Received: 18 April 2021 | Revised: 3 July 2021 | Accepted: 3 July 2021

DOI: 10.1002/fut.22250

RESEARCH ARTICLE

THE JOURNAL OF FUTURES MARKETS WILEY

Do economic variables forecast commodity futures volatility?

Loïc Maréchal

Institute of Financial Analysis, Faculty of Economics and Business, University of Neuchâtel, Neuchâtel, Switzerland

Correspondence
Loïc Maréchal, Institute of Financial Analysis, Faculty of Economics and Business, University of Neuchâtel, A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland.
Email: loic.marechal@unine.ch

Abstract

This paper explores empirically whether the supply or the demand uncertainty, the time to maturity, and the slope of the term structure (storage), explain the realized volatility of nearby commodity futures 5-min returns. I find support for the "uncertainty resolution" and the "theory of storage" hypotheses while the "time to maturity" hypothesis is rejected. These results are robust to the inclusion of autoregressive terms in the baseline model. Next, I evaluate the in- and out-of-sample forecasting ability of models including these economic variables and find mixed results. Finally, I test the validity of these forecasts in expected shortfall modeling.

KEYWORDS

commodity futures, realized volatility, risk management, Samuelson effect, supply and demand uncertainty, term structure

JEL CLASSIFICATION

C53, C58

1 | INTRODUCTION

The realized volatility (RV) framework introduced by Andersen and Bollerslev (1998a, 1998b) delivers higher forecast accuracy than generalized autoregressive conditional heteroskedasticity (GARCH) or stochastic volatility models for spot and futures prices of stocks, bonds, and currencies. Yet, the literature on the RV of commodity futures is scarce. While commodity futures share many commonalities with financial assets, the physical nature of their underlyings introduces specificities such as inventory, consumption, or decay. Hence, I ask whether the introduction of economic variables (EVs), which characterize individual commodity futures contracts, helps to model their RV.

I start the analysis with a regression model that tests whether seasonality (uncertainty resolution), time to maturity (Samuelson conjecture), and the slope of the term structure (a proxy for inventories), explain commodity futures RV. The empirical analysis is based on three groups of commodities (agriculture, energy, and metals). I select the three most liquid commodity contracts in each group, over the 2008–2019 period, and estimate the system of equations with a

Seemingly Unrelated Regression (SUR), to account for omitted variables common to commodity contracts.[1] I capture seasonal variations of RV with dummies reflecting critical months. In line with economic predictions, this variable is statistically significant at the 1% level for agricultural products (seasonal crops) and natural gas (higher heating demand in winter). I also study the time to maturity, a controversial variable related to the functioning of futures markets. The time to maturity positively determines the commodity futures RV, with statistical significance at the 1% level (for eight out of the nine commodities). Lastly, consistent with the theory of storage, I show that the slope of the term structure is related to the RV. More specifically, the absolute value of the slope shows a positive relation with RV that is statistically significant at the 1% level for the three energy contracts and copper.

Given the strong and long-lasting autocorrelation of RV, I introduce several autoregressive specifications in the baseline model. I find that the explanatory power of EVs is robust to this inclusion. I first introduce the heterogeneous autoregressive RV (HAR; see 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 exhibit. I show that the explanatory power of EVs is not altered since EVs are altogether statistically significant at the 1% level. I repeat this test with two alternative competing models: the HEXP in which the lagging terms are exponentially smoothed, and the HARQ which accounts for measurement errors (see Bollerslev et al., 2016, 2018), respectively. I find that the explanatory power of EVs remains when these alternative specifications are considered. Finally, I test the baseline model nested with autoregressive and measurement errors variables in time-varying (TV) parameter models (HAR-TV and HARQ-TV) (see, e.g., Casas et al., 2019; Chen et al., 2018). The explanatory power of EVs is also robust to these specifications. I carry on with a horse race of all competing models, that nest EVs or not, and compare their in- and out-of-sample forecast accuracy at three different horizons. In-sample, the best performing model for 1-day ahead forecasts is the HARQ-TV, whereas the HAR-TV outperforms for the 1-week and 1-month horizons. Out-of-sample, the HARQ dominates when the forecast horizon is lower than 1-week, whereas the EVHARQ-TV produces the best forecasts at the 1-month horizon. Comparing one to one the 1-day ahead forecasts of each model, I find that EVs yield lower errors at the 1% level, when they are introduced in the HEXP, HAR-TV, and HARQ-TV, compared to their constrained version. I use these out-of-sample forecasts in multiquantile VaR regressions, and extract expected shortfall for up to four coverage levels. I jointly test the parameters for the forecast bias, and uncover that the simple autoregressive models (HAR and EVHAR) provide the lowest rejection rates.

The contribution of this article is threefold. First, commodity futures are widely overlooked in the RV literature in comparison to other asset classes. To fill this void, I explore the performance of RV models that include theoretically motivated EVs. Second, I show that the addition of EVs to autoregressive models improves their explanatory power, and the accuracy of their forecasts. Since the RV provides a better approximation of the true latent volatility process than the (G)ARCH and stochastic volatility approaches, the re-examination of such EVs in this context is particularly relevant (see Andersen & Bollerslev, 1998b). Finally these tests also highlight the contracts peculiarities in terms of RV.

The remainder of the article proceeds as follows. Section 2 develops 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. Section 4 discusses the empirical results. Section 5 compares the performance of out-of-sample forecasts from both autoregressive and EV models in tail risk management. Section 6 concludes.

2 | PREVIOUS RESEARCH AND HYPOTHESES

The daily realized variance is defined as:

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

where $1/\Delta$ is the number of observations in 1 day and r_Δ^2 represents the squared change in intraday (log) prices (see, e.g., Andersen et al., 2007). Andersen and Bollerslev (1998a) show that the realized variance is the limit of the integrated variance, as the frequency increases to infinity:

[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.

$$\lim_{\Delta \to 0} RV^2_{t+1}(\Delta) \to \int_t^{t+1} \sigma^2(s)ds + \sum_{t<s\leqslant t+1} \kappa^2(s) \quad (2)$$

The right-hand side of Equation (2) is the integrated variance of the diffusion process σ, and discrete jumps of size κ. Disentangling jumps from the continuous process is an empirical issue (see, e.g., Aït-Sahalia, 2002, 2004).

Generally, RV (square root of the realized variance) is computed using constant time-intervals ranging from 1 to 30 min (see, e.g., Aït-Sahalia et al., 2005; Andersen et al., 2011b). Patton and Sheppard (2015) use consecutive transactions. When 5-min RV is taken as the benchmark, Liu et al. (2015) find little evidence that it is outperformed by any other measure. However, when using inference methods that do not require to specify a benchmark, there is some evidence that more sophisticated measures outperform. For example, Andersen et al. (2011b) propose two estimators of particular interest: (a) the average estimator and (b) the optimal estimator. The linear forecasts obtained by averaging standard sparsely sampled RV measures generally perform on par with the best alternative robust measures. Overall, 5-min RV is difficult to beat.

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:

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

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

The second difference with 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 & Rouwenhorst, 2006).

2.2 | The economic determinants of RV

Anderson and Danthine (1983) hypothesize that the key determinant of volatility is the time at which the production uncertainty is resolved. The uncertainty resolution is seasonal, for instance at the end of a crop when the supply is publicly known (see also Anderson, 1985). This seasonality should be particularly visible for agricultural products whose production are concentrated in a single annual harvest in the northern hemisphere.[2] It should also be present for the natural gas contract since the demand rises every winter in the northern hemisphere. Despite the fact that these turning points should primarily affect the cash market, Anderson and Danthine (1983) additionally show that the link between the cash and futures markets ensures 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), Khoury and Yourougou (1993), and Galloway and Kolb (1996) find a seasonal component in volatility, combined with a time to maturity effect.

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 are contradictory. On the one hand, Rutledge (1976) and

[2]See the statistics from the US Department of Agriculture: https://www.nass.usda.gov/. 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.

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 et al. (1996) develop a model in which 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 et al., 2000) and on the S&P 500 futures (Moosa & Bollen, 2001) strongly reject the Samuelson conjecture, whereas Bessembinder et al. (1996) find empirical support mainly for agricultural commodity futures.

The theory of storage states that the relation between the 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 & Laroque, 1992) also predict this link, which is confirmed empirically (see Carpantier and Samkharadze, 2012; Fama & French, 1988; Geman & Nguyen, 2005; Geman & Ohana, 2009; Ng & Pirrong, 1994). Inventories are difficult to measure at the aggregate level with a daily frequency. At the monthly frequency however, Gorton et al. (2012) find empirical evidence of a negative relation between actual inventories and the spot price volatility. More importantly, they confirm the tight link between the term structure and inventories. Thus, because the term structure is readily measurable at any frequency, this variable allows to test inventory-related hypotheses at the daily frequency. Kogan et al. (2009) extend the theory of storage prediction to a nonmonotonic and convex relation between volatility and inventories. They first document empirically that the volatility increases when the inventory levels approach their physical limits (empty or filled storage). To explain this "v-shape" pattern, they derive a model that links the volatility to investment constraints through the capacity of firms to absorb demand shocks. They introduce the slope of the term structure conditioned on its sign in GARCH models, and find that the corresponding coefficients are statistically significant at the 1% level (see also Haugom et al., 2014). To conclude, both low and high inventories lead to high volatility. Consequently, I state my hypotheses as follows. The volatility of commodities futures:

(a) is seasonal for commodities that show seasonality in the supply or the demand,
(b) increases when the time to maturity decreases, and
(c) is positively related to the intensity of both low and high inventory states.

2.3 | Endogenous determinants of RV

The main idea of Corsi (2009) is that the RV at time 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 its dynamics. Altogether, these variables give a parsimonious representation of the typical volatility exponential decay (see, e.g., Andersen et al., 2003). In the empirical part of the paper, Corsi (2009) estimates the model with the S&P 500 index, the USD/CHF exchange rate, and a 30-year US T-Bond futures. Based on the BIC criterion, the 1-day ahead in-sample performance of this model is higher than that of an AR (22), which 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 1-day, 1-week, and 1-month horizons. In addition, it is on par with an (long-memory) ARFIMA model. Lastly, the superior performance of the ARFIMA and HAR (3) increases with the forecasting horizon.

Several versions of the model have been proposed using the RV, 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. Nevertheless, the residuals of log RV are still heteroskedastic, and the parameters of the HAR are not constant over time (see Buccheri & 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 et al. (2011a) consider that RV has (i) a continuous component for the day that is well described by an HAR-GARCH model, (ii) a jump component for the day, and (iii) a GARCH component for the night, leading to the HAR-CJN model. However, the out-of-sample performance of this model is just 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" (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 when they are subject to potential mismeasurement. By the same token, this variable is supposed to capture microstructure effects and jumps. However, HARQ also shows 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, that is, the so-called "leverage effect." Casas et al. (2018) nest both models. Cipollini et al. (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 to negative shocks. Second, the coefficients are no longer a local linear function of time but a linear function of the realized quarticity (semiparametric 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 are 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 for multivariate estimations.[3]

Overall, significant progress has been made with the development of sophisticated specifications. Their main purpose is to clean the data from microstructure effects, and to account for an asymmetric reaction to negative exogenous shocks. Yet, the focus of empirical applications has been on stock indices, currencies, individual stocks, bonds, and less frequently on futures contracts. The following methodology details the implementation of the aforementioned benchmark models to test the contribution of EVs.

3 | METHODOLOGY

I start with a linear model that incorporates the EVs 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. All specifications are systems of equations (SUR), estimated with FGLS. Therefore, they account for the contemporaneous correlation of error terms across equations induced by potential omitted variables common to all contracts.

3.1 | Baseline model

To check whether the economic determinants of RV have any explanatory power, three EVs observable at the daily frequency are considered. First, I introduce monthly dummies for the corn, soybeans, wheat, and natural gas contracts, to test for seasonal effects (uncertainty resolution hypothesis). 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 United States, 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 United States, historically. This choice also matches the unconditional seasonal pattern of RV that is displayed in Figure 1.[4]

Second, I introduce the log of the time to maturity (Samuelson hypothesis), crossing the timestamp with the contract maturity information available in the full ticker.[5] I set it on a calendar basis, since I assume that the latent

[3]Multivariate RV series modeling is out of the scope of this paper.

[4]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 A1.

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

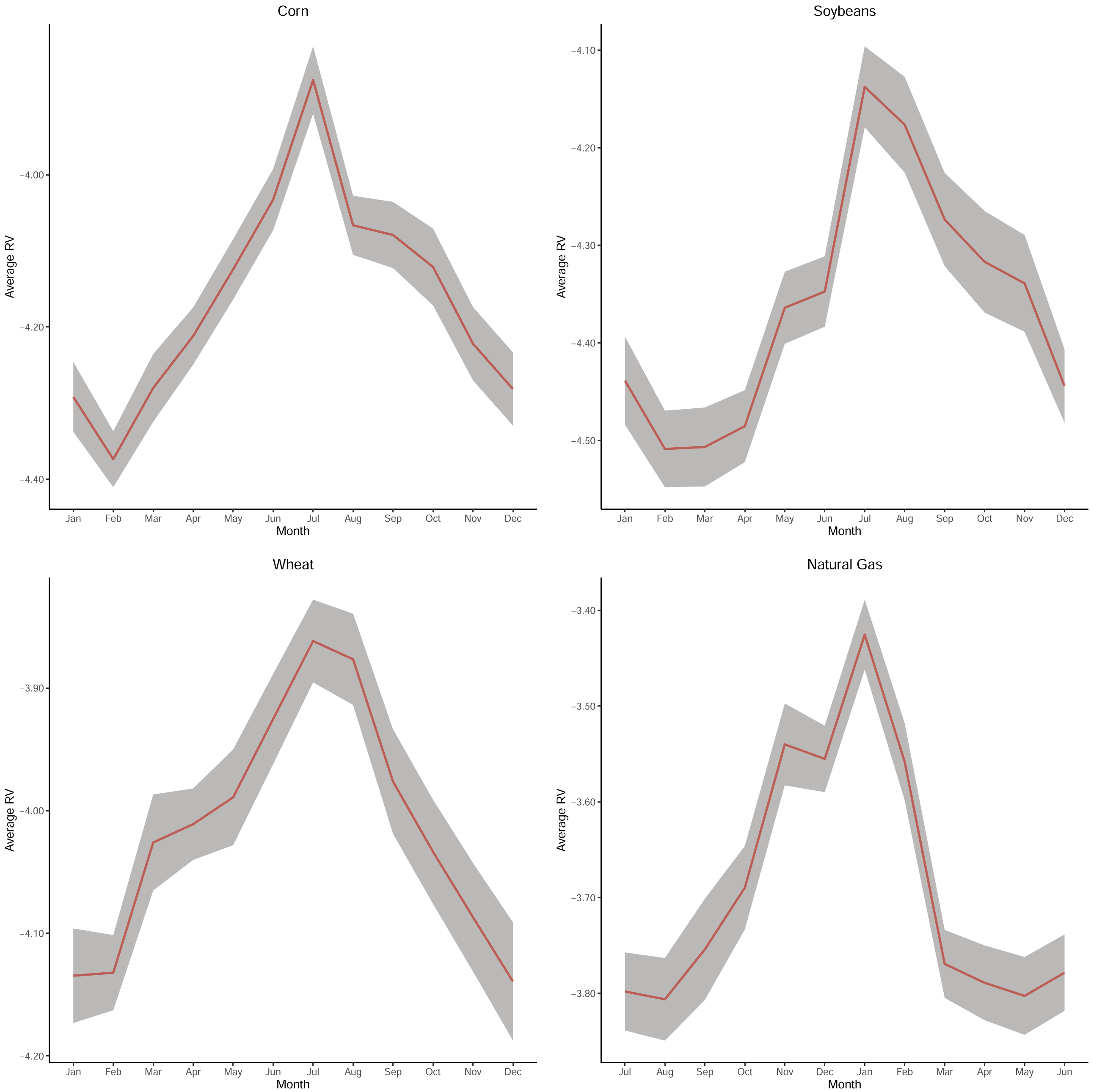

FIGURE 1 Unconditional monthly average RV. These plots display the monthly average of daily RV for the nearby futures contract written on corn, soybeans, wheat, and natural gas. The red line represents the average and the gray area represents the 90% confidence bands. For a better alignment, the plot 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. RV, realized volatility

maturity information exists even when futures are not traded.[6] Finally, consistent with the theory of storage, I consider the slope of the nearest term structure. Since the maturity gap differs across contracts, I normalize the slope measured for each contract by the maturity gap.[7] More specifically, I test the "v-shape" hypothesis (Kogan et al., 2009; Haugom et al., 2014) by adding a "backwardation" dummy set to "1" when the slope of the nearest term structure is negative, and "0" otherwise. Its interaction with the slope allows to capture simultaneously the inventory effects related to either the "v-shape" or the traditional theory of storage. The following model is estimated:

[6]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.

[7]The contracts maturities are reported in Appendix Table A1.

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

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). These variables are defined in the Appendix Table A2.

3.2 | The autoregressive components of RV

Next, I control for the autoregressive component of the RV by introducing five different specifications, that is, the HAR (Corsi, 2009), the HEXP (Bollerslev et al., 2018), the HARQ (Bollerslev et al., 2016), the HAR-TV (Chen et al., 2018), and the HARQ-TV. I choose these models because they are parsimonious and competitive in terms of explanatory power and forecasting ability.

- EVHAR

The EVHAR model is,

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

where $RV_{c,t}$ is the log RV in time t for commodity c, $\boldsymbol{EV_{c,t}}$ is the vector of EVs in Equation (3) and a constant, $RV_{c,t-n|t-p}$ is the log average 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. (2018) is similar to the HAR. It uses mixtures of exponentially smoothed past log RV. Each term is computed as, $RV_{c,t}^{CoM(\lambda)} = \sum_{i=1}^{500} \frac{e^{-i\lambda}}{e^{-\lambda} + e^{-2\lambda} + \cdots + e^{-500\lambda}}$, with $\lambda = \ln\left(1 + \frac{1}{\text{CoM}}\right)$, for decay rates $\lambda = 0.693, 0.182, 0.039$, and 0.008 corresponding to centers of mass (CoM) of 1, 5, 25, and 125 days, respectively:

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

- EVHARQ

The HARQ uses the realized quarticity to account for measurement errors. Barndorff-Nielsen and Shephard (2002) define the estimator of the log-realized quarticity (hereafter, RQ_t) as:

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

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

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

3.3 | Are the parameters time-varying?

- EVHAR-TV

To check whether parameters are time-varying, the semiparametric (local kernel) estimation approach is employed (see Casas et al., 2019; Chen et al., 2018). This method also 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,

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

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 covariance matrix of errors for the feasible generalized least squares (FGLS) is itself time-varying, with a bandwidth selected similarly. The model is,

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

where the coefficients $\alpha'(\tau_t)$ are time-varying, $\tau_t = \frac{t}{T}$ the smoothing variable for $t = 1, 2, ..., T$, and T 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,

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

3.4 | Data and descriptive statistics

3.4.1 | Data and variable definition

Nine contracts evenly spread in three main subgroups, that is, agriculture (wheat, corn, and soybeans), energy (WTI crude oil, natural gas, and heating oil), and metal (copper, gold, and silver) are selected because they have the highest open interest and turnover in their own subgroup.[8] From the Barchart API, I download 5-min closing prices of the nearby and first deferred commodity futures contracts from May 6, 2008 to January 18, 2019.[9] The data include a timestamp and the maturity date of each contract, and are restricted to the Globex session (exchange market) only. The trading hours are reported in Appendix Table A1, along with the corresponding maximum 5-min observations per day. Finally, estimations of the SUR system requires a perfect time match of the nine contracts, to avoid forward-looking bias entering the residuals covariance matrix in the FGLS. Consequently, I set a common cut-off time at 4 p.m., CT, which I define as the end of the trading day.[10]

I compute 5-min 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. The arithmetic RV (ARV) is defined as follows:

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

and the log RV (RV) as follows:

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

where $1/\Delta$ is the number of observations available given then market open hours of each contract. I choose the 5-min sampling since previous literature document its performance over alternative frequencies (see, e.g., Liu et al., 2015).[11]

[8]Contracts specifications are in the Appendix Table A1, and their liquidity characteristics in Table A4.
[9]https://www.barchart.com/
[10]In unreported robustness tests, I roll the nearby onto the first deferred 5 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.
[11]Summary statistics in the Appendix Table A3 also show that it is a reasonable choice across the nine contracts, when compared to alternative frequencies.

TABLE 1 Summary statistics: Daily RV

	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)
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
$Q_{(20)}$	4856	6723	7139	26,574	25,471	8768	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
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	6706
$Q_{(20)}$	11,283	10,703	8201	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

Note: This table reports summary statistics for four estimators of daily realized volatility. These estimators are based on 5-min log price changes of the nearby futures commodity contract. I display these statistics for the arithmetic realized volatility ARV_t and the log realized volatility RV_t. The table displays the first four moments of the distribution, the Jarque-Bera statistic JB, the Ljung-Box statistic (20th order serial correlation) $Q_{(20)}$, and the parameter d of the log-periodogram regression (Geweke & 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.

3.4.2 | Summary statistics of 5-min RV and daily market data

Table 1 compares both distributional and memory properties of the ARV and RV. The daily mean of the ARV ranges from 1.02% (gold) to 2.72% (natural gas). The ARVs of commodity futures exceed those found in previous literature for exchange rates, sovereign bonds, and stock indices. Instead, they are included in the range of typical RVs found for large traded stocks. For instance, the mean of the ARV is 0.5% for the US T-Bond, 0.67% for the DEM/USD exchange rate, and 0.93% for the S&P 500 over the 1986–2002 period (see Andersen et al., 2007). On the other hand, Bollerslev et al. (2016) find that, over the 1997–2013 period, the mean ARV of 27 Dow Jones stocks is in the 1.68%–5.42% range. Buccheri and Corsi (2019) find a slightly larger but similar range (0.95%–11.10%), for 18 NYSE stocks over the 2006–2014 period. Overall, the null hypothesis of normality is rejected for both the ARV and RV. However, in line with Andersen et al. (2003), the RV distribution is closer to normality than that of the ARV. The skewness of the ARV is at least twice that of the RV for eight contracts, and even more for the crude oil contract. The excess kurtosis of the ARV is also much larger (up to two orders of magnitude for the crude oil contract). Moreover, the persistence increases when using the RV in place of the ARV. The Ljung-Box statistics of the RV are twice those of the ARV for eight contracts. The log-periodogram parameters of the RV are also superior for agriculture and energy, but not for metal products. Overall, the memory properties of the ARV and RV are similar to previous results found for exchange rates (Andersen et al., 2003, 2007), S&P 500, and US T-Bonds (Andersen et al., 2007).

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 assets, their distributions strongly depart from normality, with an average *SD* that vastly exceeds the mean and an important excess kurtosis, from 1.93 (wheat) to 17.04 (corn). The skewness that has long been perceived as positive in commodity futures, together with positive excess returns, also range from −1.17 (soybeans) to 0.33 (natural gas) (see Gorton &

TABLE 2 Summary statistics: Daily market-level data

	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	1466.37	2151.16	827.84	16, 506.63	1658.83	2548.02	8529.10	1948.14	2, 699.74
σ	1223.27	2111.44	621.02	9594.49	1172.04	1461.08	11, 200.83	2205.74	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	8277.70	15, 145.34	4330.91	73, 000.47	6605.90	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	1201.31	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	3	1	1	1
Max	1051	1054	1051	1380	1353	1378	1380	1375	1380
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	0.00	0.01	0.04	0.09	0.15	0.00	0.00	0.00
Max	16.61	16.04	30.23	14.98	15.65	52.10	10.19	13.37	17.41

Note: 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-min 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 (20th 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.

Rouwenhorst, 2006). I additionally provide the proportion of days during which the contracts are in contango, which stands between 60.16% (soybeans) and 97.55% (wheat). The group means is 81%, 80.12%, and 67.55%, for agriculture, energy, and metal, respectively. These statistics differ from the classical view according to which agricultural products are more subject to contango, given their important storage costs (see, e.g., Fama & French, 1987; Keynes, 1930). Nonetheless, it does confirm that precious metals have the least contango, given the irrelevance of storage costs with regard to their underlying value (see, e.g., Ng & Pirrong, 1994). Finally, I report the optimal sampling frequency, to identify the best 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, the lower the bid-ask spread and the higher the optimal sampling frequency. In the remainder of the article, I define RV as the log RV computed with a 5-min sampling frequency.[12]

4 | RESULTS

4.1 | In-sample estimation

4.1.1 | EV and EVHAR

Table 3 reports the coefficient estimates of the EV (Equation 3) and the EVHAR (Equation 4) models.

In the EV specification, the monthly dummies associated with agricultural products (July) and the natural gas contract (January) are positive, and statistically significant at the 1% level. This first result is consistent with the seasonal RV structure displayed in Figure 1 and the uncertainty resolution hypothesis (see Anderson & Danthine, 1983). Second, there is a positive relation between the time to maturity and volatility. For seven out of nine contracts, the corresponding coefficients are statistically significant at the 1% level, in contradiction with the Samuelson conjecture. Third, in line with Kogan et al. (2009) and Haugom et al. (2014), I find that for the crude oil contract, the magnitude of the slope matters but not its sign. *SL* (contango) loads positively, the interacted variable $SL \times B$ (backwardation) loads negatively, and both coefficients are statistically significant at the 1% level. Thus, contango and backwardation states are positive predictors of RV, 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. This result is likely induced by the fact that the wheat contract is in contango 97.55% of the time over the sample period. The *SL* coefficient is negative and significant at the 1% level for the corn, gold, and silver contracts, thereby showing a negative contango-RV relation. If the result for the corn contract is difficult to explain (or is spurious), those for the gold and silver contracts may be related to previous empirical evidence. For instance, Ng and Pirrong (1994) document that precious and industrial metals have large differences in their inventory-volatility relations, explained by the fact that scarcity is irrelevant for the precious metals group. Finally, the explanatory power of the EV model also varies across contracts, with R^2 for individual equations ranging from 2% (copper) to 26% (crude oil).

When I include the HAR terms into the EV specification, the explanatory power of the models is improved, with adjusted R^2 ranging from 33% (soybeans) to 75% (crude oil). The likelihood ratio test (constrained vs. unconstrained model) rejects the null hypothesis at the 1% level showing that EVs continue to explain the RV beyond the autoregressive terms.[13] The HAR coefficients are statistically significant at the 1% level and closely aligned with those of Andersen et al. (2007) for the DEM/USD exchange rate, S&P 500, and US T-Bond. In the EVHAR model, the statistical significance levels of the monthly dummies decrease to 5% for the agricultural products. They remain at the 1% level for the natural gas (coefficient size in the EV model three times larger than in the EVHAR). This points to a more pronounced seasonal structure for natural gas that seems difficult to capture only with autoregressive terms (see Geman & Ohana, 2009). Similarly, the time to maturity coefficients only maintain statistical significance at the 1% level in metal products. Switching from the EV to the EVHAR model, the statistical significance of the "v-shape" coefficients is maintained at the 1% level for all energy contracts. Even if the explanatory power of EVs is partly subsumed by the HAR terms, and that memory properties capture some of the EV features, their contribution is robust from both futures markets (time to maturity), and supply and demand (seasonality and term structure slope) perspectives. Therefore, I keep EVs in the forthcoming autoregressive specifications and test further the robustness of their contribution.

[12]RV summary statistics with alternative sampling frequencies of 1-, 5-, 15-, and 60-min are reported in Appendix Table A3. It verifies the good compromise that the 5-min sampling delivers, both for 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).
[13]The constrained model estimations are reported in Appendix Table A5.

TABLE 3 EV and EVHAR estimation of RV_t

	Agriculture						Energy			
	C		S		W		CL		HO	
Constant	−4.34***	−1.25***	−4.58***	−1.51***	−4.22***	−1.09***	−4.17***	−0.55***	−4.34***	−0.72***
	(0.02)	(0.09)	(0.02)	(0.10)	(0.02)	(0.11)	(0.02)	(0.05)	(0.02)	(0.06)
$M_{c,t}$	0.31***	0.05**	0.18***	0.06**	0.21***	0.05**				
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(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
	(1.08)	(1.03)	(1.09)	(1.05)	(1.17)	(1.07)	(1.03)	(0.69)	(1.00)	(0.70)
$SL_{c,t-1}(10^{-2})$	−1.63***	−0.40	0.36	0.74	3.07***	0.61**	3.38***	0.63***	5.44***	1.53***
	(0.55)	(0.50)	(1.12)	(1.03)	(0.29)	(0.29)	(0.15)	(0.12)	(0.30)	(0.21)
$B_{c,t-1}$	0.07***	0.00	0.05**	0.02	0.16***	0.07	−0.08***	−0.02	−0.05***	−0.04***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.06)	(0.05)	(0.02)	(0.01)	(0.01)	(0.01)
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	0.63	−0.16	−1.41	−1.32	3.92	1.47	−5.05***	−1.80***	−8.70***	−3.23***
	(0.58)	(0.52)	(1.13)	(1.03)	(4.91)	(4.43)	(0.61)	(0.39)	(0.38)	(0.26)
$RV_{c,t-1}$		0.32***		0.26***		0.21***		0.42***		0.37***
		(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***
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
$RV_{c,t-6\|t-22}$		0.21***		0.22***		0.27***		0.20***		0.23***
		(0.02)		(0.02)		(0.03)		(0.02)		(0.02)
Adj. R^2	0.11	0.41	0.06	0.33	0.09	0.38	0.26	0.75	0.20	0.71
OLS R^2	0.15	0.50								
McElroy R^2	0.11	0.47								
log likelihood	−3916.70	249.93								
LR EVHAR > HAR	559.47***									

	Energy		Metal					
	NG		GC		HG		SI	
Constant	−3.96***	−0.80***	−4.83***	−1.08***	−4.57***	−0.75***	−4.34***	−1.11***
	(0.02)	(0.08)	(0.02)	(0.12)	(0.02)	(0.09)	(0.02)	(0.10)

TABLE 3 (Continued)

	Energy		Metal					
	NG		GC		HG		SI	
$M_{c,t}$	0.29***	0.11***						
	(0.03)	(0.02)						
$TM_{c,t}\,(10^2)$	5.98***	1.23	11.01***	5.85***	7.79***	6.42***	18.57***	11.89***
	(1.48)	(1.22)	(1.23)	(1.16)	(1.27)	(1.09)	(1.36)	(1.31)
$SL_{c,t-1}(10^{-2})$	1.22***	0.46***	−55.21***	−12.25***	11.91***	2.84	−43.72***	−9.92***
	(0.10)	(0.08)	(3.17)	(3.24)	(2.47)	(2.00)	(2.81)	(2.95)
$B_{c,t-1}$	0.03	−0.02	−0.06	−0.05	−0.08***	−0.06***	0.10**	0.05
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−3.27***	−1.21***	−25.92	−47.00	−20.17***	−11.96***	52.10***	23.50**
	(0.18)	(0.15)	(37.01)	(33.19)	(3.29)	(2.70)	(11.77)	(10.68)
$RV_{c,t-1}$		0.21***		0.20***		0.20***		0.16***
		(0.02)		(0.02)		(0.02)		(0.02)
$RV_{c,t-2\|t-5}$		0.37***		0.28***		0.38***		0.32***
		(0.03)		(0.03)		(0.03)		(0.03)
$RV_{c,t-6\|t-22}$		0.23***		0.32***		0.28***		0.30***
		(0.03)		(0.03)		(0.03)		(0.03)
Adj. R^2	0.19	0.55	0.16	0.42	0.02	0.48	0.17	0.42

Note: This table reports the coefficients of the EV and EVHAR. $RV_{c,t}$ is the 5-min log realized volatility. $M_{c,t}$ is one monthly dummy 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). $SL_{c,t-1}$ 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. The three lagged variables are the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$, that is, the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$, that is, 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 (LR) 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.

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

TABLE 4 EVHEXP and EVHARQ estimation of RV_t

	Agriculture						Energy			
	C		S		W		CL		HO	
Constant	−1.50***	−1.21***	−2.33***	−1.40***	−1.23***	−1.07***	−0.66***	−0.50***	−0.66***	−0.66***
	(0.12)	(0.09)	(0.19)	(0.10)	(0.14)	(0.11)	(0.08)	(0.05)	(0.08)	(0.06)
$M_{c,t}$	0.10***	0.05**	0.08***	0.05**	0.08***	0.05**				
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)				
$TM_{c,t}(10^2)$	0.50	0.11	2.22**	1.01	0.88	0.99	−0.35	−0.53	−0.17	1.19*
	(1.07)	(1.04)	(1.07)	(1.06)	(1.08)	(1.10)	(0.72)	(0.66)	(0.75)	(0.68)
$SL_{c,t-1}(10^{-2})$	0.13	−0.51	−0.78	0.99	0.45	0.56*	0.75***	0.65***	1.67***	1.43***
	(0.59)	(0.50)	(1.17)	(1.02)	(0.30)	(0.29)	(0.13)	(0.11)	(0.23)	(0.20)
$B_{c,t-1}$	0.02	0.00	0.02	0.02	0.09*	0.07	−0.01	−0.02	−0.03***	−0.04***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−0.81	−0.04	0.20	−1.56	3.21	1.39	−2.02***	−2.10***	−3.51***	−3.19***
	(0.61)	(0.52)	(1.18)	(1.03)	(4.46)	(4.45)	(0.41)	(0.38)	(0.29)	(0.26)
$RV^{CoM1}_{c,t-1}$	0.47***		0.37***		0.31***		0.64***		0.58***	
	(0.04)		(0.04)		(0.05)		(0.04)		(0.04)	
$RV^{CoM5}_{c,t-1}$	−0.17**		−0.10		0.05		−0.05		−0.06	
	(0.07)		(0.07)		(0.09)		(0.06)		(0.06)	
$RV^{CoM25}_{c,t-1}$	0.30***		0.44***		0.37***		0.25***		0.21***	
	(0.06)		(0.07)		(0.09)		(0.05)		(0.06)	
$RV^{CoM125}_{c,t-1}$	0.08		−0.22***		−0.01		−0.00		0.14***	
	(0.06)		(0.07)		(0.07)		(0.04)		(0.04)	
$RV_{c,t-1}$		0.34***		0.31***		0.22***		0.54***		0.48***
		(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***
		(0.04)		(0.03)		(0.03)		(0.05)		(0.05)
$RV_{c,t-2\|t-5}$		0.17***		0.19***		0.27***		0.17***		0.17***
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
$RV_{c,t-6\|t-22}$		0.21***		0.20***		0.27***		0.16***		0.20***

TABLE 4 (Continued)

	Agriculture						Energy			
	C		S		W		CL		HO	
		(0.02)		(0.02)		(0.03)		(0.02)		(0.02)
Adj. R^2	0.36	0.41	0.29	0.33	0.36	0.38	0.74	0.77	0.68	0.72
OLS R^2	0.49	0.51								
McElroy R^2	0.45	0.48								
Log-likelihood	−40.03	492.69								
LR EVHEXP > HEXP	638.69***									
LR EVHARQ > HARQ	459.90***									

	Energy		Metal					
	NG		GC		HG		SI	
Constant	−1.14***	−0.66***	−0.87***	−1.08***	−0.60***	−0.70***	−0.82***	−1.08***
	(0.16)	(0.07)	(0.18)	(0.12)	(0.12)	(0.09)	(0.14)	(0.10)
$M_{c,t}$	0.12***	0.09***						
	(0.02)	(0.02)						
$TM_{c,t}(10^2)$	0.05	2.48**	6.32***	3.61***	7.21***	5.08***	13.36***	10.53***
	(1.25)	(1.18)	(1.16)	(1.26)	(1.07)	(1.29)	(1.30)	(1.52)
$SL_{c,t-1}(10^{-2})$	0.48***	0.40***	−8.97**	−14.65***	5.09**	2.31	−3.20	−10.77***
	(0.08)	(0.08)	(3.82)	(3.26)	(2.03)	(2.01)	(3.25)	(2.97)
$B_{c,t-1}$	−0.03	−0.03	−0.06*	−0.05	−0.05**	−0.06***	0.05	0.05
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−1.27***	−1.04***	−64.56*	−47.00	−14.25***	−11.39***	16.34	23.98**
	(0.15)	(0.14)	(33.11)	(33.22)	(2.74)	(2.70)	(10.69)	(10.75)
$RV_{c,t-1}^{CoM1}$	0.22***		0.18***		0.23***		0.12**	
	(0.05)		(0.05)		(0.05)		(0.06)	
$RV_{c,t-1}^{CoM5}$	0.47***		0.26**		0.39***		0.44***	
	(0.08)		(0.10)		(0.09)		(0.11)	
$RV_{c,t-1}^{CoM25}$	0.10		0.41***		0.25***		0.16	

(Continues)

TABLE 4 (Continued)

	Energy		Metal					
	NG		GC		HG		SI	
	(0.07)		(0.10)		(0.08)		(0.10)	
$RV_{c,t-1}^{CoM125}$	−0.07		0.00		0.04		0.16**	
	(0.07)		(0.07)		(0.05)		(0.07)	
$RV_{c,t-1}$		0.41***		0.20***		0.22***		0.16***
		(0.02)		(0.02)		(0.02)		(0.02)
$RV_{c,t-1} \times RQ_{c,t-1}$		0.72***		0.11***		0.07**		0.04
		(0.05)		(0.02)		(0.03)		(0.03)
$RV_{c,t-2\|t-5}$		0.24***		0.26***		0.37***		0.32***
		(0.03)		(0.03)		(0.03)		(0.03)
$RV_{c,t-6\|t-22}$		0.18***		0.32***		0.28***		0.30***
		(0.02)		(0.03)		(0.03)		(0.03)
Adj. R^2	0.53	0.58	0.43	0.43	0.50	0.49	0.43	0.42

Note: This table reports the results of the EVHEXP and EVHARQ models. Realized volatility $RV_{c,t}$ is computed with 5-min (log) prices. I compute the HEXP with four terms exponentially smoothed over a backward looking 500-day window for four lagged variables, that is, 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 $RV_{c,t-2|t-5}$, that is, the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$, that is, the average of the RV from $t-6$ to $t-22$. $M_{c,t}$ is a monthly dummy 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). $TM_{c,t}$ is the log time to maturity. $SL_{c,t-1}$ is 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). 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 (LR) 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 2255 for the EVHEXP and 2733 for the EVHARQ.

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

4.1.2 | Alternative autoregressive specifications

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

Given their important lag structure (500 lags, and up to 125-day CoM), HEXP variables are likely to impact the cyclical EVs: monthly dummies and time to maturity. For instance, the 125-day CoM coefficient is negative and statistically significant at the 1% level for the soybeans contract, one of the "seasonal" commodities. Indeed, this variable is centered with a half-year lag, its negative coefficient points to a seasonal structure of the RV, in which one season is in a high state, relative to the "opposite" season. Instead, this parameter remains positive and statistically significant at the 1% level, for the heating oil and silver contracts, which was not hypothesized to be seasonal. In all contracts, the monthly dummies are robust to this deep lag structure, and remain statistically significant at the 1% level. In addition, the significance of the time to maturity and term structure-related coefficients are comparable to those of the EVHAR. Most of the variance in the EVHEXP is explained by the 1-day CoM coefficient (equivalent to a 1 day-lag) and the 25-day CoM coefficient, which display statistical significance at the 1% level for nine and seven contracts, respectively. The 5-day CoM coefficient is statistically significant at the 5% level for five contracts only. These results indicate that the (EV)HEXP may still properly explain the RV in a more parsimonious specification. Lastly, the likelihood ratio $EVHEXP/HEXP$, which tests the null that the contribution of EVs is nil in the HEXP model is statistically significant at the 1% level, which implies that the EVs' contribution is also robust to this model.[14]

The coefficients related to the quarticity are statistically significant at the 5% level for six contracts, while the inclusion of the HARQ does not affect the coefficient of the seasonal dummy (statistically significant at the 5% and 1% level for the natural gas and agricultural products, respectively), it does affect the other EVs. Seven out of nine time to maturity coefficients in the EVHARQ specification are smaller than those of the EVHAR, suggesting that measurement errors (or faster mean reversion when volatility is high) may relate to contract maturity (see Cipollini et al., 2017). The coefficients associated with the slope decrease further after the inclusion of the realized quarticity but not consistently across contracts. The explanatory power of the EVHARQ is consistently higher than those of the EVHAR and EVHEXP in terms of individual equations and system R^2. Lastly, the likelihood ratio $EVHARQ/HARQ$, which tests the null that the contribution of the EVs is nil in the HARQ specification, is also rejected at the 1% level. This supports the robustness of EVs when both autoregressive and measurement error terms are included.

4.1.3 | Time-varying coefficients

Table 5 reports the results of the EVHAR and EVHARQ models with time varying parameters (EVHAR-TV, Equation 7 and EVHARQ-TV, Equation 8). The means of the coefficients time series and their minima and maxima.

The extrema provide information about the extent to which parameters are time-varying. The largest ranges are obtained for the intercepts (RV spread of up to 23 for the corn contract), and EVs.[15] In contrast, all autoregressive parameters are more stable across contracts (minimum–maximum RV range of 4 at most). The larger range of the intercept compared to the autoregressive coefficients aligns with the results of the score-driven HAR (SHAR; see Buccheri & Corsi, 2019). The means of the parameters of the EVHAR(Q)-TV lie in the same range as those of the static EVHAR(Q), and the improvement in explanatory power when the specification accounts for measurement errors is similar. The likelihood ratio tests $EVHAR\text{-}TV/HAR\text{-}TV$ and $EVHARQ\text{-}TV/HARQ\text{-}TV$ are statistically significant at the 1% level, implying that EVs significantly improve the explanatory power of the most exhaustive autoregressive specifications, which accounts for both measurement errors and time-varying parameters. Moreover, the extrema ranges of the HAR-TV- and HARQ-TV-related parameters are considerably reduced after the inclusion of the EVs[16]. For instance, the range in RV of the intercepts of the silver contract decreases from 0.93 in the HAR-TV, to 0.12 in the EVHAR-TV, which suggests that EVs play a role in capturing the structural/long-term levels of the RV.

The cyclical components of EVs, time to maturity and monthly dummies, are partly captured by the time-varying intercepts. Yet, in both models, these parameters remain time-varying. This points to a nonlinear relation between the EVs and the level of RV (see, e.g., 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

[14]The results of the estimation of the restricted HEXP and HARQ specifications are reported in Appendix Table A5.
[15]See also Appendix Figure A2 for the time-varying pattern of the intercepts. Other coefficients are available upon request.
[16]I report the results of the estimation of the restricted HAR-TV and HARQ-TV in Appendix Table A6.

TABLE 5 EVHAR-TV and EVHARQ-TV estimation of RV_t

	Agriculture						Energy			
	C		**S**		**W**		**CL**		**HO**	
Constant	−1.27***	−1.23***	−1.53***	−1.42***	−1.10***	−1.09***	−0.58***	−0.52***	−0.75***	−0.68***
	−1.42/ − 1.19	−1.38/ − 1.15	−1.61/ − 1.49	−1.49/ − 1.38	−1.23/ − 1.02	−1.22/ − 1.02	−0.66/ − 0.50	−0.57/ − 0.47	−0.84/ − 0.67	−0.75/ − 0.62
$M_{c,t}$	0.05***	0.05***	0.06***	0.06***	0.05***	0.05***				
	0.04/0.07	0.04/0.07	0.05/0.06	0.05/0.06	0.05/0.06	0.05/0.06				
$TM_{c,t}\,(10^2)$	0.33***	0.23***	2.17***	1.09***	0.97***	1.05***	−0.45***	−0.54***	0.34***	1.11***
	−0.44/1.16	−0.61/1.13	1.83/2.62	0.75/1.54	0.74/1.27	0.84/1.31	−0.69/ − 0.24	−0.86/ − 0.22	−0.17/0.72	0.78/1.53
$SL_{c,t-1}(10^{-2})$	−0.39***	0.41	0.84***	0.34	0.59***	0.38	0.62***	0.77	1.53***	0.72
	−0.52/ − 0.14	−0.63/ − 0.30	0.56/1.06	0.86/1.31	0.52/0.64	0.46/0.60	0.61/0.64	0.63/0.66	1.50/1.61	1.41/1.49
$B_{c,t-1}$	0.00***	0.00***	0.02***	0.02***	0.07***	0.07***	−0.02***	−0.02***	−0.04***	−0.04***
	0.00/0.01	0.00/0.01	0.02/0.03	0.02/0.03	0.06/0.08	0.06/0.08	−0.02/ − 0.01	−0.02/ − 0.01	−0.05/ − 0.04	−0.04/ − 0.04
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−0.18***	−0.06***	−1.43***	−1.66***	1.50***	1.39***	−1.83***	−2.11***	−3.24***	−3.20***
	−0.41/ − 0.04	−0.24/0.07	−1.66/ − 1.15	−1.89/ − 1.44	0.46/2.40	0.67/2.10	−2.00/ − 1.72	−2.26/ − 2.03	−3.35/ − 3.16	−3.29/ − 3.14
$RV_{c,t-1}$	0.32***	0.34***	0.26***	0.31***	0.21***	0.21***	0.42***	0.54***	0.36***	0.47***
	0.31/0.33	0.32/0.35	0.25/0.26	0.30/0.31	0.20/0.22	0.20/0.22	0.40/0.43	0.52/0.55	0.34/0.38	0.46/0.48
$RV_{c,t-1} \times RQ_{c,t-1}$		0.06***		0.19***		−0.01***		0.77***		0.67***
		0.04/0.09		0.18/0.20		−0.02/0.00		0.75/0.79		0.66/0.67
$RV_{c,t-2\|t-5}$	0.18***	0.17***	0.20***	0.18***	0.26***	0.27***	0.25***	0.17***	0.24***	0.17***
	0.17/0.18	0.16/0.17	0.20/0.21	0.18/0.19	0.25/0.27	0.25/0.28	0.23/0.27	0.14/0.19	0.23/0.25	0.16/0.18
$RV_{c,t-6\|t-22}$	0.21***	0.21***	0.22***	0.20***	0.27***	0.27***	0.20***	0.16***	0.23***	0.20***
	0.20/0.22	0.20/0.21	0.21/0.22	0.20/0.20	0.27/0.28	0.27/0.27	0.19/0.20	0.16/0.17	0.22/0.23	0.20/0.20

TABLE 5 (Continued)

	Agriculture						Energy			
	C		S		W		CL		HO	
Bandwidth	0.56	0.40	0.75	0.36	0.65	0.58	6.50	0.67	20.00	20.00
Pseudo R^2	0.41	0.41	0.33	0.34	0.38	0.38	0.75	0.77	0.71	0.72
log-likelihood	231.50	477.56								
LR EVHAR-TV>HAR-TV	313.65***									
LR EVHARQ-TV>HARQ-TV	231.59***									

	Energy		Metal					
	NG		GC		HG		SI	
Constant	−0.80***	−0.66***	−1.09***	−1.09***	−0.76***	−0.70***	−1.13***	−1.09***
	−0.82/ − 0.79	−0.68/ − 0.66	−1.15/ − 1.06	−1.13/ − 1.06	−0.80/ − 0.73	−0.74/ − 0.68	−1.20/ − 1.06	−1.16/ − 1.04
$M_{c,t}$	0.11***	0.09***						
	0.10/0.12	0.08/0.09						
$TM_{c,t}(10^2)$	1.20***	2.44***	5.84***	3.65***	6.42***	5.13***	11.97***	10.62***
	0.95/1.78	2.16/2.85	5.74/5.92	3.38/4.05	6.30/6.49	4.92/5.39	11.75/12.33	9.93/11.30
$SL_{c,t-1}(10^{-2})$	0.46***	0.58	−12.35***	0.43	2.88***	0.49	−9.76***	0.42
	0.44/0.47	0.39/0.41	−14.40/ − 11.11	−16.11/ − 13.42	2.59/3.56	2.02/2.99	−10.19/ − 9.38	−11.08/ − 10.20
$B_{c,t-1}$	−0.02***	−0.03***	−0.04***	−0.05***	−0.06***	−0.06***	0.05***	0.05***
	−0.03/ − 0.02	−0.03/ − 0.02	−0.05/ − 0.03	−0.06/ − 0.03	−0.07/ − 0.06	−0.07/ − 0.06	0.03/0.05	0.04/0.05
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−1.21***	−1.03***	−44.90***	−45.53***	−12.07***	−11.48***	23.59***	24.24***
	−1.23/ − 1.20	−1.06/ − 1.03	−58.14/ − 34.02	−63.04/ − 30.51	−13.37/ − 11.63	−12.56/ − 11.04	19.72/25.79	20.56/26.38

(Continues)

TABLE 5 (Continued)

	Energy		Metal					
	NG		GC		HG		SI	
$RV_{c,t-1}$	0.21***	0.41***	0.20***	0.20***	0.20***	0.21***	0.16***	0.16***
	0.21/0.21	0.41/0.41	0.19/0.21	0.19/0.21	0.19/0.21	0.21/0.22	0.15/0.17	0.15/0.17
$RV_{c,t-1} \times RQ_{c,t-1}$		0.72***		0.11***		0.06***		0.04***
		0.72/0.73		0.10/0.12		0.05/0.07		0.02/0.06
$RV_{c,t-2\|t-5}$	0.37***	0.24***	0.28***	0.26***	0.38***	0.37***	0.32***	0.32***
	0.37/0.37	0.24/0.25	0.26/0.29	0.25/0.28	0.37/0.38	0.36/0.38	0.30/0.34	0.30/0.33
$RV_{c,t-6\|t-22}$	0.23***	0.18***	0.32***	0.32***	0.28***	0.28***	0.30***	0.30***
	0.22/0.23	0.17/0.18	0.31/0.33	0.31/0.33	0.27/0.29	0.27/0.29	0.29/0.31	0.29/0.31
Bandwidth	1.16	0.95	0.90	0.61	0.59	0.88	0.62	0.45
Pseudo R^2	0.55	0.58	0.42	0.43	0.49	0.49	0.42	0.42

Note: This table reports the results of the EVHAR-TV and EVHARQ-TV model estimated with three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$, that is, the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$, that is, the average of the RV from $t-6$ to $t-22$. $M_{c,t}$ is a dummy 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). $TM_{c,t}$ is the log time to maturity. $SL_{c,t-1}$ is 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). The reported statistics are the means of the time-varying coefficients and ***Significance of the t-test that the time-varying coefficients are not equal to zero, at the 1% level. Min/Max indicate the minima and maxima of the time-varying coefficients. 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 (LR) 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.

***$p < 0.01$.

TABLE 6 Summary of EV performance

	EV	EVHAR	EVHEXP	EVHARQ	EVHAR-TV	EVHARQ-TV
M	4^+	4^+	4^+	4^+	4^+	4^+
TM	7^+	4^+	4^+	4^+	$8^+/1^-$	$8^+/1^-$
SL	$5^+/3^-$	$4^+/2^-$	$4^+/1^-$	$3^+/2^-$	$6^+/3^-$	$6^+/3^-$
B	$4^+/3^-$	2^-	2^-	2^-	$4^+/5^-$	$4^+/5^-$
$SL \times B$	$1^+/4^-$	$1^+/4^-$	4^-	$1^+/4^-$	$2^+/7^-$	$2^+/7^-$

Note: This table summarizes the performance of the EV across the nine contracts, in the restricted and in all autoregressive specifications. For each model, I report the number of occurrence that each EV is positive at the 5% level (X^+), or negative at the 5% level (X^-). M is the monthly dummy set to "1" during the month of uncertainty resolution and to "0" otherwise, and is included in only four out of nine contracts, TM is the (log) time to maturity, SL is the log of the nearest term structure slope, B is a dummy set to "1" ("0") when the slope of the nearest term structure is negative (positive). The sample period is April 22, 2010–April 1, 2019.

negative, respectively). Finally, I find that time-varying coefficients are statistically different from zero at the 1% level. It provides evidence that even when the coefficients are small, they are mildly time-varying.

4.1.4 | EV and RV: A synthesis

In Table 6, I summarize the results for the EVs across the nine contracts. For each model and variable, I report how many coefficients are significant at the 5% level (negative or positive).

First, the coefficients for the monthly dummies are positive and significant at the 1% level for all four contracts and six models. The robustness of this seasonal term, even in autoregressive specifications, is striking. This clearly supports the contribution of EVs, beyond that of the autoregressive components and the uncertainty resolution hypothesis (see Anderson & Danthine, 1983).

Second, the results of the time to maturity are partly robust. In the restricted EV model, seven out of nine contracts have positive coefficients, significant at the 5% level. When this variable is nested in the alternative static (time-varying) models, four (eight) coefficients remain positive and significant at the 5% level, although the interpretation of the significance in the dynamic version is not straightforward. Altogether these results strongly reject the Samuelson hypothesis, since the time to maturity coefficients are positive and significant at the 1% level. Moreover, the rejection of the Samuelson hypothesis found in the precious metal, which are supposed to behave more like financial assets, also relates to the results for the NIKKEI and S&P 500 futures (see Chen et al., 2000; Moosa & Bollen, 2001). Contrary to Bessembinder et al. (1996), I do not find empirical support for this hypothesis in agricultural commodity futures. Coefficients are not statistically significant at the 1% level.

Third, the results regarding the term structure provide mixed support to the theory of storage and the "v-shape" hypotheses. Four out of nine commodity contracts have statistically significant coefficients at the 1% level. I relate these results to the contract-storage peculiarities detailed in Section 4.1.1. In particular, the contracts for which the "v-shape" hypothesis is supported are those supposed to embed important storage costs (three energy commodities and copper). However, the coefficients of the three agricultural contracts, also viewed as having significant storage costs, do not display the "v-shape" pattern. Lastly, despite the mixed results of the baseline model, they hold across autoregressive specifications, with from three (EVHARQ) to six contracts (EVHAR-TV and EVHARQ-TV) displaying the "v-shape" pattern.

To summarize the aforementioned results, EVs alone do explain the RV. When autoregressive terms are added, the explanatory power of the regressions increases and the size of the EVs coefficients decreases but their sign and statistical significance are maintained. This indicates that the information content of the EVs is captured to some extent by the various lags of the HAR(Q) and HEXP. Yet, the unrestricted versions (with EVs) improve the explanatory power of the models. Moreover, all likelihood ratios reject the null of no improvement (at the 1% level) when EVs are added. Thus, the remainder of the paper analyzes how EVs improve the in- and out-of-sample forecast accuracy of these models.

TABLE 7 In-sample model comparison

	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

Note: 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 $MSE = (\hat{\sigma} - \sqrt{h})^2$, (ii) absolute errors $MAE = |\hat{\sigma} - \sqrt{h}|$, and (iii) $QLIKE = \ln h + \frac{\hat{\sigma}^2}{h}$. When applicable, the superscripts $1, 2, \ldots, n$ indicate the ranking of the models that are included in the confidence interval $\widehat{\mathcal{M}}_{90\%}$. I display the results for 1 day, 1 week, and 1 month ahead. The sample period is April 22, 2010–April 1, 2019. The number of observations is $2255 \times 9 = 20295$.

4.2 | In-sample performance

I compute in-sample forecasts from the 11 models at the 1-day, 1-week, and 1-month horizons. I keep the time to maturity and monthly dummies contemporaneous since those are deterministic variables. Therefore, the set of information that determines their future state is fully available ex ante for every agent, such that this approach does not entail forward-looking bias. Table 7 reports the results of the Model Confidence Set (MCS) procedure with which the performance of the models is directly compared based on the three loss functions, that is, mean squared errors (MSE), mean absolute errors (MAE), and QLIKE (see Hansen et al., 2011; Patton, 2011).

Table 7 reports the tests related to the 1-day ahead forecasts. The HARQ-TV emerges as the best model with the three loss functions. The EVHARQ ranks second, although the MAE losses are excluded from the 90% confidence interval. The 1-week and 1-month ahead comparisons indicate that the HAR-TV is superior, excluding those of the 1-month ahead MSE, for which the HARQ-TV ranks first. These results are in line with those of Bollerslev et al. (2016) for the S&P 500 (MSE) and 27 Dow Jones stocks. However, for longer horizons they find the opposite. Thus, across all models, only a single EV-based model (the EVHARQ) steps up in the 90% confidence set for the MSE and QLIKE losses, at the 1-day horizon. Lastly, the losses are consistently smaller (greater) in the static (time-varying) versions of the models, when EVs are included. This points to time-varying specifications appropriately capturing the time variations of EVs through the levels (intercepts).

4.3 | Out-of-sample performance

I compute out-of-sample forecasts at the 1-day, 1-week, and 1-month horizons based on a training period from April 22, 2010 to January 30, 2013. I obtain the static model forecasts by computing $\mathbb{E}[RV_t]$. For the time-varying specifications, I use the multistage nonparametric predictor approach (see Chen et al., 2004, 2018). In this procedure, I compute the one-step ahead conditional expectations and reuse 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 et al., 1997) to benchmark these out-of-sample forecasts (see Diebold & Mariano, 1995). The results are reported in Table 8. In this analysis, I add the RiskMetrics model as a generic benchmark, given its wide use in risk management.[17] The RiskMetrics model is a parsimonious version of the HEXP with a single exponentially smoothed lagged variable. The decay rate λ is set at 6%, which corresponds to a 16-day CoM. Although the RiskMetrics is not calibrated for log variances or

[17] https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

TABLE 8 Out-of-sample model comparison

	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***	

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

Note: 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 $MSE = (\hat{\sigma} - \sqrt{h})^2$, (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 1 day, week, and month forecasts. Panel B reports the χ^2 statistics for the modified Diebold-Mariano test (see Diebold & 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 1-day ahead forecasts MSE. The in-sample 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 $1555 \times 9 = 13995$.

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

volatilities, in unreported tests, I find that this decay rate remains a good trade-off for log RVs. The RiskMetrics equation is,

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

Table 8, Panel A shows that the HARQ strictly dominates all other models at the 1-day and 1-week horizons. In addition, the benefit of including EVs vanishes over these horizons. However, at the 1-month horizon, the best performing model is the EVHARQ-TV, and its MSE (MAE) is almost a third (half) of the HARQ-TV model. The forecast improvement arising from EVs is present in almost all models and for all loss functions. This supports the benefits of including these exogenous variables when the time horizon increases. These forecasting improvements may also come 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. Finally, Table 8, Panel B reports the results of a one-to-one direct comparison of the 12 models using the modified Diebold-Mariano test, at the 1-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 not statistically significant (−1.31), hence supporting their equally superior forecasting ability at this horizon, in line with previous results. Finally, these individual comparisons allow to test the forecasting accuracy improvement of the unconstrained (with EVs) versus their constrained autoregressive counterparts. EVs inclusion is beneficial (statistically significant at the 1% level) in the HEXP, HAR-TV, and HARQ-TV, whereas it is not useful (statistically significant at the 10% level) in the HAR and HARQ-TV specifications.

TABLE 9 Forecast bias of multi-quantile regressions

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
One-day ahead												
$p = 1$												
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
One-week ahead												
$p = 4$												
J_2	0.04	0.55	0.48	0.04	0.07	0.42	0.19	0.00	0.00	0.00	0.01	0.01
One-month ahead												
$p = 4$												
J_2	0.07	0.18	0.30	0.00	0.01	0.18	0.36	0.00	0.02	0.00	0.03	0.00

Note: 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 1-day, 1-week, and 1-month ahead out-of-sample RV forecasts, and for four number of quantiles $p = 1, 2, 4$, as in Couperier and Leymarie (2020). The quantile levels always start at $\tau_1 = 0.975$, 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 $u_i = 1, 2, ..., p$. 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 1000 replications. The sample period is January 31, 2013–April 1, 2019. The number of observations is $1555 \times 9 = 13995$.

5 | TAIL-RISK MODELING

I now use the out-of-sample forecasts from the 12 models, and 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 multiquantile regression approach (see, e.g., Bayer & Dimitriadis, 2020; Couperier & Leymarie, 2020). It models the left tail of a distribution with a high granularity, since any sequence of coverage levels may be used. Table 9 reports the p-values testing the coefficients from forecast accuracy panel regressions, $\boldsymbol{RV_t} = \beta_0 + \beta_1 \widehat{\boldsymbol{RV_t}}$, where $\boldsymbol{RV_t} = [RV_{1,t}, RV_{2,t}, ..., RV_{9,t}]'$. The null hypotheses are:

- $H_{0,J_1}: \sum_{j=1}^{p} (\beta_0(\tau_j) + \beta_1(\tau_j)) = p$
- $H_{0,J_2}: \sum_{j=1}^{p} \beta_0(\tau_j) = 0$ 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 quantiles, and τ_j is the corresponding quantile level for $j = 1, 2, ..., p$. 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.

The J_2 test rejects the null hypothesis for the EV, the time-varying, and the RiskMetrics models when $p = 1$ (equivalent to the 97.5% VaR), and for finer granularities up to $p = 4$.[18] The alternative test, J_1 whose null hypothesis is that the sums of all multiquantile regressions parameters (both intercepts and coefficients) are equal to p, is only rejected for the HAR-TV and $p = 2$, at the 10% level. Similarly, I and S hypotheses are almost never rejected at the 5% level. Therefore, I focus on the rejection rates of J_2 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). The p-value decreases as the coverage (the granularity of the quantile subdivisions) increases, in line with Couperier and Leymarie (2020). In static parameter (time-varying) specifications, the inclusion of EVs decreases (increases) the expected-shortfall forecasting performance.[19] The bottom of the panel presents the J_2 statistics for $p = 4$, at the 1-week and 1-month horizon. In brief, the 1-week horizon does not reveal any dominance from the restricted or unrestricted specifications. Instead, at the 1-month horizon, all unrestricted EV models generate p-values that are higher than their restricted versions.

In Table 10, 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 horizon of the out-of-sample forecasts is of 1 day.

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 EVs inclusion. Similarly, the most complex time-varying specifications do not deliver better forecasts in terms of VaR violations. Since ES aggregates information from the left tail, the noise arising from the center and the right tail of the distribution may shadow the actual ES signal.[20] 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 the benefits of including EVs when modeling tail risk of some contracts.

6 | CONCLUSION

This study aims to test whether EV, theoretically related to the volatility of commodity futures contracts, add value to autoregressive RV models. Using joint estimations for nine commodities, my results strongly support the uncertainty resolution hypothesis (see Anderson, 1985; Anderson & Danthine, 1983). I also find that the RV is positively related to time to maturity, which rejects the Samuelson hypothesis. Lastly, the inclusion of the slope of the term structure 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

[18]The Mincer–Zarnowitz test is a particular case of the J_2, when there is a single coverage level: $p = 1$.

[19]Appendix Table A7 reports the coefficients of the predictive regressions for each coverage level τ, individually, when the number of quantiles 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 bias, and this for all coverage levels used.

[20]I thank an anonymous referee for this comment.

TABLE 10 Percentage of 97.5% VaR violations

	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

Note: 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 1-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 $1555 \times 9 = 13995$.

not identify any support for the remaining contracts. However, concerning precious metals, it is likely that the determinants of the term structure are unrelated to supply and demand, and to their inherent storage issues. Finally, the performance of the EV considered alone, lies below those of all autoregressive models, including their most parsimonious versions such as RiskMetrics. However, nesting EVs in the autoregressive specifications always brings a statistically significant improvement in explanatory power, at the 1% level. The inclusion of EVs also improves the out-of-sample forecast accuracy of three (HEXP, HAR-TV, and HARQ-TV) out of five models, at the 1-day horizon (significant at the 1% level). It additionally improves the out-of-sample forecasting accuracy for a longer time horizon (1-month ahead), even in specifications accounting for measurement errors and time-varying coefficients. These gains vanish for expected shortfall backtests from multiquantile regressions. Yet, surprisingly, for four out of nine contracts, the EV model alone generates less 97.5% VaR violations than any other model and delivers good results for the remaining contracts.

ACKNOWLEDGMENTS

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 grateful to my supervisor, Michel Dubois, for his invaluable comments and help. I thank Alex Y. Huang and Marinela Adriana Finta, for their discussions during the SFM 2019 and NZFM 2019 conferences. Finally, I am also grateful to Robert Webb (the editor) and an anonymous referee for comments that substantially improved this paper and its scope. This paper is a revised version of the second chapter of my Ph.D. thesis at the University of Neuchâtel. Open access funding provided by Universite de Neuchatel.

ORCID

Loïc Maréchal https://orcid.org/0000-0001-8039-5097

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. https://doi.org/10.1111/0022-1082.85732

Andersen, T. G., & Bollerslev, T. (1998b). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*, 885–905. https://doi.org/10.2307/2527343

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. https://doi.org/10.1162/rest.89.4.701

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*, 579–625. https://doi.org/10.1111/1468-0262.00418

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. https://doi.org/10.1016/j.jeconom.2010.03.029

Andersen, T. G., Bollerslev, T., & Meddahi, N. (2011b). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, *160*, 220–234. https://doi.org/10.1016/j.jeconom.2010.03.032

Anderson, R. W. (1985). Some determinants of the volatility of futures prices. *Journal of Futures Markets*, *5*, 331–348. https://doi.org/10.1002/fut.3990050305

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. https://doi.org/10.2307/2297415

Aït-Sahalia, Y. (2002). Telling from discrete data whether the underlying continuous-time model is a diffusion. *Journal of Finance*, *57*, 2075–2112. https://doi.org/10.1111/1540-6261.00489

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics*, *74*, 487–528. https://doi.org/10.1016/j.jfineco.2003.09.005

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. https://doi.org/10.1093/rfs/hhi016

Barchart, Inc. (2019). Commodity futures intraday data. Data retrieved from Barchart using the Java API. https://www.barchart.com/

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. https://www.jstor.org/stable/3088799

Bayer, S., & Dimitriadis, T. (2020). Regression-based expected shortfall backtesting. Journal of Financial Econometrics (forthcoming). https://doi.org/10.1093/jjfinec/nbaa013

Bekierman, J., & Manner, H. (2018). Forecasting realized variance measures using time-varying coefficient models. *International Journal of Forecasting*, *34*, 276–287. https://doi.org/10.1016/j.ijforecast.2017.12.005

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. https://doi.org/10.3905/jod.1996.407967

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies*, *31*, 2729–2773. https://doi.org/10.1093/rfs/hhy041

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*, 1–18. https://doi.org/10.1016/j.jeconom.2015.10.007

Brennan, M. J. (1958). The supply of storage. *American Economic Review*, *48*, 50–72. https://www.jstor.org/stable/1812340

Buccheri, G., & Corsi, F. (2019). HARK the SHARK: Realized volatility modeling with measurement errors and nonlinear dependencies. Journal of Financial Econometrics (forthcoming). https://doi.org/10.1093/jjfinec/nbz025

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

Carpantier, J.-F., & Dufays, A. (2012). Commodities volatility and the theory of storage. 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. https://doi.org/10.1002/fut.21566

Casas, I., Ferreira, E., & Orbe, S. (2019). Time-varying coefficient estimation in SURE models. Application to portfolio management. Journal of Financial Econometrics (forthcoming). https://doi.org/10.1093/jjfinec/nbz010

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). 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. https://www.jstor.org/stable/3647500

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. https://doi.org/10.1080/07350015.2016.1138118

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. https://doi.org/10.1002/(SICI)1096-9934(199912)19:8<895::AID-FUT3>3.0.CO;2-C

Cipollini, F., Gallo, G. M., & Otranto, E. (2017). On heteroskedasticity and regimes in volatility forecasting. https://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. https://doi.org/10.1093/jjfinec/nbp001

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. https://doi.org/10.1080/07350015.2012.663261

Couperier, O., & Leymarie, J. (2020). Backtesting expected shortfall via multi-quantile regression. 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. https://doi.org/10.2307/2297923

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263. https://doi.org/10.1198/073500102753410444

Fama, E. F., & French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of Business*, *60*, 55–73. https://www.jstor.org/stable/2352947

Fama, E. F., & French, K. R. (1988). Business cycles and the behavior of metals prices. *Journal of Finance*, *43*, 1075–1093. https://doi.org/10.2307/2328207

Galloway, T. M., & Kolb, R. W. (1996). Futures prices and the maturity effect. *Journal of Futures Markets*, *16*, 809–828. https://doi.org/10.1002/(SICI)1096-9934(199610)16:7<809::AID-FUT5>3.0.CO;2-S

Geman, H., & Nguyen, V.-N. (2005). Soybean inventory and forward curve dynamics. *Management Science*, *51*, 1076–1091. https://www.jstor.org/stable/20110399

Geman, H., & Ohana, S. (2009). Forward curves, scarcity and price volatility in oil and natural gas markets. *Energy Economics*, *31*, 576–585. https://doi.org/10.1016/j.eneco.2009.01.014

Geweke, J., Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, *4*, 221–238. https://doi.org/10.1111/j.1467-9892.1983.tb00371.x

Gorton, G., Hayashi, F., & Rouwenhorst, K. G. (2012). The fundamentals of commodity futures returns. *Review of Finance*, *17*, 35–105. https://doi.org/10.1093/rof/rfs019

Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, *62*(2), 47–68. https://doi.org/10.2469/faj.v62.n2.4083

Grammatikos, T., & Saunders, A. (1986). Futures price variability: A test of maturity and volume effects. *Journal of Business*, *59*, 319–330. https://www.jstor.org/stable/2353022

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*, 453–497. https://www.jstor.org/stable/41057463

Harvey, S., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*, 281–291. https://doi.org/10.1016/S0169-2070(96)00719-4

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. https://doi.org/10.1016/j.jbankfin.2014.05.026

Hong, H. (2000). A model of returns and trading in futures markets. *Journal of Finance*, *55*, 959–988. https://doi.org/10.1111/0022-1082.00233

Kaldor, N. (1939). Speculation and economic stability. *Review of Economic Studies*, *7*, 1–27. https://doi.org/10.2307/2967593

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* (pp. 3–46). MacMillian.

Khoury, N., & Yourougou, P. (1993). Determinants of agricultural futures price volatilities: Evidence from Winnipeg commodity exchange. *Journal of Futures Markets*, *13*, 345–356. https://doi.org/10.1002/fut.3990130403

Kogan, L., Livdan, D., & Yaron, A. (2009). Oil futures prices in a production economy with investment constraints. *Journal of Finance*, *64*, 1345–1375. https://www.jstor.org/stable/20488003

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. https://doi.org/10.1016/j.jeconom.2015.02.008

Milonas, N. T. (1986). Price variability and the maturity effect in futures markets. *Journal of Futures Markets*, *6*, 443–460. https://doi.org/10.1002/fut.3990060309

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. https://doi.org/10.1080/13504850110036355

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, *9*, 141–142. https://doi.org/10.1137/1109020

Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, *61*, 631–653. https://doi.org/10.2307/2297912

Ng, V. K., & Pirrong, S. C. (1994). Fundamentals and volatility: Storage, spreads and the dynamics of metals prices. *Journal of Business*, *67*, 203–230. https://www.jstor.org/stable/2353103

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*, 246–256. https://doi.org/10.1016/j.jeconom.2010.03.034

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. https://doi.org/10.1162/REST_a_00503

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, *23*, 1048–1072. https://www.jstor.org/stable/2242436

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, *39*, 1127–1139. https://doi.org/10.1111/j.1540-6261.1984.tb03897.x

Rutledge, D. J. S. (1976). A note on the variability of futures prices. *Review of Economics and Statistics*, *58*, 118–120. https://doi.org/10.2307/1936017

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. https://doi.org/10.2307/1936018

Watson, G. S. (1964). Smooth regression analysis. *Indian Journal of Statistics: Series A*, *26*, 359–372. https://www.jstor.org/stable/25049340

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

How to cite this article: Maréchal, L. (2021). Do economic variables forecast commodity futures volatility? *Journal of Futures Markets*, 41, 1735–1774. https://doi.org/10.1002/fut.22250

APPENDIX A

TABLE A1 Description of futures contracts

Ticker	Trading venue	Underlying	Unit	Maturity	Trading hours (CT)	# 5-min obs.
Agriculture						
C	CBT	Corn	bu (5,000)	HKNUZ	Sunday-Friday, 7:00 p.m.-7:45 a.m. and Monday-Friday, 8:30 a.m.-1:20 p.m.	211
S	CBT	Soybeans	bu (5,000)	FHKNQUX	Sunday-Friday, 7:00 p.m.-7:45 a.m. and Monday-Friday, 8:30 a.m.-1:20 p.m.	211
W	CBT	Chicago wheat	bu (5,000)	HKNUZ	Sunday-Friday, 7:00 p.m.-7:45 a.m. and Monday-Friday, 8:30 a.m.-1:20 p.m.	211
Energy						
CL	NYMEX/ICE	WTI crude oil	bbl (1,000)	FGHJKMNQUVXZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
HO	NYMEX	Heating oil	gal (42,000)	FGHJKMNQUVXZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
NG	NYMEX/ICE	Natural gas	MMBtu (10,000)	FGHJKMNQUVXZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
Metal						
GC	CMX	Gold	oz (100)	GJMQVZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
HG	COMEX	Copper	lb (25,000)	FGHJKMNQUVXZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
SI	CMX	Silver	oz (5,000)	FHKNUZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276

Note: This table reports the specifications of the futures contracts written on the nine selected commodities. The specifications include the ticker, underlying commodity, unit, maturity, the trading hours used for the sample (Globex) and the maximum number of 5-min trade observations per day.

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

TABLE A2 Variable definition

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
$M_{c,t}$	Monthly dummy	$M_{c,t} = \begin{cases} 1 & \text{if} \quad t \in \quad \text{critical month} \\ 0 & \text{if} \quad t \notin \quad \text{critical month} \end{cases}$	–	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.
$SL_{c,t}$	(Annualized log) term structure	$SL_{c,t} = \frac{\left(f_{c,t}^{N}\right) - \left(f_{c,t}^{FD}\right)}{\delta} \times 250$	–	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	$B_{c,t} = \begin{cases} 1 & \text{if} \quad SL_{c,t} < 0 \\ 0 & \text{if} \quad SL_{c,t} > = 0 \end{cases}$	–	D	–
Δ	The temporal period = 300 s	Reciprocal of 10/3 mHz in SI	s	s	With a day composed of 211 and 276 5-min (300 s) observations for agriculture contracts and energy and metal contracts, respectively, so that $1/\Delta = [211; 276]$.
$ARV_{c,t}$	Arithmetic realized volatility	$ARV_{c,t} = \sqrt{\sum_{j=1}^{1/\Delta} r_{t,\Delta \times j}^{2}}$	–	D	
$RV_{c,t}$	Log realized volatility	$RV_{c,t} = \frac{1}{2} \times \ln \sum_{j=1}^{1/\Delta} r_{t,\Delta \times j}^{2}$	–	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-min periods elapsed each day, i.e., $j = 1, 2, ..., n$.
$RV_{c,t-k\mid t-n}$	Average of the RV from $t - k$ to $t - n$	$RV_{c,t-k\mid t-n} = \frac{1}{n-k+1} \sum_{i=k}^{n} RV_{c,t-i}$	–	D	–
$RV_{c,t}^{CoM(\lambda)}$	Exponential average of the RV	$RV_{c,t}^{CoM(\lambda)} = \sum_{i=1}^{500} \frac{e^{-i\lambda}}{e^{-\lambda} + e^{-2\lambda} + \cdots + e^{-500\lambda}}$	–	D	With $\lambda = \ln\left(1 + \frac{1}{\mathrm{CoM}}\right)$

Note: 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.

TABLE A3 Summary statistics: Daily RV with alternative sampling frequency

	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-min 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	79,586	71,432	91,174	481	1916	1379	28, 121	1784	10, 245
$Q(20)$	7995	9645	7340	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-min 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	23, 626	35, 914	41, 036	161	95	259	21, 286	343	6706
$Q(20)$	11, 283	10, 703	8201	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-min 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	13, 738	28, 629	31, 470	132	77	223	18, 283	236	5, 197
$Q(20)$	10, 848	9618	7178	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-min 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	6549	15, 229	14, 596	72	33	86	11, 567	132	3854
$Q(20)$	8606	7318	5406	21, 352	20, 053	10, 619	8700	16, 626	8127
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-min 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

TABLE A3 (Continued)

	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)
JB	25, 942	36, 111	43, 102	154	85	268	20, 329	289	5695
$Q(20)$	10, 643	10, 475	8160	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

Note: This table reports statistics on daily log realized volatility sampled at 1, 5, 15, and 60 min-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 X^2), the Ljung-Box statistic for the 20th order serial correlation ($Q(20)$), and the parameter of the log-periodogram regression based on a bandwidth 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.

TABLE A4 Average daily turnover and open interest over 5 years before the sample

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	1192.12	2208.12	6597.15	2756.31	10, 155.23	5678.65

Metal	Gold (GC)	Copper (HG)	Platinum (PL)	Silver (SI)
Turnover	4, 252.86	1227.43	227.20	1300.46
Open interest	13, 488.65	3792.76	625.33	4339.29

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	1644.18	3395.23	4724.07	433.67	4592.96

Note: This table reports the average daily turnover and open interest in millions USD for 20 contracts components of the SP-GSCI/BCOM, before the sample period selection. The pre-sample period is January 1, 2003–April 30, 2008, and the contracts selected for the study are in bold font.

TABLE A5 HAR, HEXP, and HARQ estimation of RV_t

	Agriculture									Energy					
	C			S			W			CL			HO		
Constant	−1.09***	−1.05***	−1.36***	−1.20***	−1.11***	−1.83***	−0.86***	−0.85***	−1.04***	−0.36***	−0.31***	−0.44***	−0.43***	−0.37***	−0.36***
	(0.08)	(0.08)	(0.10)	(0.09)	(0.09)	(0.17)	(0.09)	(0.09)	(0.13)	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)	(0.07)
$RV_{c,t-1}$	0.33***	0.35***		0.27***	0.32***		0.22***	0.22***		0.44***	0.55***		0.40***	0.51***	
	(0.02)	(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*			0.20***			−0.00			0.76***			0.68***	
		(0.04)			(0.03)			(0.03)			(0.05)			(0.06)	
$RV_{c,t-2\|t-5}$	0.20***	0.19***		0.23***	0.20***		0.28***	0.28***		0.27***	0.19***		0.27***	0.20***	
	(0.02)	(0.02)		(0.02)	(0.02)		(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***		0.20***	0.17***		0.23***	0.20***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.03)	(0.03)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-1}^{CoM1}$			0.47***			0.33***			0.31***			0.67***			0.61***
			(0.04)			(0.04)			(0.05)			(0.04)			(0.04)
$RV_{c,t-1}^{CoM5}$			−0.12**			0.06			0.08			−0.03			0.00
			(0.06)			(0.07)			(0.09)			(0.06)			(0.06)
$RV_{c,t-1}^{CoM25}$			0.32***			0.35***			0.42***			0.26***			0.15***
			(0.06)			(0.07)			(0.09)			(0.05)			(0.05)
$RV_{c,t-1}^{CoM125}$			0.02			−0.14**			−0.05			0.01			0.16***
			(0.05)			(0.06)			(0.07)			(0.03)			(0.04)
Adj. R^2	0.41	0.41	0.36	0.32	0.33	0.29	0.38	0.38	0.36	0.75	0.76	0.74	0.70	0.71	0.67
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			Metal								
	NG			GC			HG			SI		
Constant	−0.46***	−0.36***	−0.95***	−0.71***	−0.74***	−0.42***	−0.58***	−0.56***	−0.42***	−0.70***	−0.69***	−0.46***
	(0.07)	(0.06)	(0.16)	(0.10)	(0.10)	(0.13)	(0.08)	(0.08)	(0.11)	(0.09)	(0.09)	(0.11)

TABLE A5 (Continued)

	Energy			Metal								
	NG			GC			HG			SI		
$RV_{c,t-1}$	0.25***	0.46***		0.22***	0.22***		0.23***	0.23***		0.20**	0.17***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-1} \times RQ_{c,t-1}$		0.77***			0.12***			0.13***			0.11***	
		(0.05)			(0.02)			(0.03)			(0.02)	
$RV_{c,t-2\|t-5}$	0.41***	0.27***		0.29***	0.28***		0.40***	0.39***		0.34***	0.33***	
	(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)	
$RV_{c,t-6\|t-22}$	0.22***	0.16***		0.34***	0.36***		0.25***	0.26***		0.32***	0.34***	
	(0.03)	(0.02)		(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)	
$RV^{CoM1}_{c,t-1}$			0.28***			0.21***			0.27***			0.21***
			(0.05)			(0.05)			(0.06)			(0.06)
$RV^{CoM5}_{c,t-1}$			0.49***			0.29***			0.39***			0.33***
			(0.08)			(0.10)			(0.09)			(0.11)
$RV^{CoM25}_{c,t-1}$			0.11			0.34***			0.20**			0.20*
			(0.07)			(0.10)			(0.08)			(0.10)
$RV^{CoM125}_{c,t-1}$			−0.13*			0.09			0.06			0.18***
			(0.07)			(0.06)			(0.05)			(0.06)
Adj. R^2	0.53	0.57	0.51	0.42	0.43	0.42	0.47	0.48	0.48	0.41	0.42	0.42
OLS R^2												
McElroy R^2												
Log-likelihood												

Note: This table reports the results of the restricted HAR, HEXP, and HARQ estimated with 5-min log realized volatility $RV_{c,t}$. The HAR and HARQ have three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ that is the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$, that is, 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 & Shephard, 2002). The HEXP has four terms exponentially smoothed over a backward looking 500-day window for four lagged variables, that is, 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.

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

TABLE A6 HAR-TV and HARQ-TV estimation of RV_t

	Agriculture						Energy			
	C		S		W		CL		HO	
Constant	−1.29***	−1.25***	−1.27***	−1.16***	−0.78***	−0.79***	−0.39***	−0.34***	−0.45***	−0.37***
	−1.77/ − 0.88	−1.62/ − 0.93	−1.67/ − 0.81	−1.43/ − 0.79	−1.05/ − 0.57	−1.07/ − 0.55	−0.46/ − 0.31	−0.44/ − 0.26	−0.52/ − 0.38	−0.41/ − 0.33
$RV_{c,t-1}$	0.31***	0.32***	0.26***	0.31***	0.23***	0.23***	0.45***	0.56***	0.40***	0.51***
	0.22/0.38	0.20/0.41	0.23/0.30	0.24/0.38	0.20/0.26	0.19/0.27	0.44/0.46	0.53/0.59	0.39/0.41	0.49/0.52
$RV_{c,t-1} \times RQ_{c,t-1}$		0.03***		0.20***		−0.02***		0.75***		0.65***
		−0.15/0.20		0.05/0.34		−0.12/0.07		0.68/0.82		0.62/0.67
$RV_{c,t-2\|t-5}$	0.18***	0.18***	0.22***	0.20***	0.29***	0.29***	0.27***	0.19***	0.27***	0.21***
	0.13/0.21	0.13/0.22	0.14/0.31	0.13/0.30	0.27/0.32	0.27/0.31	0.25/0.28	0.16/0.21	0.26/0.28	0.19/0.23
$RV_{c,t-6\|t-22}$	0.21***	0.21***	0.24***	0.22***	0.30***	0.30**	0.20***	0.16***	0.22***	0.19***
	0.12/0.32	0.12/0.33	0.18/0.31	0.16/0.31	0.23/0.34	0.25/0.33	0.19/0.20	0.14/0.17	0.22/0.23	0.18/0.19
Bandwidth	0.23	0.25	0.29	0.30	0.63	0.63	4.03	0.61	20.00	20.00
Pseudo R^2	0.42	0.42	0.34	0.34	0.38	0.39	0.75	0.77	0.70	0.71
log likelihood	74.80	244.74								

	Energy		Metal					
	NG		GC		HG		SI	
Constant	−0.48***	−0.38***	−0.82***	−0.84***	−0.71***	−0.59***	−1.14***	−1.11***
	−0.52/ − 0.45	−0.47/ − 0.31	−1.01/ − 0.72	−1.06/ − 0.73	−1.03/ − 0.44	−0.86/ − 0.45	−1.69/ − 0.76	−1.70/ − 0.77
$RV_{c,t-1}$	0.24***	0.45***	0.23***	0.22***	0.23***	0.24***	0.18***	0.16***
	0.21/0.29	0.39/0.52	0.19/0.28	0.17/0.27	0.18/0.29	0.21/0.27	0.12/0.29	0.08/0.29
$RV_{c,t-1} \times RQ_{c,t-1}$		0.77***		0.14***		0.15***		0.12***
		0.70/0.84		0.12/0.17		0.13/0.15		0.07/0.16
$RV_{c,t-2\|t-5}$	0.41***	0.27***	0.29***	0.28***	0.39***	0.38***	0.29***	0.28***
	0.39/0.43	0.24/0.29	0.22/0.37	0.21/0.34	0.36/0.43	0.36/0.41	0.21/0.37	0.20/0.36
$RV_{c,t-6\|t-22}$	0.23***	0.17***	0.32***	0.33***	0.23***	0.24***	0.27***	0.30**
	0.20/0.25	0.14/0.19	0.21/0.42	0.23/0.43	0.20/0.24	0.22/0.25	0.15/0.35	0.16/0.39

TABLE A6 (Continued)

	Energy		Metal					
	NG		**GC**		**HG**		**SI**	
Bandwidth	0.74	0.72	0.53	0.57	0.40	0.81	0.25	0.25
Pseudo R^2	0.53	0.57	0.42	0.43	0.48	0.49	0.43	0.44
log likelihood								

Note: This table reports the results of the joint estimation (SUR) of the restricted HAR-TV (Equation 7) and HARQ-TV (Equation 8) models, computed with 5-min log realized volatility $RV_{c,t}$ and with three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$, that is, the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$, that is, 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 & Shephard, 2002). The reported statistics are the means of the time-varying coefficients and "***" indicates the significance of the *t*-test that the time-varying coefficients are not equal to zero, at the 1% level. Min/Max indicate the minima and maxima of the time-varying coefficients. 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.

$^{***}p < 0.01$.

TABLE A7 Forecast bias of individual quantile regressions

	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***

Note: This table reports the coefficients β_0 and β_1 of individual predictive quantile regressions, for the nine commodities and for four 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 1-day-ahead daily RV forecasts from all models tested. The sample period is January, 2013–April 1, 2019. The number of observations is $1555 \times 9 = 13995$.

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

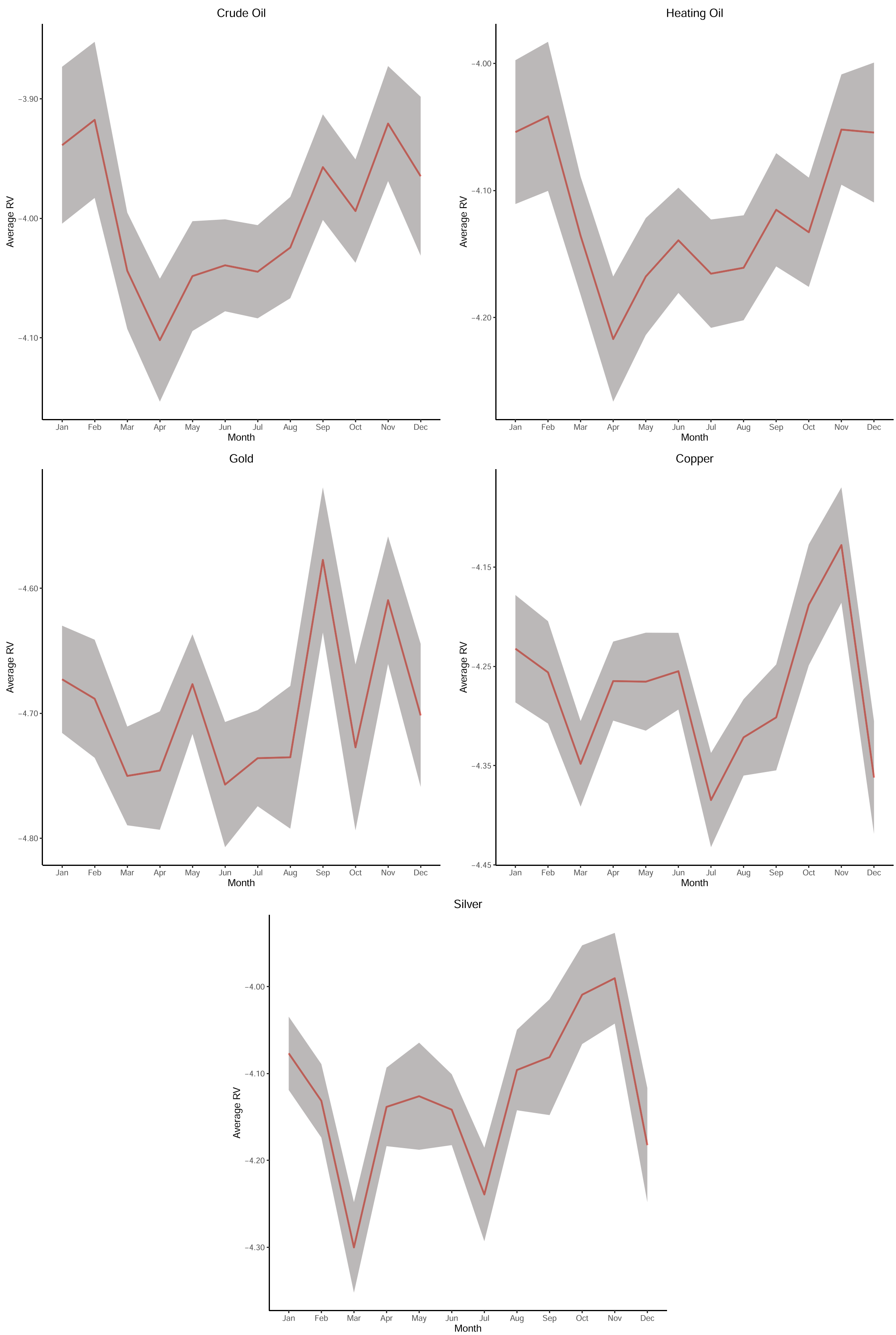

FIGURE A1 Unconditional monthly average RV. These plots display the monthly average of daily RV for the nearby futures contract written on crude oil, heating oil, gold, copper, and silver. The red line represents the average and the gray area represents the 90% confidence bands. The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755. RV, realized volatility [Color figure can be viewed at wileyonlinelibrary.com]

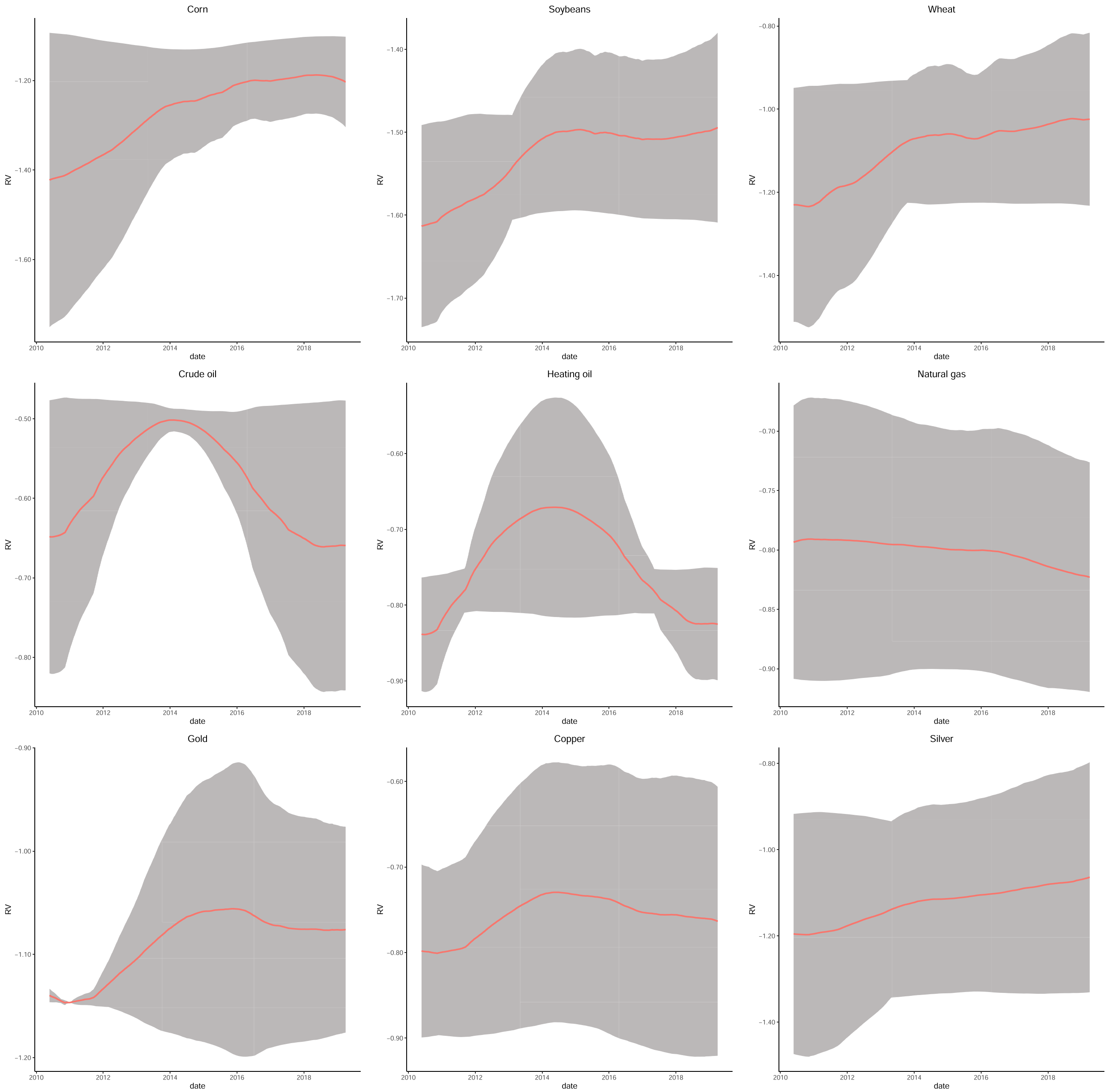

FIGURE A2 Time-varying intercepts in the EVHAR-TV model. This plot displays the pattern of time-varying intercepts in the joint estimation of the EVHAR-TV model (SUR). The intercept unit is RV for the nine nearby futures contracts. The sample period is May 5, 2010–April 1, 2019. The 90% confidence intervals are displayed with the shaded area and the number of observations per contract is 2755 [Color figure can be viewed at wileyonlinelibrary.com]