & W. K. Newey, eds, 'Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress', Econometric Society Monographs, Cambridge University Press, pp. 328–372.

Basak, S. & Pavlova, A. (2014), A model of financialization of commodities, Working paper, London Business School.

Bekaert, G., Harvey, C. R., Lundblad, C. T. & Siegel, S. (2011), 'What Segments Equity Markets?', *Review of Financial Studies* **24**(12), 3841–3890.

Bessembinder, H. (1992), 'Systematic risk, hedging pressure, and risk premiums in futures markets', *Review of Financial Studies* **5**(4), 637–67.

Bessembinder, H. & Chan, K. (1992), 'Time-varying risk premia and forecastable returns in futures markets', *Journal of Financial Economics* **32**(2), 169–193.

Boons, M., Roon, F. & Szymanowska, M. (2013), 'The stock market price of commodity risk', *Working paper, Tilburg University* .

Büyüksahin, B., Haigh, M. S. & Robe, M. A. (2010), 'Commodities and equities: A "market of one"?', *Journal of Alternative Investments* **12**.

Carrieri, F., Errunza, V. & Hogan, K. (2007), 'Characterizing world market integration through time', *Journal of Financial and Quantitative Analysis* **42**, 915–940.

Chen, Z. & Petkova, R. (2012), 'Does idiosyncratic volatility proxy for risk exposure?', *Review of Financial Studies* **25**(9), 2745–2787.

Cheng, I.-H. & Xiong, W. (2013), The financialization of commodity markets, NBER working papers, National Bureau of Economic Research, Inc.

Daskalaki, C., Kostakis, A. & Skiadopoulos, G. (2014), 'Are there common factors in individual commodity futures returns?', *Journal of Banking & Finance* **40**(C), 346–363.

Duarte, J., Kamara, A., Siegel, S. & Sun, C. (2014), 'The systematic risk of idiosyncratic volatility', *SSRN Working Paper* .

Engle, R. F. (2002), 'Dynamic conditional correlation - a simple class of multivariate GARCH models,', *Journal of Business & Economic Statistics* **20**(3), 339–350.

Gorton, G. B., Hayashi, F. & Rouwenhorst, G. K. (2012), 'The fundamentals of commodity futures returns', *Yale ICF Working paper* **No. 07-08**.

Gorton, G. & Rouwenhorst, K. G. (2006), 'Facts and fantasies about commodity futures', *Financial Analysts Journal* pp. 47–68.

Hamilton, J. D. & Wu, J. C. (2014), Effects of index-fund investing on commodity futures prices, NBER working paper, National Bureau of Economic Research, Inc.

Hansen, P. R. & Lunde, A. (2005), 'A realized variance for the whole day based on intermittent high-frequency data', *Journal of Financial Econometrics* **3**, 525–544.

Hansen, P. R. & Lunde, A. (2011), Forecasting volatility using high-frequency data, *in* M. Clements & D. Hendry, eds, 'The Oxford Handbook of Economic Forecasting', Oxford: Blackwell, chapter 19, pp. 525–556.

Henderson, B., Pearson, N. & Wang, L. (2013), New evidence on the financialization of commodity markets, Working paper, University of Illinois at Urbana-Champaign.

Herskovic, B., Kelly, B., Lustig, H. & Van Nieuwerburgh, S. (2014), The common factor in idiosyncratic volatility: Quantitative asset pricing implications, Working paper, University of Chicago, Booth School of Business.

Kim, T.-H. & White, H. (2004), 'On more robust estimation of skewness and kurtosis', *Finance Research Letters* **1**(1), 56–73.

Longin, F. & Solnik, B. (2001), 'Extreme correlation of international equity markets', *The Journal of Finance* **56**(2), 649–676.

Maheu, J. M. & McCurdy, T. H. (2011), 'Do high-frequency measures of volatility improve forecasts of return distributions?', *Journal of Econometrics* **160**(1), 69–76.

Patton, A. J. & Verardo, M. (2012), 'Does beta move with news? firm-specific information flows and learning about profitability', *Review of Financial Studies* **25**(9), 2789–2839.

Pukthuanthong, K. & Roll, R. (2009), 'Global market integration: An alternative measure and its application', *Journal of Financial Economics* **94**(2), 214–232.

Singleton, K. J. (2014), 'Investor flows and the 2008 boom/bust in oil prices', *Management Science* **60**(2), 300–318.

Szymanowska, M., De Roon, F., Nijman, T. & Van Den Goorbergh, R. (2014), 'An anatomy of commodity futures risk premia', *The Journal of Finance* **69**(1), 453–482.

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

Zhang, L., Mykland, P. A. & Aït-Sahalia, Y. (2005), 'A tale of two time scales: Determining integrated volatility with noisy high frequency data', *Journal of the American Statistical Association* **100**, 1394–1411.

Figure 1: Daily Closing Prices for 15 Commodities and Spyder. Normalized.

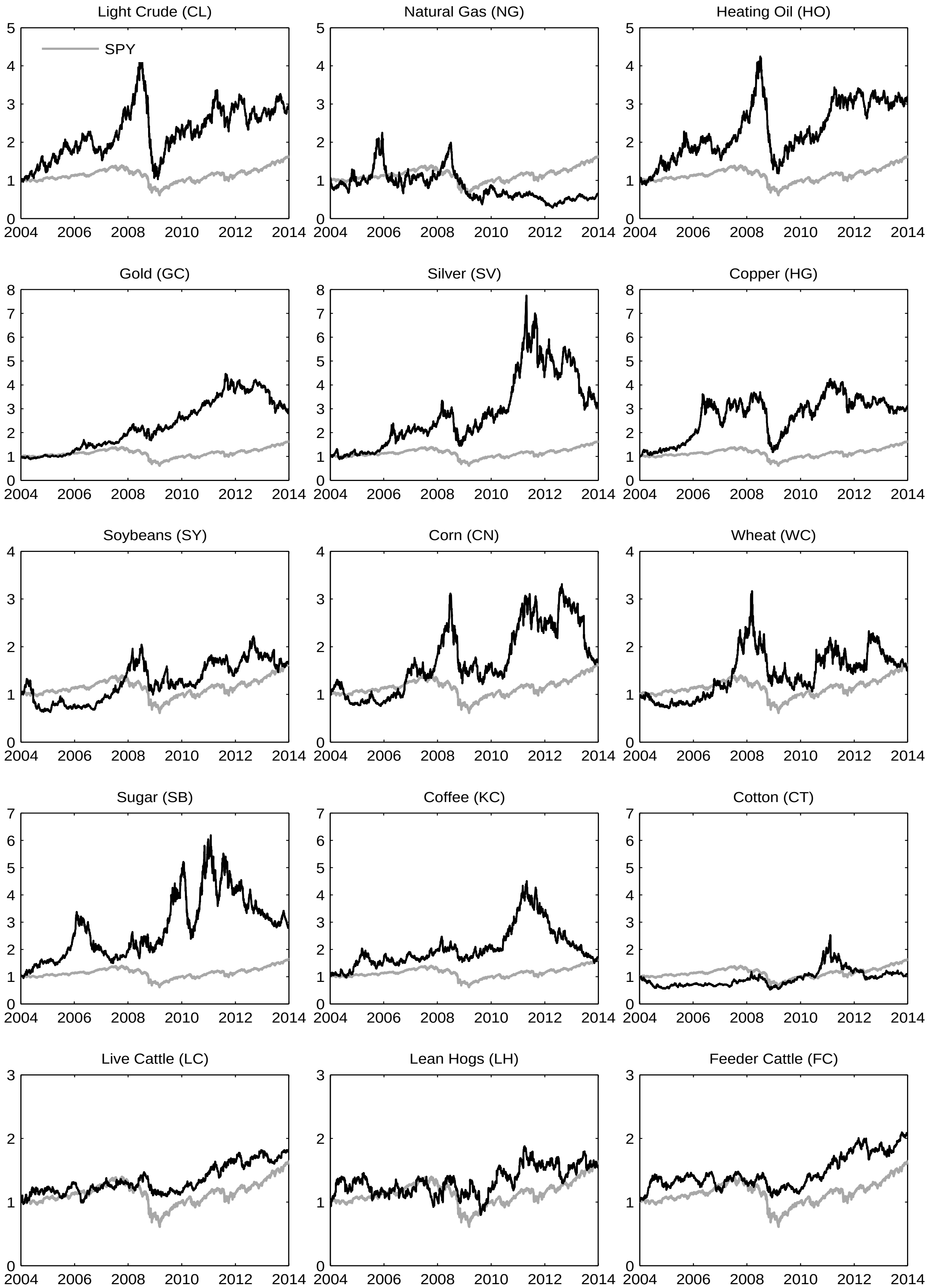
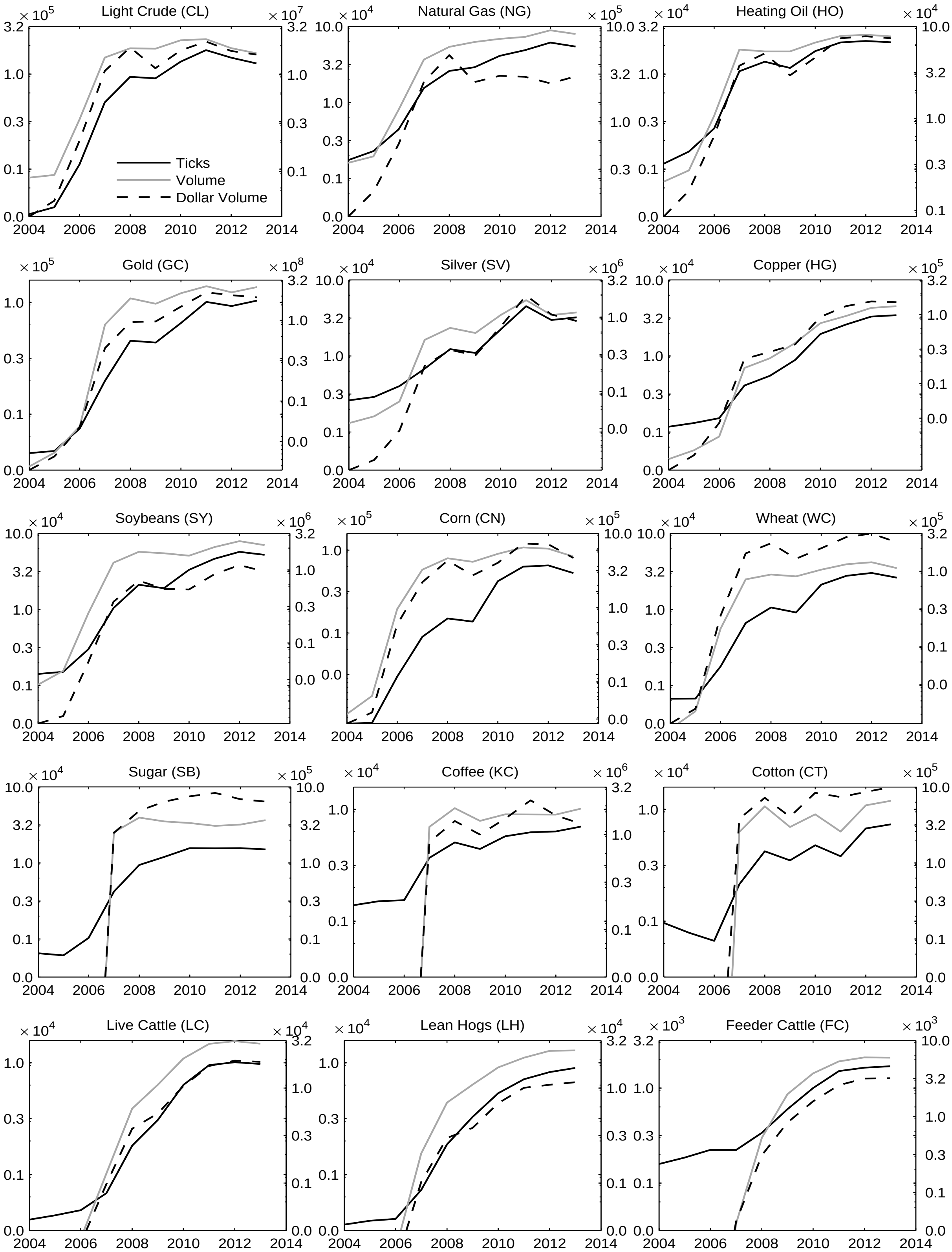


Figure 2: Number of Ticks, Volume and Dollar Volume for 15 Commodities. Annual Averages.



Notes: Ticks (left axis) denote the number of transactions, which may consist of one or more futures contracts (pit and electronic). Volume (left axis) denotes the number of futures contracts traded (electronic only), Dollar Volume (right axis) denotes the dollar value of the futures contracts traded using the closing price each day (electronic only).

Figure 3: Daily Trading Hours for 15 Commodities and Spyder.

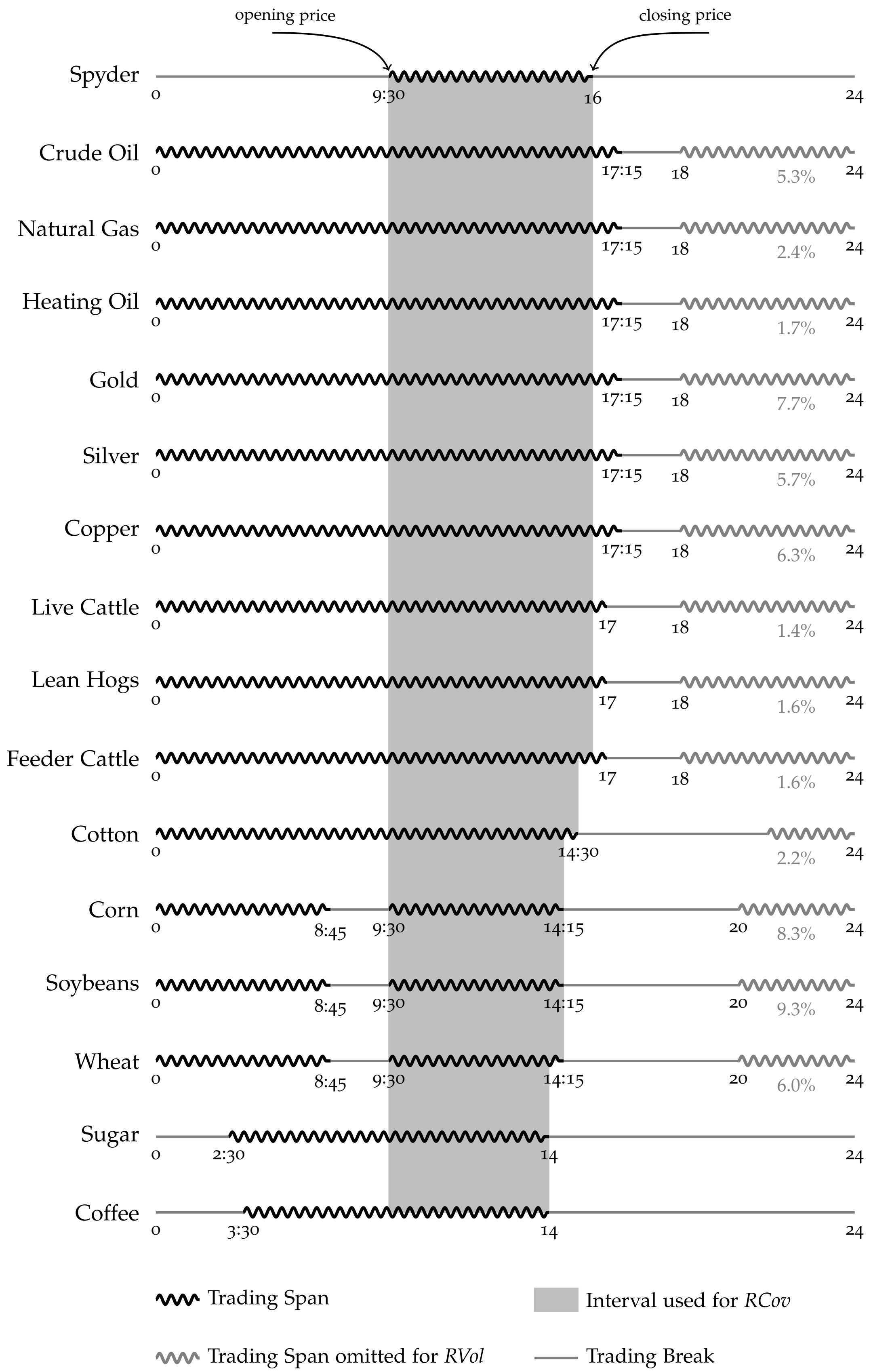
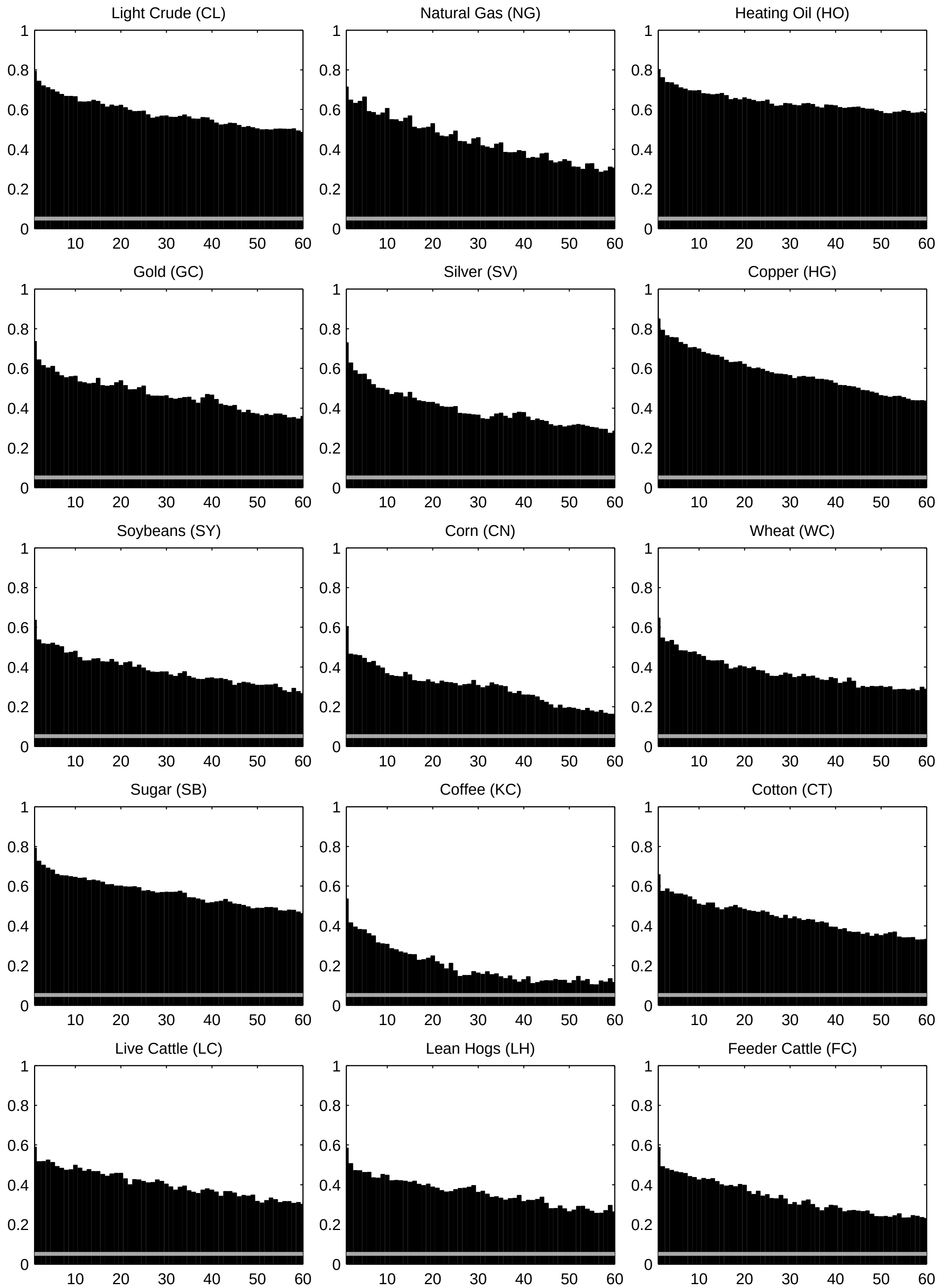
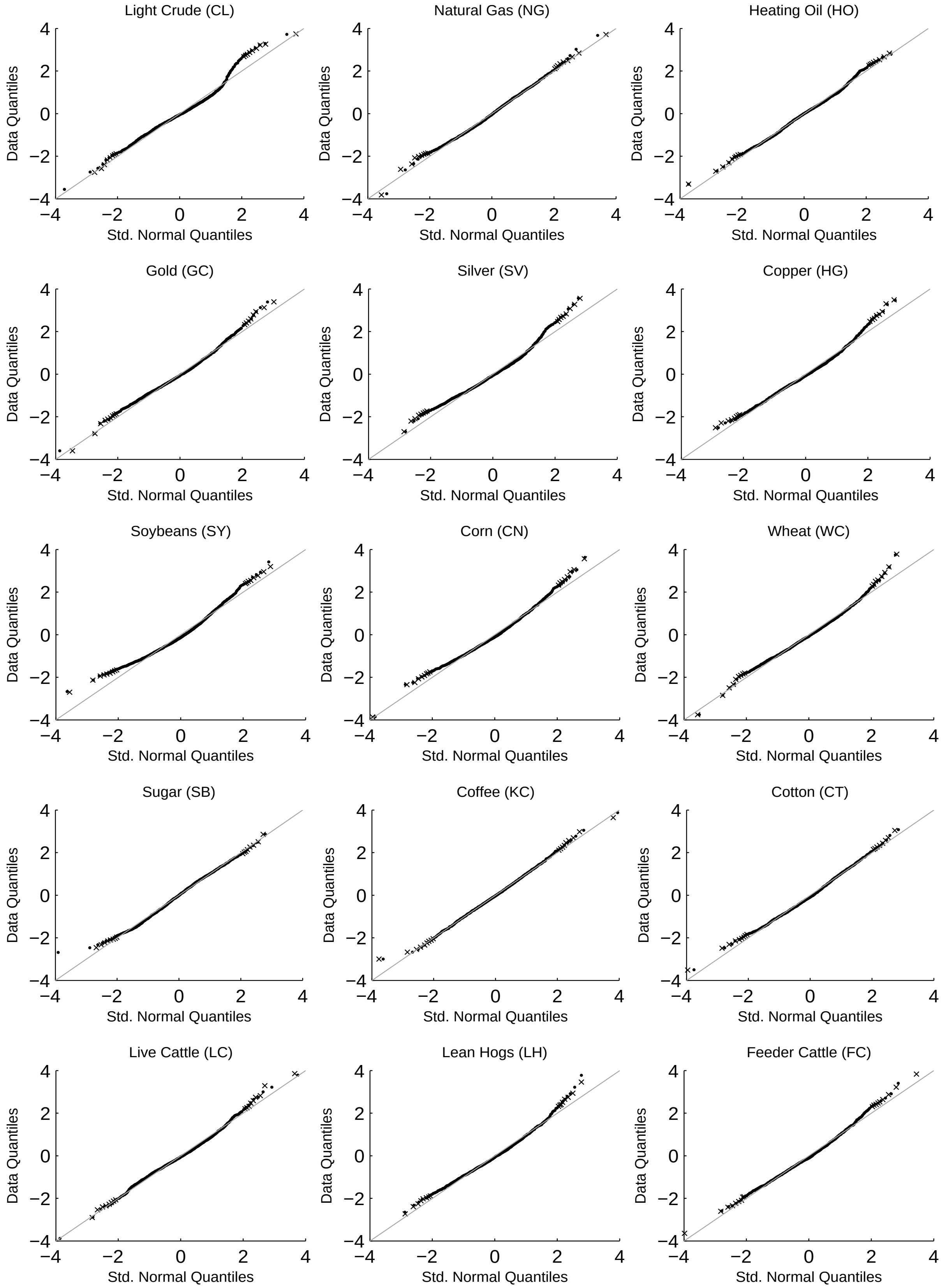


Figure 4: Autocorrelation Function of Daily Log-Realized Volatility.



Notes: The grey line indicates the upper bound on the 99% confidence interval under the assumption that the series are Gaussian white noise. The horizontal axis indicates the lag order in trading days.

Figure 5: QQ-Plot of Daily Log-Realized Volatility.



Notes: All observations are marked by dots. Crosses indicate large quantiles that do not correspond to rollover returns.

Figure 6: Daily Expected Log-Realized Volatility.

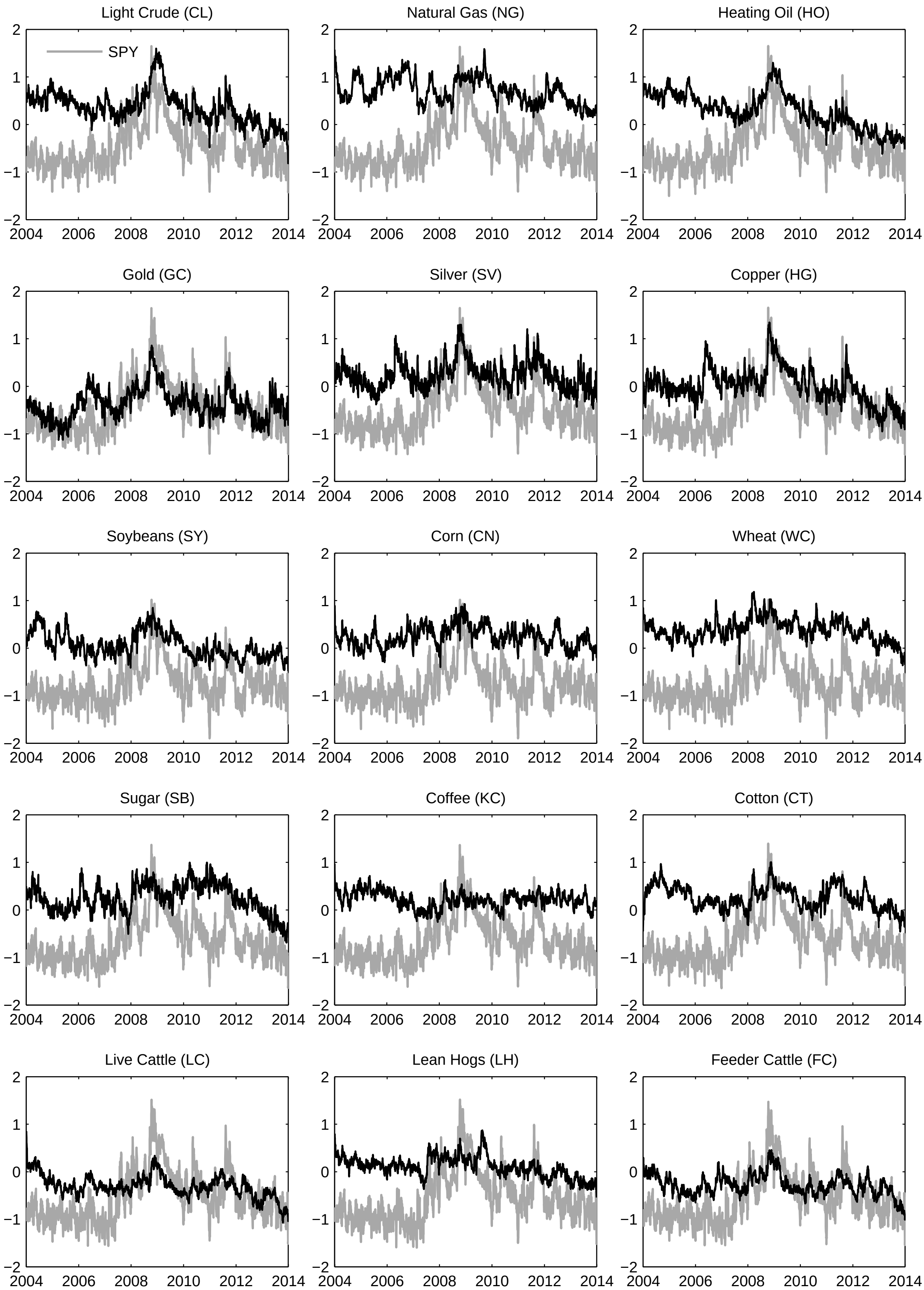


Figure 7.a: First Principal Component of Log-Returns. Cumulative Sum.

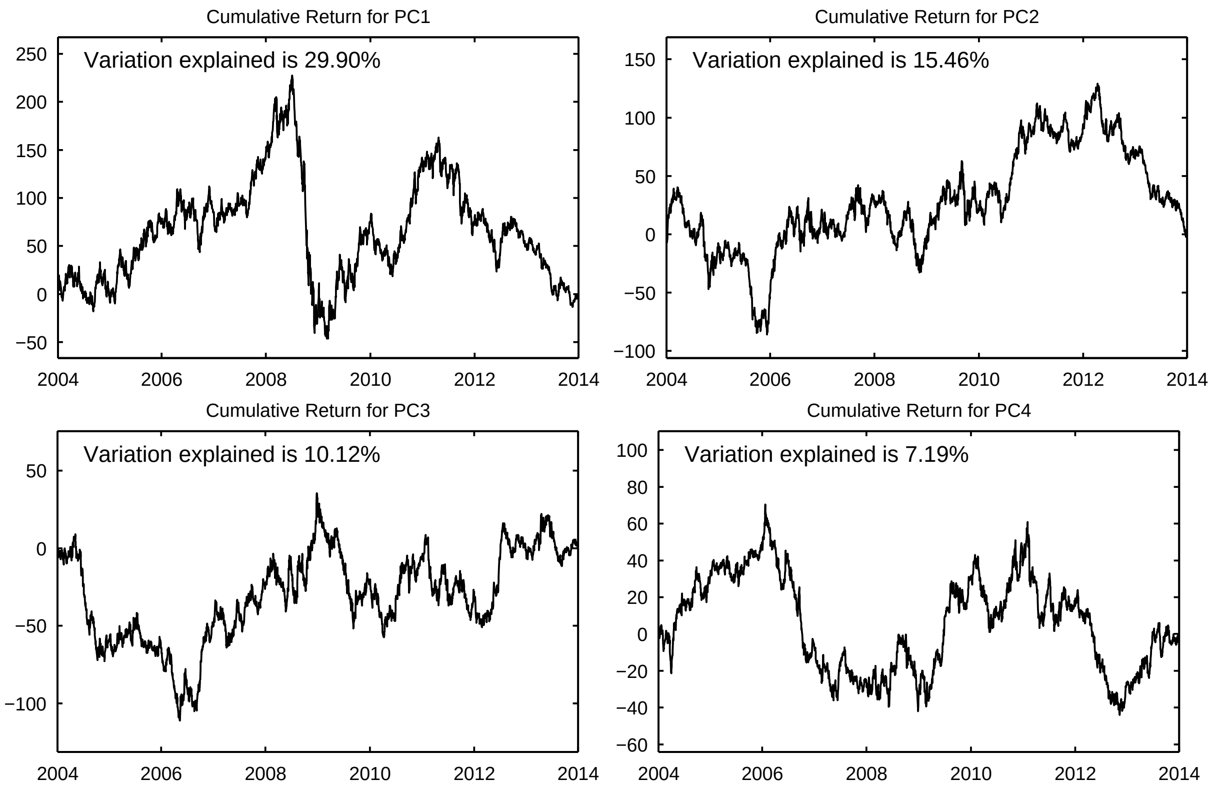


Figure 7.b: First Principal Component of Log-Realized Volatility.

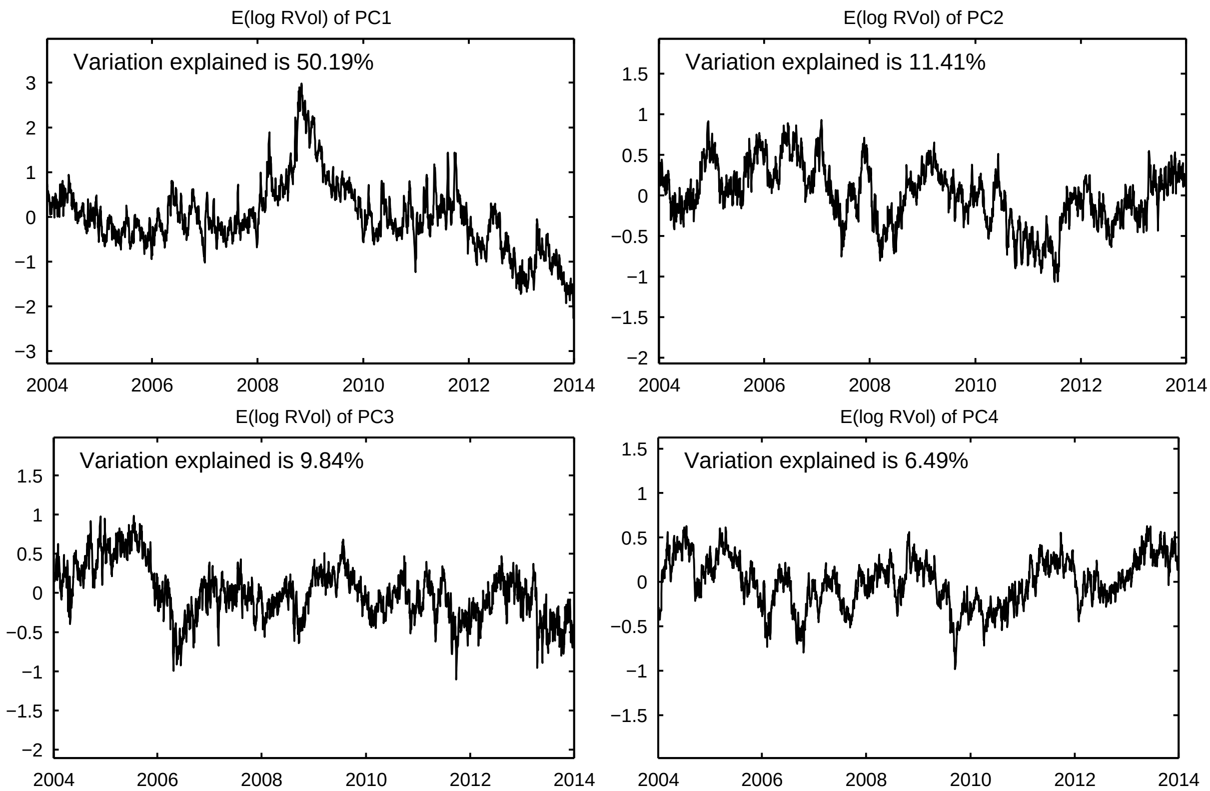


Figure 8: Time-Varying Market Integration in Returns and Volatility.

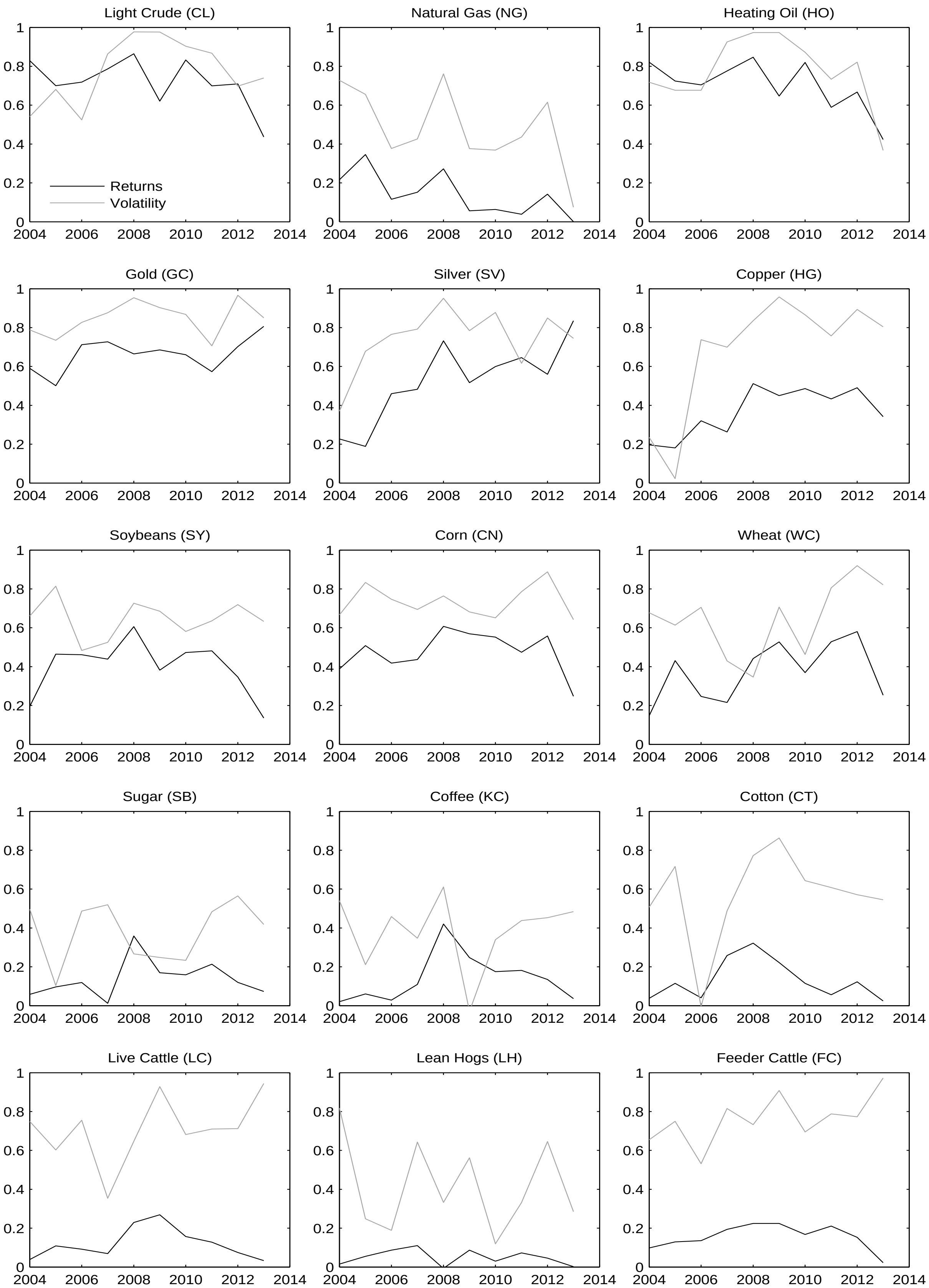
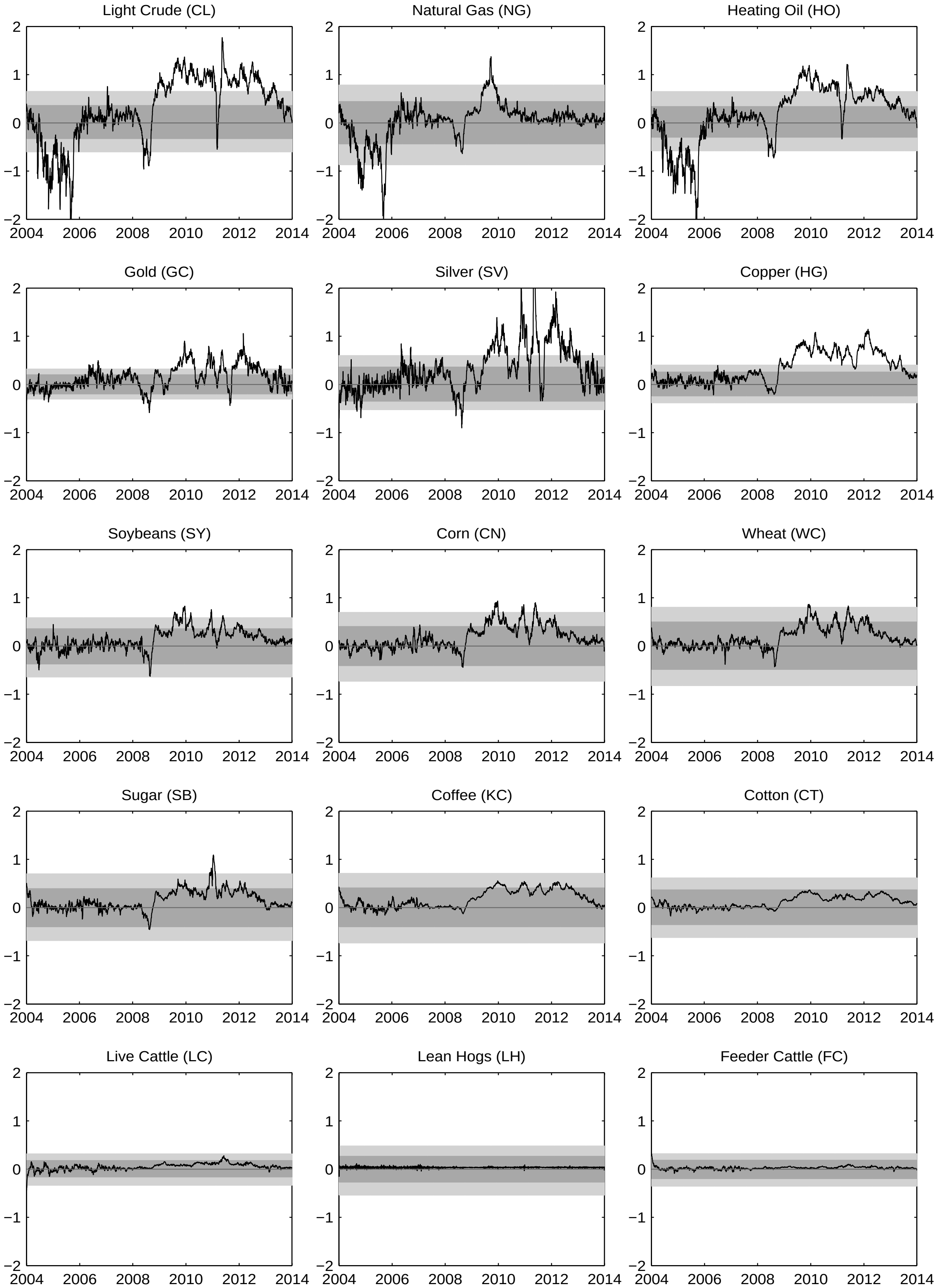
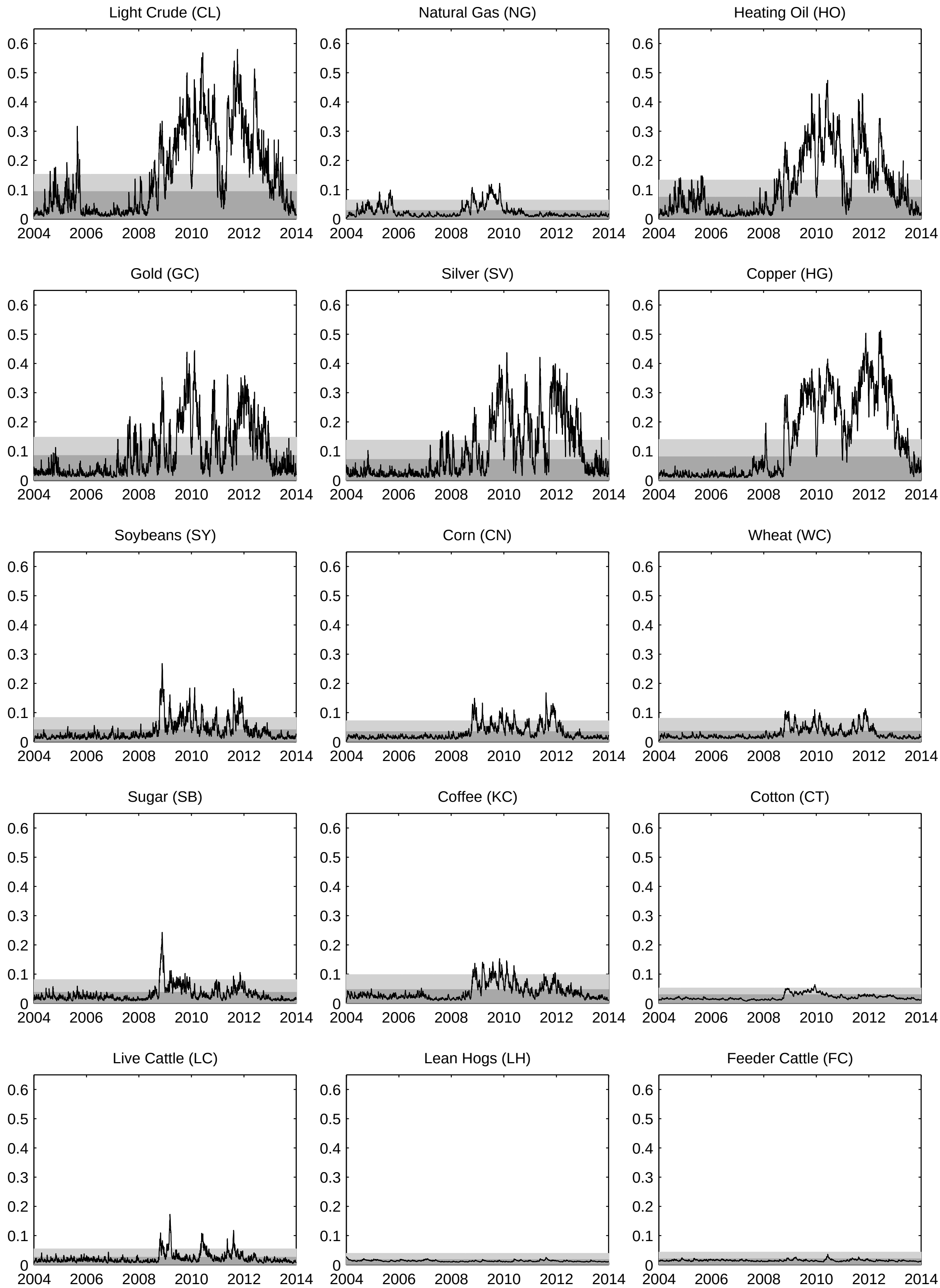


Figure 9: Expected Daily Beta with the Stock Market.



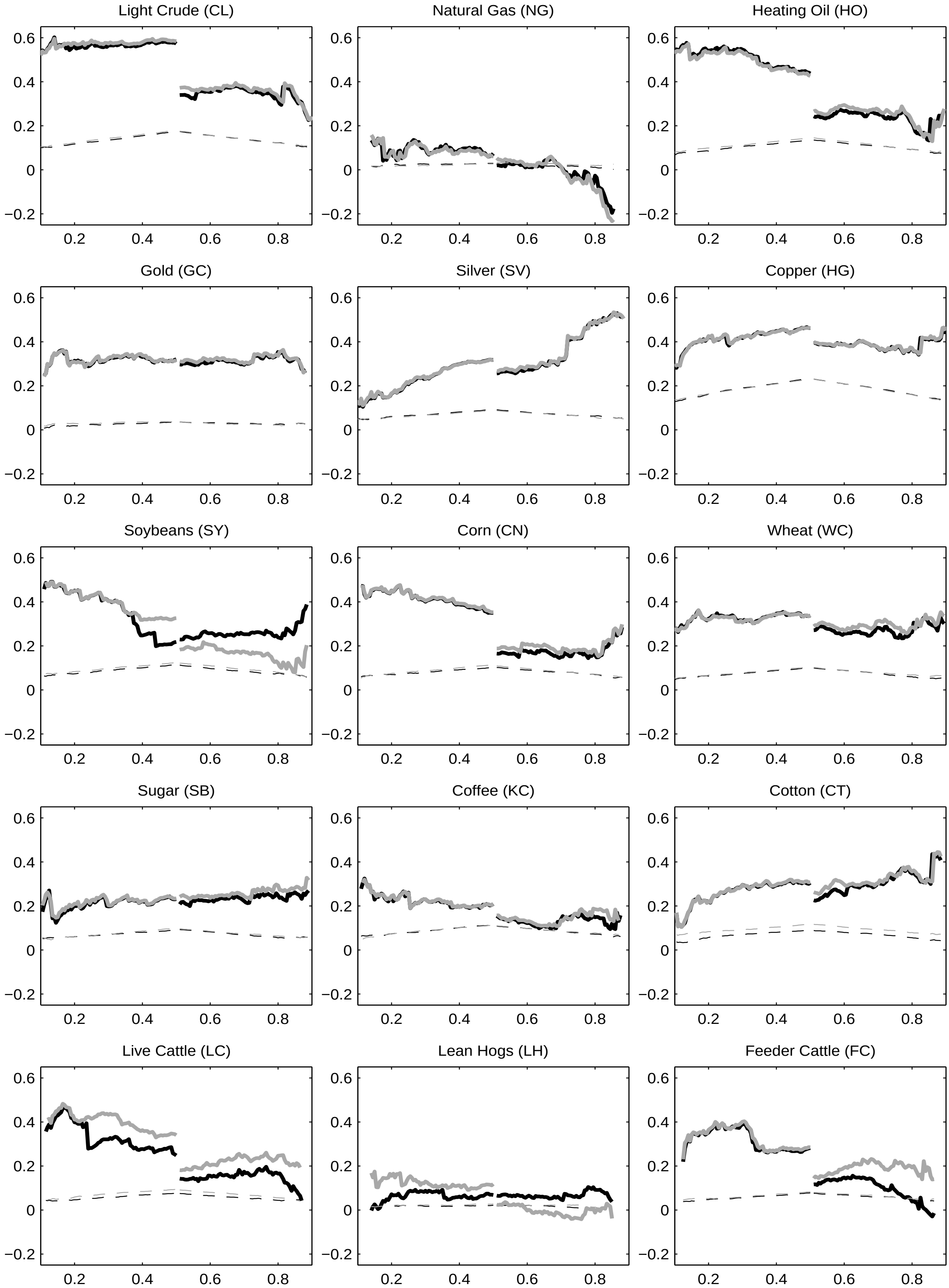
Notes: The shaded areas denote 75% (light grey) and 90% (dark grey) bootstrapped confidence intervals for the stock market betas.

Figure 10: Expected Stock Market Systematic Risk Ratio for 15 Commodities.



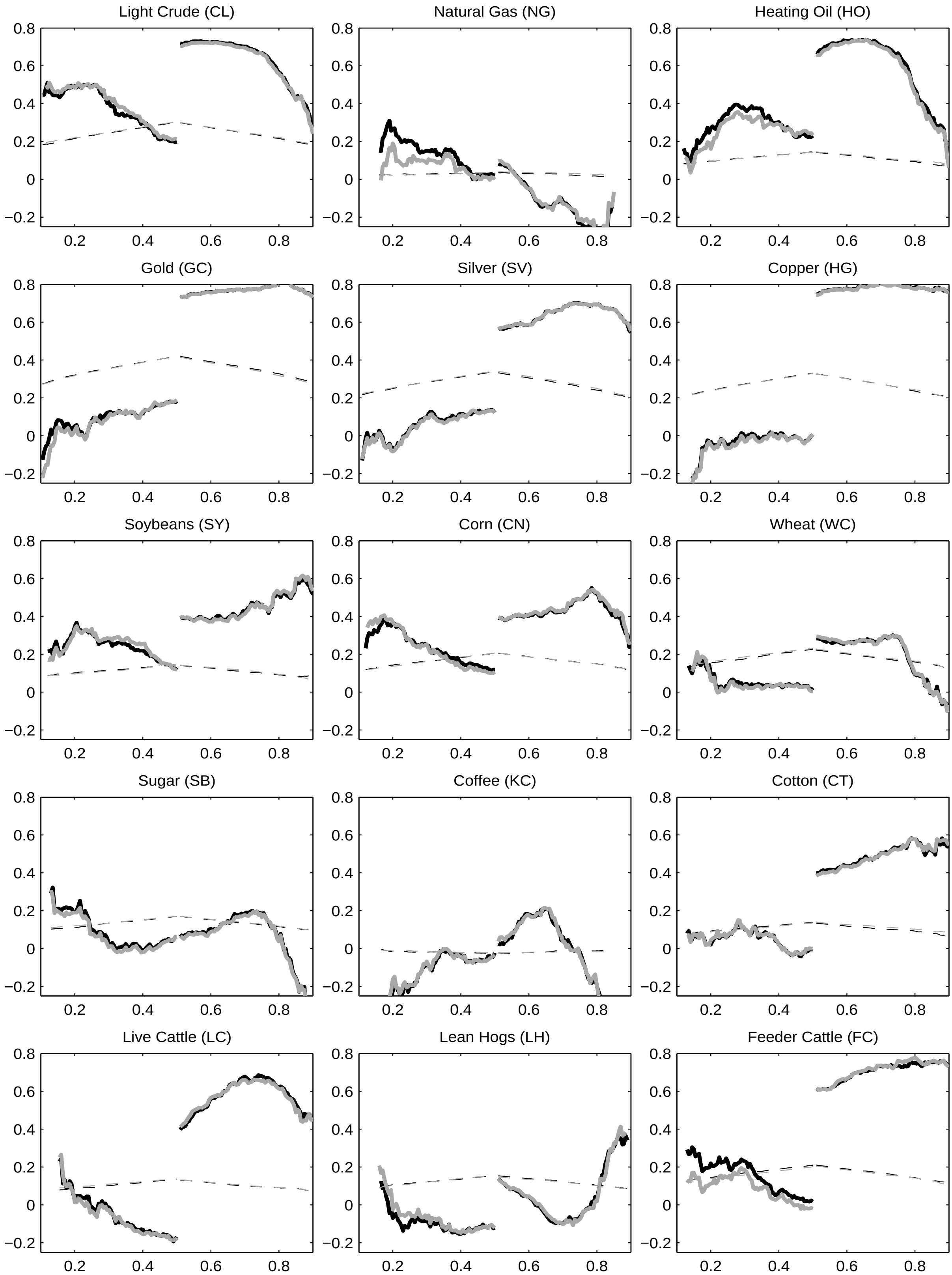
Notes: The shaded areas denote 75% (light grey) and 90% (dark grey) bootstrapped confidence intervals for the systematic risk ratio.

Figure 11: Threshold Correlation Between Commodity and Stock Market Returns.



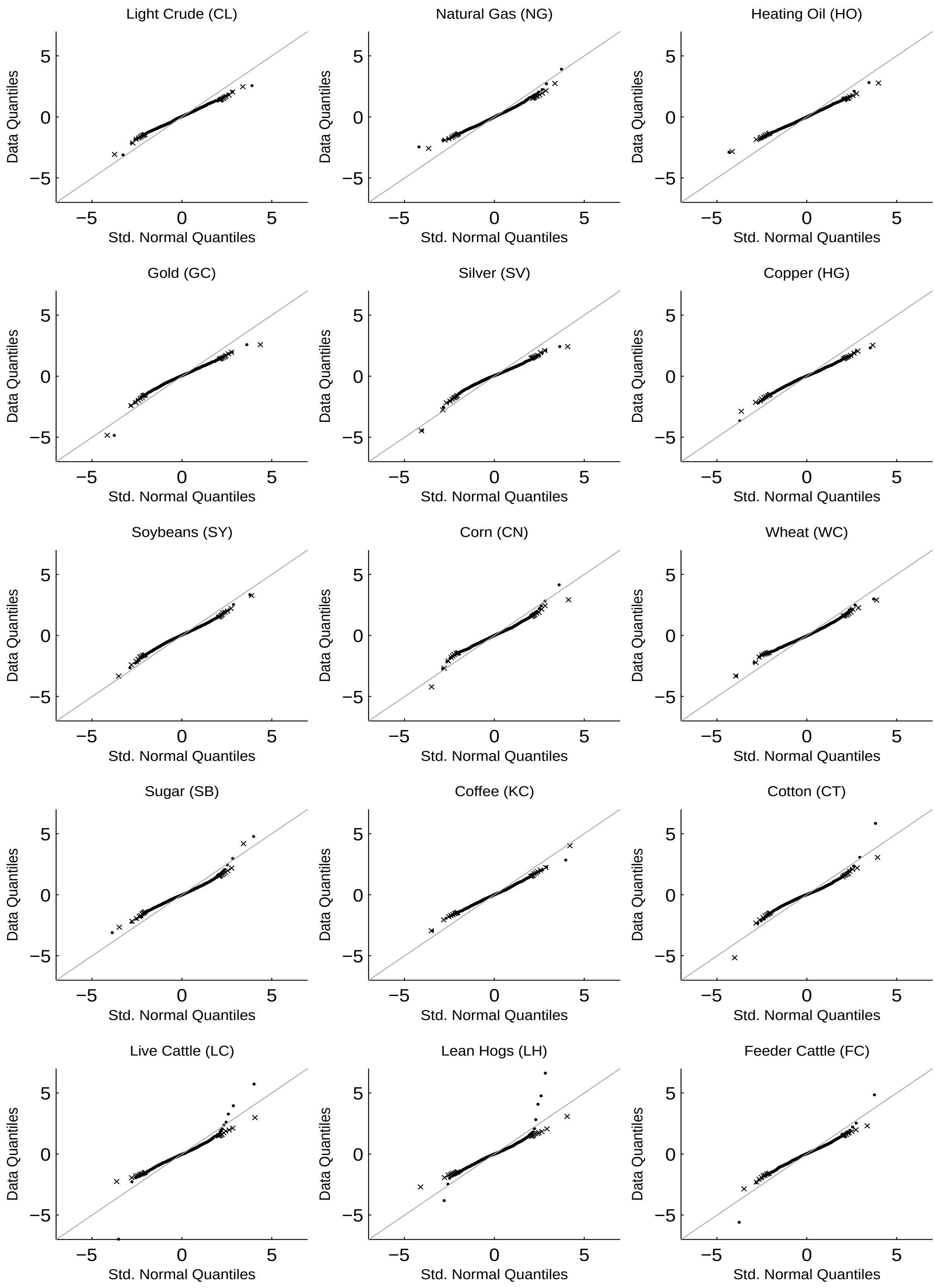
Notes: Threshold correlations with rollover days included (black lines) and with rollover days excluded (grey lines). The dashed lines denote threshold correlations for the bivariate Gaussian distribution.

Figure 12: Threshold Correlation Between Expected Log Realized Volatility for Spyder and 15 Commodities.



Notes: Threshold correlations with rollover days included (black lines) and with rollover days excluded (grey lines). The dashed lines denote threshold correlations for the bivariate Gaussian distribution. The horizontal axis indicates thresholds in percentiles.

Figure 13: QQ-Plot of Daily Return Shocks.



Notes: All observations are marked by dots. Crosses indicate large quantiles that do not correspond to rollover returns.

Table 2.a: Sample Statistics for Daily Futures Returns (in percent).

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
Mean	0.044	−0.015	0.047	0.042	0.047	0.046	0.021	0.022	0.020	0.042	0.024	0.005	0.024	0.020	0.030
Std. Dev.	2.220	3.301	1.986	1.248	2.318	1.974	1.866	2.121	2.238	2.227	1.872	2.100	1.053	1.902	0.983
Min	−12.961	−14.362	−7.605	−8.461	−22.012	−11.294	−27.708	−19.222	−18.419	−11.950	−8.233	−39.341	−8.019	−14.466	−7.687
Max	19.211	25.694	8.979	9.524	12.365	11.359	8.628	12.604	12.460	13.680	8.255	12.195	7.840	25.923	6.075
Skewness	0.044	0.710	0.033	−0.499	−1.197	−0.290	−2.135	−0.113	0.025	0.057	0.060	−2.523	0.200	2.023	−0.095
Kurtosis	7.371	7.181	4.122	7.931	11.490	6.999	29.815	8.651	6.790	6.856	3.892	54.194	10.301	33.220	5.899
$ACF(1)$	−0.056*	−0.046	−0.011	0.011	−0.038	−0.038	−0.009	−0.027	−0.071*	−0.020	−0.032	0.005	0.021	0.044	0.045
$Q(5)$	13.775	27.980*	5.989	7.230	5.128	11.875	3.006	4.005	13.363	5.449	3.990	6.255	3.027	10.133	8.921
$Q(21)$	44.135*	60.454*	25.878	31.191	33.486	48.812*	39.285*	30.927	36.268	22.976	20.604	21.471	38.243	16.706	42.496*
#Obs	2560	2558	2558	2558	2558	2554	2519	2518	2518	2510	2510	2510	2515	2517	2516

Table 2.b: Sample Statistics for Daily Futures Return Shocks.

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
Mean	0.010	0.000	0.007	0.011	0.020	0.009	0.011	−0.002	0.003	0.005	0.002	0.006	−0.001	−0.005	−0.002
Std. Dev.	0.705	0.744	0.692	0.708	0.703	0.697	0.799	0.782	0.737	0.736	0.719	0.772	0.810	0.970	0.762
Min	−3.121	−2.461	−2.894	−4.848	−4.461	−3.668	−10.078	−9.706	−3.313	−3.114	−2.954	−11.170	−6.977	−10.266	−5.596
Max	2.561	3.900	2.806	2.574	2.420	2.320	3.320	4.148	2.978	4.774	2.836	5.841	5.721	9.341	4.838
Skewness	−0.046	0.343	0.056	−0.381	−0.398	−0.216	−1.444	−0.570	0.224	0.383	0.133	−0.887	0.406	0.666	−0.052
Kurtosis	3.301	3.859	3.038	4.420	4.696	3.720	19.047	13.683	3.703	5.302	3.395	23.861	9.963	27.968	5.269
$ACF(1)$	−0.036	−0.025	0.012	0.023	0.014	−0.020	0.018	−0.020	−0.047	−0.003	−0.031	0.012	0.022	0.044	0.040
$Q(5)$	5.894	12.814	4.578	6.407	2.350	3.512	3.459	2.517	8.187	4.618	4.959	6.963	2.066	13.240	5.588
$Q(21)$	16.677	28.140	17.366	19.530	18.457	26.268	26.665	23.182	24.304	23.994	22.006	22.588	35.286	18.973	36.361

Notes: The sample period covers 5 January 2004 to 30 December 2013. ACF(1) denotes the first-order autocorrelation. $Q(L)$ is the Ljung-Box test of zero autocorrelation in lags 1 through L. An asterisk indicates rejection of the null of jointly zero autocorrelations at the 1% level.

Table 3.a: Sample Statistics for Daily Realized Volatility.

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
Mean	2.717	3.955	2.544	1.522	2.792	2.385	2.065	2.481	2.718	2.685	2.484	2.409	1.227	1.882	1.229
Std. Dev.	1.308	1.502	1.112	0.728	1.343	1.164	0.966	1.101	1.214	1.121	0.716	1.116	0.425	0.682	0.452
Min	0.602	0.968	0.615	0.321	0.400	0.532	0.675	0.535	0.664	0.827	1.040	0.531	0.335	0.472	0.344
Max	11.026	13.704	13.261	8.616	16.707	11.505	14.757	11.708	28.304	9.483	6.997	17.341	3.925	8.684	4.478
Skewness	2.362	1.335	1.706	2.430	2.714	2.426	2.770	2.325	5.489	1.148	1.098	2.622	1.550	2.269	1.691
Kurtosis	10.550	6.016	8.950	13.831	16.027	12.420	20.711	12.782	88.282	5.332	5.469	22.013	7.335	13.535	8.352
$ACF(1)$	0.822*	0.701*	0.779*	0.760*	0.745*	0.850*	0.555*	0.518*	0.547*	0.758*	0.509*	0.576*	0.559*	0.527*	0.584*
$Q(5)$	7741*	5227*	6969*	5463*	4951*	7954*	2731*	1889*	2278*	5524*	1858*	3562*	3017*	2548*	3033*
$Q(21)$	28206*	15622*	25122*	17875*	13619*	25875*	8698*	4908*	6314*	17787*	4263*	11986*	10612*	8152*	9527*

Table 3.b: Sample Statistics for Log Daily Realized Volatility.

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
Mean	0.913	1.310	0.851	0.331	0.942	0.779	0.643	0.832	0.931	0.905	0.871	0.792	0.152	0.579	0.148
Std. Dev.	0.400	0.356	0.403	0.408	0.395	0.410	0.387	0.378	0.357	0.407	0.277	0.408	0.321	0.317	0.335
Min	−0.508	−0.032	−0.486	−1.136	−0.917	−0.630	−0.393	−0.626	−0.409	−0.189	0.039	−0.634	−1.095	−0.751	−1.067
Max	2.400	2.618	2.585	2.154	2.816	2.443	2.692	2.460	3.343	2.250	1.945	2.853	1.367	2.161	1.499
Skewness	0.541	0.200	0.159	0.413	0.594	0.481	0.668	0.545	0.512	−0.001	0.109	0.295	0.202	0.506	0.305
Kurtosis	4.188	2.902	3.134	3.736	4.195	3.804	3.554	3.635	4.580	2.603	3.140	3.248	3.659	4.040	3.401
$ACF(1)$	0.793*	0.714*	0.801*	0.735*	0.729*	0.849*	0.635*	0.603*	0.646*	0.792*	0.536*	0.658*	0.589*	0.584*	0.588*
$Q(5)$	6905*	5580*	7241*	5294*	4924*	7867*	3767*	3022*	3885*	6516*	2274*	4386*	3570*	3162*	3167*
$Q(21)$	23915*	17910*	26090*	17182*	14035*	26182*	11942*	8337*	11581*	22481*	5407*	14887*	12413*	10303*	10219*

Notes: The sample period covers 5 January 2004 to 30 December 2013. ACF(1) denotes the first-order autocorrelation. $Q(L)$ is the Ljung-Box test of zero autocorrelation in lags 1 through L. An asterisk indicates rejection at the 1% level.

Table 4: Unconditional Correlations for Daily Returns (upper diagonal) and Log-Volatility (lower diagonal).

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*	
Light Crude		29.2	85.2	31.6	38.2	47.0	29.2	28.5	25.4	25.0	23.9	19.8	16.0	6.0	11.7	
Natural Gas	38.0		32.2	11.9	12.4	13.4	13.5	12.9	10.8	13.3	10.9	8.0	7.0	−3.2	4.4	
Heating Oil	91.8	44.5		29.5	35.0	41.4	26.5	25.1	22.4	21.9	23.3	17.6	12.8	6.5	7.6	
Gold	49.5	24.4	39.2		81.0	41.6	18.8	22.9	18.5	17.6	16.6	9.0	2.5	2.2	−1.6	
Silver	50.5	19.4	40.0	83.9		49.0	24.5	26.5	22.8	22.6	22.6	14.3	7.4	2.9	3.1	
Copper	71.8	40.2	69.3	66.2	65.2		32.1	29.4	26.0	25.3	24.9	21.8	13.3	5.2	10.6	
Soybeans	59.7	17.9	61.1	31.5	36.8	46.7		58.0	44.7	20.9	19.4	27.2	9.5	4.6	−2.8	
Corn	47.7	20.5	36.8	41.7	42.1	49.8	60.6		61.8	24.7	21.1	26.2	14.2	4.2	−10.0	
Wheat	49.0	21.1	40.4	40.5	45.3	49.2	54.6	76.9		22.3	20.9	24.7	11.8	4.7	−6.1	
Sugar	35.1	15.1	25.7	32.9	35.6	34.0	28.4	50.8	62.8		23.1	23.8	10.7	2.1	4.4	
Coffee	23.9	7.9	21.5	4.5	15.9	3.2	29.1	13.3	24.4	10.1		20.4	11.8	3.9	7.2	
Cotton	49.8	0.7	38.2	20.1	38.6	32.3	49.8	48.2	55.9	47.2	36.4		9.6	3.1	4.9	
Live Cattle	60.3	20.4	58.6	34.8	48.8	55.4	52.9	43.5	53.8	41.4	24.0	55.4		17.5	61.1	
Lean Hogs	56.1	43.0	56.4	30.1	33.2	49.6	58.1	45.8	55.3	41.8	19.1	42.4	52.1		16.3	
Feeder Cattle	58.9	10.8	49.6	39.4	45.4	47.4	51.8	50.1	56.2	40.8	17.2	48.2	79.8	43.6		
Average Correlations																*Average*
For Returns	34.4	18.5	32.5	26.8	30.8	32.1	28.4	29.7	27.4	23.8	23.3	22.0	20.3	11.7	14.1	25.1
For $E(\log RVol)$	56.1	28.3	51.5	42.6	46.7	52.0	49.3	48.5	52.4	40.1	23.4	44.2	52.1	48.4	49.3	45.7
Average Correlations																*Average*
With SPY Returns	36.7	7.3	30.0	9.2	22.0	45.2	26.5	23.5	23.1	21.6	25.5	21.2	18.1	5.6	18.7	22.3
With SPY $E(\log RVol)$	53.9	9.4	31.6	66.4	58.0	57.1	31.5	41.6	44.2	35.9	−7.5	30.0	30.1	33.2	42.1	37.2

Notes: The sample period is 5 January 2004 to 30 December 2013.

Table 5.a: Regression of Daily Futures Returns on Principal Components.

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
$PC1$	0.335	0.212	0.296	0.137	0.255	0.247	0.208	0.242	0.226	0.187	0.148	0.151	0.046	0.023	0.015
$PC2$	0.008	0.126	−0.026	0.071	0.127	0.129	0.134	0.172	−0.163	−0.073	−0.065	0.093	0.014	0.040	−0.011
$PC3$	−0.197	0.204	−0.187	−0.158	−0.189	−0.178	0.201	0.309	−0.293	−0.024	0.003	0.091	0.003	−0.014	−0.049
$PC4$	0.012	−0.061	0.012	−0.032	−0.016	−0.027	−0.056	−0.031	0.033	−0.013	0.044	−0.012	0.040	0.014	0.060
R^2	0.427	0.098	0.425	0.349	0.271	0.360	0.342	0.403	0.319	0.136	0.127	0.124	0.044	0.008	0.038

Table 5.b: : Regression of Daily Expected Log Realized Volatility on Principal Components.

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
$PC1$	0.379	0.131	0.343	0.256	0.264	0.379	0.234	0.209	0.229	0.219	0.050	0.217	0.210	0.181	0.197
$PC2$	0.124	0.156	0.119	0.062	0.034	0.209	−0.057	−0.185	−0.264	−0.141	−0.062	−0.185	−0.077	−0.032	−0.105
$PC3$	−0.316	0.087	−0.534	−0.430	0.375	0.264	0.181	−0.052	0.074	−0.268	0.087	−0.269	0.084	−0.123	−0.039
$PC4$	0.115	−0.307	0.074	0.495	0.365	−0.042	0.124	−0.056	0.123	−0.296	0.074	0.239	0.101	−0.138	0.121
R^2	0.794	0.256	0.782	0.697	0.685	0.687	0.499	0.502	0.639	0.379	0.124	0.481	0.553	0.496	0.500

Notes: The sample period covers 5 January 2004 to 30 December 2013.

Regression of Daily Futures Returns on Spyder and Spyder-Orthogonal Principal Components.

Table 6.a: Results for Returns: $r_{Com,t} = \alpha + \beta_{SPY} r_{SPY,t} + \beta_{\hat{PC}1} P\hat{C}1 + \beta_{\hat{PC}2} P\hat{C}2 + \beta_{\hat{PC}3} P\hat{C}3 + \beta_{\hat{PC}4} P\hat{C}4 + \epsilon_t$

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
Constant	0.028	−0.024	0.035	0.038	0.034	0.026	0.013	0.009	0.009	0.028	0.010	−0.004	0.018	0.013	0.025
r_{SPY}	0.656*	0.203*	0.502*	0.147*	0.545*	0.702*	0.404*	0.409*	0.424*	0.389*	0.418*	0.365*	0.148*	0.140*	0.149*
$P\hat{C}1$	0.312*	0.217*	0.299*	0.135*	0.207*	0.194*	0.178*	0.199*	0.188*	0.157*	0.118*	0.126*	0.039*	0.040*	0.005
$P\hat{C}2$	0.023	−0.149*	−0.015	0.117*	−0.167*	−0.120*	0.158*	−0.185*	0.190*	0.094*	0.071*	0.112*	0.016	−0.026	−0.021*
$P\hat{C}3$	−0.176*	−0.240*	−0.167*	−0.167*	0.166*	0.189*	0.201*	−0.294*	0.294*	0.034	0.017	0.062*	−0.003	−0.011	−0.044*
$P\hat{C}4$	0.063*	−0.078	0.065*	−0.098*	0.034	−0.030	0.045*	−0.113*	−0.082*	0.033	0.024	0.052*	0.016	0.003	0.029*
R^2	0.440	0.108	0.420	0.375	0.265	0.379	0.349	0.373	0.328	0.137	0.140	0.129	0.052	0.015	0.058

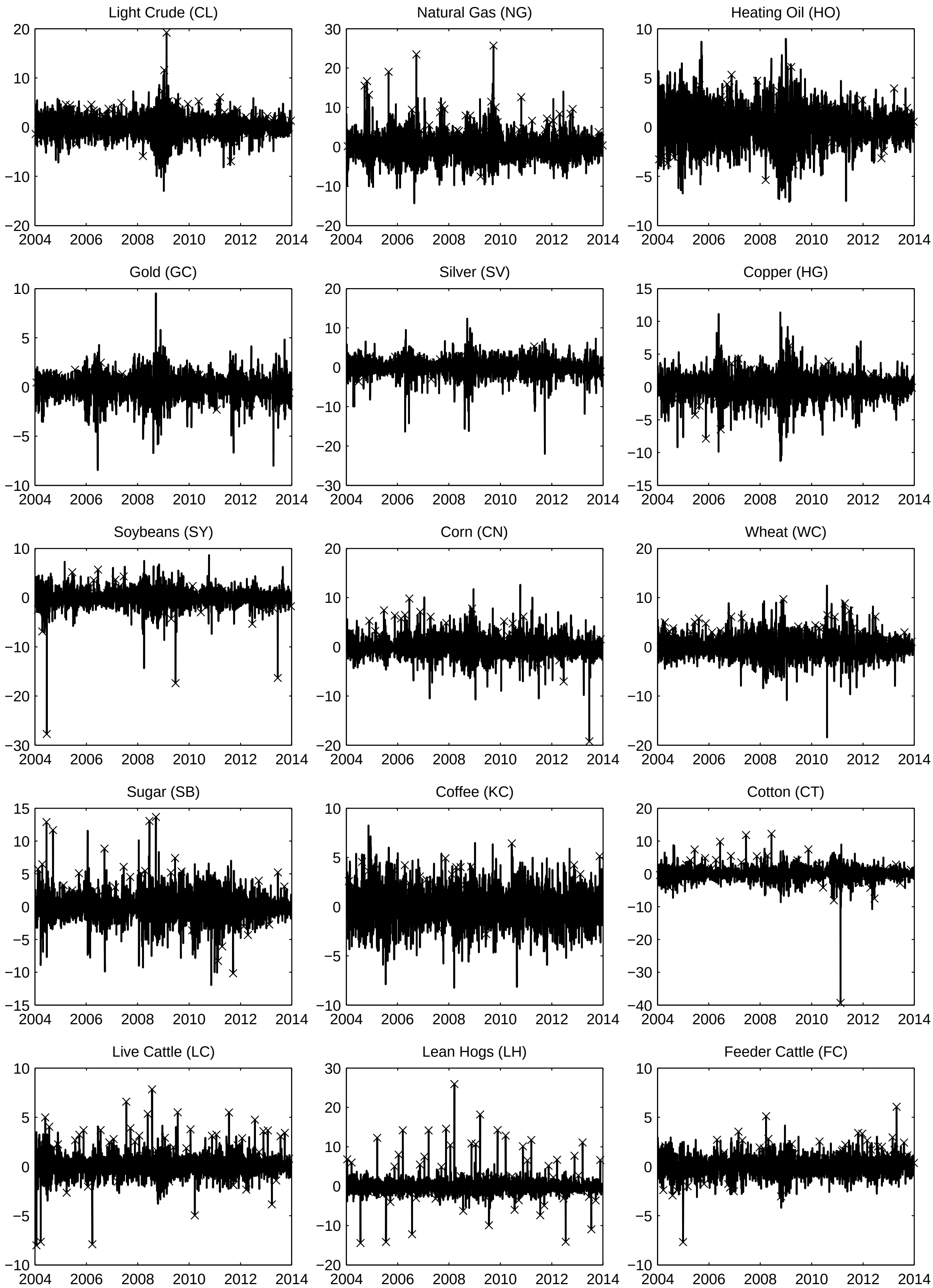
Table 6.b: Results for Volatility: $E\left(\log RVol_{Com,t}\right) = \alpha + \beta_{SPY} E\left(\log RVol_{SPY,t}\right) + \beta_{\hat{PC}1} P\hat{C}1 + \beta_{\hat{PC}2} P\hat{C}2 + \beta_{\hat{PC}3} P\hat{C}3 + \beta_{\hat{PC}4} P\hat{C}4 + \epsilon_t$

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
Constant	0.691*	0.956*	0.588*	0.102*	0.666*	0.583*	0.375*	0.556*	0.651*	0.601*	0.522*	0.553*	−0.125*	0.322*	−0.107*
E($\log RVol_{SPY}$)	0.464*	0.044	0.310*	0.470*	0.389*	0.477*	0.227*	0.241*	0.273*	0.219*	0.014	0.206*	0.176*	0.165*	0.226*
$P\hat{C}1$	0.368*	0.232*	0.443*	0.122*	0.168*	0.329*	0.257*	0.144*	0.182*	0.192*	0.075*	0.206*	0.245*	0.207*	0.192*
$P\hat{C}2$	0.202*	−0.016	0.132*	0.096*	−0.117*	0.084*	−0.094	0.110*	−0.109*	0.094	−0.114*	−0.034	0.005	0.079*	0.027
$P\hat{C}3$	−0.249*	−0.222*	−0.515*	0.470*	0.347*	−0.319*	0.182*	0.165*	0.274*	0.283*	0.104*	0.327*	0.129*	0.025	0.147*
$P\hat{C}4$	0.038	0.123	0.007	0.390*	−0.180*	0.098	0.100	0.019	0.045	−0.116	−0.006	0.064	0.125*	−0.106	0.145*
R^2	0.831	0.334	0.824	0.736	0.644	0.717	0.520	0.448	0.607	0.330	0.163	0.474	0.590	0.508	0.532

Notes: Commodity returns and expected log volatility are regressed on Spyder returns and expected log volatility. To avoid endogeneity the principal components are based on the unconditional covariance matrix of the demeaned regression residuals for the 14 other commodities. The principal components are constructed as the matrix of residuals multiplied by the eigenvectors of the covariance matrix. Significance at the 1% level when using HAC standard errors is indicated by an asterisk.

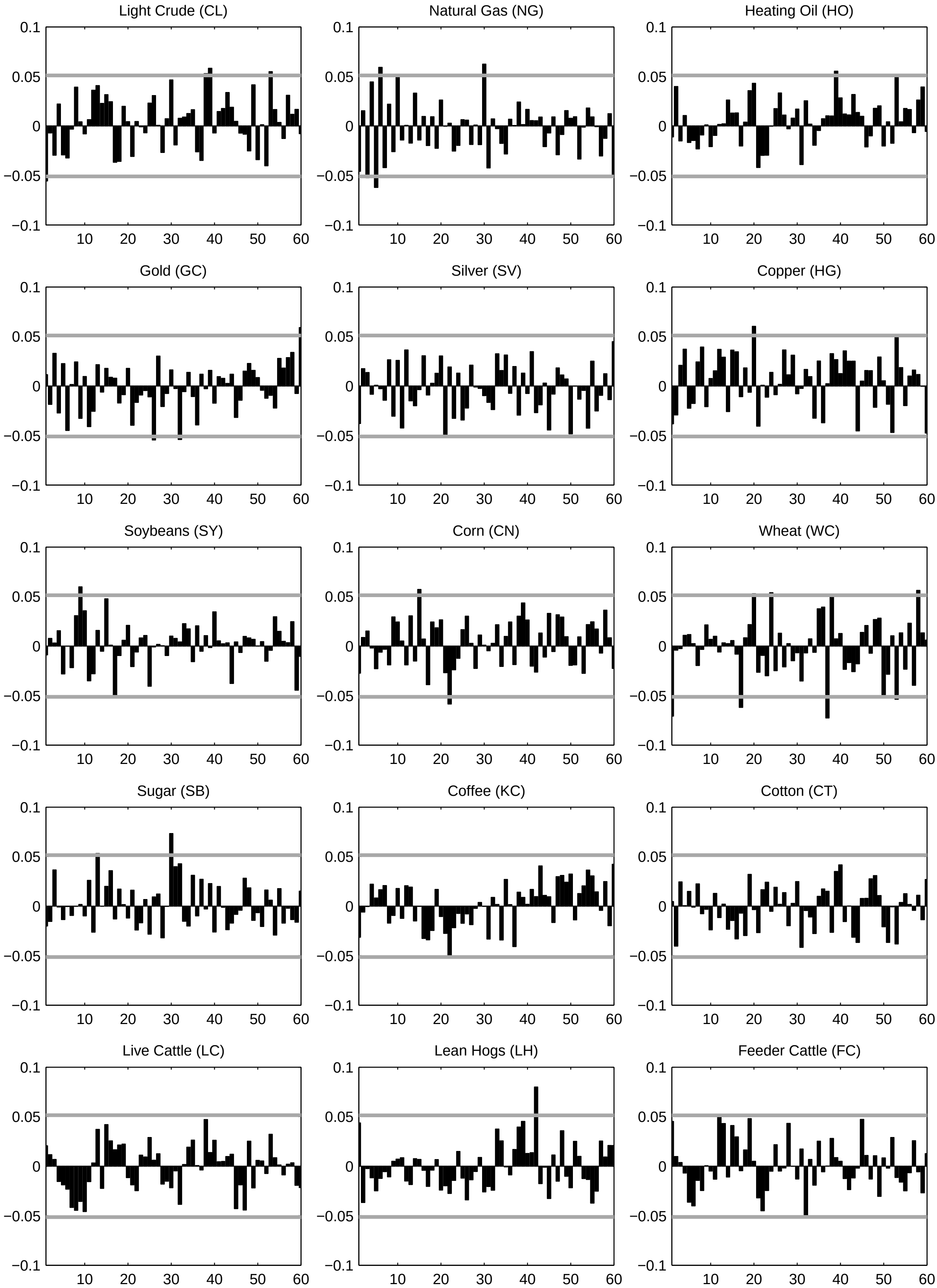
7 Appendix

Figure A.1: Daily Futures Returns.



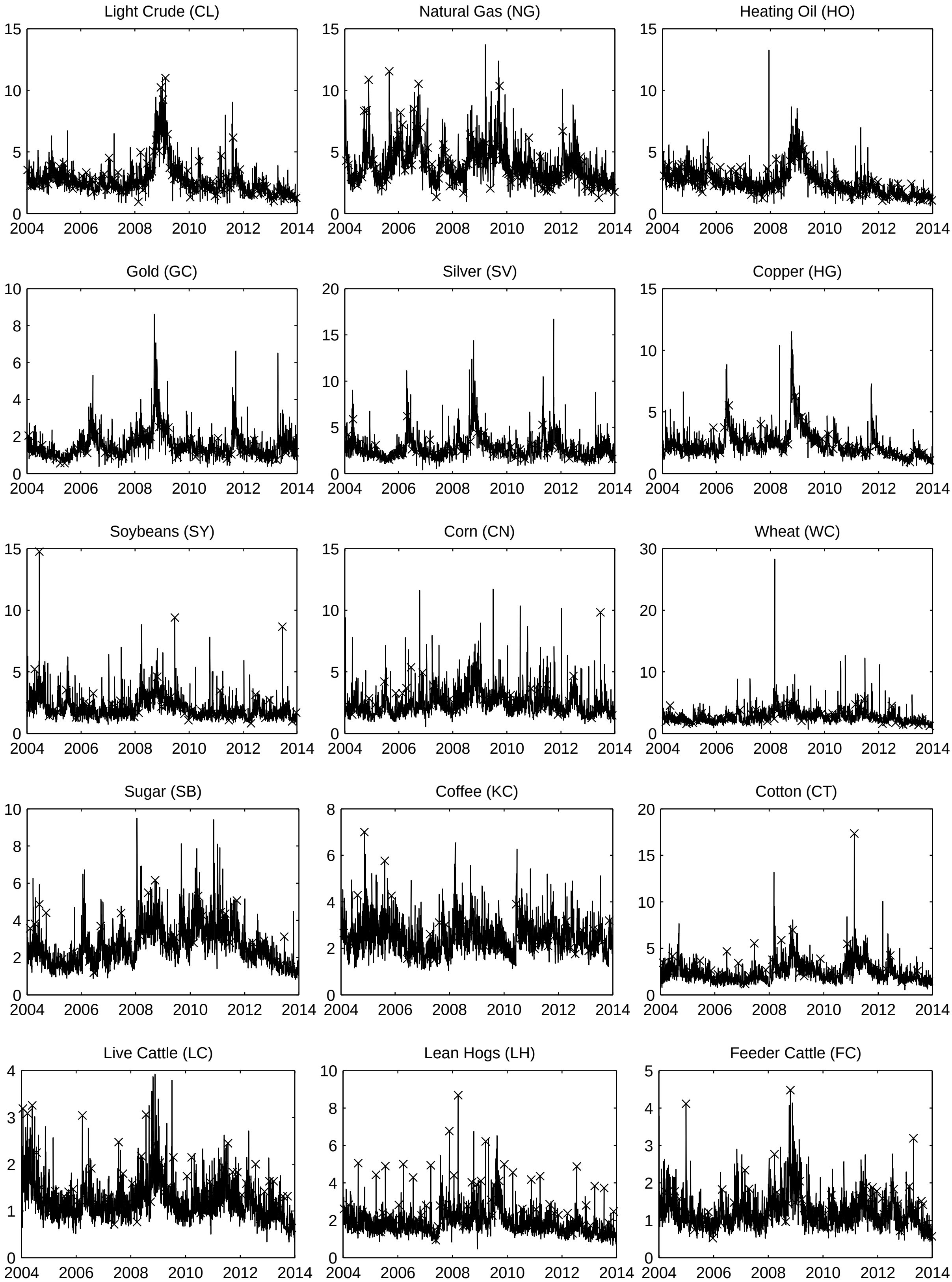
Notes: The crosses denote rollover days.

Figure A.2: Autocorrelation Function of Daily Futures Returns.



Notes: Grey lines indicate 99% confidence bounds assuming that the series are Gaussian white noise. The horizontal axis indicates the lag order in days.

Figure A.3: Daily Realized Volatility.



Notes: The crosses denote rollover days.

Figure A.4: Daily Log-Realized Volatility.

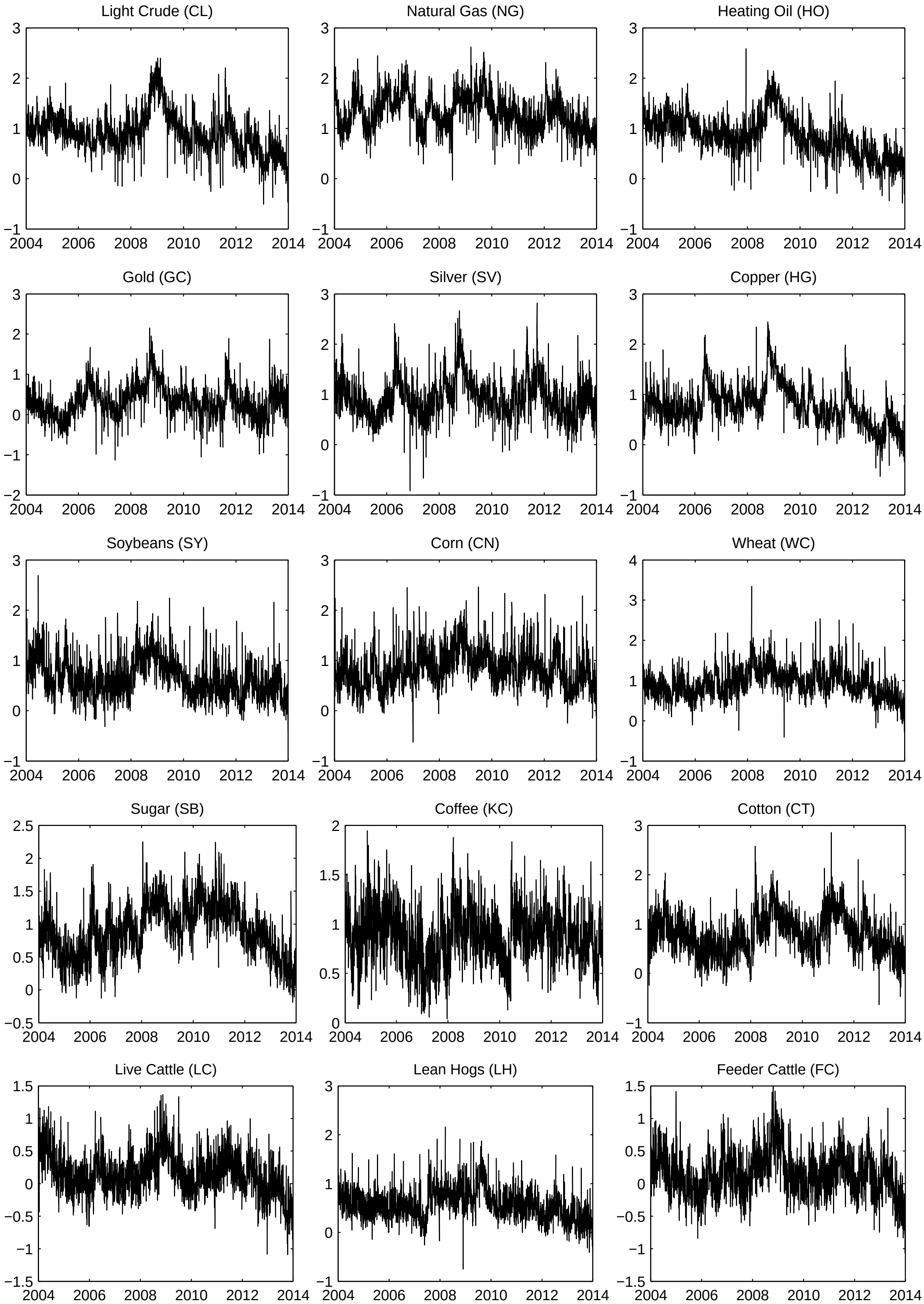


Figure A.5: Realized Stock Market Beta for 15 Commodities.

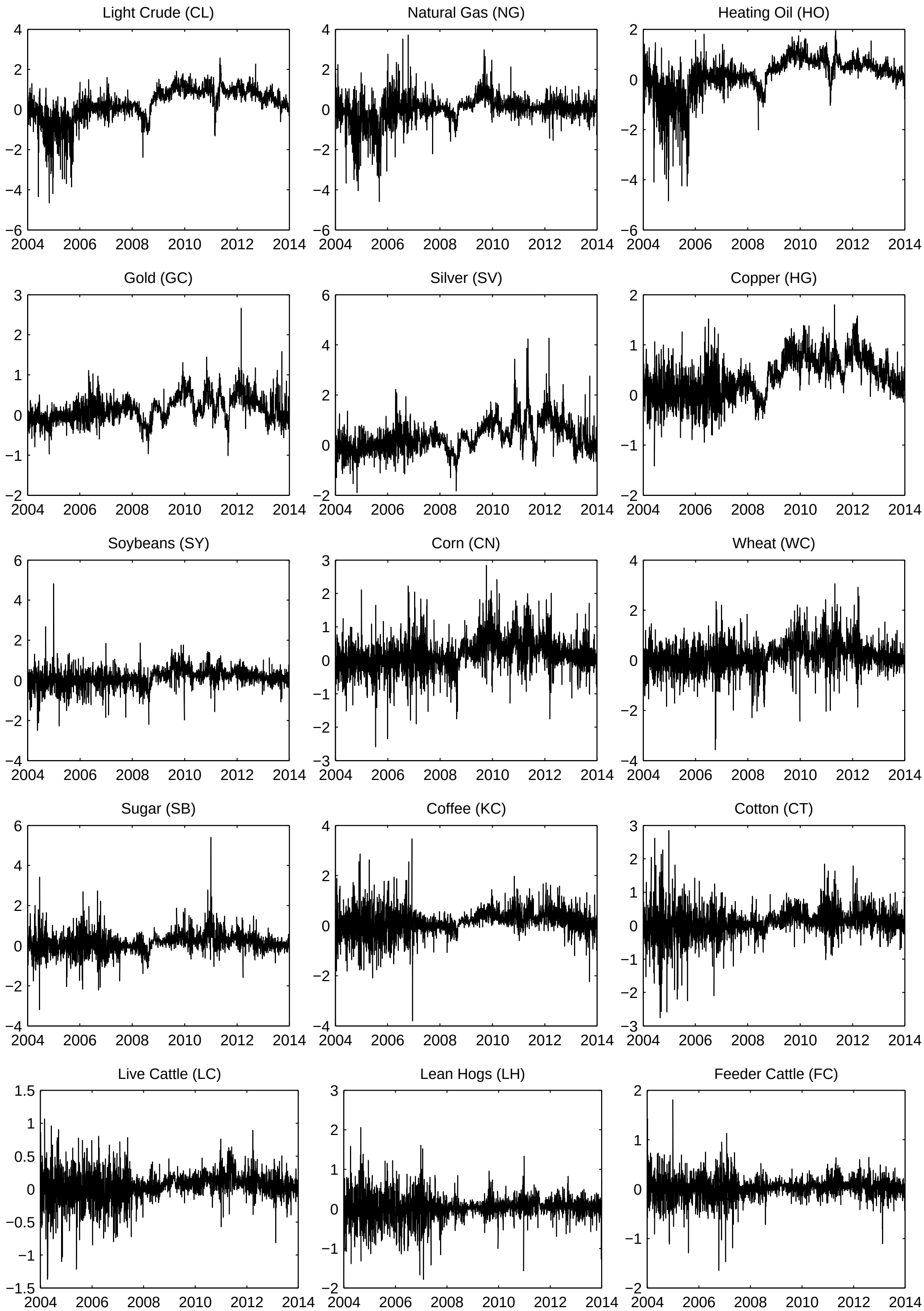
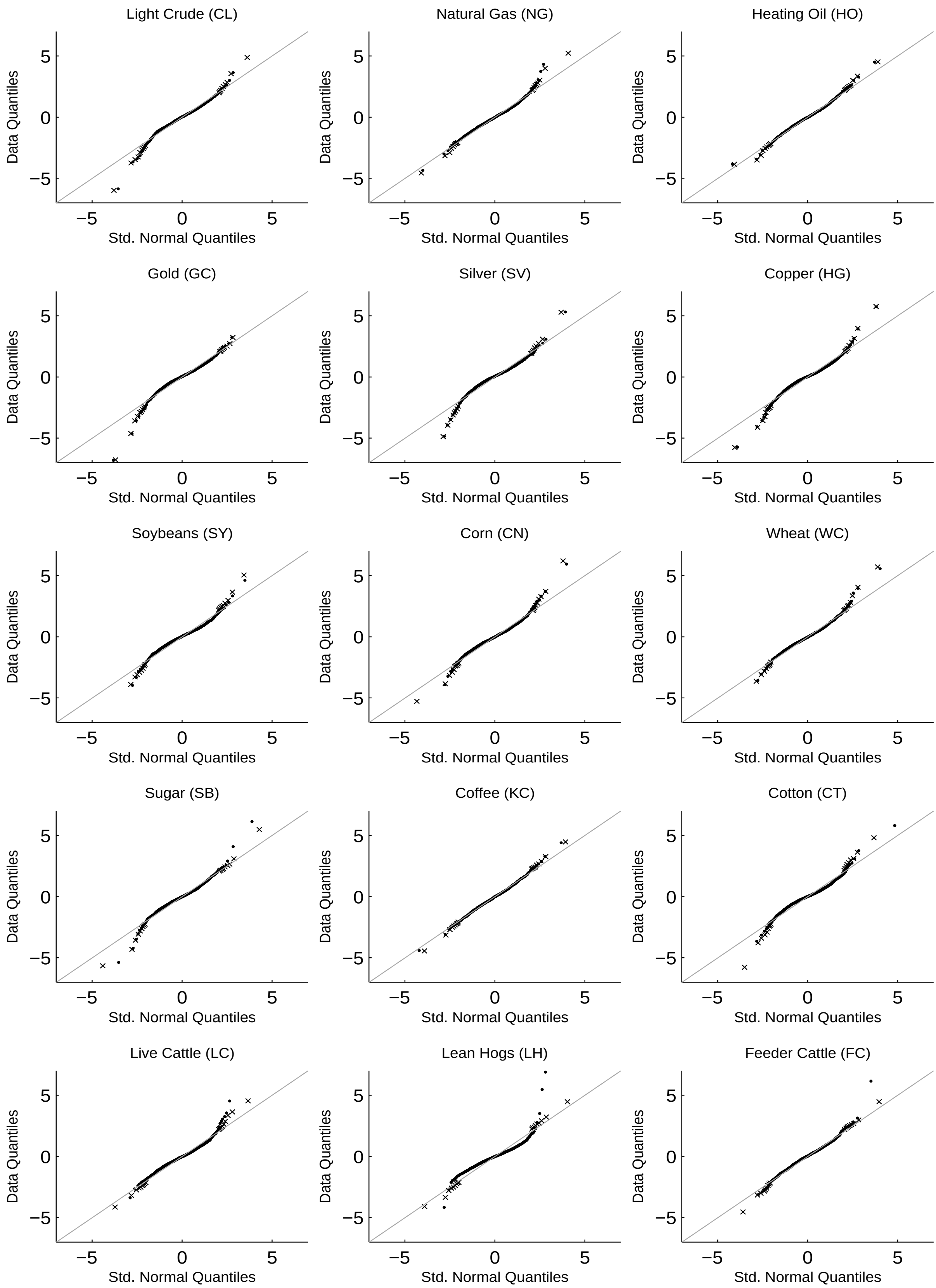
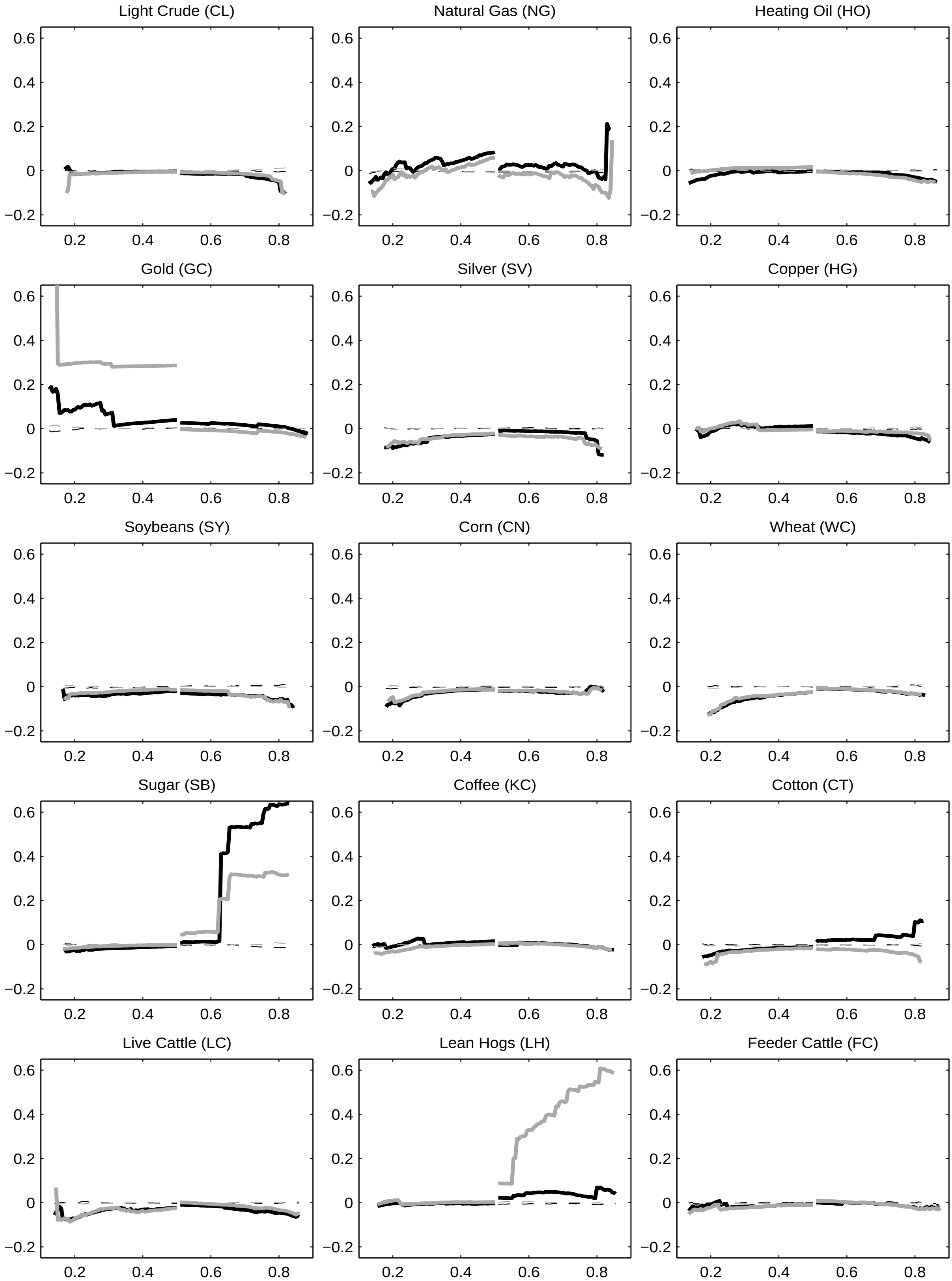


Figure A.6: QQ-Plots of Daily Futures Returns.



Notes: All observations are marked by dots. Crosses indicate large quantiles that do not correspond to rollover returns.

Figure A.7: Threshold Correlation Between Daily Commodity Return Shocks and Spyder.



Notes: Threshold correlations with rollover days included (black lines) and with rollover days excluded (grey lines). The dashed lines denote threshold correlations for the bivariate Gaussian distribution. The horizontal axis indicates thresholds in percentiles.

Table A.1.a: Sample Statistics for Daily Futures Returns (in percent). Rollover Days Excluded.

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
Mean	−0.008	−0.167	0.031	0.037	0.046	0.046	0.042	−0.012	−0.042	0.003	−0.009	−0.009	−0.005	−0.038	0.014
Std.Dev.	2.163	3.108	1.979	1.253	2.328	1.965	1.700	2.033	2.192	2.112	1.852	1.865	0.915	1.352	0.929
Min	−12.961	−14.362	−7.605	−8.461	−22.012	−11.294	−14.337	−10.731	−18.419	−11.950	−8.233	−10.794	−3.799	−5.582	−4.206
Max	10.552	16.059	8.979	9.524	12.365	11.359	8.628	12.604	12.460	11.574	8.255	8.943	4.148	5.998	4.166
Skewness	−0.284	0.199	0.029	−0.495	−1.208	−0.278	−0.351	0.090	−0.058	−0.272	0.035	−0.142	0.150	0.003	−0.113
Kurtosis	5.404	4.449	4.192	7.929	11.495	7.045	7.187	6.429	6.994	5.634	3.941	5.856	4.400	3.969	3.840
ACF(1)	−0.030	−0.055*	−0.015	0.010	−0.035	−0.035	−0.013	−0.023	−0.060*	−0.035	−0.022	0.018	0.014	0.078*	0.045
Q(5)	4.022	22.212*	5.737	7.121	5.038	10.448	3.320	5.426	10.917	8.089	3.768	8.212	4.518	21.525*	8.945
Q(21)	47.762*	32.954	32.833	23.821	21.727	50.120*	41.859*	24.598	22.232	36.752	27.772	34.494	39.280*	53.795*	47.043*

Table A.1.b: Sample Statistics for Daily Return Shocks. Rollover Days Excluded.

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
Mean	0.014	0.007	0.006	0.010	0.019	0.009	0.008	−0.002	0.005	0.005	0.002	0.002	−0.001	0.002	−0.003
Std.Dev.	0.697	0.715	0.687	0.708	0.707	0.692	0.734	0.732	0.722	0.698	0.713	0.700	0.697	0.690	0.722
Min	−3.076	−2.580	−2.850	−4.840	−4.472	−2.877	−3.334	−4.218	−3.320	−2.659	−2.952	−5.164	−2.252	−2.696	−2.853
Max	2.470	2.731	2.773	2.576	2.419	2.526	3.264	2.918	2.909	4.186	4.007	3.044	2.973	3.073	2.299
Skewness	−0.087	0.103	0.037	−0.385	−0.433	−0.165	−0.075	−0.023	0.131	0.097	0.131	−0.167	0.072	−0.002	−0.161
Kurtosis	3.253	3.043	2.999	4.429	4.852	3.528	3.820	4.153	3.725	3.950	3.680	5.054	3.368	3.325	3.322
ACF(1)	−0.024	−0.022	0.016	0.021	0.016	−0.018	0.007	−0.016	−0.031	−0.011	−0.025	0.020	0.012	0.087*	0.045
Q(5)	2.942	7.808	3.406	6.186	2.697	4.490	4.238	3.037	5.142	5.630	5.821	8.505	2.172	23.770*	7.338
Q(21)	17.901	14.986	13.218	15.299	14.033	33.582	29.865	17.510	10.458	32.783	27.708	25.551	33.346	50.415*	36.249

Notes: The sample period covers 5 January 2004 to 30 December 2013. ACF(1) is the first-order autocorrelation. $Q(L)$ is the Ljung-Box test of zero auto-correlation in lags 1 through L. An asterisk indicates rejection at the 1 percent level.

Table A.2: Optimal Weights on Open-to-Close Realized Volatility and Overnight Returns.

	Light Crude	*Natural Gas*	*Heating Oil*	*Gold*	*Silver*	*Copper*	*Soy-beans*	*Corn*	*Wheat*	*Sugar*	*Coffee*	*Cotton*	*Live Cattle*	*Lean Hogs*	*Feeder Cattle*
$\hat{E}\left[\left(r_{i,t}^{co}\right)^2 + RV_{i,t}\right]$	8.110	16.926	7.136	2.539	8.471	6.293	4.668	6.717	7.837	7.981	6.449	6.422	1.600	3.712	1.614
$\hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]$	3.995	8.879	3.449	1.249	4.045	3.128	2.502	3.644	4.160	3.971	3.118	3.101	0.853	1.986	0.831
$\hat{E}\left[RV_{i,t}\right]$	4.115	8.047	3.687	1.289	4.426	3.164	2.165	3.073	3.677	4.010	3.331	3.321	0.747	1.726	0.782
$\frac{\hat{E}\left[RV_{i,t}\right]}{\hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]}$	1.030	0.906	1.069	1.032	1.094	1.012	0.865	0.844	0.884	1.010	1.068	1.071	0.875	0.869	0.941
$\hat{E}\left[\left(\left(r_{i,t}^{co}\right)^2 - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\right)^2\right]$	43.501	188.279	26.970	4.453	51.291	31.895	19.028	40.099	45.252	43.315	20.498	30.954	2.210	14.476	1.625
$\hat{E}\left[\left(RV_{i,t} - \hat{E}\left[RV_{i,t}\right]\right)^2\right]$	18.563	36.166	10.501	1.466	17.769	9.901	4.163	7.992	8.427	10.224	3.626	8.759	0.268	1.417	0.339
$\frac{\hat{E}\left[\left(RV_{i,t} - \hat{E}\left[RV_{i,t}\right]\right)^2\right]}{\hat{E}\left[\left(\left(r_{i,t}^{co}\right)^2 - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\right)^2\right]}$	0.427	0.192	0.389	0.329	0.346	0.310	0.219	0.199	0.186	0.236	0.177	0.283	0.121	0.098	0.209
$\frac{\hat{E}\left[RV_{i,t}\left(\left(r_{i,t}^{co}\right)^2 - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\right)\right]}{\hat{E}\left[\left(r_{i,t}^{co}\right)^2 - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\right]\cdot\hat{E}\left[RV_{i,t} - \hat{E}\left[RV_{i,t}\right]\right]}$	0.437	0.356	0.404	0.385	0.400	0.455	0.265	0.278	0.302	0.382	0.198	0.331	0.236	0.204	0.230
$\hat{\varphi}$	0.852	0.930	0.879	0.892	0.913	0.934	0.852	0.865	0.904	0.945	0.923	0.910	0.933	0.943	0.877
$\hat{\omega}_1^\star$	0.300	0.133	0.250	0.220	0.181	0.133	0.276	0.249	0.181	0.111	0.159	0.186	0.125	0.107	0.238
$\hat{\omega}_2^\star$	1.680	1.957	1.702	1.756	1.748	1.857	1.837	1.890	1.926	1.881	1.787	1.760	2.000	2.028	1.809

Notes: The table contains the weighting parameters that are used to compute the optimal measure of daily volatility in equation.

$$\hat{\varphi} \equiv \frac{\hat{E}\left[RV_{i,t}\right]^2 \hat{E}\left[\left(\left(r_{i,t}^{co}\right)^2 - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\right)^2\right] - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\hat{E}\left[RV_{i,t}\right]\hat{E}\left[RV_{i,t}\left(\left(r_{i,t}^{co}\right)^2 - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\right)\right]}{\hat{E}\left[RV_{i,t}\right]^2 \hat{E}\left[\left(\left(r_{i,t}^{co}\right)^2 - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\right)^2\right] + \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]^2 \hat{E}\left[\left(RV_{i,t} - \hat{E}\left[RV_{i,t}\right]\right)^2\right] - 2\hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\hat{E}\left[RV_{i,t}\right]\hat{E}\left[RV_{i,t}\left(\left(r_{i,t}^{co}\right)^2 - \hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]\right)\right]}$$

$$\hat{\omega}_1^\star \equiv (1-\hat{\varphi})\frac{\hat{E}\left[\left(r_{i,t}^{co}\right)^2 + RV_{i,t}\right]}{\hat{E}\left[\left(r_{i,t}^{co}\right)^2\right]}, \qquad \hat{\omega}_2^\star \equiv \hat{\varphi}\frac{\hat{E}\left[\left(r_{i,t}^{co}\right)^2 + RV_{i,t}\right]}{\hat{E}\left[RV_{i,t}\right]}.$$

Note that, as expected, the importance factor, $\hat{\varphi}$, is close to one in all cases, which causes the realized variance for the "active" period, $RV_{i,t}$, to be scaled upward ($\hat{\omega}_2 > 1$), whereas the squared overnight return, $\left(r_{i,t}^{co}\right)^2$, is scaled downward ($\hat{\omega}_1 < 1$).

Table A.3.a: ARMA(1,1) on Log Realized Volatility.

	Light Crude	Natural Gas	Heating Oil	Gold	Silver	Copper	Soy-beans	Corn	Wheat	Sugar	Coffee	Cotton	Live Cattle	Lean Hogs	Feeder Cattle
ϕ_0	0.016	0.026	0.011	0.013	0.060	0.022	0.016	0.040	0.029	0.024	0.059	0.014	0.001	0.012	0.003
ϕ_1	0.982	0.980	0.986	0.961	0.936	0.972	0.975	0.952	0.968	0.974	0.933	0.982	0.987	0.979	0.980
θ_1	−0.684	−0.738	−0.718	−0.629	−0.529	−0.547	−0.755	−0.689	−0.715	−0.637	−0.671	−0.782	−0.839	−0.798	−0.801
R^2	0.696	0.603	0.715	0.590	0.570	0.756	0.484	0.418	0.492	0.680	0.346	0.535	0.479	0.440	0.441
σ_e	0.220	0.225	0.215	0.261	0.259	0.203	0.278	0.288	0.254	0.230	0.224	0.278	0.232	0.237	0.251
$ACF_e(1)$	0.097*	0.093*	0.101*	0.119*	0.089*	0.097*	0.126*	0.127*	0.116*	0.110*	0.090*	0.109*	0.093*	0.117*	0.124*
$Q_e(5)$	50.305*	82.743*	48.632*	98.360*	83.465*	95.605*	73.091*	90.471*	64.568*	68.131*	49.283*	51.811*	29.714*	53.743*	58.806*
$Q_e(21)$	96.087*	236.965*	91.653*	170.809*	122.133*	132.838*	98.654*	111.602*	87.282*	94.699*	67.392*	76.690*	55.795*	70.773*	76.115*

Table A.3.b: ARMA(1,1) on Realized Beta.

	Light Crude	Natural Gas	Heating Oil	Gold	Silver	Copper	Soy-beans	Corn	Wheat	Sugar	Coffee	Cotton	Live Cattle	Lean Hogs	Feeder Cattle
ϕ_0	0.002	0.000	0.001	0.003	0.005	0.001	0.002	0.002	0.001	0.001	0.000	0.000	0.001	0.041	0.001
ϕ_1	0.994	0.991	0.995	0.983	0.984	0.996	0.987	0.991	0.993	0.991	0.997	0.997	0.978	−0.194	0.955
θ_1	−0.820	−0.895	−0.863	−0.766	−0.759	−0.889	−0.902	−0.918	−0.940	−0.930	−0.965	−0.972	−0.915	0.234	−0.921
R^2	0.704	0.356	0.618	0.586	0.623	0.602	0.220	0.230	0.158	0.169	0.132	0.075	0.065	0.007	0.015
σ_e	0.435	0.543	0.447	0.204	0.377	0.254	0.402	0.440	0.520	0.449	0.465	0.412	0.202	0.305	0.215
$ACF_e(1)$	0.013	0.019	−0.005	0.014	0.004	0.030	0.014	0.008	−0.002	−0.005	−0.001	0.024	−0.004	0.010	0.027
$Q_e(5)$	22.122*	5.013	9.668	4.294	2.906	4.085	4.277	0.682	5.757	9.908	0.955	10.921	3.073	15.593*	7.543*
$Q_e(21)$	119.697*	33.585	57.005*	46.251*	20.316	34.037	36.098	13.283	32.961	32.312	13.972	57.411*	29.016	56.466*	15.491*

Table A.3.c: ARMA(1,1) on Realized Systematic Risk Ratio.

	Light Crude	Natural Gas	Heating Oil	Gold	Silver	Copper	Soy-beans	Corn	Wheat	Sugar	Coffee	Cotton	Live Cattle	Lean Hogs	Feeder Cattle
ϕ_0	0.002	0.000	0.001	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.001
ϕ_1	0.988	0.984	0.987	0.976	0.980	0.994	0.972	0.978	0.981	0.975	0.981	0.994	0.965	0.972	0.962
θ_1	−0.726	−0.865	−0.749	−0.710	−0.709	−0.799	−0.781	−0.848	−0.879	−0.830	−0.869	−0.953	−0.798	−0.944	−0.924
R^2	0.744	0.304	0.687	0.592	0.650	0.753	0.397	0.284	0.218	0.297	0.250	0.115	0.291	0.013	0.020
σ_e	0.082	0.033	0.071	0.076	0.072	0.077	0.044	0.041	0.038	0.039	0.048	0.027	0.029	0.020	0.021
$ACF_e(1)$	0.055*	−0.001	0.023	0.070*	0.052*	0.067*	−0.008	0.051	0.000	0.058*	0.025	0.065*	−0.006	0.009	0.006
$Q_e(5)$	11.657	14.881	2.848	25.814*	13.187	40.153*	8.759	25.814*	3.884	12.557	7.419	16.209*	9.811	2.607	1.563
$Q_e(21)$	51.207*	60.715*	40.756*	40.528*	39.704*	67.605*	30.559	47.957*	24.130	62.616*	31.123	47.733*	52.623*	21.712	17.240

Notes: The sample period covers 5 January 2004 to 30 December 2013.

Table A.4: Intervals of Trading for 15 Commodities.

Commodity	Period	Trading Interval(s)
Crude Oil, Natural Gas, and Heating Oil	5/1-2004 to 9/6-2006	00.00 - 14.30
		15.15 - 12.00
	12/6-2006 to 30/12-2013	00.00 - 17.15
		18.00 - 12.00
Gold, Silver	5/1-2004 to 28/5-2004	00.00 - 13.30
		15.15 - 12.00
	1/6/2004 to 1/12-2006	00.00 - 13.30
		14.00 - 12.00
	4/12-2006 to 30/12-2013	00.00 - 17.15
		18.00 - 12.00
Copper	5/1-2004 to 4/6-2004	00.00 - 13.00
		15.15 - 12.00
	7/6-2004 to 1/12-2006	00.00 - 13.00
		14.00 - 12.00
	4/12-2006 to 30/12-2013	00.00 - 17.15
		18.00 - 12.00
Live Cattle, Lean Hogs, and Feeder Cattle	5/1-2004 to 1/6-2007	10.05 - 14.00
	4/6-2007 to 30/12-2013	00.00 - 17.00
		18.00 - 12.00
Corn, Soybeans, and Wheat	5/1-2004 to 7/10-2005	00.00 - 7.00
		10.30 - 14.15
		20.30 - 12.00
	10/10-2005 to 11/1-2008	00.00 - 7.00
		10.30 - 14.15
		19.30 - 12.00
	11/1-2008 to 30/6-2009	00.00 - 7.00
		10.30 - 14.15
		19.00 - 12.00
	1/7-2009 to 18/5-2012	00.00 - 8.15
		10.30 - 14.15
		19.00 - 12.00
	21/5-2012 to 5/4-2013	00.00 - 15.00
		18.00 - 12.00
	8/4-2013 to 30/12-2013	00.00 - 8.45
		9.30 - 14.15
		20.00 - 12.00
Sugar	5/1-2004 to 1/2-2007	9.00 - 00.00
	2/2-2007 to 30/12-2013	2.30 - 14.00*
Coffee	5/1-2004 to 1/2-2007	9.15 - 12.30
	2/2-2007 to 30/12-2013	3.30 - 14.00*
Cotton	5/1-2004 to 1/2-2007	10.30 - 14.15
	2/2-2007 to 27/3-2009	3.30 - 14.45
	30/3-2009 to 30/12-2013	00.00 - 14.30*
		21.00 - 12.00

Notes: All times are Eastern Time. The list indicates for each commodity the intraday trading intervals available in our dataset in the sample period .

* Minor changes in trading interval occured during this period.

Research Papers
2013

2014-15: Federico Carlini and Katarzyna Łasak: On an Estimation Method for an Alternative Fractionally Cointegrated Model

2014-16: Mogens Bladt, Samuel Finch and Michael Sørensen: Simulation of multivariate diffusion bridges

2014-17: Markku Lanne and Henri Nyberg: Generalized Forecast Error Variance Decomposition for Linear and Nonlinear Multivariate Models

2014-18: Dragan Tevdovski: Extreme negative coexceedances in South Eastern European stock markets

2014-19: Niels Haldrup and Robinson Kruse: Discriminating between fractional integration and spurious long memory

2014-20: Martyna Marczak and Tommaso Proietti: Outlier Detection in Structural Time Series Models: the Indicator Saturation Approach

2014-21: Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Discretization of Lévy semistationary processes with application to estimation

2014-22: Giuseppe Cavaliere, Morten Ørregaard Nielsen and A.M. Robert Taylor: Bootstrap Score Tests for Fractional Integration in Heteroskedastic ARFIMA Models, with an Application to Price Dynamics in Commodity Spot and Futures Markets

2014-23: Maggie E. C. Jones, Morten Ørregaard Nielsen and Michael Ksawery Popiel: A fractionally cointegrated VAR analysis of economic voting and political support

2014-24: Sepideh Dolatabadim, Morten Ørregaard Nielsen and Ke Xu: A fractionally cointegrated VAR analysis of price discovery in commodity futures markets

2014-25: Matias D. Cattaneo and Michael Jansson: Bootstrapping Kernel-Based Semiparametric Estimators

2014-26: Markku Lanne, Jani Luoto and Henri Nyberg: Is the Quantity Theory of Money Useful in Forecasting U.S. Inflation?

2014-27: Massimiliano Caporin, Eduardo Rossi and Paolo Santucci de Magistris: Volatility jumps and their economic determinants

2014-28: Tom Engsted: Fama on bubbles

2014-29: Massimiliano Caporin, Eduardo Rossi and Paolo Santucci de Magistris: Chasing volatility - A persistent multiplicative error model with jumps

2014-30: Michael Creel and Dennis Kristensen: ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models

2014-31: Peter Christoffersen, Asger Lunde and Kasper V. Olesen: Factor Structure in Commodity Futures Return and Volatility