Do economic variables forecast commodity futures volatility?

Loïc Maréchal*

January 8, 2021

Abstract

This research combines recent advances in the realized volatility literature and three economically motivated variables, related to well-known hypotheses of commodity volatility determinants, to improve the volatility forecast of commodity futures contracts. The three forecasting variables are the term structure slope (theory of storage), the time to maturity (Samuelson hypothesis), and a seasonal measure that proxies for the supply and demand uncertainty (uncertainty resolution hypothesis). I first assess the empirical contribution of these variables to explain realized volatility and find support for the theory of storage and uncertainty resolution. I uncover a positive relationship between time to maturity and volatility, in contradiction with the Samuelson hypothesis. Second, I compare the performance of the HAR (Corsi, 2009), HEXP (Bollerslev, Hood, Huss, and Pedersen, 2018a), HARQ (Bollerslev, Patton, and Quaedvlieg, 2016), and time-varying parameter HAR-TV (Chen, Gao, Li, and Silvapulle, 2018) models, with and without the above-mentioned economic variables. I evaluate the in- and out-of-sample performance of these forecasts in econometric tests and find that the inclusion of economic variables is beneficial for long-term horizons only. Finally, I test the validity of these forecasts to improve expected shortfall modeling. The inclusion of economic variables provides mixed results.

JEL classification: C53, C58, *Keywords*: Commodity futures, realized volatility, term structure, Samuelson effect, supply and demand uncertainty, risk management.

*University of Neuchâtel loic.marechal@unine.ch

I thank Tim Bollerslev for his advice in the context of the 2018 Volatility Modeling Summer School in Gerzensee. His comments on the first research proposal were of great help. I am also grateful to my supervisor, Michel Dubois, for his invaluable comments. Finally, I thank Alex Y. Huang and Marinela Adriana Finta, for their discussions during the SFM 2019 and NZFM 2019 conferences.

1. Introduction

The realized volatility framework introduced by Andersen and Bollerslev (1998a,b) delivers higher forecast accuracy than generalized autoregressive conditional heteroskedasticity (GARCH) or stochastic volatility (SV) models for spot and futures prices of stocks, bonds and currencies. However, the literature on realized volatility for commodity futures is scarce. While commodities share many commonalities with financial assets, the real nature of the underlying assets introduces some specificities such as the existence of physical inventories. Hence, I ask whether introducing economic variables that characterize individual commodity futures contracts helps modeling their realized volatility.

I start the analysis with a regression model where I test the explanatory power of the slope of the term structure (a proxy for inventories), seasonality (uncertainty resolution), and time to maturity (Samuelson conjecture), on the realized volatility. The empirical analysis is based on three groups of commodities (agriculture, energy, and metals). I select three commodity contracts in each group, over the 2008–2019 period.[1] Consistent with the theory of storage, I show that the slope of the term structure is related to the realized volatility. I separate the contango and backwardation states, and find that the corresponding coefficients are statistically significant at the 1% level for all contracts but soybeans. Yet, the sign of the slope coefficients support a weak form of the theory of storage for five futures contracts. More specifically the full term structure-volatility relationship is present in four out of the nine contracts. I capture seasonal variations of the realized volatility in futures contracts, with dummies reflecting critical months. In line with economic predictions, this variable is economically and statistically significant at the 1% level for agricultural products (seasonal crops) and natural gas (typically higher heating demand in winter). Finally, I find that the time to maturity, a controversial variable related to the functioning of futures markets themselves rather than commodity fundamentals, is a positive determinant of commodity futures volatility, statistically significant at the 1% level for eight out of the nine commodities (crude oil contract statistically insignificant).

Given the strong and long-lasting autocorrelation of realized volatility, I introduce several time-series specifications in the baseline model. I find that all economic variables continue to marginally improve future realized volatility (RV) forecasts. First, I introduce the heterogeneous autoregressive RV (HAR) of Corsi (2009). The motivation for using the HAR is based on the econometric performance (both in- and out-of-sample), and justified by the general long-term memory that commodity futures RV also exhibit. I find that this combination

[1] The agriculture contracts are corn, soybeans, and wheat. The energy contracts are crude oil, heating oil, and natural gas. The metal contracts are gold, copper, and silver.

reduces the in-sample error and weakly improves the out-of-sample forecast accuracy. I additionally include the economic variables in competing models, the HEXP of Bollerslev et al. (2018a), where the lagging terms are exponentially smoothed, and the HARQ of Bollerslev et al. (2016), which accounts for measurement errors, approximated by the realized quarticity. The inclusion of the economic variables in these models produces a similar, marginal, improvement of the explanatory power. Finally, I test the baseline model with the inclusion of autoregressive variables and measurement errors in time-varying (TV) parameter models (HAR-TV and HARQ-TV see, *e.g.*, Casas, Ferreira, and Orbe, 2019, Chen et al., 2018). The forecasting power improvement resulting from the inclusion of the economic variables in this context is negligible. In sample, the best performing model for one-day ahead forecasts is the HARQ-TV, whereas the HAR-TV outperforms all other specifications for the one-week and one-month horizons. Out-of-sample tests however, show that the HARQ dominates when the forecast horizon is lower than one-week, whereas the EVHARQ-TV (nesting economic variables) yields the most accurate forecasts for the one-month horizon. Finally, I use the one-day-ahead out-of-sample forecasts in multi-quantile VaR regressions to extract expected shortfall for up to six coverage levels. I jointly test the parameters for the forecast bias and find that the simple autoregressive models (HAR and EVHAR) provide the lowest rejection and therefore, the higher forecast accuracy.

This research departs from the existing literature on two aspects. First, commodity futures are widely overlooked in the RV literature, in comparison with other asset classes (stocks, indices, bonds, and currencies). In particular, the research looking at non-energy products, agricultural and to a lesser-extent metals, is virtually nonexistent. Second, I introduce economically-motivated, exogenous, variables in the RV framework which is non-standard, as most of the existing research aims to improve RV modeling using autoregressive specifications.

The contribution of this article is threefold. First, I scrutinize the performance of standard autoregressive volatility models for commodity futures. I verify that, as for other asset classes, the RV approach is entirely valid for all commodities. Second, I introduce non-autoregressive variables in the RV framework and find that the addition of exogenous information improves the explanatory power of RV models in-sample as well as out-of-sample forecasts. Finally, I reconsider the determinants of commodity futures volatility. As the RV provides a better approximation of the true latent volatility process than the (G)ARCH and stochastic volatility approaches, the reexamination of such variables in this context is particularly relevant; see, Andersen and Bollerslev (1998b). In the following tests, I highlight the contracts peculiarities, and find strong and medium support for the uncertainty resolution and theory of storage, and strongly reject the Samuelson hypothesis.

The remainder of the article proceeds as follows. Section 2 presents the literature review on RV and, more specifically, its economic determinants. Section 3 presents the research design, including the data organization and models to test. In section 4, I discuss the empirical results. Section 5 compares the performance of out-of-sample forecasts from both autoregressive and economic variable models in tail risk management. Section 6 concludes.

2. Previous research and hypotheses

Currently, there are three main econometric approaches for volatility modeling. First, the (generalized) auto-regressive conditional heteroscedasticity ((G)ARCH) of Engle (1982) and Bollerslev (1986) expresses the conditional volatility as a linear function of its own past realizations and squared innovations. Second, the Stochastic Volatility (SV), where the volatility follows a diffusion process (see, *e.g.*, Heston, 1993), and possibly jumps (see, *e.g.* Bates, 1996)[2]. The daily realized variance is defined as:

$$RV_{t+1}^{2}\left(\Delta\right) \equiv \sum_{j=1}^{1/\Delta} \left(r_{\Delta,t+j\times\Delta}^{2}\right), \tag{1}$$

where $1/\Delta$ is the number of observations in one day and r_{Δ}^{2} represents the squared change in intra-day (log) prices; see, *e.g.* Andersen, Bollerslev, and Diebold (2007). Andersen and Bollerslev (1998a) show that the realized variance is the limit of the integrated variance, as the frequency increases to infinity:

$$\lim_{\Delta\to 0} RV_{t+1}^{2}\left(\Delta\right) \to \int_{t}^{t+1} \sigma^{2}\left(s\right) ds + \sum_{t<s\leq t+1} \kappa^{2}\left(s\right) \tag{2}$$

The right hand side of Eq. (2) is the integrated variance, with diffusion process σ, and discrete jumps of size κ. Therefore, disentangling jumps from the continuous process is an empirical issue; see, *e.g.*, Aït-Sahalia (2002, 2004). The realized volatility is the square root of the realized variance.

Generally, RV is computed using equally-spaced observations, the time intervals ranging from 1 to 30 minutes; see, *e.g.*, Andersen, Bollerslev, and Meddahi (2011b), Aït-Sahalia, Mykland, and Zhang (2005). Patton and Sheppard (2015) use consecutive transactions. When 5-minute RV is taken as the benchmark, Liu, Patton, and Sheppard (2015) find little evidence that it is outperformed by any other measure. When using inference methods

[2]Recent reviews of the literature are available for both the GARCH and SV models; see Bauwens, Laurent, and Rombouts (2006), Bollerslev (2008), Broto and Ruiz (2004)

that do not require to specify a benchmark, there is some evidence that more sophisticated measures outperform. For example, Andersen et al. (2011b) propose several estimators. Two are of particular interest: a) the average estimator and b) the optimal estimator. The linear forecasts obtained by averaging standard sparsely sampled realized volatility measures generally perform on par with the best alternative robust measures. Overall, 5-minute RV is difficult to beat, and the most classic definitions of the realized volatility are either the standard deviation of intraday (log) price changes or its log.

2.1. Stylized facts

Changes in consecutive (log) prices of financial assets, including stock, bonds and currencies, present common characteristics that are also shared by commodities futures. These stylized facts can be summarized as follows:

a. Standard deviation completely dominates the mean over daily and weekly return horizons.
b. Daily, weekly, and monthly horizons show excess kurtosis (with respect to a normal distribution).
c. Squared (and absolute) returns are strongly auto-correlated.
d. There are periods of high volatility (volatility clustering).
e. Outliers and jumps are more frequent than they should be (with respect to a normal distribution).

Despite the above-mentioned commonalities, three main differences between commodity futures and futures on stock indices have been documented in the literature. The first noticeable discrepancy is the inverse asymmetric reaction between commodity futures price and volatility or the "inverse leverage effect", arising from shocks on inventories. Typically, when the resources are scarce, the supply on the corresponding market becomes inelastic, and a decrease in one unit of inventory leads to a dramatic price upward revision; see Carpantier (2010), Carpantier and Dufays (2012), Carpantier and Samkharadze (2012), Ng and Pirrong (1994).

The second difference with respect to other financial assets has to do with the underlying stochastic process that generates price changes. Commodity futures price changes are positively skewed and, contrary to stock returns, this skewness strongly shows up at the contract level; see, e.g., Gorton and Rouwenhorst (2006). The third difference is that, at least as a first approximation, changes in (log) prices do not show any significant trend; see the one-factor model presented in Schwartz (1997, p. 926).[3]

[3]Commodity (log) prices follow a (mean reverting) Ornstein-Uhlenbeck process.

2.2. *The economic determinants of RV*

The theory of storage states that the relation between volatility of storable commodities and the level of inventories is convex and negative; see, *e.g.*, Brennan (1958), Kaldor (1939), Working (1933). More recent versions of the theory of storage in equilibrium (*e.g.* Deaton and Laroque, 1992) also predict this link, which is confirmed empirically; see Carpantier and Samkharadze (2012), Fama and French (1988), Geman and Nguyen (2005), Geman and Ohana (2009), Ng and Pirrong (1994). Interestingly, Kogan, Livdan, and Yaron (2009) extend the prediction of a non-monotonic, convex, relationship between volatility and inventories, due to investment constraints. They confirm empirically the existence of a "v-shape" relation; see also Haugom, Langeland, Molnar, and Westgaard (2014). To summarize, both low and high inventories lead to high volatility. Inventories are difficult to measure at the aggregate level with a daily frequency. A large strand of literature highlights the positive/negative relation between the slope of the term structure and inventories, and find strong empirical support; see, *e.g.*, Gorton, Hayashi, and Rouwenhorst (2012).

Samuelson (1965, 1976) conjectures that the volatility of commodity contracts is higher when the remaining time to maturity is lower. Despite many empirical tests, the results concerning this conjecture are contradictory. On the one hand, Rutledge (1976) and Grammatikos and Saunders (1986) do not find evidence of any increase in volatility. On the other hand, Milonas (1986) and Galloway and Kolb (1996) find support for all commodities. Consistent with the Samuelson hypothesis, Bessembinder, Coughenour, Seguin, and Monroe-Smoller (1996) develop a model where the spot price has negative covariance with the slope of the term structure. This implies a temporary price change, which is more likely to occur in real assets than in financial assets. Indeed, recent empirical tests on the NIKKEI (Chen, Duan, and Hung, 2000) and on the SP-500 futures (Moosa and Bollen, 2001) strongly reject the Samuelson conjecture, whereas Bessembinder et al. (1996) find empirical support mainly for the agricultural commodity futures.

Anderson and Danthine (1983) hypothesize that the key determinant of volatility is not the time to maturity but, instead, the time at which the production uncertainty is resolved. The uncertainty resolution, which occurs seasonally, for instance at the end of a crop when the supply is publicly known, theoretically commands more volatility. Conversely, the volatility should be lower when the uncertainty is resolved; see also Anderson (1985). This seasonality should be particularly visible for agricultural products whose production is concentrated in a single annual harvest in the northern hemisphere; see the statistics from the US Department of Agriculture.[4][5] It should also be present for the natural gas contract as its term structure has also a strong seasonal component due to the demand rising every winter in the northern hemisphere. Despite the fact that these turning points should firstly affect the cash market, Anderson and Danthine (1983) additionally show that the link between the cash and futures markets ensure the volatility diffusion from the former to the latter. The research also shows that intangible commodities like electricity or those whose exchange value is higher than their consumption value, such as gold or silver, behave more like traditional financial assets. Anderson (1985), Galloway and Kolb (1996), Khoury and Yourougou (1993) find a seasonal component in volatility, combined with a time to maturity effect.

To summarize, I state my hypotheses as follows. The volatility of commodities futures:

a. is positively related to the absolute value of the slope of the term structure.
b. increases when the time to maturity decreases.
c. is seasonal for commodities that show seasonality in the supply or the demand.

2.3. Endogenous determinants of RV

2.3.1. Univariate series

The introduction of a parsimonious autoregressive model dates back to Corsi (2009).[6] The main idea of this paper is that the RV on day t depends on past values of the RV at time $t-1$, $t-2, \ldots, t-p$ where p can be very high (20 or more) suggesting a long-memory process. However, this process is mean reverting toward a long-term component. Therefore, the transitory component of the daily variance relates to the RV at $t-1$ and the introduction of two additional components (weekly and monthly RVs) smooths the dynamics of RV. Altogether, these variables give a parsimonious representation (HAR) of the typical volatility exponential decay (see, *e.g.* Andersen, Bollerslev, Diebold, and Labys, 2003). In the empirical part of the paper, Corsi (2009) estimates the model with the SP-500, the USD/CHF exchange rate, and a T-Bond. Based on the BIC criterion, the one-day ahead in-sample performance of this model is higher than that of an AR (22). This clearly shows that the HAR (3) model is parsimonious. Out-of-sample, the model steadily outperforms the short-memory models (AR (1) and AR (3)) at the one-day, one-week, and one-month horizons. In addition, it is on par with an (long-memory) ARFIMA model. Moreover, the superior performance of the ARFIMA and HAR (3) increases with the forecasting horizon.

[4] `https://www.nass.usda.gov/`

[5] Although soybeans production is also very large in the southern hemisphere, the underlying product specifications and delivery locations of the contract studied are all in the northern hemisphere.

[6] Note that the first version of the paper was available in 2003.

Several versions of the model have been proposed using the realized volatility, its log, and its square. Andersen et al. (2007) show that the log of RV is the closest to normality and that jumps are negligible in terms of RV forecasting. Microstructure effects could introduce measurement errors, which lead to biased coefficients when a linear model is used. Nevertheless, the residuals of log RV are still heteroskedastic, and the parameters of the HAR are not stable over time; see Buccheri and Corsi (2019).

Using a simple linear process could be insufficient for at least three reasons: a) jumps, b) measurement errors, and c) time-varying parameters. Andersen, Bollerslev, and Huang (2011a) consider that RV has a continuous part that is well described by the HAR model, two jump components (day/night) and a night component that follows a GARCH (1,1), leading to the HAR-CJN model. The out-of-sample performance of the HAR-CJN model is slightly higher than that of the HAR.

Given the measurement error that plagues the estimation of RV, Bollerslev et al. (2016) introduce the "realized quarticity" (variance of the RV) in the HAR model. The authors write an extension (HARQ model) where the coefficients are a linear function of the quarticity. The idea is to put less weight on past high values of RV, that is when RV is subject to potential mismeasurement. By the same token, this variable is supposed to capture microstructure effects, and jumps as well. However, HARQ shows also signs of misspecification. As an alternative, Corsi and Reno (2012) and Patton and Sheppard (2015) examine whether the RV reacts symmetrically to positive and negative shocks that affect prices, *i.e.*, the so-called "leverage effect". Casas, Mao, and Veiga (2018) nest both models. Cipollini, Gallo, and Otranto (2017) show that HARQ is observationally equivalent to another model where a quadratic term in RV accounts for a faster mean reversion when volatility is high. They argue that the realized quarticity and a time-varying mean seem to play a more important role than measurement errors. In these models, the time-varying coefficients are linear functions of the realized quarticity (parametric specification). Chen et al. (2018) generalize this approach by considering a log HAR model with time-varying coefficients of unspecified functional forms (HAR-TV). These coefficients are approximated with a local linear function of time. Casas et al. (2018) extend Chen et al. (2018) in two directions. First, they consider a potential asymmetric reaction of RV to negative shocks. Second, the coefficients are no longer a local linear function of time but a linear function of the realized quarticity (semi-parametric approach). Unfortunately, the forecasting performance of the RV is not examined specifically since the main purpose of the paper is to forecast the stock market.

Bekierman and Manner (2018) take a different stance. They propose a state-space representation of the HAR model that can be augmented by functions of the realized quarticity. They attribute the higher performance of the state-space HAR models to the fact that the realized quarticity is a noisy proxy for the true measurement error, which is likely to be greater in periods of high volatility. Furthermore, their state-space models are able to capture other sources of time variation in the parameters that is not explained by the measurement error. Buccheri and Corsi (2019) generalize the state-space representation approach in several directions. Their state-space model allows for a time-varying error, and considers that the updated parameters depend on the level of uncertainty that is based on the score function. This model (SHARK) appears to perform very well, both in- and out-of-sample, but there is no straightforward extension when several assets are considered simultaneously.

2.3.2. Multivariate series

When several assets are considered simultaneously as in portfolio optimization, the covariance matrix must be accounted for. As already documented with GARCH models, when left unconstrained, the matrix requires the estimation of a large number of parameters. If K parameters are estimated for each equation, the number of equations being $N \times (N-1)/2$, the total number of parameters is $K \times (N \times (N-1)/2)$. The idea is to impose constraints so that the number of parameters is reduced dramatically; see Bollerslev et al. (2018a), Bollerslev, Patton, and Quaedvlieg (2018b), Buccheri, Bormetti, Corsi, and Lillo (2020), Chiriac and Voev (2011). Chiriac and Voev (2011) use HAR for each term of the matrix so four parameters are estimated. They apply their methodology to six stocks, and estimate a total of 60 parameters (15 linear regressions with four parameters each). The extension of this approach to more general HAR models (HARQ) is straightforward, and the estimation is easy as soon as the number of assets is not too big. Bollerslev et al. (2018b) apply this procedure to 10 stocks to construct the minimum variance and minimum tracking error portfolios. The turnover is reduced and economic gains are around 170 basis points per year, under realistic assumptions in terms of risk aversion.

Buccheri et al. (2020) assume that log prices follow a random walk and asynchronicity is treated as a missing value problem. The structure of the covariance matrix warrants it is semi-definite positive. The time-varying matrix is assumed to be the same across assets; see Engle (2002a). Therefore, all available data are used when filtering the covariances, and market microstructure noise is taken into account. The full dynamics of the process is described with a Kalman filter and the estimation is performed through maximum likelihood. When applied to 10 NYSE stocks, the authors show that opening hours are dominated by idiosyncratic risk, and that a market factor progressively emerges in the second part of the day.

Bollerslev et al. (2018a) propose a model in which the realized volatility in excess of the long-term volatility is a linear function of a set of exponentially smoothed transforms of lagged values of the RV. In one version of the model, coefficients relative to the same asset class (*i.e.*, commodities, stocks, bonds, currencies) are constrained to be equal. In terms of forecasting, the performance of their model is remarkable.

The alternative is to consider a latent factor model that spans the RV space as in Bollerslev, Meddahi, and Nyawa (2019). The authors provide a new factor-based estimator of the realized covolatility matrix, applicable when the number of assets is large and the high-frequency data are contaminated with microstructure noises. From the covolatility matrix of SP-500 stocks, they derive the minimum variance portfolio. Compared to other practically feasible competing covolatility estimators, including $1/N$, this method produces the lowest ex-post variation; see, *e.g.*, DeMiguel, Garlappi, and Uppal (2009).

Significant progress has been made in two distinct directions. First, in a univariate setting, more sophisticated specifications have been developed. Their main purpose is to clean the data from microstructure effects, and to account for an asymmetric reaction to positive and negative exogenous shocks. The focus of empirical applications has been on futures contracts, stock indices, currencies, individual stocks, and bonds. While the speed at which parameters are estimated is not a problem for univariate series, this issue becomes serious when covariance matrices are estimated. Second, the appropriate model depends very much on the financial application. When dealing with portfolios, whose composition is changing over time, a model for the covariance matrix is required since the financial series are correlated. Typically, this is the case when optimizing portfolios (minimum variance portfolio). For pricing application, univariate series are the ones that matter to forecast the implied volatility (pricing derivatives) or compute the value at risk (expected shortfall) for a single asset. For risk management, the composition of the portfolio is known ex-ante. Therefore, a single series of volatilities is constructed, and the corresponding univariate model is estimated.

3. Methodology

I start with a linear model that incorporates the economic variables discussed in Section 2.2. Then, I test whether these economic determinants have explanatory power beyond that of the past realizations of RV. Finally, I test several specifications allowing coefficients to vary over time.

3.1. Baseline model

To check whether the economic determinants of RV have any explanatory power beyond that of its own past realizations, I introduce three economic variables observable at a daily frequency. First, consistent with the theory of storage, I consider the slope of the term structure. Since there is a maturity gap across contracts, I normalize the slope. Second, to test the "v-shape" hypothesis (Kogan et al., 2009), I use a methodology similar to Haugom et al. (2014) and add a "backwardation" dummy. Third, to test the Samuelson hypothesis, I compute the log of the time to maturity in seconds for each daily observation, crossing the time stamp with the contract maturity information available in the full ticker.[7] I set it on a calendar basis, as I assume that the latent maturity information exists even when futures are not traded.[8]

[**Insert Figure 1 here**]

Finally, to test the uncertainty resolution hypothesis, I introduce monthly dummies for the corn, soybeans, wheat, and natural gas contracts to control for seasonal effects. The monthly dummies are set in July for the agricultural products, which corresponds to the harvest month of the soft red winter wheat contract in the US (and more generally for winter wheat in the northern hemisphere). It also corresponds to the "filling" month for corn and soybeans in the northern hemisphere, which is more critical than the subsequent harvesting months. For the natural gas, I select January which corresponds to the coldest month and to the highest consumption month in the US, historically. This choice also matches the unconditional seasonal pattern of RV that is displayed in Figure 1.[9]

$$\begin{aligned} RV_{c,t} = \alpha_{0,c} + \alpha_{1,c} M_{c,t} + \alpha_{2,c} TM_{c,t} + \alpha_{3,c} SL_{c,t-1} + \alpha_{4,c} B_{c,t-1} + \\ \alpha_{5,c} B_{c,t-1} \times SL_{c,t-1} + \epsilon_{c,t}, \end{aligned} \tag{3}$$

where $M_{c,t}$ is a dummy equal to "1" during the critical month of the corresponding contract and to "0" otherwise, $TM_{c,t}$ is the (log) time to maturity, $SL_{c,t-1}$ is the annualized (log) term structure slope between the nearby and first deferred contract, and $B_{c,t-1}$ is a dummy equal to "1" ("0") when this slope is in backwardation (contango). For the detailed computation procedure of the variables, see Appendix, Table A2.

[7] Bessembinder et al. (1996) consider the square root of time-to-maturity instead of the log.

[8] This is also the methodology underlying the VIX computation. In unreported robustness tests I also use business time with virtually no differences in the results.

[9] For supply-related information of agricultural products see, https://ipad.fas.usda.gov/countrysummary/. For demand-related information of the natural gas see, https://www.eia.gov/outlooks/steo/report/natgas.php. For the unconditional level of historical RV of other contracts see Appendix, Figure A8.

3.2. The autoregressive components of RV

Next, I control for the autoregressive component of the log realized volatility (RV) by introducing five different models, *i.e.*, the HAR (Corsi, 2009), the HEXP (Bollerslev et al., 2018a), the HARQ (Bollerslev et al., 2016), the HAR-TV (Chen et al., 2018), and the HARQ-TV, which nests the two latter. I choose these models because they are parsimonious and are competitive in terms of explanatory power and forecasting ability. These models are estimated with seemingly unrelated regressions (SUR), which make them easily comparable. Therefore, I first estimate the following models,

- EVHAR

The EVHAR model is,

$$RV_{c,t} = \boldsymbol{\alpha' EV_{c,t}} + \beta_{1,c} RV_{c,t-1} + \beta_{2,c} RV_{c,t-2|t-5} + \beta_{3,c} RV_{c,t-6|t-22} + \epsilon_{c,t}, \tag{4}$$

where $RV_{c,t}$ is the log RV in time t for commodity c, $\boldsymbol{EV_{c,t}}$ is the vector of the economic variables in Eq. 3 and a constant, $RV_{c,t-n|t-p}$ is the average log RV computed over the days $t-n$ to $t-p$ (previous week and month); see Corsi (2009).

- EVHEXP

The Heterogeneous Exponential of Bollerslev et al. (2018a) is similar to the HAR as it uses mixtures of exponentially smoothed past log RV. Each term is computed as, $RV_{c,t}^{CoM(\lambda)} = \sum_{i=1}^{500} \frac{e^{-i\lambda}}{e^{-\lambda}+e^{-2\lambda}+\ldots+e^{-500\lambda}}$, with $\lambda = \ln\left(1+\frac{1}{\mathrm{CoM}}\right)$, for decay rates $\lambda = 0.693, 0.182, 0.039,$ and 0.008 corresponding to centers of mass (CoM) of 1, 5, 22, and 125 days, respectively. The HEXP is,

$$RV_{c,t} = \boldsymbol{\alpha' EV_{c,t}} + \gamma_{1,c} RV_{c,t-1}^{CoM_1} + \gamma_{2,c} RV_{c,t-1}^{CoM_5} + \gamma_{3,c} RV_{c,t-1}^{CoM_{25}} + \gamma_{4,c} RV_{c,t-1}^{CoM_{125}} + \epsilon_{c,t}, \tag{5}$$

- EVHARQ

The HARQ model uses the realized quarticity to account for measurement errors. I use the following estimator for the log realized quarticity (hereafter, RQ_t),[10]

$$RQ_t = \frac{2}{3} \frac{\sum_{i=1}^{1/\Delta} r_{i,t}^4}{\left(\sum_{i=1}^{1/\Delta} r_{i,t}^2\right)^2}$$

[10]See Barndorff-Nielsen and Shephard (2002)

I use the parsimonious version where only the first coefficient of the HAR is penalized for measurement errors,

$$RV_{c,t} = \boldsymbol{\alpha' EV_{c,t}} + (\delta_{1,c} + \delta_{1Q,c} RQ_{c,t-1}) \times RV_{c,t-1} + \delta_{2,c} RV_{c,t-2|t-5} + \delta_{3,c} RV_{c,t-6|t-22} + \epsilon_{c,t}, \quad (6)$$

3.3. Are the parameters time-varying?

- EVHAR-TV

To check whether parameters are time-varying, I use the semi-parametric, local kernel, estimation approach; see Chen et al. (2018) and Casas et al. (2019). This method allows for an estimation in system (SUR), which makes the comparison with the competing models consistent. I use the Nadaraya-Watson (Nadaraya, 1964, Watson, 1964) estimator, or local constant specification. The estimator is,

$$\hat{\alpha_h}(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i) y_i}{\sum_{j=1}^{n} K_h(x - x_j)},$$

where K is the Epanechnikov kernel for a bandwidth h. The procedure uses "leave-one-out cross-validation" to select the optimal bandwidth. An additional advantage of the kernel regression is that the variance covariance matrix of errors for the feasible generalized least squares (FGLS) is itself time-varying, with a bandwith similarly selected. The model is,

$$RV_{c,t} = \boldsymbol{\alpha'(\tau_t) EV_{c,t}} + \theta_{1,c}(\tau_t) RV_{c,t-1} + \theta_{2,c}(\tau_t) RV_{c,t-2|t-5} + \theta_{3,c}(\tau_t) RV_{c,t-6|t-22} + \epsilon_{c,t}, \quad (7)$$

where the coefficients α are now time-varying, and dependent on the smoothing variable $\tau_t = \frac{t}{T}, t = 1, 2, \ldots, T$, where T is the sample size.

- EVHARQ-TV

The following model nests time-varying parameters and measurement errors, penalizing the first term of the HAR-TV with the (log) realized quarticity. The model is,

$$\begin{aligned} RV_{c,t} = \boldsymbol{\alpha'(\tau_t) EV_{c,t}} + (\phi_{1,c}(\tau_t) + \phi_{1Q,c}(\tau_t) RQ_{c,t-1}) \times RV_{c,t-1} + \phi_{2,c}(\tau_t) RV_{c,t-2|t-5} + \\ \phi_{3,c}(\tau_t) RV_{c,t-6|t-22} + \epsilon_{c,t}, \end{aligned} \quad (8)$$

3.4. *Data and descriptive statistics*

3.4.1. *Data and variable definition*

From the Barchart API, I download 5-minute closing prices of the nearby and first deferred commodity futures contracts from May 6$^{\text{th}}$, 2008 to January 18$^{\text{th}}$, 2019.[11] I choose a cross-section of nine contracts evenly spread in three main subgroups, *i.e.*, agriculture (wheat, corn, and soybeans), energy (WTI crude oil, natural gas, and heating oil), and metal (copper, gold, and silver).[12] These contracts have the highest open interest and turnover in their own subgroup.[13] Contrary to daily computations of futures price changes, that must account for the regular contract expiry, the intraday volatility computation does not require to roll the position from the nearby to the first deferred contract the day prior to maturity. In unreported robustness tests, I roll the nearby onto the first deferred five business days before maturity, with virtually the same results. Previous research on futures price changes justify this procedure because of possible market squeezes and thinly traded contracts immediately before the maturity. The data includes a time-stamp and the maturity date of each contract.

I compute 5-minute log price changes for each nearby futures contract available as $r_{c,t,j} = f^N_{c,t,j} - f^N_{c,t,j-1}$ where f is the log of the futures price and the subscripts c, t and j stand for commodity, day, and time of the observation, respectively. I compute the arithmetic RV (ARV) as,

$$ARV_{c,t} = \sqrt{\sum_{j=1}^{1/\Delta} r^2_{t,\Delta\times j}}$$

and the log RV (RV) as,

$$RV_{c,t} = \ln ARV_{c,t}$$

where $1/\Delta$ is the number of observations available given the market open hours of each contract.[14] I choose the 5-minute sampling given that the previous literature documents its performance over alternative frequencies (see, *e.g.*, Liu et al., 2015).[15]

[Insert Table 1 here]

[11] https://www.barchart.com/

[12] I report the contracts specifications in Appendix, Table A1.

[13] See Appendix, Table A4.

[14] There are 288 observations per day.

[15] Summary statistics in Appendix, Table A3 also show that it is a good compromise across the nine contracts, when compared to alternative frequencies.

3.4.2. *Summary statistics of 5-minute RV and daily market data*

Table 1 compares both distributional and memory properties of the arithmetic (Panel A) and log RV (Panel B). The daily mean of the arithmetic RV of commodity futures ranges from 1.02% (gold) to 2.72% (natural gas). The RV of commodity futures exceed those found in the previous literature for exchange rates, sovereign bonds, and stock indices. Instead, they are included in the range of typical RV found for large traded stocks. For instance, Andersen et al. (2007) find a mean arithmetic RV for the period 1986–2002 of 0.5% for the US T-Bond, of 0.67% for the Deutsche-Mark/USD exchange rate, and of 0.93% for the SP-500. On the other hand, Bollerslev et al. (2016) find that over the 1997–2013 period, the mean arithmetic RV of 27 Dow-Jones stocks is in the 1.68%–5.42% range. Buccheri and Corsi (2019) find a similar range, albeit larger (0.95%–11.10%), for 18 NYSE stocks over the 2006–2014 period. Overall, the null of normality is rejected for both arithmetic and log RV. However, in line with Andersen et al. (2003), the log RV distribution is much closer to normality than that of the arithmetic RV. The skewness of log RV is at least twice smaller compared to that of the arithmetic RV, and up to eight times smaller for the crude oil contract. The excess kurtosis of the log RV is also much smaller (by up to two orders of magnitude for the crude oil contract). Moreover, the persistence increases when using the log in place of arithmetic RV. The Ljung-Box statistic raises by up to twofold, except for the silver contract where it is slightly reduced. The log-periodogram parameter of log RV is also superior for agriculture and energy, but not for the metal products. Overall, the memory properties of the arithmetic and log RV are similar to the previous results found for exchange rates (see Andersen et al., 2007, 2003), SP-500, and US T-Bond; see Andersen et al. (2007). The bottom part of Panel B reports the distribution and memory-related statistics for the "significant" jumps J_t and the diffusion component C_t (residual) of log RV, computed for a critical value of 0.1%. In line with Andersen et al. (2007), the memory component of the log RV (Ljung-Box statistic and log-periodogram parameter) almost fully lies in the diffusion component. Also, the Ljung-Box test-statistic for the nine commodities has the same order of magnitude than that of the SP-500, albeit lower. Finally, the distribution of J_t is leptokurtic, whereas that of C_t is similar to the log RV distribution.

[Insert Table 2 here]

Table 2 reports the four moments of the distribution for the nine nearby futures contracts. The average log price changes (Panel A) is close to zero over the sample period, for all contracts. Similar to financial securities, their distributions strongly departs from normality, with an average standard deviation greatly exceeding the mean, and with an important excess kurtosis, from 1.93 (wheat), and up to 17.04 (corn). Interestingly, the skewness that has long been perceived as positive in commodity futures along with positive excess returns (Gorton and Rouwenhorst, 2006), also range from -1.17 (soybeans) to 0.33 (natural gas). I additionally provide the proportion of days during which the contracts are in contango (positive slope between nearby and first deferred contracts) which stands between 60% (soybeans) and 98% (wheat). These statistics differ from the classical view (see, *e.g.*, Keynes, 1930), according to which agricultural products are more subject to contango, because of their important storage costs. Finally, I report the optimal sampling frequency, to identify the optimal trade-off between resolution and market microstructure noise; see Aït-Sahalia et al. (2005). This optimal sampling is roughly related to the market trading volume (Panel B), minutes per day that have at least a transaction (Panel C), and bid-ask spread (Panel D). In brief, these results tend to indicate that the higher the trading activity (turnover and transactions), the lower the bid-ask spread and the optimal sampling frequency. In the remainder of the article, I define RV as the log RV computed with a 5-minute sampling frequency.

In Appendix, Table A3, I report the summary statistics of the RV, with alternative sampling frequencies of 1-, 5-, 15-, and 60-min. It verifies the good compromise that provides the 5-min sampling, both in terms of distributional and memory properties. The 5-min frequency remains also superior to the aggregated measure that averages all the aforementioned variables which is reported at the bottom of the Table; see Andersen et al. (2011b).

4. Results

4.1. In-sample estimation

4.1.1. EV and EVHAR

I estimate all models with SUR (using FGLS) for the nine commodities. Table 3 reports the coefficient estimates of the EV (Eq. 3) and the EVHAR (Eq. 4) models.

[Insert Table 3 here]

In the EV specification, the monthly dummies associated with agricultural products (July) and natural gas (January) contracts are positive and statistically significant at the 1% level.

This result is consistent with the uncertainty resolution hypothesis; see Anderson and Danthine (1983). Second, I find clear evidence of a positive relationship between the time to maturity and volatility. For seven out of the nine contracts, the corresponding coefficients are statistically significant at the 1% level. The wheat and crude oil contracts do not show statistical significance at the 10% level. Third, in line with Kogan et al. (2009) and Haugom et al. (2014), I find that the magnitude of the slope matters, but not its sign, for the crude oil contract. SL loads positively and significantly, in concurrence with a significantly negative interaction term with the backwardation dummy, $SL \times B$. This indicates that both contango and backwardation market are positive predictors of RV, thus supporting the "v-shape" hypothesis. This pattern is present for the crude oil, heating oil, natural gas, and copper contracts. The contango slope coefficient SL of the wheat contract is also positive and significant at the 1% level, but not the backwardation slope coefficients. Interestingly, the coefficient for SL is negative and significant at the 1% level for the corn, gold, and silver contract, thereby showing a negative contango-RV relationship. The explanatory power of the EV model also varies across contracts, with adjusted R^2 for individual equations ranging from 2% (copper) to 26% (crude oil).

When I include the EV into the HAR, I find that the explanatory power of RV is vastly improved, with adjusted R^2 ranging from 33% (soybeans) to 75% (crude oil). Consistently, the likelihood ratio ${}^{EVHAR}\!/_{HAR}$ that tests whether the EV coefficients are nil is rejected at the 1% level, which points to a significant benefit for including these variables.[16] The HAR coefficients closely align with those of Andersen et al. (2007) for the DEM/USD exchange rate, SP-500, and US T-Bond. Autoregressive terms are statistically significant at the 1% level. The magnitude of the monthly dummies decreases by at least threefold. Their statistical significance is reduced, at the 5% level for the agricultural products but remain statistically significant at the 1% level for the natural gas. Similarly, the time to maturity coefficients only maintain statistical significance at the 1% level across metal products (and for the soybeans contract, at the 5% level). Interestingly, the size of the slope-related coefficients decreases, are still inconsistent across metal products, but remain of the same sign. The EV coefficients decline in the EVHAR specification. It indicates that the memory properties already capture most of EV-related effects, from both futures markets (time to maturity) and supply and demand (seasonality and term structure slope) perspectives.

4.1.2. Alternative autoregressive specifications

Table 4 reports the joint estimation of the EVHEXP (Eq. 5) and EVHARQ (Eq. 6) models.

[16] I report the results of the estimation of the restricted HAR specification in Appendix, Table A5.

[Insert Table 4 here]

The inclusion of the realized quarticity affects further the EV coefficients. First, the coefficient of the monthly dummy is statistically significant at the 1% (5%) level for the natural gas (agricultural products). The coefficients corresponding to the time to maturity decline more drastically, by twofold for the silver contract, by threefold for the gold contract. This suggests that measurement errors are also related to the contract maturity. The coefficient related to quarticity is also statistically significant at the 1% level for all contracts subject to a (positive) time to maturity-RV relationship in the simple EV model (except for the silver contract). Finally, the coefficients associated to the slope decrease further after the inclusion of the realized quarticity but not consistently across contracts. This makes it difficult to relate the realized quarticity to the level of the term structure. Last, the explanatory power induced by the realized quarticity inclusion increases further. Adjusted R^2 of individual equations are up to three percentage points with respect to the EVHAR version. The likelihood ratio ${}^{EVHARQ}/_{HARQ}$, which tests whether the EV coefficients are nil in the HARQ specification, is rejected at the 1% level. This points again to a significant benefit for the EV inclusion.[17]

As expected, given their lags, HEXP variables particularly impact the cyclical EVs, monthly dummies and time to maturity, but do not decrease the coefficients as much as the HARQ. In particular for the metal groups, the time to maturity coefficients are statistically significant at the 1% level. Similarly, the monthly dummies of the four contracts remain larger and more statistically significant for the EVHEXP than for the EVHARQ. The HEXP coefficients themselves show that most of the variance is explained by the first lag (1-day CoM). The coefficients of 5-days CoM are statistically insignificant except for the metal products and the natural gas contracts, whereas the one of 25-days CoM is statistically significant at the 1% level for seven of the nine contracts. The coefficients for the 125-days CoM is negative for all "seasonal" commodities (agricultural products and natural gas contracts), but only statistically significant at the 1% level for the soybeans contract. Indeed, this variable is centered with a lag of half a year, and thereby its negative coefficient confirms the seasonal structure of volatility. This parameter remains instead positive (and statistically significant at the 1% level) for the heating oil and silver contracts, which I did not hypothesize to be seasonal. Finally, the EVHEXP underperforms both EVHAR and

[17] I report the results of the estimation of the restricted HARQ specification in Appendix, Table A5.

EVHARQ specifications, with adjusted R^2 of individual equations that are five percentage points lower (except for metal products). Along with these results, the OLS and McElroy R^2 for the EVHEXP system remain lower than those of EVHAR and EHVARQ specifications. Finally, the likelihood ratio test is statistically significant at the 1% level, which implies a larger benefit of the EV inclusion in the HEXP specification.[18]

4.1.3. Time-varying coefficients

Table 5 reports the results of the estimations of the EVHAR and EVHARQ specifications, in which the parameters are allowed to be time-varying (EVHAR-TV, Eq. 7 and EVHARQ-TV, Eq. 8). The fact that all coefficients vary permits to make inference on their stability over time, but also to analyze whether the previously documented cyclical effects are linear. I report the averages of the parameters across time and their standard error in square brackets.

[Insert Table 5 here]

In the time-varying specifications, I find that the coefficient variations are the largest for the intercept in all nine equations, with up to a 29% standard deviation for the crude oil contract in the EVHAR-TV.[19] The variation intensity for the first autoregressive parameter RV_{t-1} is much lower on average across contracts, with a maximum of 5% for the silver and a minimum of 1% for the crude and heating oil contract. The coefficients for the week and month lags lie in the same range. This difference in time-variation intensity is in line with the results of Buccheri and Corsi (2019) in their score-driven HAR (SHAR).[20] The means of the parameters of the EVHAR(Q)-TV are closely aligned on those of the static EVHAR(Q) version. The EVHARQ-TV explanatory power over the EVHAR-TV is improved, similar to the static comparison. The likelihood ratio tests ${}^{EVHAR-TV}/_{HAR-TV}$ and ${}^{EVHARQ-TV}/_{HARQ-TV}$ are both significant at the 1% level.[21] The pseudo R^2 of the individual equations are mostly unchanged, with upward or downward revision by one percentage point at most. Yet, the standard deviations of HAR-TV and HARQ-TV- related parameters are considerably reduced when the EVs are included. For instance, the standard deviation of the

[18] I report the results of the estimation of the restricted HEXP specification in Appendix, Table A5.

[19] See Appendix, Figure A9.

[20] Given the low variations of both autoregressive variables and EV, I do not report the related plots. These results are available upon request.

[21] I report the results of the estimation of the restricted HAR-TV and HARQ-TV in Appendix, Table A6.

intercept of the silver contract decreases from 29% in the HAR-TV, to 4% in the EVHAR-TV.

Given the high variation of the intercept, it is likely that the cyclical components of the EV (time to maturity and monthly dummies) are captured by the time-varying specifications. However, in both models, the EV parameters remain time-varying. The time to maturity coefficients, for instance, have standard deviations of up to 53% for the corn contract. This points to a non-linear relationship between the time to maturity and the level of RV, as documented by Hong (2000). The term structure-related coefficients remain heterogeneous, with only five contracts that display slope and interaction parameters following the "v-shape" pattern (jointly positive and negative, respectively).

Finally, I report in square brackets the standard errors of the time-varying coefficients and the corresponding statistical significance level of a t-test that the coefficient significantly departs from zero. For all time-varying specifications and coefficients, I find statistical significance at the 1% level. This points to the fact that even when the coefficients are small, they are mildly time-varying. The pseudo R^2 of the individual equations increase by up to four percentage points (natural gas contract).

To summarize the aforementioned results, I find that the EV do explain the RV but with a much lower explanatory power than any autoregressive specification. The signs of the coefficients and their statistical significance strongly support the uncertainty resolution. On the other hand, I uncover a positive time to maturity-RV relationship, both statistically and economically significant, thereby rejecting the Samuelson hypothesis. Finally, the results for the term structure variables support both the theory of storage and the "v-shape" hypotheses, but in only four of the nine commodity contracts. When the autoregressive specifications are introduced, these relationships are significantly reduced regarding time to maturity and monthly dummies coefficients. They vanish in the case of the term structure slope. This indicates that the information content of the EV is already captured by the various lags of the HAR(Q) and HEXP specifications. Yet, the unrestricted versions, which include EV, improve the explanatory power of the restricted autoregressive models in all cases, with likelihood ratio tests always statistically significant at the 1% level. The following in- and out-of-sample forecast analysis disentangles further the contribution of the EV in the RV modeling significant both statistically and economically.

4.2. In-sample performance

I compute in-sample forecasts at the one-day, one-week, and one-month horizons, jointly for the 11 models; see, *e.g.*, Andersen et al. (2007). I keep the time to maturity and monthly dummies contemporaneous because of their deterministic characteristics. Table 6 reports the results of the Model Confidence Set (MCS) procedure, which allows for direct comparison performance of all models at once, and for different losses: mean squared errors (MSE), mean absolute errors (MAE) and QLIKE; see Hansen, Lunde, and Nason (2011), Patton (2011).

[**Insert Table 6 here**]

Table 6, Panel A shows the tests on the one-day ahead forecasts. The HARQ-TV is superior for the three losses considered, and the EVHARQ ranks second, although the MAE losses are excluded from the 90% confidence interval. The one-week and one-month ahead comparisons indicate that the HAR-TV is superior, excluding the one-month ahead MSE, for which the HARQ-TV ranks first. These results are in line with Bollerslev et al. (2016) who find that the restricted HARQ (with the realized quarticity applied only to the one-day lag) is superior at the one-day horizon over the "full" HARQ, for the SP-500 (MSE) and for 27 Dow-Jones stocks (MSE and QLIKE). However, for longer horizons they find the converse. Thus, across all models and in-sample configurations, only a single EV-based model (the EVHARQ) steps up in the 90% confidence set for the MSE and QLIKE losses, at the one-day horizon. Finally, the losses are consistently smaller (greater) in the static (time-varying) parameters versions of the models, when the EV is included. This points to the fact that time-varying specifications appropriately capture the time variations in EV.

4.3. Out-of-sample performance

I compute out-of-sample forecasts based on a calibration window that spans the April 22, 2010–January 30, 2013 period for the one-day, one-week, and one-month horizon. I obtain the static models forecasts computing the expectations of RV_t given the parameters estimated from the calibration window. For the time-varying specifications, I use the multi-stage non parametric predictor approach; see Chen, Yang, and Hafner (2004) and Chen et al. (2018). In this procedure, I first compute the one-step ahead conditional expectations and re-use them to select the new conditional optimal bandwidth for the next step(s), iteratively. Next, I use the MCS procedure and the modified Diebold-Mariano test (Harvey, Leybourne, and Newbold, 1997) to benchmark these out-of-sample forecasts; see Diebold and Mariano (1995). These results are reported in Table 7. In this analysis, I add the RiskMetrics model as a generic benchmark, given its wide use in risk management.[22] The RiskMetrics model

[22]The RiskMetrics methodology can be found here: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a

can be seen as a parsimonious version of the HEXP of Bollerslev et al. (2018a), with a single exponentially smoothed lagged variable. The decay rate λ is set at 6%, which corresponds to a 16-days CoM. Although the RiskMetrics is not calibrated for log variances or volatilities, in unreported test, I find that this decay rate remains a good trade-off for log realized volatilities. The RiskMetrics equation is,

$$RV_{c,t} = \mu_{0,c} + \mu_{1,c} RV_{c,t-1}^{CoM_{16}} + \epsilon_{c,t},$$

[**Insert Table 7 here**]

Table 7, Panel A shows that the HARQ strictly dominates all other models at the one-day and one-week horizon. Additionally, I find that the benefit of the EV inclusion vanishes over these two horizons. However, at the one-month horizon, the best performing model is the EVHARQ-TV, with an MSE (MAE) almost three (two) times lower than the one of the HARQ-TV. This forecast improvement, arising from the EV inclusion, is present in almost all models and for all losses. This supports the benefits of including these exogenous variables when the time horizon increases. These forecasting improvements may also arise from the fact that the time to maturity and monthly dummies are deterministic variables. As they are readily available for the n-ahead periods, they could improve the forecasting power at longer horizons. However, such interpretation departs from the concept of market efficiency. Finally, Table 7, Panel B reports the results of a one-to-one direct comparison of the 12 models using the modified Diebold-Mariano test, at the one-day ahead horizon and for the MSE loss. These results indicate a strict dominance of the HARQ and EVHARQ-TV models. When these two models are compared, the χ^2 statistic is insignificant (-1.31), hence supporting their equally superior forecasting ability at this horizon, in line with previous results.

5. Tail-risk modeling

I now use the out-of-sample forecasts from the 12 models to compare their ability to forecast the (left) tail risk. A large strand of literature adopts the expected shortfall (ES) methodology, in place of the value at risk (VaR). I adopt the multi-quantile regression approach; see, *e.g.*, Bayer and Dimitriadis (2020), Couperier and Leymarie (2020). It allows to model the left tail of a distribution with higher granularity as any sequence of coverage levels may be used. Table 8 reports the p-values of tests on coefficients biases from forecast accuracy panel regressions, $\boldsymbol{RV_t} = \alpha + \beta\widehat{\boldsymbol{RV_t}}$, where $\boldsymbol{RV_t} = [RV_{1,t}, RV_{2,t}, \ldots, RV_{9,t}]'$. The tests are for the following four null hypotheses:

- $H_{0,J_1} : \sum_{j=1}^{p} (\beta_0(\tau_j)) + (\beta_1(\tau_j)) = p$
- $H_{0,J_2} : \sum_{j=1}^{p} \beta_0(\tau_j) = 0$ and $\sum_{j=1}^{p} \beta_1(\tau_j) = p$
- $H_{0,I} : \sum_{j=1}^{p} \beta_0(\tau_j) = 0$
- $H_{0,S} : \sum_{j=1}^{p} \beta_1(\tau_j) = p$

Where p is the number of coverage levels selected, and τ is the corresponding level for $j = 1, 2, \ldots, p$. In other words, J_1 tests the null hypothesis that the sum of the intercepts and slopes sum to p, J_2 that the sum of the intercepts and slopes sum to zero and p, respectively, and I (S) that the sum of the intercepts (slopes) sum to zero (p), individually.

[Insert Table 8 here]

The J_2 test, which is similar to a Mincer-Zarnowitz regression, albeit less restrictive, rejects the null hypothesis for the EV, the time-varying models and the RiskMetrics.[23] This is the case when the coverage is set at $p = 1$, equivalent to the 97.5% VaR, and for higher granularities up until $p = 6$. The alternative Wald test, $J1$ whose null hypothesis is that the sums of all multi-quantile regressions parameters (both intercepts and coefficients) are equals to p, is only rejected for the HAR-TV and $p = 2$, at the 10% level. Similarly, the Wald tests I and S for which the null hypotheses is that the sums of the intercepts and slopes are jointly equals to zero and p, respectively are almost significant at the 5% level, excluding the EVHARQ in its I test. Therefore, I focus on the rejection rates of the J_2 test, to compare the relative performance of the models. First, the static parameter version of the HAR delivers the highest p-values (minimum of 0.25). Interestingly, the p-value decreases as the coverage increases, in line with Couperier and Leymarie (2020) and their original test

[23]The Mincer-Zarnowitz test is a particular case of the J_2, when there is a single coverage level, *i.e.*, $p = 1$.

on AR (1) and GARCH (1,1) models. In static parameter (time-varying) specifications, the inclusion of the EV decreases (increases) the expected-shortfall forecasting performance.[24]

I conclude the tail risk section with the analysis of the results of Table 9, where I present the percentage of violations occurring over the sample for each contract, for a single quantile regression, with a coverage level set at the 97.5% VaR. The horizons of the out-of-sample forecasts is of one day.

[**Insert Table 9 here**]

On average, the EVHARQ provides the lowest violation percentage at the 97.5% VaR level, but there are important disparities across contracts. There is no systematic benefit arising from the EV inclusion. Similarly, the most complex time-varying specifications do not deliver better forecasts in terms of VaR violations. Finally, in four contracts, the single EV model yields the lowest violation percentages. Despite the fact that these results stand for a single coverage level and do not encompass an entire expected shortfall violation, they point to some benefits of EV inclusion when modeling tail risk.

6. Conclusion

This research aims to test whether economic variables, theoretically related to the volatility of commodity futures contracts, add value to autoregressive RV models, selected for their forecasting performance in other asset classes. Using joint estimations for nine commodities, my results strongly support the uncertainty resolution hypothesis; see Anderson and Danthine (1983) and Anderson (1985). Second, I find a strong rejection of the Samuelson hypothesis (Samuelson, 1976) with results indicating that the RV is positively related to time to maturity. Third, the inclusion of the term structure slope in the regressions yields mixed results. On the one hand, I find support for the theory of storage and "v-shape" hypothesis of Kogan et al. (2009) for the heating oil, natural gas, copper, and crude oil contracts; see also Haugom et al. (2014). On the other hand, I do not identify any support for the soybeans, wheat, and gold contracts and even opposite results in the cases of corn and silver contracts, for which contango commands a lower RV. However, in the case of metals, it is likely that

[24] Appendix, Table A7 reports the coefficients of the predictive regressions for each coverage level τ, individually, when the number of quantile is set to $p = 4$. It also shows that while all specifications yield a significant forecasting bias, the static HAR and EVHAR models generate the lowest, and this for all coverage levels used.

the determinants of the term structure are mostly unrelated to supply and demand and their inherent storage issues. Finally, the performance of the economic variables considered alone, lies far below those of all autoregressive models, including their most parsimonious versions such as RiskMetrics. Similarly, when I nest the EV in autoregressive models, I find marginal gains in explanatory power, both in terms of R^2 and losses. Moreover, all likelihood ratio tests of unrestricted EV *vs.* restricted specifications are significant at the 1% level. In out-of-sample tests, the economic variables improve the forecast accuracy at longer time horizon (one-month ahead), even in specifications accounting for measurement errors and time-varying coefficients together. These gains vanish for expected shortfall backtests from multi-quantile regressions. Yet, surprisingly, in four out of the nine contracts, the EV model alone generates less 97.5% VaR violations than all other models, and provide good results for five other contracts. I leave for future research the exploration of whether exogenous, economically-motivated, variables related to volatility theories provide similar results in other asset classes.

References

Andersen, T. G., Bollerslev, T., 1998a. Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements and long run dependencies. Journal of Finance 53, 219–265.

Andersen, T. G., Bollerslev, T., 1998b. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Review of Economics and Statistics 89, 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Andersen, T. G., Bollerslev, T., Huang, X., 2011a. A reduced form framework for modeling volatility of speculative prices based on realized variation measures. Journal of Econometrics 160, 176–189.

Andersen, T. G., Bollerslev, T., Meddahi, N., 2011b. Realized volatility forecasting and market microstructure noise. Journal of Econometrics 160, 220–234.

Anderson, R. W., 1985. Some determinants of the volatility of futures prices. Journal of Futures Markets 5, 331–348.

Anderson, R. W., Danthine, J. P., 1983. The time pattern of hedging and the volatility of futures prices. Review of Economic Studies 50, 249–266.

Aït-Sahalia, Y., 2002. Telling from discrete data whether the underlying continuous-time model is a diffusion. Journal of Finance 57, 2075–2112.

Aït-Sahalia, Y., 2004. Disentangling diffusion from jumps. Journal of Financial Economics 74, 487–528.

Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B 64, 253–280.

Bates, D. S., 1996. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. Review of Financial Studies 9, 69–107.

Bauwens, L., Laurent, S., Rombouts, J. V. K., 2006. Multivariate GARCH models: A survey. Journal of Applied Econometrics 21, 79–109.

Bayer, S., Dimitriadis, T., 2020. Regression-based expected shortfall backtesting. Journal of Financial Econometrics, forthcoming.

Bekierman, J., Manner, H., 2018. Forecasting realized variance measures using time-varying coefficient models. International Journal of Forecasting 34, 276–287.

Bessembinder, H., Coughenour, J. F., Seguin, P. J., Monroe-Smoller, M., 1996. Is there a term structure of futures volatilities? Reevaluating the Samuelson hypothesis. Journal of Derivatives 4 (2), 45–58.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., 2008. Glossary to arch (garch). CREATES Research Paper 2008-49, available at: `http://dx.doi.org/10.2139/ssrn.1263250` .

Bollerslev, T., Hood, B., Huss, J., Pedersen, L. H., 2018a. Risk everywhere: Modeling and managing volatility. Review of Financial Studies 31, 2729–2773.

Bollerslev, T., Meddahi, N., Nyawa, S., 2019. High-dimensional multivariate realized volatility estimation. Journal of Econometrics 212, 116–136.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192, 1–18.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2018b. Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. Journal of Econometrics 207, 71–91.

Brennan, M. J., 1958. The supply of storage. American Economic Review 48, 50–72.

Broto, C., Ruiz, E., 2004. Estimation methods for stochastic volatility models: A survey. Journal of Economic Surveys 18, 613–649.

Buccheri, G., Bormetti, G., Corsi, F., Lillo, F., 2020. A score-driven conditional correlation model for noisy and asynchronous data: An application to high-frequency covariance dynamics. Journal of Business and Economic Statistics, forthcoming.

Buccheri, G., Corsi, F., 2019. HARK the SHARK: Realized volatility modeling with measurement errors and nonlinear dependencies. Journal of Financial Econometrics, forthcoming.

Carpantier, J.-F., 2010. Commodities inventory effect. Available at: https://econpapers.repec.org/RePEc:cor:louvco:2010040 .

Carpantier, J.-F., Dufays, A., 2012. Commodities volatility and the theory of storage. Available at: https://EconPapers.repec.org/RePEc:cor:louvco:2012037 .

Carpantier, J.-F., Samkharadze, B., 2012. The asymmetric commodity inventory effect on the optimal hedge ratio. Journal of Futures Markets 33, 868–888.

Casas, I., Ferreira, E., Orbe, S., 2019. Time-varying coefficient estimation in SURE models. Application to portfolio management. Journal of Financial Econometrics, forthcoming.

Casas, I., Mao, X., Veiga, H., 2018. Reexamining financial and economic predictability with new estimators of realized variance and variance risk premium. CREATES Research Paper 2018-10, available at https://ideas.repec.org/p/aah/create/2018-10.html .

Chen, R., Yang, L., Hafner, C., 2004. Nonparametric multistep-ahead prediction in time series analysis. Journal of the Royal Statistical Society: Series B 66, 669–686.

Chen, X. B., Gao, J., Li, D., Silvapulle, P., 2018. Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. Journal of Business and Economic Statistics 36, 88–100.

Chen, Y.-J., Duan, J.-C., Hung, M.-W., 2000. Volatility and maturity effects in the NIKKEI index futures. Journal of Futures Markets 19, 895–909.

Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realized volatility. Journal of Applied Econometrics 26, 922–947.

Cipollini, F., Gallo, G. M., Otranto, E., 2017. On heteroskedasticity and regimes in volatility forecasting. Available at http://dx.doi.org/10.2139/ssrn.3037550 .

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Corsi, F., Reno, R., 2012. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. Journal of Business and Economic Statistics 30, 368–380.

Couperier, O., Leymarie, J., 2020. Backtesting expected shortfall via multi-quantile regression. Available at: https://halshs.archives-ouvertes.fr/halshs-01909375v5/file/Backtesting%20ES%20via%20Multi-Quantile%20Regression.pdf .

Deaton, A., Laroque, G., 1992. On the behaviour of commodity prices. Review of Economic Studies 59, 1–23.

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies 22, 1915–1953.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R. F., 2002a. Dynamic conditional correlation. Journal of Business and Economic Statistics 20, 339–350.

Fama, E. F., French, K. R., 1988. Business cycles and the behavior of metals prices. Journal of Finance 43, 1075–1093.

Galloway, T. M., Kolb, R. W., 1996. Futures prices and the maturity effect. Journal of Futures Markets 16, 809–828.

Geman, H., Nguyen, V.-N., 2005. Soybean inventory and forward curve dynamics. Management Science 51, 1076–1091.

Geman, H., Ohana, S., 2009. Forward curves, scarcity and price volatility in oil and natural gas markets. Energy Economics 31, 576–585.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.

Gorton, G., Hayashi, F., Rouwenhorst, K. G., 2012. The fundamentals of commodity futures returns. Review of Finance 17, 35–105.

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

Grammatikos, T., Saunders, A., 1986. Futures price variability: A test of maturity and volume effects. Journal of Business 59, 319–330.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79, 453–497.

Harvey, S., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13, 281–291.

Haugom, E., Langeland, H., Molnar, P., Westgaard, S., 2014. Forecasting volatility of the U.S. oil market. Journal of Banking and Finance 47, 1–14.

Heston, S. L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327–343.

Hong, H., 2000. A model of returns and trading in futures markets. Journal of Finance 55, 959–988.

Kaldor, N., 1939. Speculation and economic stability. Review of Economic Studies 7, 1–27.

Keynes, J. M., 1930. A treatise on money: The applied theory of money. In: MacMillian (ed.), *A Treatise on Money: The Applied Theory of Money*, MacMillian, London, pp. 3–46.

Khoury, N., Yourougou, P., 1993. Determinants of agricultural futures price volatilities: Evidence from Winnipeg commodity exchange. Journal of Futures Markets 13, 345–356.

Kogan, L., Livdan, D., Yaron, A., 2009. Oil futures prices in a production economy with investment constraints. Journal of Finance 64, 1345–1375.

Liu, L. Y., Patton, A. J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Journal of Econometrics 187, 293–311.

Milonas, N. T., 1986. Price variability and the maturity effect in futures markets. Journal of Futures Markets 6, 443–460.

Moosa, I. A., Bollen, B., 2001. Is there a maturity effect in the price of the S&P 500 futures contract? Applied Economic Letters 8, 693–695.

Nadaraya, E. A., 1964. On estimating regression. Theory of Probability and its Applications 9, 141–142.

Newey, W. K., West, K. D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–653.

Ng, V. K., Pirrong, S. C., 1994. Fundamentals and volatility: Storage, spreads and the dynamics of metals prices. Journal of Business 67, 203–230.

Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

Patton, A. J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Robinson, P. M., 1995. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048–1072.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Rutledge, D. J. S., 1976. A note on the variability of futures prices. Review of Economics and Statistics 58, 118–120.

Samuelson, P. A., 1965. Proof that properly anticipated prices fluctuate randomly. Industrial Management Review 6, 41–49.

Samuelson, P. A., 1976. Is real-world price a tale told by the idiot of chance? Review of Economics and Statistics 58, 120–123.

Schwartz, E. S., 1997. The stochastic behavior of commodity prices: Implications for valuation and hedging. Journal of Finance 52, 923–973.

Watson, G. S., 1964. Smooth regression analysis. Indian Journal of Statistics: Series A 26, 359–372.

Working, H., 1933. Wheat studies. Food Research Institute 9 (6), 185–240.

Table 1: **Summary statistics: Daily RV**

This table reports summary statistics for four estimators of daily realized volatility. These estimators are based on 5-minute log price changes of the nearby futures commodity contract. I display these statistics for the arithmetic realized volatility ARV_t (Panel A), and the log realized volatility RV_t (Panel B). In Panel B, I also split the RV in log "significant" jumps $\frac{1}{2} \times \ln J_t$ and in log continuous diffusion $\frac{1}{2} \times \ln C_t$ components. I compute J_t and C_t as in Andersen et al. (2007, p. 710), with a critical value of 0.1%. Each panel displays the first four moments of the distribution, the Jarque-Bera statistic JB, the Ljung-Box statistic (20[th] order serial correlation) $Q_{(20)}$, and the parameter d of the log-periodogram regression (Geweke and Porter-Hudak, 1983, Robinson, 1995) based on a bandwidth exponent of 4/5 as in Andersen et al. (2003). The sample period is May 5, 2008–April 1, 2019 for nine commodity futures contracts. The number of observations per contract is 2755.

	Agriculture			Energy			Metal		
	Corn (C)	Soybeans (S)	Wheat (W)	WTI crude oil (CL)	Heating oil (HO)	Natural Gas (NG)	Gold (GC)	Copper (HG)	Silver (SI)
					Panel A Daily arithmetic realized volatility, ARV_t				
Mean%	1.72	1.41	1.94	2.06	1.80	2.72	1.02	1.57	1.83
σ%	1.01	0.80	0.84	1.21	0.93	1.36	0.58	0.93	1.06
Skewness	7.50	4.86	2.41	4.13	2.82	4.35	2.94	2.84	2.96
Kurtosis	145.87	52.33	11.81	49.42	21.66	46.60	16.78	11.57	16.52
JB	2,471,897	325,629	18,710	288,677	57,598	258,315	36,358	19,095	35,395
$Q_{(20)}$	4,856	6,723	7,139	26,574	25,471	8,768	18,971	27,365	15,973
d	0.24	0.35	0.32	0.50	0.50	0.41	0.44	0.59	0.49
					Panel B Daily logarithmic realized volatility, RV_t				
Mean%	−415.13	−434.90	−400.54	−400.03	−412.02	−369.11	−468.93	−427.66	−412.04
σ%	46.48	46.19	41.54	46.69	43.93	39.85	50.87	46.46	52.02
Skewness	2.06	2.37	2.16	0.51	0.42	0.55	1.61	0.65	0.58
Kurtosis	13.73	17.02	18.39	0.61	0.36	1.01	13.22	1.14	7.55
JB	23,627	35,914	41,036	161	95	259	21,286	343	6,706
$Q_{(20)}$	11,283	10,703	8,201	32,443	31,598	18,330	12,106	23,773	11,329
d	0.34	0.40	0.32	0.52	0.53	0.51	0.33	0.51	0.38
					Daily logarithmic significant jumps at 0.1%, $\frac{1}{2} \times \ln J_t$				
Mean%	−176.68	−186.85	−153.55	−85.97	−140.97	−113.07	−239.81	−210.52	−162.31
σ%	226.56	232.20	211.24	179.99	215.66	185.86	249.65	229.00	214.20
Skewness	−0.55	−0.48	−0.69	−1.65	−0.91	−1.08	−0.12	−0.21	−0.62
Kurtosis	−1.62	−1.70	−1.45	0.82	−1.09	−0.73	−1.91	−1.87	−1.51
JB	439	436	458	1,334	519	600	426	421	437
$Q_{(20)}$	120	273	73	100	220	84	1,839	1,817	582
d	0.10	0.08	0.08	0.05	0.02	0.01	0.13	0.15	0.11
					Daily logarithmic continuous component at 0.1%, $\frac{1}{2} \times \ln C_t$				
Mean%	−428.25	−451.39	−416.08	−403.84	−418.74	−376.33	−507.84	−468.77	−442.42
σ%	48.63	50.24	50.06	46.74	44.33	38.04	80.93	81.57	77.21
Skewness	1.34	1.32	0.36	0.47	0.21	0.31	−0.56	−0.93	−0.83
Kurtosis	12.98	14.51	12.26	0.45	0.37	0.27	4.18	1.73	3.41
JB	20,217	25,027	17,345	124	37	53	2,156	745	1,652
$Q_{(20)}$	10,277	7,173	3,969	33,615	29,694	21,764	4,083	7,481	3,931
d	0.37	0.37	0.34	0.54	0.48	0.52	0.35	0.45	0.33

Table 2: **Summary statistics: Daily market-level data**

This table reports summary statistics of daily data for nine nearby commodity futures contracts. I display the statistics for the log price changes (Panel A), the trading volume (Panel B), the number of minutes with at least one transaction (Panel C), and the bid-ask spread estimated on 1-minute data with the Roll (1984) methodology (Panel D). I report the first four moments of the distributions, the minima, and the maxima. Panel A also reports the Ljung-Box statistic (20^{th} order serial correlation for the log price changes) $Q_{(20)}$ and the proportion of days during which the nearest term structure was in contango. Finally, I report the optimal sampling in minutes based on Aït-Sahalia et al. (2005, p. 361), assuming that the microstructure noise is gaussian and only driven by the bid-ask spread. The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755.

	Agriculture			Energy			Metal		
	Corn (C)	Soybeans (S)	Wheat (W)	WTI crude oil (CL)	Heating oil (HO)	Natural Gas (NG)	Gold (GC)	Copper (HG)	Silver (SI)
	Panel A: Daily log price changes of nearby contracts								
Mean%	−0.02	−0.01	−0.02	−0.02	−0.02	−0.05	0.01	−0.01	−0.004
σ%	1.92	1.60	2.07	2.38	1.93	2.96	1.12	1.73	1.99
Skewness	−1.07	−1.17	0.16	0.05	−0.17	0.33	−0.07	−0.16	−0.99
Kurtosis	17.04	9.30	1.93	4.19	3.10	3.25	7.52	4.01	7.55
$Q_{(20)}$	26.54	31.48	31.37	52.41	30.09	60.09	17.29	74.53	22.54
Contango%	85.30	60.16	97.55	80.50	74.52	85.36	71.62	62.39	68.68
Optimal sampling (min)	43	20	30	21	17	29	14	18	25
	Panel B: Trading volume (in million USD)								
Mean	1,466.37	2,151.16	827.84	16,506.63	1,658.83	2,548.02	8,529.10	1,948.14	2,699.74
σ	1,223.27	2,111.44	621.02	9,594.49	1,172.04	1,461.08	11,200.83	2,205.74	3,003.31
Skewness	0.86	1.06	0.60	0.54	0.57	1.31	1.34	1.42	2.13
Kurtosis	0.82	1.81	0.64	0.89	−0.39	5.47	2.01	2.96	11.25
Min	0.03	0.06	0.04	0.13	0.08	0.60	0.08	0.07	0.07
Max	8,277.70	15,145.34	4,330.91	73,000.47	6,605.90	15,031.34	94,868.16	16,387.95	35,581.24
	Panel C: Minutes with at least one transaction								
Mean	528.16	518.13	509.41	1,201.31	675.81	898.16	679.75	704.80	744.90
σ	254.11	316.54	264.43	308.60	260.95	258.02	604.09	556.00	552.40
Skewness	−0.47	−0.17	−0.61	−0.99	−0.60	−0.52	0.22	−0.15	−0.31
Kurtosis	−0.70	−1.28	−0.67	2.71	0.49	2.50	−1.48	−1.54	−1.50
Min	1	1	1	1	1	1	1	1	1
Max	1,158	21,226	1,051	1,395	1,413	1,416	1,394	1,427	1,392
	Panel D: Bid-ask spread (in bps)								
Mean	2.67	1.27	2.19	1.93	1.33	2.99	0.68	1.26	1.82
σ	1.73	1.31	1.86	1.73	1.46	2.42	0.85	1.43	1.80
Skewness	0.65	2.35	2.54	1.55	1.85	3.91	2.65	2.39	2.00
Kurtosis	3.62	13.64	28.30	4.92	7.16	63.38	14.26	10.30	8.48
Min	0.01	0.004	0.01	0.04	0.09	0.15	0.002	0.002	0.0002
Max	16.61	16.04	30.23	14.98	15.65	52.10	10.19	13.37	17.41

Figure 1: **Unconditional monthly average RV**

These plots display the monthly average of daily RV for the nearby futures contract written on corn, soybeans, wheat, and natural gas. The red line represents the average and the gray area represents the 90% confidence bands. For a better alignment, the plot of the natural gas contract is centered in January. The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755.

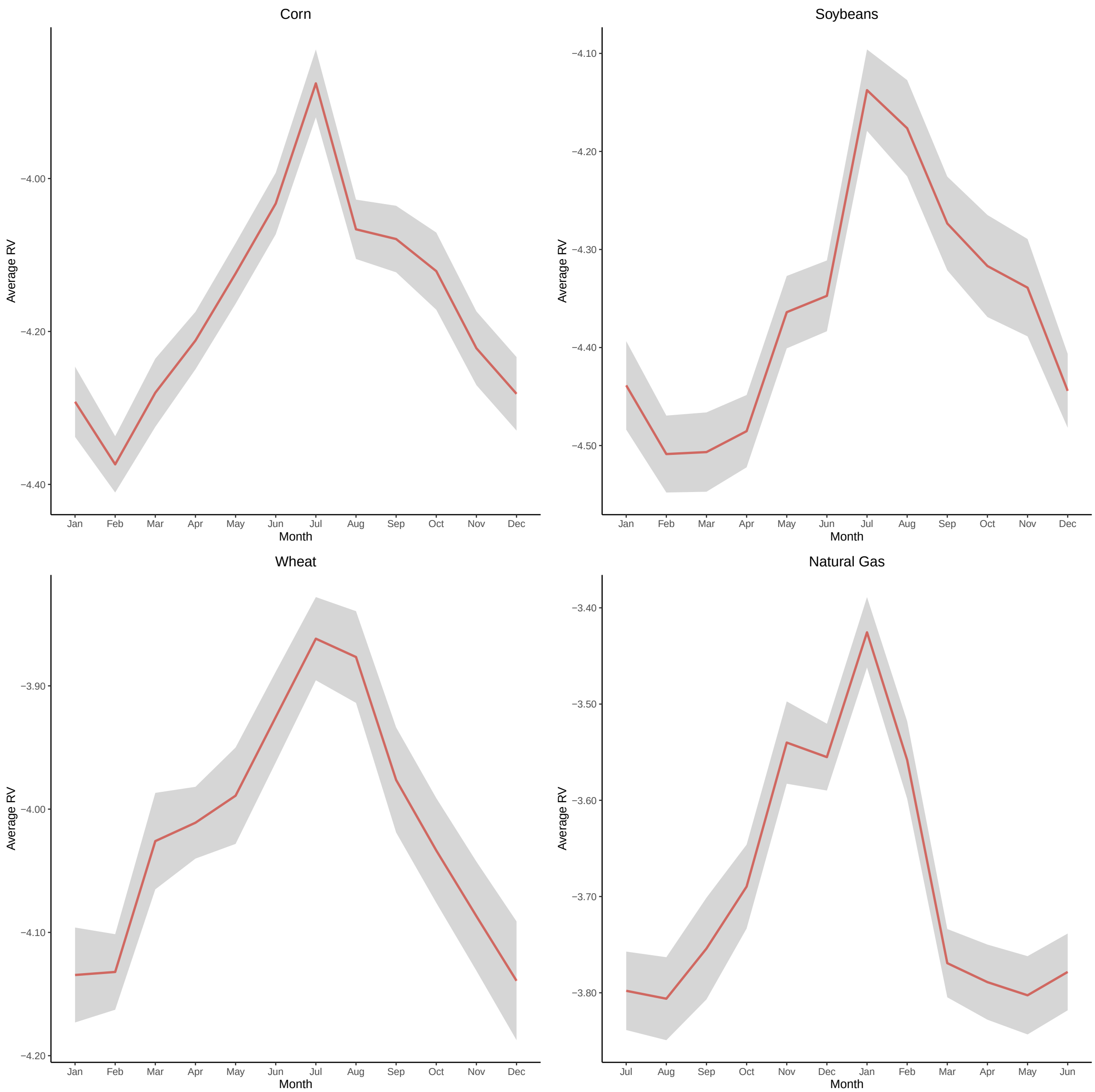

Table 3: **EV and EVHAR estimation of RV_t**

This table reports the coefficients of the EV (Eq. 3) and EVHAR (Eq. 4) jointly estimated (SUR). $RV_{c,t}$ is the 5-minute log realized volatility, $SL_{c,t-1}$ is the log of the nearest term structure slope. $B_{c,t-1}$ is a dummy set to "1" ("0") when the slope of the nearest term structure is negative (positive). $TM_{c,t}$ is the (log) time to maturity, and there is one monthly dummy $M_{c,t}$, set to "1" during the month of uncertainty resolution and to "0" otherwise (July for the three agriculture products and January for the natural gas contract). The three lagged variables are the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. I report Newey and West (1994) standard errors with automatic lag selection in parenthesis. I report the adjusted R^2 for the individual equations of the system, the overall OLS R^2, McElroy R^2, the log likelihood, and the likelihood ratio test that compares the unrestricted model EVHAR with the restricted HAR. The sample period is May 5, 2008–April 1, 2019. The number of observations per equation is 2733.

	Agriculture						Energy						Metal					
	C		S		W		CL		HO		NG		GC		HG		SI	
Constant	-4.34^{***}	-1.25^{***}	-4.58^{***}	-1.51^{***}	-4.22^{***}	-1.09^{***}	-4.17^{***}	-0.55^{***}	-4.34^{***}	-0.72^{***}	-3.96^{***}	-0.80^{***}	-4.83^{***}	-1.08^{***}	-4.57^{***}	-0.75^{***}	-4.34^{***}	-1.11^{***}
	(0.02)	(0.09)	(0.02)	(0.10)	(0.02)	(0.11)	(0.02)	(0.05)	(0.02)	(0.06)	(0.02)	(0.08)	(0.02)	(0.12)	(0.02)	(0.09)	(0.02)	(0.10)
$M_{c,t}$	0.31^{***}	0.05^{**}	0.18^{***}	0.06^{**}	0.21^{***}	0.05^{**}					0.29^{***}	0.11^{***}						
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)					(0.03)	(0.02)						
$TM_{c,t}(10^2)$	4.81^{***}	0.27	4.46^{***}	2.08^{**}	0.95	0.93	0.31	-0.44	3.52^{***}	0.44	5.98^{***}	1.23	11.01^{***}	5.85^{***}	7.79^{***}	6.42^{***}	18.57^{***}	11.89^{***}
	(1.08)	(1.03)	(1.09)	(1.05)	(1.17)	(1.07)	(1.03)	(0.69)	(1.00)	(0.70)	(1.48)	(1.22)	(1.23)	(1.16)	(1.27)	(1.09)	(1.36)	(1.31)
$SL_{c,t-1}(10^{-2})$	-1.63^{***}	-0.40	0.36	0.74	3.07^{***}	0.61^{**}	3.38^{***}	0.63^{***}	5.44^{***}	1.53^{***}	1.22^{***}	0.46^{***}	-55.21^{***}	-12.25^{***}	11.91^{***}	2.84	-43.72^{***}	-9.92^{***}
	(0.55)	(0.50)	(1.12)	(1.03)	(0.29)	(0.29)	(0.15)	(0.12)	(0.30)	(0.21)	(0.10)	(0.08)	(3.17)	(3.24)	(2.47)	(2.00)	(2.81)	(2.95)
$B_{c,t-1}$	0.07^{***}	0.00	0.05^{**}	0.02	0.16^{***}	0.07	-0.08^{***}	-0.02	-0.05^{***}	-0.04^{***}	0.03	-0.02	-0.06	-0.05	-0.08^{***}	-0.06^{***}	0.10^{**}	0.05
	(0.02)	(0.02)	(0.02)	(0.02)	(0.06)	(0.05)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	0.63	-0.16	-1.41	-1.32	3.92	1.47	-5.05^{***}	-1.80^{***}	-8.70^{***}	-3.23^{***}	-3.27^{***}	-1.21^{***}	-25.92	-47.00	-20.17^{***}	-11.96^{***}	52.10^{***}	23.50^{**}
	(0.58)	(0.52)	(1.13)	(1.03)	(4.91)	(4.43)	(0.61)	(0.39)	(0.38)	(0.26)	(0.18)	(0.15)	(37.01)	(33.19)	(3.29)	(2.70)	(11.77)	(10.68)
$RV_{c,t-1}$		0.32^{***}		0.26^{***}		0.21^{***}		0.42^{***}		0.37^{***}		0.21^{***}		0.20^{***}		0.20^{***}		0.16^{***}
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
$RV_{c,t-2\|t-5}$		0.18^{***}		0.21^{***}		0.26^{***}		0.25^{***}		0.24^{***}		0.37^{***}		0.28^{***}		0.38^{***}		0.32^{***}
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)
$RV_{c,t-6\|t-22}$		0.21^{***}		0.22^{***}		0.27^{***}		0.20^{***}		0.23^{***}		0.23^{***}		0.32^{***}		0.28^{***}		0.30^{***}
		(0.02)		(0.02)		(0.03)		(0.02)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)
Adj. R^2	0.11	0.41	0.06	0.33	0.09	0.38	0.26	0.75	0.20	0.71	0.19	0.55	0.16	0.42	0.02	0.48	0.17	0.42
OLS R^2	0.15	0.50																
McElroy R^2	0.11	0.47																
log likelihood	-3916.70	249.93																
LR test EVHAR > HAR	559.47^{***}																	

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 4: **EVHARQ and EVHEXP estimation of** RV_t

This table reports the results of the joint estimation (SUR) of the EVHEXP (Eq. 5) and EVHARQ models (Eq. 6) with inclusion of the EV computed with 5-minute log realized volatility $RV_{c,t}$. I compute the HEXP with four terms exponentially smoothed over a backward looking 500-day window for four lagged variables, *i.e.*, centers of mass (CoM) of 1, 5, 25, and 125 days. I compute the HARQ with three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. I implement the parsimonious version of the model, adding the product of the first autoregressive term $RV_{c,t-1}$ with the log realized quarticity measure $RQ_{c,t-1}$ (see, Barndorff-Nielsen and Shephard, 2002). I introduce $SL_{c,t-1}$, the log difference of the nearest term structure slope, $B_{c,t-1}$ a dummy set to "1" ("0") when the slope of the nearest term structure is negative (positive), $TM_{c,t}$ the log time to maturity, and there is one montly dummy $M_{c,t}$, set to "1" during the months of uncertainty resolution and to "0" otherwise (July for the three agriculture products and January for the natural gas contract). I report Newey and West (1994) corrected standard errors with automatic lag selection in parenthesis. I report the adjusted R^2 for each individual equation of the system, the overall OLS R^2, McElroy R^2, the log likelihood, and the likelihood ratio tests that compares the unrestricted EVHARQ and EVHEXP models with the HARQ and HEXP, respectively. The sample period is April 22, 2010–April 1, 2019. The number of observations per equation is 2733 for the EVHARQ and 2255 for the EVHEXP.

	Agriculture						Energy						Metal					
	C		S		W		CL		HO		NG		GC		HG		SI	
Constant	−1.21***	−1.50***	−1.40***	−2.33***	−1.07***	−1.23***	−0.50***	−0.66***	−0.66***	−0.66***	−0.66***	−1.14***	−1.08***	−0.87***	−0.70***	−0.60***	−1.08***	−0.82***
	(0.09)	(0.12)	(0.10)	(0.19)	(0.11)	(0.14)	(0.05)	(0.08)	(0.06)	(0.08)	(0.07)	(0.16)	(0.12)	(0.18)	(0.09)	(0.12)	(0.10)	(0.14)
$M_{c,t}$	0.05**	0.10***	0.05**	0.08***	0.05**	0.08***					0.09***	0.12***						
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)					(0.02)	(0.02)						
$TM_{c,t}(10^2)$	0.11	0.50	1.01	2.22**	0.99	0.88	−0.53	−0.35	1.19*	−0.17	2.48**	0.05	3.61***	6.32***	5.08***	7.21***	10.53***	13.36***
	(1.04)	(1.07)	(1.06)	(1.07)	(1.10)	(1.08)	(0.66)	(0.72)	(0.68)	(0.75)	(1.18)	(1.25)	(1.26)	(1.16)	(1.29)	(1.07)	(1.52)	(1.30)
$SL_{c,t-1}(10^{-2})$	−0.51	0.13	0.99	−0.78	0.56*	0.45	0.65***	0.75***	1.43***	1.67***	0.40***	0.48***	−14.65***	−8.97**	2.31	5.09**	−10.77***	−3.20
	(0.50)	(0.59)	(1.02)	(1.17)	(0.29)	(0.30)	(0.11)	(0.13)	(0.20)	(0.23)	(0.08)	(0.08)	(3.26)	(3.82)	(2.01)	(2.03)	(2.97)	(3.25)
$B_{c,t-1}$	0.00	0.02	0.02	0.02	0.07	0.09*	−0.02	−0.01	−0.04***	−0.03***	−0.03	−0.03	−0.05	−0.06*	−0.06***	−0.05**	0.05	0.05
	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−0.04	−0.81	−1.56	0.20	1.39	3.21	−2.10***	−2.02***	−3.19***	−3.51***	−1.04***	−1.27***	−47.00	−64.56*	−11.39***	−14.25***	23.98**	16.34
	(0.52)	(0.61)	(1.03)	(1.18)	(4.45)	(4.46)	(0.38)	(0.41)	(0.26)	(0.29)	(0.14)	(0.15)	(33.22)	(33.11)	(2.70)	(2.74)	(10.75)	(10.69)
$RV_{c,t-1}$	0.34***		0.31***		0.22***		0.54***		0.48***		0.41***		0.20***		0.22***		0.16***	
	(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)	
$RV_{c,t-1} \times RQ_{c,t-1}$	0.07		0.19***		−0.01		0.77***		0.67***		0.72***		0.11***		0.07**		0.04	
	(0.04)		(0.03)		(0.03)		(0.05)		(0.05)		(0.05)		(0.02)		(0.03)		(0.03)	
$RV_{c,t-2\|t-5}$	0.17***		0.19***		0.27***		0.17***		0.17***		0.24***		0.26***		0.37***		0.32***	
	(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)	
$RV_{c,t-6\|t-22}$	0.21***		0.20***		0.27***		0.16***		0.20***		0.18***		0.32***		0.28***		0.30***	
	(0.02)		(0.02)		(0.03)		(0.02)		(0.02)		(0.02)		(0.03)		(0.03)		(0.03)	
$RV_{c,t-1}^{CoM1}$		0.47***		0.37***		0.31***		0.64***		0.58***		0.22***		0.18***		0.23***		0.12**
		(0.04)		(0.04)		(0.05)		(0.04)		(0.04)		(0.05)		(0.05)		(0.05)		(0.06)
$RV_{c,t-1}^{CoM5}$		−0.17**		−0.10		0.05		−0.05		−0.06		0.47***		0.26**		0.39***		0.44***
		(0.07)		(0.07)		(0.09)		(0.06)		(0.06)		(0.08)		(0.10)		(0.09)		(0.11)
$RV_{c,t-1}^{CoM25}$		0.30***		0.44***		0.37***		0.25***		0.21***		0.10		0.41***		0.25***		0.16
		(0.06)		(0.07)		(0.09)		(0.05)		(0.06)		(0.07)		(0.10)		(0.08)		(0.10)
$RV_{c,t-1}^{CoM125}$		0.08		−0.22***		−0.01		−0.00		0.14***		−0.07		0.00		0.04		0.16**
		(0.06)		(0.07)		(0.07)		(0.04)		(0.04)		(0.07)		(0.07)		(0.05)		(0.07)
Adj. R^2	0.41	0.36	0.33	0.29	0.38	0.36	0.77	0.74	0.72	0.68	0.58	0.53	0.43	0.43	0.49	0.50	0.42	0.43
OLS R^2	0.51	0.49																
McElroy R^2	0.48	0.45																
log likelihood	492.69	−40.03																
LR test EVHARQ > HARQ	459.90***																	
LR test EVHEXP > HEXP	638.69***																	

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table 5: **EVHAR-TV and EVHARQ-TV estimation of** RV_t

This table reports the results of the joint estimation (SUR) of the EVHAR-TV (Eq. 7) and EVHARQ-TV model (Eq. 8) computed with 5-minute log realized volatility $RV_{c,t}$ and with three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. I implement the parsimonious version of the HARQ-TV model, adding the product of the first autoregressive term $RV_{c,t-1}$ with the log realized quarticity measure $RQ_{c,t-1}$ (see, Barndorff-Nielsen and Shephard, 2002). I introduce $SL_{c,t-1}$, the log difference of the nearest term structure slope, $B_{c,t-1}$ a dummy set to "1" ("0") when the slope of the nearest term structure is negative (positive), $TM_{c,t}$ the log time to maturity, and there is one montly dummy $M_{c,t}$, set to "1" during the months of uncertainty resolution and to "0" otherwise (July for the three agriculture products and January for the natural gas contract). I report the means of the time-varying coefficients and their standard errors in square brackets. I report the optimal bandwidth selected by "leave-one-out cross-validation", the pseudo R^2 for each individual equation, the log likelihood, and the likelihood ratio tests that compares the unrestricted EVHAR-TV and EVHARQ-TV models with the HAR-TV and HARQ-TV, respectively. The sample period is April 22, 2010–April 1, 2019. The number of observations per equation is 2733.

	Agriculture						Energy						Metal					
	C		S		W		CL		HO		NG		GC		HG		SI	
Constant	−1.27***	−1.23***	−1.53***	−1.42***	−1.10***	−1.09***	−0.58***	−0.52***	−0.75***	−0.68***	−0.80***	−0.66***	−1.09***	−1.09***	−0.76***	−0.70***	−1.13***	−1.09***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$M_{c,t}$	0.05***	0.05***	0.06***	0.06***	0.05***	0.05***					0.11***	0.09***						
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]					[0.00]	[0.00]						
$TM_{c,t}(10^2)$	0.33***	0.23***	2.17***	1.09***	0.97***	1.05***	−0.45***	−0.54***	0.34***	1.11***	1.20***	2.44***	5.84***	3.66***	6.42***	5.13***	11.97***	10.62***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]	[0.01]
$SL_{c,t-1}(10^{-2})$	−0.39***	−0.50***	0.84***	1.09***	0.59***	0.56***	0.62***	0.64***	1.53***	1.43***	0.46***	0.40***	−12.35***	−14.59***	2.88***	2.34***	−9.76***	−10.60***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.02]	[0.02]	[0.01]	[0.01]	[0.01]	[0.01]
$B_{c,t-1}$	0.00***	0.00***	0.02***	0.02***	0.07***	0.07***	−0.02***	−0.02***	−0.04***	−0.04***	−0.02***	−0.03***	−0.04***	−0.05***	−0.06***	−0.06***	0.05***	0.05***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−0.18***	−0.06***	−1.43***	−1.66***	1.50***	1.39***	−1.83***	−2.11***	−3.24***	−3.20***	−1.21***	−1.03***	−44.91***	−45.54***	−12.07***	−11.48***	23.59***	24.23***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.21]	[0.28]	[0.01]	[0.01]	[0.04]	[0.04]
$RV_{c,t-1}$	0.32***	0.34***	0.26***	0.31***	0.21***	0.21***	0.42***	0.54***	0.36***	0.47***	0.21***	0.41***	0.20***	0.20***	0.20***	0.21***	0.16***	0.16***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$RV_{c,t-1} \times RQ_{c,t-1}$		0.06***		0.19***		−0.01***		0.77***		0.67***		0.72***		0.11***		0.06***		0.04***
		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]
$RV_{c,t-2\|t-5}$	0.18***	0.17***	0.20***	0.18***	0.26***	0.27***	0.25***	0.17***	0.24***	0.17***	0.37***	0.24***	0.28***	0.26***	0.38***	0.37***	0.32***	0.32***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$RV_{c,t-6\|t-22}$	0.21***	0.21***	0.22***	0.20***	0.27***	0.27***	0.20***	0.16***	0.23***	0.20***	0.23***	0.18***	0.32***	0.32***	0.28***	0.28***	0.30***	0.30***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Bandwidth	0.56	0.40	0.75	0.36	0.65	0.58	6.50	0.67	20.00	20.00	1.16	0.95	0.90	0.61	0.59	0.88	0.62	0.45
Pseudo R^2	0.41	0.41	0.33	0.34	0.38	0.38	0.75	0.77	0.71	0.72	0.55	0.58	0.42	0.43	0.49	0.49	0.42	0.42
log likelihood	231.50	477.56																
LR test EVHAR-TV > HAR-TV	313.65***																	
LR test EVHARQ-TV > HARQ-TV	231.59***																	

$^{***}p < 0.01; ^{**}p < 0.05; ^{*}p < 0.1$

Table 6: **In-sample model comparison**

This table reports the results of the model confidence set procedure; see Hansen et al. (2011). The reported values for each model are the mean of the losses of the three functions: (i) squared errors $MSE = \left(\hat{\sigma} - \sqrt{h}\right)^2$, (ii), absolute errors $MAE = |\hat{\sigma} - \sqrt{h}|$, and (iii) $QLIKE = \ln h + \frac{\hat{\sigma}^2}{h}$. When applicable, the superscripts 1, 2, ..., n indicate the ranking of the models that are included in the confidence interval $\widehat{\mathcal{M}}_{90\%}$. I display the results for one day, one week, and one month ahead. The sample period is April 22, 2010–April 1, 2019. The number of observations is $2255 \times 9 = 20295$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV
One day ahead											
MSE	0.1316	0.0785	0.0773	0.0806	0.0792	0.0768	^{2}*0.0759*	0.0774	0.0774	^{1}**0.0756**	0.0760
MAE	0.2764	0.1984	0.1981	0.2018	0.2014	0.1962	0.1959	0.1971	0.1982	^{1}**0.1948**	0.1960
QLIKE	3.8969	3.8909	3.8908	3.8913	3.8911	3.8907	^{2}*3.8906*	3.8908	3.8908	^{1}**3.8906**	3.8906
One week ahead											
MSE	0.1331	0.0925	0.0902	0.0941	0.0919	0.0916	0.0895	^{1}**0.0746**	0.0906	0.0792	0.0900
MAE	0.2783	0.2194	0.2180	0.2217	0.2204	0.2187	0.2173	^{1}**0.1931**	0.2183	0.2018	0.2178
QLIKE	3.8971	3.8926	3.8923	3.8928	3.8925	3.8924	3.8922	^{1}**3.8905**	3.8924	3.8910	3.8923
One month ahead											
MSE	0.1361	0.1131	0.1078	0.1138	0.1086	0.1127	0.1075	^{2}*0.0823*	0.0950	^{1}**0.0823**	0.0942
MAE	0.2808	0.2486	0.2430	0.2497	0.2442	0.2485	0.2427	^{1}**0.2073**	0.2243	^{2}*0.2074*	0.2234
QLIKE	3.8975	3.8949	3.8943	3.8950	3.8944	3.8949	3.8943	^{1}**3.8915**	3.8929	^{2}*3.8915*	3.8928

Table 7: **Out-of-sample model comparison**

Panel A reports the results of the model confidence set procedure; see Hansen et al. (2011). The reported values for each model are the mean of the losses of the three functions: (i) squared errors $MSE = \left(\hat{\sigma} - \sqrt{h}\right)^2$, (ii), absolute errors $MAE = |\hat{\sigma} - \sqrt{h}|$, and (iii) $QLIKE = \ln h + \frac{\hat{\sigma}^2}{h}$. When applicable, the superscripts 1, 2, ..., n indicate the ranking of the models that are included in the confidence interval $\widehat{\mathcal{M}}_{90\%}$. I report the results for the one day, week, and month forecasts. Panel B reports the χ^2 statistics for the modified Diebold-Mariano test; see Diebold and Mariano (1995) and Harvey et al. (1997). The null hypothesis is that the model in the row-entry is equal to the one of the column-entry. I report the statistics for the one-day ahead forecasts MSE. The calibration period is April 22, 2010–January 30, 2013 and the out-of-sample period is January 31, 2013–April 1, 2019. The number of observation in the out-of-sample period is $1555 \times 9 = 13995$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
						Panel A: Model confidence set						
One day ahead												
MSE	0.2220	0.0810	0.0808	0.0927	0.0905	^{1}**0.0788**	^{2}*0.0791*	0.0863	0.0819	0.0815	^{3}0.0793	0.0930
MAE	0.3800	^{5}0.2038	0.2059	0.2228	0.2217	^{1}**0.2012**	^{3}0.2032	0.2113	0.2075	^{4}0.2043	^{2}*0.2027*	0.2230
QLIKE	3.9293	3.9124	3.9124	3.9138	3.9136	^{1}**3.9121**	^{2}*3.9121*	3.9130	3.9125	3.9124	^{3}3.9122	3.9139
One week ahead												
MSE	0.2209	^{4}0.0963	^{3}0.0953	0.1140	0.1097	^{2}*0.0950*	^{1}**0.0943**	0.9771	0.1000	0.1593	^{5}0.0973	0.1049
MAE	0.3798	^{2}*0.2284*	^{4}0.2301	0.2515	0.2481	^{1}**0.2277**	^{3}0.2291	0.5166	0.2351	0.3022	^{5}0.2316	0.2395
QLIKE	3.9292	3.9143	3.9141	3.9163	3.9158	^{2}*3.9141*	^{1}**3.9140**	3.9588	3.9147	3.9209	3.9143	3.9152
One month ahead												
MSE	0.2245	0.1302	0.1278	0.1639	0.1543	0.1300	0.1274	0.3099	0.1050	0.2908	^{1}**0.1038**	0.1340
MAE	0.3833	0.2782	0.2776	0.3121	0.3035	0.2783	0.2771	0.4125	0.2421	0.4095	^{1}**0.2406**	0.2789
QLIKE	3.9297	3.9183	3.9180	3.9222	3.9210	3.9182	3.9179	3.9451	3.9152	3.9452	^{1}**3.9151**	3.9186
						Panel B: Modified Diebold-Mariano test						
One day ahead												
EV												
HAR	59.35***											
EVHAR	60.11***	0.83										
HEXP	57.79***	−18.95***	−17.28***									
EVHEXP	58.98***	−14.50***	−17.16***	5.51***								
HARQ	58.08***	6.88***	5.14***	18.37***	15.74***							
EVHARQ	59.52***	5.12***	7.85***	17.79***	17.73***	−0.87						
HAR-TV	55.17***	−7.00***	−7.03***	8.51***	5.14***	−8.99***	−8.63***					
EVHAR-TV	60.01***	−2.26*	−3.90***	16.35***	16.76***	−6.98***	−8.05***	6.15***				
HARQ-TV	57.46***	−1.26	−1.53	16.57***	12.41***	−7.76***	−5.47***	7.54***	0.79			
EVHARQ-TV	59.43***	3.97***	4.20***	18.42***	19.04***	−1.31	−0.87	9.14***	11.95***	5.83***		
RiskMetrics	56.51***	−16.65***	−15.61***	−0.47	−3.45***	−17.05***	−16.34***	−9.65***	−15.65***	−16.25***	−17.36***	

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 8: **Forecast bias of multi-quantile regressions**

This table reports the p-values of the Wald tests achieved on the parameters from multi-quantile predictive panel regressions estimated with the daily log price changes and one-day ahead out-of-sample RV forecasts, and for four number of quantiles $p = 1, 2, 4, 6$, as in Couperier and Leymarie (2020). The quantile levels start at $\tau = 0.975$, which is the Basel Committee on Banking Supervision (Basel III) regulatory level, and each level u_i within is determined as: $\tau + (u_i - 1) \times {}^{(1-\tau)}/_{p}$, for each $u_i = 1, 2, \ldots, p$. The tests are for the following null hypotheses: (i) $H_{0,J_1} : \sum_{j=1}^{p} (\beta_0(\tau_j)) + (\beta_1(\tau_j)) = p$, (ii) $H_{0,J_2} : \sum_{j=1}^{p} \beta_0(\tau_j) = 0$ and $\sum_{j=1}^{p} \beta_1(\tau_j) = p$, (iii) $H_{0,I} : \sum_{j=1}^{p} \beta_0(\tau_j) = 0$, and (iv) $H_{0,S} : \sum_{j=1}^{p} \beta_1(\tau_j) = p$. The p-values are obtained with a bootstrap of 1,000 replications. The sample period is January 31, 2013–April 1, 2019. The number of observations is $1555 \times 9 = 13995$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
$p = 1$												
J_1	0.82	0.61	0.82	0.68	0.39	0.29	0.67	0.18	0.27	0.55	0.68	0.23
J_2	0.01	0.42	0.31	0.15	0.05	0.22	0.26	0.00	0.04	0.02	0.04	0.02
I	0.23	0.29	0.59	0.74	0.78	0.07	0.24	0.87	0.87	0.46	0.50	0.93
S	0.80	0.60	0.85	0.69	0.40	0.28	0.65	0.19	0.27	0.57	0.70	0.24
$p = 2$												
J_1	0.21	0.74	0.35	0.61	0.83	0.32	0.20	0.09	0.74	0.41	0.93	0.18
J_2	0.04	0.35	0.19	0.08	0.04	0.18	0.13	0.00	0.03	0.01	0.04	0.01
I	0.05	0.31	0.06	0.82	0.38	0.08	0.03	0.96	0.51	0.67	0.20	0.98
S	0.20	0.72	0.33	0.62	0.86	0.31	0.19	0.10	0.76	0.43	0.90	0.19
$p = 4$												
J_1	0.17	0.63	0.23	0.74	0.98	0.31	0.18	0.17	0.75	0.47	0.87	0.21
J_2	0.04	0.33	0.16	0.12	0.04	0.17	0.16	0.00	0.05	0.01	0.04	0.01
I	0.05	0.26	0.06	0.70	0.28	0.08	0.06	0.92	0.45	0.70	0.20	0.98
S	0.16	0.62	0.22	0.76	0.95	0.30	0.17	0.18	0.77	0.49	0.85	0.22
$p = 6$												
J_1	0.22	0.69	0.24	0.68	0.65	0.55	0.28	0.24	0.95	0.37	0.87	0.20
J_2	0.04	0.25	0.19	0.06	0.03	0.18	0.19	0.00	0.05	0.01	0.03	0.01
I	0.05	0.25	0.09	0.73	0.55	0.20	0.11	0.90	0.32	0.84	0.21	0.84
S	0.21	0.67	0.23	0.69	0.67	0.54	0.27	0.26	0.98	0.39	0.84	0.21

Table 9: **Percentage of 97.5% VaR violations**

This table reports the violation percentage of the 97.5% VaR which corresponds to a single quantile regression of coverage $\tau = 0.975$ from the one-day ahead forecasts and the average violation percentage for the nine commodities. The sample period is January, 2013–April 1, 2019. The number of observations is $1555 \times 9 = 13995$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
C	1.35	1.04	1.04	1.04	0.96	1.12	1.04	1.04	1.04	1.20	1.12	1.20
S	1.20	1.35	1.20	1.20	1.27	1.12	1.12	1.35	1.20	1.12	1.27	1.04
W	1.04	1.20	1.12	1.20	1.12	1.12	1.12	1.20	1.20	1.20	1.12	1.27
CL	1.35	1.20	1.20	1.12	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.35
HO	0.96	1.59	1.27	1.27	1.20	1.51	1.20	1.51	1.27	1.27	1.20	1.51
NG	1.12	1.35	1.51	1.20	1.20	1.27	1.12	1.27	1.20	1.35	1.20	1.27
GC	1.27	1.12	1.12	1.12	1.27	1.12	1.04	1.12	1.27	1.04	1.27	1.43
HG	1.20	1.35	1.35	1.35	1.20	1.43	1.43	1.35	1.35	1.43	1.43	1.43
SI	1.27	0.96	1.20	1.27	1.43	1.12	1.12	1.35	1.12	0.96	1.12	1.20
Average	1.20	1.24	1.22	1.20	1.20	1.22	1.15	1.27	1.20	1.20	1.21	1.30

Appendix

Table A1: **Description of futures contracts**

This table reports the specifications of the futures contracts written on the nine selected commodities. The specifications include their trading venue, ticker, underlying commodity and unit. I also report their maturity months with the appropriate letter code.

Ticker	Trading venue	Underlying	Unit	Maturity
		Agriculture		
C	CBT	Corn	bu (5,000)	HKNUZ
S	CBT	Soybeans	bu (5,000)	FHKNQUX
W	CBT	Chicago wheat	bu (5,000)	HKNUZ
		Energy		
CL	NYMEX/ICE	WTI crude oil	bbl (1,000)	FGHJKMNQUVXZ
HO	NYMEX	Heating oil	gal (42,000)	FGHJKMNQUVXZ
NG	NYMEX/ICE	Natural gas	MMBtu (10,000)	FGHJKMNQUVXZ
		Metal		
GC	CMX	Gold	oz (100)	GJMQVZ
HG	COMEX	Copper	lb (25,000)	FGHJKMNQUVXZ
SI	CMX	Silver	oz (5,000)	FHKNUZ

Letter code: F = January, G = February, H = Mars, J = April, K = May, M = June, N = July, Q = August, U = September, V = October, X = November, Z = December.

Table A2: **Variable definition**

This table defines all variables used throughout the article, with their symbol, plain definition, computation, unit, frequency, and additional descriptions when applicable. For all variables, the sample period is: May 5, 2008–April 1, 2019.

Symbol	Title	Computation	Unit	Frequency	Description
$F_{c,t}^m$	Futures price	–	USD	5 min	m: maturity, c: commodity, t: time
$f_{c,t}^m$	Log futures price	$f_{c,t}^m = \ln F_{c,t}^m$	–	5 min	m: maturity, c: commodity, t: time
$r_{c,t}$	Log price change	$r_{c,t} = f_{c,t} - f_{c,t-1}$	%	5 min	m: maturity, c: commodity, t: time
$SL_{c,t}$	(Annualized log) term structure	$SL_{c,t} = \frac{(f_{c,t}^N)-(f_{c,t}^{FD})}{\delta} \times 250$	–	D	Where N is the nearby contract, FD is the first deferred contract, and δ is the time difference in days between two consecutive maturities. I annualize the variable for ease of interpretation.
$B_{c,t}$	Backwardation dummy	$B_{c,t} = \begin{cases} 1 & \text{if } SL_{c,t} < 0 \\ 0 & \text{if } SL_{c,t} >= 0 \end{cases}$	–	D	–
$M_{c,t}$	Month dummy	$M_{c,t} = \begin{cases} 1 & \text{if } t \in \text{critical month} \\ 0 & \text{if } t \notin \text{critical month} \end{cases}$	–	D	"Critical month": July (corn, soybeans, and wheat) and January (natural gas).
$TM_{c,t}$	(Log) time to maturity	$TM_{c,t} = \ln T - t$	s	D	T and t are the time at maturity and current time in POSIX time format.
$ARV_{c,t}$	Arithmetic realized volatility	$ARV_{c,t} = \sqrt{\sum_{j=1}^{1/\Delta} r_{t,\Delta\times j}^2}$	–	D	In amount of 5-minute periods elapsed each day, *i.e.*, $j = 1, 2, \ldots, 288$.
$RV_{c,t}$	Arithmetic realized volatility	$RV_{c,t} = \frac{1}{2} \times \ln \sum_{j=1}^{1/\Delta} r_{t,\Delta\times j}^2$	–	D	In amount of 5-minute periods elapsed each day, *i.e.*, $j = 1, 2, \ldots, 288$.
$J_{c,t}$	Significant jump (1%)	see, Andersen et al. (2007, p. 710)	–	D	–
$C_{c,t}$	Continuous diffusion component (1%)	see, Andersen et al. (2007, p. 710)	–	D	–
$RQ_{c,t}$	(Log realized quarticity)	$RQ_t = \frac{2}{3} \frac{\sum_{i=1}^{1/\Delta} r_{i,t}^4}{\left(\sum_{i=1}^{1/\Delta} r_{i,t}^2\right)^2}$	–	D	In amount of 5-minute periods elapsed each day, *i.e.*, $j = 1, 2, \ldots, 288$.
$RV_{c,t-k\|t-n}$	Average of the RV from $t-k$ to $t-n$	$RV_{c,t-k\|t-n} = \frac{1}{n-k+1} \sum_{i=k}^{n} RV_{c,t-i}$	–	D	–
$RV_{c,t}^{CoM(\lambda)}$	Exponential average of the RV	$RV_{c,t}^{CoM(\lambda)} = \sum_{i=1}^{500} \frac{e^{-i\lambda}}{e^{-\lambda}+e^{-2\lambda}+\ldots+e^{-500\lambda}}$	–	D	With $\lambda = \ln\left(1 + \frac{1}{\text{CoM}}\right)$

Table A3: **Summary statistics: Daily RV with alternative sampling frequency**

This table reports statistics on daily log realized volatility sampled at 1, 5, 15, and 60 minute-intervals in panels A to D, respectively. In panel E, I add the statistics for the average of these four measures, as in Andersen et al. (2011b). For each panel, I report the four moments of the distribution, the Jarque-Bera statistic (JB χ^2), the Ljung-Box statistic for the 20^{th} order serial correlation ($Q(20)$), and the parameter of the log-periodogram regression based on a bandwith exponent of $4/5$ (d), as in Andersen et al. (2003). The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755.

	Agriculture			Energy			Metal		
	Corn (C)	Soybeans (S)	Wheat (W)	WTI crude oil (CL)	Heating oil (HO)	Natural Gas (NG)	Gold (GC)	Copper (HG)	Silver (SI)
					Panel A: RV_t 1-minute sampling				
Mean%	−402.74	−430.24	−393.64	−396.43	−409.49	−363.35	−466.62	−424.40	−406.93
σ%	42.72	46.16	40.80	46.28	44.28	38.46	50.95	46.48	51.43
Skewness	3.24	3.15	3.08	0.29	0.51	0.30	1.83	0.89	0.67
Kurtosis	25.50	24.11	27.48	1.96	3.95	3.41	15.20	3.52	9.34
$JB\chi^2$	79,586	71,432	91,174	481	1,916	1,379	28,121	1,784	10,245
$Q(20)$	7,995	9,645	7,340	32,279	29,356	17,552	11,641	23,501	10,770
d	0.35	0.38	0.30	0.53	0.45	0.48	0.34	0.50	0.40
					Panel B: RV_t 5-minute sampling				
Mean%	−415.13	−434.90	−400.54	−400.03	−412.02	−369.11	−468.93	−427.66	−412.04
σ%	46.48	46.19	41.54	46.69	43.93	39.85	50.87	46.46	52.02
Skewness	2.06	2.37	2.16	0.51	0.42	0.55	1.61	0.65	0.58
Kurtosis	13.73	17.02	18.39	0.61	0.36	1.01	13.22	1.14	7.55
$JB\chi^2$	23,626	35,914	41,036	161	95	259	21,286	343	6,706
$Q(20)$	11,283	10,703	8,201	32,443	31,598	18,330	12,106	23,773	11,329
d	0.34	0.40	0.32	0.52	0.53	0.51	0.33	0.51	0.38
					Panel C: RV_t 15-minute sampling				
Mean%	−421.47	−437.96	−404.39	−402.19	−414.20	−373.01	−470.73	−429.22	−414.93
σ%	50.33	47.83	43.27	47.97	45.23	41.54	52.02	47.27	52.95
Skewness	1.64	2.17	1.93	0.47	0.37	0.47	1.55	0.55	0.67
Kurtosis	10.42	15.17	16.09	0.53	0.37	1.02	12.22	0.91	6.59
$JB\chi^2$	13,738	28,629	31,470	132	77	223	18,283	236	5,197
$Q(20)$	10,848	9,618	7,178	29,330	28,077	15,463	11,447	21,689	10,863
d	0.35	0.39	0.31	0.51	0.48	0.44	0.31	0.49	0.37
					Panel D: RV_t 60-minute sampling				
Mean%	−427.89	−441.96	−409.07	−406.34	−418.90	−378.32	−475.29	−432.69	−419.19
σ%	56.01	52.20	48.10	51.65	49.39	47.09	55.49	50.68	56.73
Skewness	1.23	1.67	1.45	0.34	0.23	0.32	1.34	0.39	0.81
Kurtosis	7.13	11.01	10.89	0.39	0.26	0.58	9.67	0.74	5.56
$JB\chi^2$	6,549	15,229	14,596	72	33	86	11,567	132	3,854
$Q(20)$	8,606	7,318	5,406	21,352	20,053	10,619	8,700	16,626	8,127
d	0.31	0.35	0.31	0.42	0.40	0.42	0.30	0.41	0.32
					Panel E: RV_t average of 1-, 5-, 15-, and 60-minute sampling				
Mean%	−414.66	−435.20	−400.61	−400.22	−412.68	−369.68	−469.49	−427.72	−412.20
σ%	45.93	46.20	41.28	46.88	44.33	40.14	51.17	46.66	51.57
Skewness	2.14	2.38	2.20	0.50	0.40	0.55	1.62	0.60	0.77
Kurtosis	14.40	17.07	18.86	0.56	0.34	1.05	12.90	1.04	6.87
$JB\chi^2$	25,942	36,111	43,102	154	85	268	20,329	289	5,695
$Q(20)$	10,643	10,475	8,160	31,440	30,408	17,361	11,800	22,889	11,387
d	0.35	0.39	0.32	0.52	0.52	0.50	0.33	0.50	0.39

Table A4: **Average daily turnover and open interest over five years prior to the sample**

This table reports the average daily turnover and open interest in millions USD for 20 contracts components of the SP-GSCI / BCOM, prior to the sample period selection. The pre-sample period is January 1, 2003–April 30, 2008.

Energy	WTI crude oil (CL)	Heating oil (HO)	Brent crude oil (LCO)	Gasoil (LGO)	Natural gas (NG)	RBOB gasoline (RB)
Turnover	15,126.82	3,476.32	7,860.83	2,804.78	5,093.58	3,969.98
Open interest	28,209.41	8,962.40	15,468.90	7,704.32	16,167.50	3,528.27

Agriculture	Corn (C)	Feeder cattle (FC)	Kansas wheat (KW)	Live cattle (LC)	Lean hogs (LH)	Soybeans (S)	Wheat (W)
Turnover	1,891.39	203.31	329.13	955.63	478.49	3,056.77	1,111.43
Open interest	10,752.81	1,192.12	2,208.12	6,597.15	2,756.31	10,155.23	5,678.65

Metal	Gold (GC)	Copper (HG)	Platinum (PL)	Silver (SI)
Turnover	4,252.86	1227.43	227.20	1,300.46
Open interest	13,488.65	3792.76	625.33	4,339.29

Soft	Cocoa (CC)	Cotton (CT)	Coffee (KC)	Orange juice (OJ)	Raw sugar (SB)
Turnover	177.56	467.97	602.04	58.25	626.54
Open interest	1,644.18	3,395.23	4,724.07	433.67	4,592.96

Table A5: **HAR, HEXP, and HARQ estimation of RV_t**

This table reports the results of the joint estimation (SUR) of the restricted HAR (Eq. 4), HEXP (Eq. 5), and HARQ (Eq. 6) computed with 5-minute log realized volatility $RV_{c,t}$. The HAR and HARQ have three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. I implement the parsimonious version of the HARQ, adding the product of the first autoregressive term $RV_{c,t-1}$ with the log realized quarticity measure $RQ_{c,t-1}$ (Barndorff-Nielsen and Shephard, 2002). The HEXP has four terms exponentially smoothed over a backward looking 500-day window for four lagged variables, *i.e.* centers of mass (CoM) of 1, 5, 25, and 125 days. I report Newey and West (1994) corrected standard errors with automatic lag selection in parenthesis. I report the adjusted R^2 for each individual equation of the system, the overall OLS R^2, McElroy R^2, and the log likelihood. The sample period is June 3, 2008–April 1, 2019. The number of observations per equation is 2733 for the HAR and HARQ and 2255 for the HEXP.

	Agriculture								
	C			S			W		
Constant	−1.09***	−1.05***	−1.36***	−1.20***	−1.11***	−1.83***	−0.86***	−0.85***	−1.04***
	(0.08)	(0.08)	(0.10)	(0.09)	(0.09)	(0.17)	(0.09)	(0.09)	(0.13)
$RV_{c,t-1}$	0.33***	0.35***		0.27***	0.32***		0.22***	0.22***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-1} \times RQ_{c,t-1}$		0.07*			0.20***			−0.00	
		(0.04)			(0.03)			(0.03)	
$RV_{c,t-2\|t-5}$	0.20***	0.19***		0.23***	0.20***		0.28***	0.28***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-6\|t-22}$	0.22***	0.22***		0.24***	0.23***		0.30***	0.30***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.03)	(0.03)	
$RV_{c,t-1}^{CoM1}$			0.47***			0.33***			0.31***
			(0.04)			(0.04)			(0.05)
$RV_{c,t-1}^{CoM5}$			−0.12**			0.06			0.08
			(0.06)			(0.07)			(0.09)
$RV_{c,t-1}^{CoM25}$			0.32***			0.35***			0.42***
			(0.06)			(0.07)			(0.09)
$RV_{c,t-1}^{CoM125}$			0.02			−0.14**			−0.05
			(0.05)			(0.06)			(0.07)
Adj. R^2	0.41	0.41	0.36	0.32	0.33	0.29	0.38	0.38	0.36
OLS R^2	0.49	0.50	0.48						
McElroy R^2	0.45	0.47	0.43						
log likelihood	−29.81	244.74	−359.38						

	Energy								
	CL			HO			NG		
Constant	−0.36***	−0.31***	−0.44***	−0.43***	−0.37***	−0.36***	−0.46***	−0.36***	−0.95***
	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)	(0.07)	(0.07)	(0.06)	(0.16)
$RV_{c,t-1}$	0.44***	0.55***		0.40***	0.51***		0.25***	0.46***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-1} \times RQ_{c,t-1}$		0.76***			0.68***			0.77***	
		(0.05)			(0.06)			(0.05)	
$RV_{c,t-2\|t-5}$	0.27***	0.19***		0.27***	0.20***		0.41***	0.27***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.03)	(0.03)	
$RV_{c,t-6\|t-22}$	0.20***	0.17***		0.23***	0.20***		0.22***	0.16***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.03)	(0.02)	
$RV_{c,t-1}^{CoM1}$			0.67***			0.61***			0.28***
			(0.04)			(0.04)			(0.05)
$RV_{c,t-1}^{CoM5}$			−0.03			0.00			0.49***
			(0.06)			(0.06)			(0.08)
$RV_{c,t-1}^{CoM25}$			0.26***			0.15***			0.11
			(0.05)			(0.05)			(0.07)
$RV_{c,t-1}^{CoM125}$			0.01			0.16***			−0.13*
			(0.03)			(0.04)			(0.07)
Adj. R^2	0.75	0.76	0.74	0.70	0.71	0.67	0.53	0.57	0.51

	Metal								
	GC			HG			SI		
Constant	−0.71***	−0.74***	−0.42***	−0.58***	−0.56***	−0.42***	−0.70***	−0.69***	−0.46***
	(0.10)	(0.10)	(0.13)	(0.08)	(0.08)	(0.11)	(0.09)	(0.09)	(0.11)
$RV_{c,t-1}$	0.22***	0.22***		0.23***	0.23***		0.20***	0.17***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-1} \times RQ_{c,t-1}$		0.12***			0.13***			0.11***	
		(0.02)			(0.03)			(0.02)	
$RV_{c,t-2\|t-5}$	0.29***	0.28***		0.40***	0.39***		0.34***	0.33***	
	(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)	
$RV_{c,t-6\|t-22}$	0.34***	0.36***		0.25***	0.26***		0.32***	0.34***	
	(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)	
$RV_{c,t-1}^{CoM1}$			0.21***			0.27***			0.21***
			(0.05)			(0.06)			(0.06)
$RV_{c,t-1}^{CoM5}$			0.29***			0.39***			0.33***
			(0.10)			(0.09)			(0.11)
$RV_{c,t-1}^{CoM25}$			0.34***			0.20**			0.20*
			(0.10)			(0.08)			(0.10)
$RV_{c,t-1}^{CoM125}$			0.09			0.06			0.18***
			(0.06)			(0.05)			(0.06)
Adj. R^2	0.42	0.43	0.42	0.47	0.48	0.48	0.41	0.42	0.42

$^{***}p < 0.01; ^{**}p < 0.05; ^{*}p < 0.1$

Table A6: **HAR-TV and HARQ-TV estimation of RV_t**

This table reports the results of the joint estimation (SUR) of the restricted HAR-TV (Eq. 7) and HARQ-TV (Eq. 8) models, computed with 5-minute log realized volatility $RV_{c,t}$ and with three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. I implement the parsimonious version of the model, adding the product of the first autoregressive term $RV_{c,t-1}$ with the log realized quarticity measure $RQ_{c,t-1}$ (Barndorff-Nielsen and Shephard, 2002). I report the means of the time-varying coefficients and their standard errors in square brackets. I report the optimal bandwidth selected by "leave-one-out cross-validation", the pseudo R^2 for each individual equation, and the log likelihood of the system. The sample period is June 3, 2008–April 1, 2019. The number of observations per equation is 2733.

	Agriculture						Energy						Metal					
	C		S		W		CL		HO		NG		GC		HG		SI	
Constant	-1.29^{***}	-1.25^{***}	-1.27^{***}	-1.16^{***}	-0.78^{***}	-0.79^{***}	-0.39^{***}	-0.34^{***}	-0.45^{***}	-0.37^{***}	-0.48^{***}	-0.38^{***}	-0.82^{***}	-0.84^{***}	-0.71^{***}	-0.59^{***}	-1.14^{***}	-1.11^{***}
	[0.01]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]
$RV_{c,t-1}$	0.31^{***}	0.32^{***}	0.26^{***}	0.31^{***}	0.23^{***}	0.23^{***}	0.45^{***}	0.56^{***}	0.40^{***}	0.51^{***}	0.24^{***}	0.45^{***}	0.23^{***}	0.22^{***}	0.23^{***}	0.24^{***}	0.18^{***}	0.16^{***}
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$RV_{c,t-1} \times RQ_{c,t-1}$		0.03^{***}		0.20^{***}		-0.02^{***}		0.75^{***}		0.65^{***}		0.77^{***}		0.14^{***}		0.15^{***}		0.12^{***}
		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]
$RV_{c,t-2\|t-5}$	0.18^{***}	0.18^{***}	0.22^{***}	0.20^{***}	0.29^{***}	0.29^{***}	0.27^{***}	0.19^{***}	0.27^{***}	0.21^{***}	0.41^{***}	0.27^{***}	0.29^{***}	0.28^{***}	0.39^{***}	0.38^{***}	0.29^{***}	0.28^{***}
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$RV_{c,t-6\|t-22}$	0.21^{***}	0.21^{***}	0.24^{***}	0.22^{***}	0.30^{***}	0.30^{***}	0.20^{***}	0.16^{***}	0.22^{***}	0.19^{***}	0.23^{***}	0.17^{***}	0.32^{***}	0.33^{***}	0.23^{***}	0.24^{***}	0.27^{***}	0.30^{***}
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Bandwidth	0.23	0.25	0.29	0.30	0.63	0.63	4.03	0.61	20.00	20.00	0.74	0.72	0.53	0.57	0.40	0.81	0.25	0.25
Pseudo R^2	0.42	0.42	0.34	0.34	0.38	0.39	0.75	0.77	0.70	0.71	0.53	0.57	0.42	0.43	0.48	0.49	0.43	0.44
log likelihood	74.80	244.74																

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A7: **Forecast bias of individual quantile regressions**

This table reports the coefficients β_0 and β_1 of individual predictive quantile regressions, for the nine commodities and for six levels $u_i = 1, 2, \ldots, 4$, when $p = 4$. I also report the χ^2 of the joint test that $\beta_0 = 0\%$ and $\beta_1 = 100\%$, as in Couperier and Leymarie (2020). The data are the negative daily log price changes of the nine commodities and the one-day-ahead daily RV forecasts from all models tested. The sample period is January, 2013–April 1, 2019. The number of observations is $1555 \times 9 = 13995$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
u_1												
β_0	−0.34	0.15	0.08	0.05	0.04	0.23	0.17	0.03	−0.02	0.12	0.11	0.01
	(0.23)	(0.14)	(0.15)	(0.14)	(0.16)	(0.14)	(0.14)	(0.16)	(0.15)	(0.16)	(0.16)	(0.16)
β_1	103.14	97.41	100.65	102.09	103.58	95.25	97.82	107.88	105.76	103.16	101.69	105.93
	(7.47)	(5.31)	(5.54)	(5.45)	(5.67)	(5.31)	(5.18)	(6.43)	(5.21)	(6.22)	(5.88)	(6.01)
χ^2	203.30***	20.85***	34.88***	55.32***	85.17***	41.65***	40.38***	273.62***	106.60***	171.90***	100.48***	153.49***
u_2												
β_0	−0.43	0.20	0.19	0.11	0.12	0.27	0.27	0.05	0.13	0.12	0.21	0.10
	(0.26)	(0.16)	(0.14)	(0.16)	(0.16)	(0.15)	(0.13)	(0.16)	(0.16)	(0.16)	(0.15)	(0.18)
β_1	105.54	96.26	96.35	100.78	101.50	94.59	93.90	106.94	101.29	103.60	99.22	103.16
	(8.63)	(5.79)	(4.84)	(5.99)	(5.74)	(5.58)	(4.67)	(6.74)	(5.39)	(6.42)	(5.12)	(6.38)
χ^2	205.10***	33.51***	29.46***	65.73***	108.98***	53.73***	45.76***	283.63***	114.44***	207.34***	132.52***	162.51***
u_3												
β_0	−0.50	0.26	0.27	0.12	0.10	0.31	0.28	−0.02	0.15	0.08	0.24	0.07
	(0.33)	(0.19)	(0.17)	(0.19)	(0.16)	(0.18)	(0.17)	(0.18)	(0.17)	(0.16)	(0.14)	(0.19)
β_1	106.93	94.83	92.91	101.00	101.83	93.54	93.21	110.34	100.59	105.74	97.21	103.99
	(9.75)	(6.56)	(5.13)	(6.94)	(5.43)	(6.49)	(5.35)	(7.21)	(5.97)	(6.17)	(5.27)	(7.53)
χ^2	202.61***	48.07***	40.65***	94.36***	106.41***	61.90***	44.49***	425.66***	105.65***	297.59***	88.79***	164.69***
u_4												
β_0	−0.88	0.22	0.48	0.04	0.21	0.34	0.45	−0.04	0.22	0.02	0.26	−0.01
	(0.35)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.18)	(0.22)	(0.19)	(0.19)	(0.18)	(0.20)
β_1	118.61	97.70	88.44	105.12	97.91	93.85	89.37	110.85	98.52	106.92	96.95	109.73
	(9.92)	(7.09)	(6.69)	(6.82)	(6.22)	(7.01)	(6.25)	(7.52)	(6.22)	(6.66)	(6.24)	(7.73)
χ^2	184.76***	72.15***	128.22***	206.99***	72.35***	86.68***	115.86***	503.51***	103.06***	305.72***	91.30***	497.78***

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Figure A8: **Unconditional monthly average RV**

These plots display the monthly average of daily RV for the nearby futures contract written on crude oil, heating oil, gold, copper, and silver. The red line represents the average and the gray area represents the 90% confidence bands. The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755.

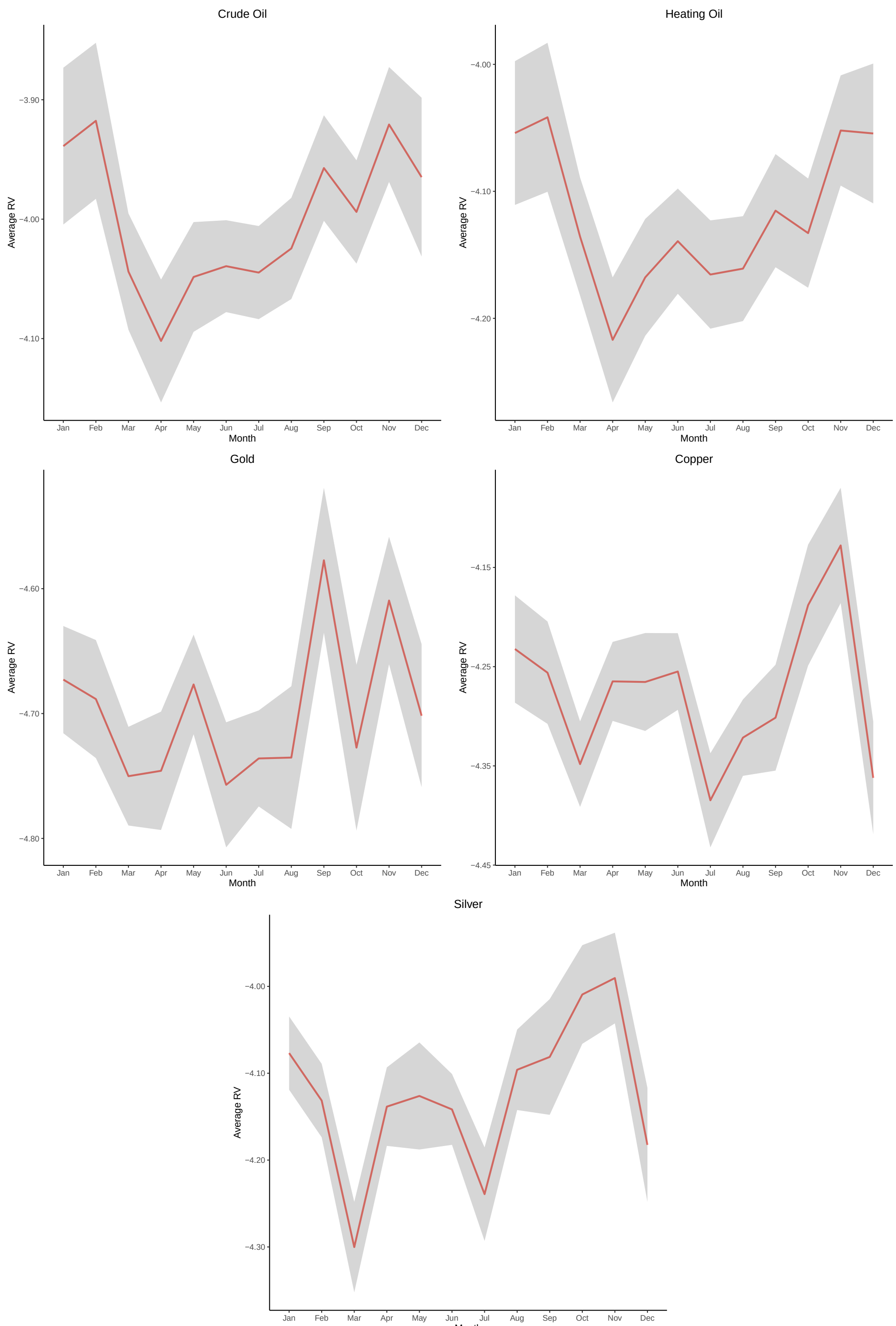

Figure A9: **Time-varying intercepts in the EVHAR-TV model**

This plot displays the pattern of time-varying intercepts in the joint estimation of the EVHAR-TV model (SURE). The intercept unit is RV for the nine nearby futures contracts. The sample period is May 5, 2010–April 1, 2019. The number of observations per contract is 2755.