

DO JUMPS MATTER FOR VOLATILITY FORECASTING? EVIDENCE FROM ENERGY MARKETS

MARCEL PROKOPCZUK, LAZAROS SYMEONIDIS* and CHARDIN WESE SIMEN

This paper characterizes the dynamics of jumps and analyzes their importance for volatility forecasting. Using high-frequency data on four prominent energy markets, we perform a model-free decomposition of realized variance into its continuous and discontinuous components. We find strong evidence of jumps in energy markets between 2007 and 2012. We then investigate the importance of jumps for volatility forecasting. To this end, we estimate and analyze the predictive ability of several Heterogenous Autoregressive (HAR) models that explicitly capture the dynamics of jumps. Conducting extensive in-sample and out-of-sample analyses, we establish that explicitly modeling jumps does not significantly improve forecast accuracy. Our results are broadly consistent across our four energy markets, forecasting horizons, and loss functions. © 2015 Wiley Periodicals, Inc. Jrl Fut Mark

1. INTRODUCTION

The theory of quadratic variation (QV) posits that the total variation of an asset return can be decomposed into continuous and discontinuous components. The aim of this paper is to advance our understanding of the dynamics of each of these components and investigate their importance for volatility forecasting.

We make three important contributions to the literature. First, we identify and characterize the dynamics of jumps in four leading energy markets, namely crude oil, gasoline, heating oil, and natural gas. Using intraday transaction prices, we implement a non-parametric jump detection test to identify jumps. We then rigorously analyze the time-series behavior of jumps, thus shedding more light on their dynamics. Our analysis shows that jumps are rare events that affect only a small proportion of our sample. Moreover, we find important asymmetries in the intensity of positive and negative jumps, suggesting that it may be important to separately model the dynamics of positive and negative jumps.

Second, we investigate the importance of disentangling continuous volatility from jumps for volatility forecasting. We present and thoroughly assess the predictive ability of several

Marcel Prokopczuk is at the Leibniz University Hannover, Koenigsworther Platz 1, 30167 Hannover, Germany. He is also at ICMA Centre, Henley Business School, University of Reading, RG6 6BA, United Kingdom. Lazaros Symeonidis is at the Norwich Business School, University of East Anglia, Norwich NR4 7TJ, United Kingdom. Chardin Wese Simen is at the University of Liverpool, Chatham Street, Liverpool, L69 7ZH, United Kingdom. We thank Chris Brooks, Michael Clements, Richard Harris, an anonymous reviewer, participants at the 2015 FMA European Conference (Venice), and the 14th Conference on Research on Economic Theory and Econometrics (Chania) for helpful comments.

JEL Classification: C1, C53, C58, G1, G13

*Correspondence author, Norwich Business School, University of East Anglia, Norwich NR4 7TJ, United Kingdom. Tel: +44 1603 597332, Fax: +44 1603 593343, e-mail: l.symeonidis@uea.ac.uk

Received December 2014; Accepted August 2015

models of the Heterogenous Autoregressive (HAR) family that seek to explicitly capture the dynamics of jumps.¹ We begin by analyzing the in-sample predictive power of all models, which we compare to our benchmark HAR–RV model, which does not account for the impact of jumps. To this end, we perform simple ordinary least squares (OLS) regressions of realized volatility on the in-sample forecasts obtained from the various HAR models. We find that all models yield adjusted R^2 that are very close to each other, indicating that the benefits of explicitly modeling jumps are likely to be small. This is true for all forecast horizons, i.e., 1-, 5-, and 22-day ahead.

Third, we go beyond the in-sample analysis and rigorously analyze the out-of-sample performance of competing models. We use a rolling window of 600 days to estimate the parameters of forecasting models. Equipped with these, we then forecast the volatility of the next period, which we compare to realized volatility (observed ex-post). We employ six distinct loss functions to analyze the accuracy of competing forecasts. Our results establish that all models yield forecast errors that are of the same order of magnitude, indicating that explicitly modeling jumps does not noticeably improve forecast accuracy. We analyze the question of statistical significance by implementing the test of Giacomini and White (2006), hereafter GW. This analysis suggests that explicitly modeling jumps does not significantly improve forecast accuracy. This result generally holds for all horizons, loss functions, and markets.

We conduct several robustness checks. First, one may wonder whether our conclusions change depending on whether we forecast variance and log variance rather than volatility. Focusing on the task of forecasting realized variance and log realized variance, we repeat our analyses and obtain broadly similar conclusions. Second, we investigate the robustness of our findings to the jump detection methodology. In particular, we draw on recent theoretical results by Andersen et al. (2012), who introduce novel jump–robust estimators of integrated variance. Repeating our analysis with the new estimator does not change our main insights. Third, one may argue that the width of the rolling window may impact our analysis. We consider alternative windows of 400, 800, and 1,000 observations. Our core message is the same: explicitly modeling jumps does not significantly improve volatility forecasts. Finally, we assess the robustness of our results to the estimation methodology. Because our dependent variable, i.e., volatility, varies substantially over time, our OLS estimates may be driven by a highly volatile pocket of data. To address this concern, we repeat our analyses by estimating all models using a **weighted least squares (WLS)** approach (rather than OLS) and reach similar conclusions.

Our study relates to the literature on the econometrics of jumps. Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen et al. (2004), Barndorff-Nielsen and Shephard (2006), Tauchen and Zhou (2011), and Andersen et al. (2012) propose a number of non-parametric tests to identify jumps. Eraker et al. (2003) and Eraker (2004) rely on tightly parameterized continuous-time models to estimate jumps. Aït-Sahalia et al. (2014) and Maneesoonthorn et al. (2014) model jumps as processes that are self-exciting and explore the implications of this modeling framework for derivatives prices.² We contribute to this literature by presenting a thorough and comprehensive model-free study on the dynamics of jumps in four leading energy markets. Furthermore, our finding that explicitly modeling jumps does not significantly improve volatility forecasts may have important implications for continuous-time models that are needed for energy prices. If jumps are self-exciting then historical jump

¹ See Corsi (2009) for an excellent treatment of the HAR model.

² By “self-exciting” jumps, the financial modeling literature typically refers to the clustering of extreme events: a jump event tends to trigger another jump event in the same market.

returns should contain information about the discontinuous component of the realized variation of asset prices. Hence, we would expect to obtain more accurate volatility forecasts by explicitly modeling the dynamics of jumps. Our results reveal that this is not the case, thus indicating that models with self-exciting jumps are unlikely to successfully match the dynamics of energy prices.

Our paper also connects with the growing literature that uses intraday data to obtain more accurate volatility forecasts (Andersen and Bollerslev, 1998; Areal and Taylor, 2002; Martens and Zein, 2004; Giot and Laurent, 2007; Corsi and Reno, 2009; Corsi et al., 2010; Chung et al., 2011; Patton and Sheppard, 2015; Sévi, 2014). Our paper relates to the important works of Andersen et al. (2007) and Andersen et al. (2011), who study the importance of jumps for volatility forecasting in the equity, fixed-income, and foreign exchange markets. We complement these studies by presenting fresh evidence from energy markets. Different from the evidence from the other asset classes, we find that modeling jumps does not noticeably improve the accuracy of volatility forecasts in commodity markets. Thus, our results suggest that findings from other asset classes do not necessarily extend to commodities. We conjecture that the different results may be due to the underlying determinants of jumps in commodity prices. It is possible that the unpredictable nature of important events, such as political unrest in major oil producing countries and natural disasters, that trigger jumps in energy prices explains why past jumps are not necessarily informative about future volatility.

The remainder of the paper is organized as follows. Section 2 introduces our methodology and the dataset. Section 3 presents our empirical results. Section 4 discusses various robustness checks. Finally, Section 5 concludes.

2. METHODOLOGY AND DATA

This section begins with a brief overview of jump detection tests. We then introduce the competing models. Finally, we present our dataset of intraday transaction prices.

2.1. Jump Detection Test

Consider the logarithmic price process, p_t , defined on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, P)$, where \mathcal{F}_t is the information set available up to time t , such that p_t is \mathcal{F}_t -measurable and evolves in continuous-time as a jump-diffusion process:

$$dp_t = \mu_t dt + \sigma_t dW_t + \eta_t dN_t \quad (1)$$

where dp_t denotes the change in log price. μ_t is the drift, which is a locally bounded and predictable process of finite variance. dt is an increment of time. σ_t is the instantaneous (or spot) volatility, which is a càdlàg process. W_t refers to the Brownian motion. η_t is a random variable capturing the jump size. Finally, N_t is a Poisson jump process. If a jump occurs during the increment dt , then $dN_t = 1$. Otherwise, $dN_t = 0$. The probability of a jump occurring in the time interval dt is $P[dN_t = 1] = \lambda_t dt$, where λ_t is the (time-varying) jump intensity.

The quadratic variation, QV_t , of the above return process can then be expressed as the sum of a continuous and a discontinuous component (Barndorff-Nielsen and Shephard, 2004, 2006). More formally, we have the following:

$$QV_t = \underbrace{\int_{t-1}^t \sigma_s^2 ds}_{\text{continuous}} + \underbrace{\sum_{t-1 \leq \tau_i \leq t} \eta_{\tau_i}^2}_{\text{discontinuous}} \quad (2)$$

4 Prokopczuk, Symeonidis, and Wese Simen

where τ_i are the times corresponding to jump occurrences (with $i = 1, 2, \dots, N_t$) and all other variables are as previously defined. The first term on the right hand side of the equality sign is the “integrated variance”; it is the continuous component of the quadratic variation. The second term is the discontinuous component of the quadratic variation.

We now explain in detail how to empirically compute each of the quantities shown in Equation (2). Suppose that on a trading day t , we observe $M + 1$ prices at times: t_0, t_1, \dots, t_M . If p_{t_j} is the logarithmic price at time t_j , then the corresponding return, r_{t_j} , for the j^{th} intraday interval of day t is defined as follows:

$$r_{t_j} = p_{t_j} - p_{t_{j-1}} \quad (3)$$

Andersen et al. (2001, 2003) and Barndorff-Nielsen and Shephard (2002a,b) propose the “realized variance” estimator, defined as the sum of squared intraday returns:

$$RV_t = \sum_{j=1}^M r_{t_j}^2 \quad (4)$$

The authors show that, as the sampling frequency increases ($M \rightarrow \infty$), realized variance is a consistent estimator of daily quadratic variation:

$$\lim_{M \rightarrow \infty} RV_t \equiv QV_t \quad (5)$$

Barndorff-Nielsen and Shephard (2004, 2006) introduce the “bipower variation” (BPV), which is a consistent estimator of the continuous component of QV_t :³

$$BPV_t = \mu_1^{-2} \left(\frac{M}{M-2} \right) \sum_{j=2}^M |r_{t_{j-1}}| \cdot |r_{t_j}| \quad (6)$$

where $\mu_1 = \mathbb{E}(|Z|) = \sqrt{2/\pi}$ is the first moment of the absolute value of a standard normal random variable. The term $M/(M-2)$ corresponds to a finite sample bias correction. BPV_t consistently estimates the continuous sample path of quadratic variation as $M \rightarrow \infty$ (Barndorff-Nielsen and Shephard, 2004, 2006; Barndorff-Nielsen et al., 2006):

$$\lim_{M \rightarrow \infty} BPV_t \equiv \int_{t-1}^t \sigma_s^2 ds \quad (7)$$

where all variables are as previously defined.

Huang and Tauchen (2005), Andersen et al. (2007), and others document that using staggered returns alleviates many of the microstructure biases inherent in high-frequency data. Instead of computing the product of adjacent returns, i.e., $|r_{t_{j-1}}| \cdot |r_{t_j}|$, they consider $|r_{t_{j-(k+1)}}| \cdot |r_{t_j}|$, where k is a positive integer indicating the number of returns to skip. When staggered returns are employed, the realized bipower variation of Equation (6) is modified

³Andersen et al. (2012) point out that because of microstructure effects, the contribution of jumps does not vanish asymptotically, leading to an upward bias in the BPV estimator. We address this issue in our robustness analysis by replacing the BPV estimator with the MedRV estimator based on nearest neighbor truncation (Andersen et al., 2012).

as follows:

$$\text{BPV}_t = \mu_1^{-2} \left(\frac{M}{M - (k + 1)} \right) \sum_{j=k+2}^M |r_{tj-(k+1)}| \cdot |r_{tj}| \quad (8)$$

Throughout our study, we work with staggered returns to allay concerns related to the microstructure noise. We always skip 1 return observation, i.e., setting $k = 1$.

Since the quadratic variation is the sum of continuous and discontinuous components (see Equation (2)), one can express the discontinuous component as the difference between the quadratic variation and the continuous component. A direct implication of this is that we can infer the discontinuous component from the realized variance and the bipower variation:

$$\text{RV}_t - \text{BPV}_t \xrightarrow{p} \sum_{t-1 \leq \tau_i \leq t} \eta_{\tau_i}^2 \quad (9)$$

This intuition lies at the heart of most jump detection tests. Huang and Tauchen (2005) show that the ratio statistic with a maximum adjustment (Barndorff-Nielsen and Shephard, 2006) has good size and power properties. This result motivates us to use the following test to detect significant jumps:

$$z_{\text{TQ},t} = \Delta^{-1/2} \frac{(\text{RV}_t - \text{BPV}_t) / \text{RV}_t}{\sqrt{((\frac{\pi}{2})^2 + \pi - 5) \max(1, \frac{\text{TQ}_t}{\text{BPV}_t^2)}}} \quad (10)$$

TQ_t above is the realized tripower quarticity:

$$\text{TQ}_t = M \left(\frac{M}{M - 2(k + 1)} \right) \mu_{4/3}^{-3} \sum_{j=3}^M |r_{tj-2(k+1)}|^{4/3} \cdot |r_{tj-(k+1)}|^{4/3} \cdot |r_{tj}|^{4/3} \quad (11)$$

where $\mu_{4/3} = 2^{2/3} [\Gamma(7/6) / \Gamma(1/2)]$.

Using the test statistic of Equation (10) and a significance level α , which we set equal to 0.1%, we extract the significant jumps, J_t , as follows:⁴

$$J_t = \mathbb{I}_{\{z_{\text{TQ},t} > \Phi_{1-\alpha}\}} \cdot (\text{RV}_t - \text{BPV}_t) \quad (12)$$

where: $\mathbb{I}_{\{z_{\text{TQ},t} > \Phi_{1-\alpha}\}}$ is an indicator function which is equal to 1 when a significant jump occurs and zero otherwise, $\Phi_{1-\alpha}$ is the corresponding critical value from the cumulative standard normal distribution at confidence level $1 - \alpha$. Since RV_t is equal to the sum of the continuous component (C_t) plus jumps (J_t), the continuous path of realized variance can be identified as follows:

$$C_t = \mathbb{I}_{\{z_{\text{TQ},t} > \Phi_{1-\alpha}\}} \cdot \text{BPV}_t + \mathbb{I}_{\{z_{\text{TQ},t} \leq \Phi_{1-\alpha}\}} \cdot \text{RV}_t \quad (13)$$

where all variables are as previously defined.

⁴This stringent choice is mainly motivated by the theoretical results of Bajgrowicz et al. (2014), who forcefully show that multiple testing issues could result in spurious jumps. The authors recommend using stringent significance levels such as 0.1% to allay concerns that the results of jump tests may be driven by any false positives. We follow their recommendation. As a further robustness check, we also consider a significance level of 1% and reach very similar conclusions (see Tables A.1–A.5 of the online Appendix). We are very grateful to the reviewer for suggesting this analysis.

2.2. Volatility Forecasting Models

1. **HAR–RV**: Our benchmark econometric model is the HAR–RV recently implemented in Patton and Sheppard (2015). The simple structure of this model enables it to parsimoniously capture the long-memory behavior of realized volatility. This is achieved by combining historical estimates of realized volatility computed over various non-overlapping horizons. Patton and Sheppard (2015) emphasize that non-overlapping horizons are important to: (i) allay any concerns about the strong correlation between the components of the model and (ii) facilitate the interpretation of the coefficient estimates. We follow their recommendation and estimate the following volatility forecasting model:^{5,6}

$$RV_{t:t+h}^{1/2} = \omega + \beta_d RV_t^{1/2} + \beta_w RV_{t-5:t-1}^{1/2} + \beta_m RV_{t-22:t-5}^{1/2} + e_{t+h} \quad (14)$$

As mentioned above, each component in the HAR–RV model is computed over different horizons. Therefore, if RV_t is the realized variance of day t (from time $t - 1$ to t), then the h -day annualized realized variance is expressed as follows:⁷

$$RV_{t:t+h} = \frac{252}{h} (RV_{t+1} + RV_{t+2} + \dots + RV_{t+h}) \quad (15)$$

In a similar manner, the weekly and monthly components are computed as follows:

$$RV_{t-5:t-1} = \frac{252}{4} \sum_{i=2}^5 RV_{t-i+1} \quad (16)$$

$$RV_{t-22:t-5} = \frac{252}{17} \sum_{i=6}^{22} RV_{t-i+1} \quad (17)$$

2. **HAR–J**: Andersen et al. (2007) propose the HAR–J, which is a simple extension of the HAR–RV model that seeks to capture the dynamics of jumps. The main feature of the HAR–J model is that it replaces the most recent realized volatility ($RV_{t-1}^{1/2}$) with two components: $C_{t-1}^{1/2}$ and $J_{t-1}^{1/2}$. Each of these components has its own coefficient estimate:

$$RV_{t:t+h}^{1/2} = \omega + \beta_d C_t^{1/2} + \beta_w RV_{t-5:t-1}^{1/2} + \beta_m RV_{t-22:t-5}^{1/2} + \gamma J_t^{1/2} + e_{t+h} \quad (18)$$

where all variables are as previously defined.

3. **HAR–RJ**: The previous model can be criticized on the grounds that it ignores the sign of jumps. The HAR–RJ addresses this limitation (Tauchen and Zhou, 2011). We identify significant realized jumps as follows:

$$RJ_t = \text{sign}(r_t) \cdot \sqrt{J_t} \quad (19)$$

⁵We focus on the task of forecasting volatility (rather than variance) because volatility plays a key role in modern finance theory. For instance, it is a key variable for option pricing and asset allocation. We also consider the task of forecasting variance and log variance. See Section 4 for further results.

⁶Strictly speaking, the daily realized volatility should be written as $RV_{t-1:t}$. However, to simplify our notation, we write it as RV_t .

⁷Similar to Busch et al. (2011), among others, we focus on *annualized* variance. This explains the presence of the factor 252 in Equation (15). Naturally, this multiplicative factor does not affect the statistical and economic interpretation of our findings.

where RJ_t is the realized jump on day t , r_t is the daily return and $\text{sign}(\cdot)$ is the sign operator. The HAR–RJ model is then defined as follows:

$$RV_{t:t+h}^{1/2} = \omega + \beta_d C_t^{1/2} + \beta_w RV_{t-5:t-1}^{1/2} + \beta_m RV_{t-22:t-5}^{1/2} + \gamma_{RJ} RJ_t + e_{t+h} \quad (20)$$

where all components are as previously defined.⁸

4. **HAR–ARJ:** It may be that positive and negative observations of RJ exert an asymmetric impact on volatility. As a result, it is interesting to investigate which of positive and negative jumps is more important for volatility forecasting. We advance in this direction by further decomposing RJ_t into components due to positive and negative jumps:

$$RJ_t^+ = \max(RJ_t; 0) \quad (21)$$

$$RJ_t^- = \min(RJ_t; 0) \quad (22)$$

The HAR–ARJ specification is employed to test whether the variation from negative jumps has a more pronounced impact on future volatility than that of positive jumps:

$$RV_{t:t+h}^{1/2} = \omega + \beta_d C_t^{1/2} + \beta_w RV_{t-5:t-1}^{1/2} + \beta_m RV_{t-22:t-5}^{1/2} + \gamma_{RJ^+} RJ_t^+ + \gamma_{RJ^-} RJ_t^- + e_{t+h} \quad (23)$$

5. **HAR–C–J:** Finally, we consider a more general specification, similar to that of Andersen et al. (2007), which fully decomposes each realized variance component (i.e., daily, weekly, monthly) into its continuous and jump parts:

$$RV_{t:t+h}^{1/2} = \omega + \beta_d C_t^{1/2} + \beta_w C_{t-5:t-1}^{1/2} + \beta_m C_{t-22:t-5}^{1/2} + \gamma_{Jd} J_t^{1/2} + \gamma_{Jw} J_{t-5:t-1}^{1/2} + \gamma_{Jm} J_{t-22:t-5}^{1/2} + e_{t+h} \quad (24)$$

2.3. Data

Our dataset consists of tick-by-tick transaction prices on four energy futures contracts traded at NYMEX, namely WTI crude oil, gasoline (RBOB), heating oil, and natural gas. The data come from TickData and spans the period from January 2, 2007 to June 29, 2012.⁹ Energy futures contracts trade on two venues: pit and electronic. Trading hours on both platforms have no overlap and collectively span 22:45 hours. Pit trading takes place between 9:30 AM (ET) and 4:15 PM (ET). Electronic trading starts at 4:30 PM (ET), pauses at 5:15 PM (ET) for 45 minutes, resumes at 6:00 PM (ET), and stops the following day at 9:15 AM (ET).

We use both pit and electronic transaction records and process the dataset as follows. First, we discard all transactions with prices lower than or equal to zero. Second, we expunge all trades with time-stamps that are inconsistent with the exchange's trading hours. Third,

⁸Notice that the superscript 1/2 is omitted from RJ_t , since it is already expressed in volatility form (see Equation (19)).

⁹The Gasoline RBOB futures contracts started trading in October 2005. This means that we can only have a common sample period from that point onwards. While we could consider the sample period from 2005 onwards, we feel that it is important to allow for about a year to elapse to ensure that the gasoline futures contracts are actively traded.

TABLE I
Summary Statistics for Variation Measures

	RV	BPV	J	\sqrt{RV}	\sqrt{BPV}	\sqrt{J}
<i>A. Crude Oil</i>						
Mean	0.150	0.137	0.092	0.344	0.328	0.250
St. dev.	0.200	0.185	0.156	0.178	0.171	0.169
Skewness	3.885	3.770	3.751	2.187	2.224	2.138
Kurtosis	23.395	20.217	19.279	8.786	8.875	8.132
Min	0.010	0.008	0.004	0.101	0.091	0.059
Max	2.288	1.655	1.027	1.513	1.286	1.014
<i>B. Heating Oil</i>						
Mean	0.114	0.098	0.069	0.307	0.285	0.225
St. dev.	0.125	0.103	0.093	0.141	0.128	0.134
Skewness	3.230	2.802	2.806	1.732	1.615	1.532
Kurtosis	17.444	12.079	11.709	6.682	5.937	5.345
Min	0.006	0.005	0.004	0.079	0.070	0.064
Max	1.273	0.752	0.540	1.128	0.867	0.735
<i>C. Natural Gas</i>						
Mean	0.224	0.189	0.152	0.443	0.408	0.337
St. dev.	0.204	0.164	0.202	0.167	0.152	0.189
Skewness	4.264	4.227	3.515	1.669	1.433	1.820
Kurtosis	32.985	37.412	17.786	8.576	7.758	7.287
Min	0.013	0.002	0.005	0.116	0.047	0.070
Max	2.295	2.164	1.331	1.515	1.471	1.154
<i>D. Gasoline</i>						
Mean	0.151	0.123	0.109	0.343	0.311	0.271
St. dev.	0.221	0.169	0.209	0.183	0.160	0.188
Skewness	4.991	4.556	5.613	2.540	2.392	2.615
Kurtosis	37.737	30.949	42.919	11.745	10.714	12.661
Min	0.003	0.001	0.003	0.055	0.023	0.050
Max	2.559	1.831	1.946	1.600	1.353	1.395

Note. This table presents summary statistics of realized variance (RV), bipower variation (BPV), significant jumps (J), and their square root (volatility) counterparts (\sqrt{RV} , \sqrt{BPV} , \sqrt{J}). Panels A–D report statistics for crude oil, heating oil, natural gas, and gasoline, respectively. The dataset covers the period from January 2, 2007 to June 29, 2012. Daily variance (volatility) series are computed using 15-min returns and are annualized by multiplying by (the square root of) 252. Each panel reports the mean, standard deviation, skewness, kurtosis, minimum, and maximum, respectively.

we retain the futures contract with the highest number of transactions only (usually the first or second nearest contract). Following existing studies, e.g., Lee and Mykland (2008) and Bradley et al. (2014), we sample our data at the 15-min frequency.¹⁰

Table I presents summary statistics for the different (annualized) measures of variance. Columns 2–4 relate to realized variance (RV), bipower variation (BPV), and significant jumps (J), i.e., RV – BPV , respectively. Columns 5–7 present results for the square root of RV , BPV , and J , respectively. A comparison of \sqrt{RV} across the four energy futures markets reveals that on average natural gas exhibits the highest volatility (44.3% per year), followed by crude oil (34.4%), gasoline (34.3%), and heating oil (30.7%). These numbers are broadly consistent with the results of Thomakos and Wang (2003) and Wang et al. (2008).

¹⁰We also analyze the volatility signature plot, which plots realized volatility as a function of sampling frequency (Andersen et al., 1999). The plots support the choice of the 15-min sampling frequency.

TABLE II
Summary Statistics of Significant Daily Jumps

	Crude Oil	Heating Oil	Natural Gas	Gasoline
#Jumps	90	122	142	163
#Positive Jumps	58	65	61	88
#Negative Jumps	32	57	81	75
Intensity	0.063	0.085	0.100	0.115
Intensity ⁺	0.041	0.045	0.043	0.062
Intensity ⁻	0.022	0.040	0.057	0.053
Mean	0.250	0.225	0.337	0.271
Mean ⁺	0.254	0.227	0.338	0.281
Mean ⁻	-0.243	-0.225	-0.337	-0.260
St. Dev.	0.169	0.134	0.189	0.188
St. Dev. ⁺	0.180	0.142	0.224	0.190
St. Dev. ⁻	0.149	0.127	0.160	0.186

Note. This table presents summary statistics for the series of significant daily jumps. The first row of the table presents the total number of jump days. We detect statistically significant jumps by using the $z_{TQ,t}$ statistic shown in Equation (10) and a confidence level of 99.9%. The second and third rows show the total number of jumps positive and negative jumps, respectively. The row labeled “Intensity” shows the jump intensity, that is the ratio of jump days over the total number of days. The next two rows further decompose jump intensity into its positive and negative parts following Tauchen and Zhou (2011). The table also reports the mean (“Mean”) and standard deviation (“St.Dev.”) of the series of significant jumps, as well as the corresponding statistics for positive and negative jumps.

3. EMPIRICAL RESULTS

This section presents our main results. We begin by characterizing the dynamics of jumps. Next, we compare the predictive ability of these models in an in-sample setting. Finally, we present a comprehensive and rigorous analysis of the performance of these models.

3.1. The Dynamics of Jumps

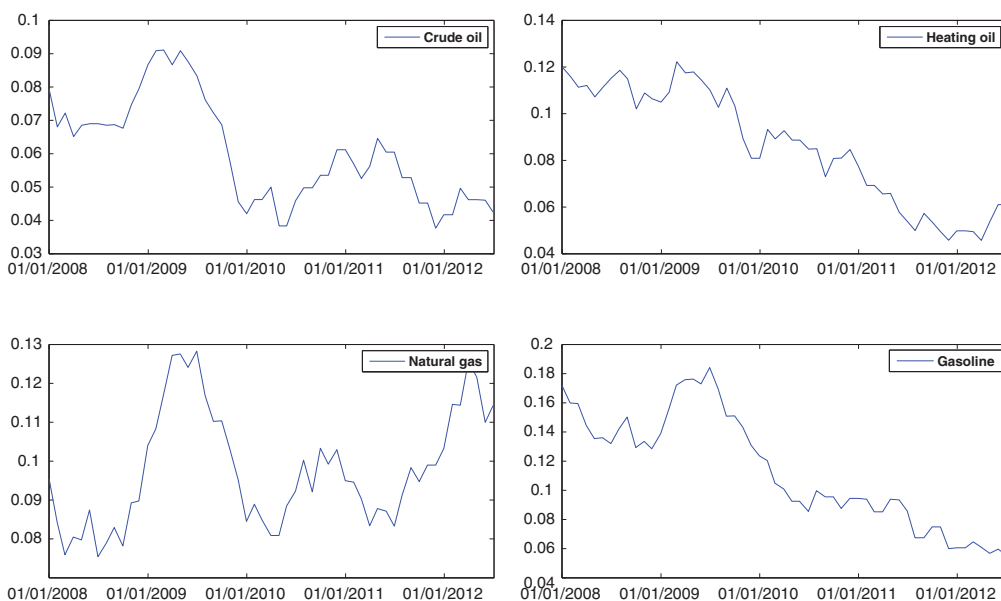
Table II displays summary statistics of significant jumps. We use a conservative significance level of 0.1%. We observe that the proportion of jump days (*Intensity*) is highest for gasoline (11.5%), followed by natural gas (10%), heating oil (8.5%), and crude oil (6.3%).

Following Andersen et al. (2007) and Tauchen and Zhou (2011), we compute the geometric average of monthly jump intensity to obtain a smoothed time-series. We define the jump intensity of a specific month as the number of jump days in that month over the total number of trading days in that particular month. Figure 1 reveals important time variations in the intensity of jumps, which peaks between 2008 and 2009. We also observe interesting differences across markets. While the jump intensities of heating oil and gasoline both steadily decline post-2010, the jump intensity of crude oil displays much more variation.

The second and third rows of Table II present some evidence of asymmetries in the time-series of jumps. This is particularly visible by looking at the intensities of positive and negative jumps, reported under *Intensity*⁺ and *Intensity*⁻, respectively. For example, the proportion of positive jumps (4.1%) is almost twice as high as that of negative jumps in the crude oil market. Another interesting observation is that the average positive jump return (*Mean*⁺) is very similar in magnitude to that of negative jumps (*Mean*⁻). Remarkably, this pattern holds for all markets.

3.2. In-Sample Analysis

We begin by analyzing the in-sample predictive power of the competing models introduced in the previous section. To this end, we use all daily observations to estimate the models using

**FIGURE 1**

Time-varying jump intensity

Note. This figure presents the series of monthly jump intensities for the four energy markets. Monthly jump intensity is the ratio of the number of days associated with jumps in a given month over the total number of days in that particular month. The series of monthly intensities are smoothed by taking a rolling 12-month geometric average. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

OLS.¹¹ We consider three forecasting horizons, namely 1-, 5-, and 22-day. Tables III–VI report our results. We report in brackets, the Newey–West corrected t-statistics with 5, 10, and 44 lags for the 1, 5, and 22 day forecasting horizons, respectively. We highlight in bold, all significant estimates at the 5% level. Looking first at the benchmark HAR–RV model, we see that this model predicts realized volatility with adjusted R^2 up to 76% in the crude oil market.

Turning to the HAR–J specification, we observe that the coefficient of the jump component is generally positive and statistically significant. This is true for all forecast horizons. This result indicates that volatility increases following a jump event. The magnitude of the jump component differs across markets. In particular, the coefficient estimate of the jump component at the 22-day horizon takes the value 0.098, 0.074, and 0.038 in the crude oil, heating oil, and natural gas markets, respectively.

Focusing on the HAR–ARJ model, we see that the negative jump component generally dominates its positive counterpart. This is true in terms of both economic magnitude and statistical significance. Interestingly, the negative jump component enters the regression with a negative loading, indicating that negative jumps predict increases in future volatility. These results generally hold across all horizons.

The finding that only the negative jump component of the HAR–ARJ model is statistically significant helps understand why the HAR–RJ typically yields an insignificant jump

¹¹One may argue that volatility itself fluctuates significantly over time, raising concerns that the OLS estimation may be driven by a small pocket of data. To address this concern, we also use a weighted least squares (WLS) estimation. See Section 4 for further details.

TABLE III
In-Sample Predictability: Crude Oil

	β_d	β_w	β_m	β_{C_d}	β_{C_w}	β_{C_m}	γ_{J_d}	γ_{J_w}	γ_{J_m}	γ_{RJ}	γ_{RJ+}	γ_{RJ-}	\bar{R}^2	Obs.
<i>Panel A: 1-Day Horizon</i>														
HAR-RV	0.334 (7.115)	0.396 (7.560)	0.226 (4.561)	—	—	—	—	—	—	—	—	—	0.758	1392
HAR-J	—	0.397 (7.396)	0.230 (4.506)	0.334 (6.222)	—	—	0.053 (0.916)	—	—	—	—	—	0.757	1392
HAR-RJ	—	0.397 (7.401)	0.234 (4.362)	0.331 (6.055)	—	—	—	—	—	0.037 (0.780)	—	—	0.757	1392
HAR-ARJ	—	0.396 (7.377)	0.231 (4.480)	0.334 (6.174)	—	—	—	—	—	—	0.069 (0.925)	−0.022 (−0.359)	0.757	1392
HAR-C-J	—	—	—	0.326 (6.164)	0.409 (7.531)	0.220 (4.194)	0.053 (0.935)	0.041 (1.905)	0.018 (1.361)	—	—	—	0.759	1392
<i>Panel B: 5-Day Horizon</i>														
HAR-RV	0.264 (7.896)	0.431 (8.683)	0.257 (5.446)	—	—	—	—	—	—	—	—	—	0.852	1388
HAR-J	—	0.433 (8.635)	0.258 (5.416)	0.261 (6.846)	—	—	0.080 (2.178)	—	—	—	—	—	0.851	1388
HAR-RJ	—	0.437 (8.580)	0.263 (5.275)	0.256 (6.737)	—	—	—	—	—	−0.031 (−1.207)	—	—	0.850	1388
HAR-ARJ	—	0.435 (8.638)	0.256 (5.384)	0.261 (6.946)	—	—	—	—	—	—	0.039 (1.103)	− 0.160 (−2.610)	0.852	1388
HAR-C-J	—	—	—	0.250 (6.780)	0.446 (8.682)	0.245 (4.905)	0.082 (2.133)	0.038 (2.011)	0.030 (1.665)	—	—	—	0.854	1388

Continue

TABLE III
Continued

	β_d	β_w	β_m	β_{C_d}	β_{C_w}	β_{C_m}	γ_{J_d}	γ_{J_w}	γ_{J_m}	γ_{RJ}	γ_{RJ^+}	γ_{RJ^-}	\bar{R}^2	<i>Obs.</i>
	<i>Panel C: 22-Day Horizon</i>													
HAR-RV	0.233 (7.501)	0.395 (4.604)	0.270 (3.559)	—	—	—	—	—	—	—	—	—	0.817	1371
HAR-J	—	0.399 (4.577)	0.271 (3.587)	0.227 (7.030)	—	—	0.098 (4.027)	—	—	—	—	—	0.817	1371
HAR-RJ	—	0.403 (4.538)	0.278 (3.597)	0.222 (7.092)	—	—	—	—	—	−0.010 (−0.567)	—	—	0.815	1371
HAR-ARJ	—	0.400 (4.554)	0.270 (3.544)	0.228 (7.054)	—	—	—	—	—	—	0.068 (3.156)	−0.154 (−3.365)	0.817	1371
HAR-C-J	—	—	—	0.213 (7.461)	0.401 (4.662)	0.239 (3.131)	0.097 (4.459)	0.071 (2.382)	0.064 (1.671)	—	—	—	0.825	1371

Note. This table assesses the predictive ability of several models in the crude oil market. Each of the three panels shows results for a different forecasting horizon (1-, 5-, and 22-day horizon). The jump components are computed based on Equation (12) using the test statistic of Equation (10) and a significance level of 0.1%. All regressions are estimated using Newey–West (1987) corrected standard errors with 5, 10, and 44 lags for the 1-, 5-, and 22-day horizon, respectively. T-statistics are reported in parentheses. The intercepts are not reported to save space. Significant coefficients at the 5% level are highlighted in bold. The second last column reports the adjusted \bar{R}^2 of each regression. The last column shows the number of observations. The sample period is from January 2, 2007 to June 29, 2012.

TABLE IV
In-Sample Predictability: Heating Oil

	β_d	β_w	β_m	β_{Cd}	β_{Cw}	β_{Cm}	γ_{Jd}	γ_{Jw}	γ_{Jm}	γ_{RJ}	γ_{RJ+}	γ_{RJ-}	\bar{R}^2	Obs.
<i>Panel A: 1-Day Horizon</i>														
HAR-RV	0.296 (5.952)	0.394 (7.608)	0.265 (5.558)	—	—	—	—	—	—	—	—	—	0.711	1395
HAR-J	—	0.396 (7.265)	0.267 (5.492)	0.294 (5.137)	—	—	0.097 (2.590)	—	—	—	—	—	0.710	1395
HAR-RJ	—	0.406 (7.359)	0.276 (5.422)	0.283 (4.825)	—	—	—	—	—	0.018 (0.443)	—	—	0.708	1395
HAR-ARJ	—	0.396 (7.285)	0.267 (5.508)	0.294 (5.117)	—	—	—	—	—	—	0.107 (2.151)	−0.086 (−1.462)	0.710	1395
HAR-C-J	—	—	—	0.291 (5.168)	0.399 (7.768)	0.250 (5.057)	0.096 (2.591)	0.039 (1.866)	0.027 (1.758)	—	—	—	0.710	1395
<i>Panel B: 5-Day Horizon</i>														
HAR-RV	0.240 (8.999)	0.377 (7.694)	0.331 (7.460)	—	—	—	—	—	—	—	—	—	0.823	1391
HAR-J	—	0.377 (7.508)	0.332 (7.478)	0.239 (7.616)	—	—	0.085 (3.444)	—	—	—	—	—	0.823	1391
HAR-RJ	—	0.385 (7.558)	0.341 (7.513)	0.230 (7.040)	—	—	—	—	—	−0.020 (−0.719)	—	—	0.821	1391
HAR-ARJ	—	0.376 (7.549)	0.332 (7.534)	0.240 (7.729)	—	—	—	—	—	—	0.061 (2.184)	− 0.114 (−2.887)	0.823	1391
HAR-C-J	—	—	—	0.233 (7.558)	0.395 (7.822)	0.303 (6.496)	0.083 (3.261)	0.026 (1.464)	0.033 (1.823)	—	—	—	0.825	1391

Continue

TABLE IV
Continued

	β_d	β_w	β_m	β_{C_d}	β_{C_w}	β_{C_m}	γ_{J_d}	γ_{J_w}	γ_{J_m}	γ_{RJ}	γ_{RJ^+}	γ_{RJ^-}	\bar{R}^2	Obs.
	Panel C: 22-Day Horizon													
HAR-RV	0.180 (9.561)	0.304 (5.443)	0.425 (6.593)	—	—	—	—	—	—	—	—	—	0.821	1374
HAR-J	—	0.306 (5.474)	0.425 (6.577)	0.178 (8.980)	—	—	0.074 (4.167)	—	—	—	—	—	0.821	1374
HAR-RJ	—	0.312 (5.364)	0.433 (6.531)	0.170 (8.806)	—	—	—	—	—	−0.034 (−0.966)	—	—	0.819	1374
HAR-ARJ	—	0.304 (5.433)	0.426 (6.566)	0.179 (9.123)	—	—	—	—	—	—	0.038 (1.521)	−0.118 (−2.543)	0.821	1374
HAR-C-J	—	—	—	0.173 (8.705)	0.307 (5.096)	0.405 (5.769)	0.072 (4.196)	0.041 (2.597)	0.035 (1.245)	—	—	—	0.823	1374

Note. This table assesses the predictive ability of several models in the heating oil market. Each of the three panels shows results for a different forecasting horizon (1-, 5-, and 22-day horizon). The jump components are computed based on Equation (12) using the test statistic of Equation (10) and a significance level of 0.1%. All regressions are estimated using Newey–West (1987) corrected standard errors with 5, 10, and 44 lags for the 1-, 5-, and 22-day horizon, respectively. T-statistics are reported in parentheses. The intercepts are not reported to save space. Significant coefficients at the 5% level are highlighted in bold. The second last column reports the adjusted \bar{R}^2 of each regression. The last column shows the number of observations. The sample period is from January 2, 2007 to June 29, 2012.

TABLE V
In-Sample Predictability: Natural Gas

	β_d	β_w	β_m	β_{C_d}	β_{C_w}	β_{C_m}	γ_{C_d}	γ_{C_w}	γ_{C_m}	γ_{J_d}	γ_{J_w}	γ_{J_m}	γ_{RJ}	γ_{RJ^+}	γ_{RJ^-}	\bar{R}^2	Obs.
Panel A: 1-Day Horizon																	
HAR-RV	0.253 (6.970)	0.438 (9.440)	0.174 (3.771)	—	—	—	—	—	—	—	—	—	—	—	—	0.444	1396
HAR-J	—	0.419 (8.988)	0.169 (3.686)	0.284 (6.494)	—	—	0.053 (1.990)	—	—	—	—	—	—	—	—	0.447	1396
HAR-RJ	—	0.418 (8.923)	0.171 (3.670)	0.286 (6.444)	—	—	—	—	—	—	—	—	—0.002 (−0.088)	—	—	0.446	1396
HAR-ARJ	—	0.419 (9.021)	0.169 (3.682)	0.284 (6.417)	—	—	—	—	—	—	—	—	—	0.054 (1.571)	−0.052 (−1.353)	0.447	1396
HAR-C-J	—	—	—	0.269 (6.334)	0.439 (8.434)	0.181 (3.632)	0.054 (2.029)	0.057 (2.998)	−0.009 (−0.873)	—	—	—	—	—	—	0.451	1396
Panel B: 5-Day Horizon																	
HAR-RV	0.237 (7.202)	0.400 (9.344)	0.200 (3.947)	—	—	—	—	—	—	—	—	—	—	—	—	0.602	1392
HAR-J	—	0.390 (8.971)	0.198 (3.927)	0.252 (6.669)	—	—	0.079 (3.005)	—	—	—	—	—	—	—	—	0.604	1392
HAR-RJ	—	0.388 (8.854)	0.201 (3.880)	0.256 (6.575)	—	—	—	—	—	—	—	—	−0.015 (−0.640)	—	—	0.600	1392
HAR-ARJ	—	0.390 (9.006)	0.198 (3.917)	0.253 (6.514)	—	—	—	—	—	—	—	—	—	0.068 (1.850)	−0.089 (−2.390)	0.604	1392
HAR-C-J	—	—	—	0.236 (6.664)	0.423 (7.902)	0.198 (3.655)	0.080 (2.993)	0.040 (2.001)	−0.004 (−0.282)	—	—	—	—	—	—	0.611	1392

Continue

TABLE V
Continued

	β_d	β_w	β_m	β_{C_d}	β_{C_w}	β_{C_m}	γ_{J_d}	γ_{J_w}	γ_{J_m}	γ_{RJ}	γ_{RJ}^+	γ_{RJ}^-	\bar{R}^2	$Obs.$
	<i>Panel C: 22-Day Horizon</i>													
HAR-RV	0.170 (8.527)	0.291 (8.189)	0.266 (3.075)	—	—	—	—	—	—	—	—	—	0.540	1375
HAR-J	—	0.278 (7.383)	0.263 (3.033)	0.191 (7.796)	—	—	0.038 (2.110)	—	—	—	—	—	0.542	1375
HAR-RJ	—	0.276 (7.210)	0.264 (3.019)	0.195 (7.526)	—	—	—	—	—	-0.043 (-2.218)	—	—	0.543	1375
HAR-ARJ	—	0.276 (7.333)	0.262 (3.034)	0.194 (7.626)	—	—	—	—	—	—	-0.006 (-0.249)	-0.076 (-2.414)	0.543	1375
HAR-C-J	—	—	—	0.177 (7.825)	0.315 (7.015)	0.245 (2.998)	0.038 (2.129)	0.022 (1.421)	0.011 (0.419)	—	—	—	0.547	1375

Note. This table assesses the predictive ability of several models in the natural gas market. Each of the three panels shows results for a different forecasting horizon (1-, 5-, and 22-day horizon). The jump components are computed based on Equation (12) using the test statistic of Equation (10) and a significance level of 0.1%. All regressions are estimated using Newey–West (1987) corrected standard errors with 5, 10, and 44 lags for the 1-, 5-, and 22-day horizon, respectively. T-statistics are reported in parentheses. The intercepts are not reported to save space. Significant coefficients at the 5% level are highlighted in bold. The second last column reports the adjusted R^2 of each regression. The last column shows the number of observations. The sample period is from January 2, 2007 to June 29, 2012.

TABLE VI
In-Sample Predictability: Gasoline

	β_d	β_w	β_m	β_{C_d}	β_{C_w}	β_{C_m}	γ_{J_d}	γ_{J_w}	γ_{J_m}	γ_{RJ}	γ_{RJ^+}	γ_{RJ^-}	\bar{R}^2	Obs.
Panel A: 1-Day Horizon														
HAR-RV	0.210 (3.463)	0.475 (7.898)	0.265 (4.687)	—	—	—	—	—	—	—	—	—	0.730	1394
HAR-J	—	0.476 (7.665)	0.267 (4.645)	0.222 (3.361)	—	—	0.021 (0.431)	—	—	—	—	—	0.731	1394
HAR-RJ	—	0.490 (7.687)	0.266 (4.686)	0.213 (3.216)	—	—	—	—	—	−0.049 (−1.533)	—	—	0.732	1394
HAR-ARJ	—	0.484 (7.629)	0.261 (4.532)	0.220 (3.362)	—	—	—	—	—	—	−0.023 (−0.433)	−0.080 (−1.166)	0.732	1394
HAR-C-J	—	—	—	0.204 (2.979)	0.491 (6.821)	0.332 (4.893)	0.023 (0.474)	0.032 (1.453)	−0.011 (−0.674)	—	—	—	0.734	1394
Panel B: 5-Day Horizon														
HAR-RV	0.215 (7.984)	0.458 (8.558)	0.285 (5.616)	—	—	—	—	—	—	—	—	—	0.858	1390
HAR-J	—	0.459 (8.495)	0.286 (5.466)	0.227 (8.145)	—	—	0.034 (1.222)	—	—	—	—	—	0.859	1390
HAR-RJ	—	0.473 (8.576)	0.289 (5.446)	0.216 (7.475)	—	—	—	—	—	−0.031 (−1.522)	—	—	0.859	1390
HAR-ARJ	—	0.464 (8.437)	0.282 (5.344)	0.226 (8.183)	—	—	—	—	—	—	0.004 (0.105)	−0.073 (−2.231)	0.859	1390
HAR-C-J	—	—	—	0.206 (7.050)	0.493 (8.918)	0.323 (6.165)	0.038 (1.479)	0.020 (1.141)	0.000 (0.005)	—	—	—	0.864	1390

Continue

TABLE VI
Continued

	β_d	β_w	β_m	β_{C_d}	β_{C_w}	β_{C_m}	γ_d	γ_w	γ_m	γ_{RJ}	γ_{RJ^+}	γ_{RJ^-}	\bar{R}^2	<i>Obs.</i>
	<i>Panel C: 22-Day Horizon</i>													
HAR-RV	0.198 (6.142)	0.420 (4.621)	0.296 (3.081)	—	—	—	—	—	—	—	—	—	0.835	1373
HAR-J	—	0.421 (4.611)	0.297 (3.049)	0.209 (6.113)	—	—	0.029 (1.195)	—	—	—	—	—	0.836	1373
HAR-RJ	—	0.432 (4.683)	0.300 (3.018)	0.200 (5.627)	—	—	—	—	—	−0.020 (−1.054)	—	—	0.836	1373
HAR-ARJ	—	0.424 (4.602)	0.295 (3.018)	0.208 (6.079)	—	—	—	—	—	—	0.009 (0.281)	−0.056 (−1.869)	0.836	1373
HAR-C-J	—	—	—	0.191 (6.784)	0.444 (4.815)	0.329 (3.895)	0.032 (1.346)	0.025 (1.324)	0.005 (0.133)	—	—	—	0.840	1373

Note. This table assesses the predictive ability of several models in the gasoline market. Each of the three panels shows results for a different forecasting horizon (1-, 5-, and 22-day horizon). The jump components are computed based on Equation (12) using the test statistic of Equation (10) and a significance level of 0.1%. All regressions are estimated using Newey–West (1987) corrected standard errors with 5, 10, and 44 lags for the 1-, 5-, and 22-day horizon, respectively. T-statistics are reported in parentheses. The intercepts are not reported to save space. Significant coefficients at the 5% level are highlighted in bold. The second last column reports the adjusted \bar{R}^2 of each regression. The last column shows the number of observations. The sample period is from January 2, 2007 to June 29, 2012.

component. Since positive and negative jumps are mixed together, this blurs the information content of jumps and buries any evidence of predictability.

The last row of each panel reveals that the jump components of the HAR-C-J model are generally significant. In spite of the statistical significance of the jump components exhibited by the more sophisticated models, we can see that there is very little to distinguish between the explanatory power of all models. This is true for all markets and forecast horizons. For instance, in the crude oil market, all models yield adjusted R^2 roughly equal to 82%. The upshot of this is that the benefits of explicitly modeling the dynamics of jumps are small.

3.3. Out-of-Sample Analysis

We now turn our attention to the out-of-sample performance of the competing models. To do this, we adopt a simple procedure that allows us to generate forecasts using parameters estimated on a rolling windows basis. Each day, we use the most recent 600 observations to estimate the forecasting models.¹² Equipped with the parameter estimates, we generate out-of-sample volatility forecasts for a given horizon, e.g., 22-day, which we then compare with realized volatility (computed ex-post). We roll our window forward by one day and repeat all the steps above, yielding a time-series of volatility forecasts that are compared with the corresponding realized volatility. We do this for each market, model, and forecasting horizon.

We consider the following six loss functions: the mean squared error (MSE), the mean squared percentage error (MSPE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the logarithmic loss (LL), and the quasi-likelihood loss (QLIKE). These loss functions are defined as follows:

$$\begin{aligned}
 \text{MSE} &= \frac{1}{N} \sum_{t=1}^N (RV_{t:t+h}^{1/2} - F_{t:t+h}^{1/2})^2 & \text{MSPE} &= \frac{1}{N} \sum_{t=1}^N \left(\frac{RV_{t:t+h}^{1/2} - F_{t:t+h}^{1/2}}{F_{t:t+h}^{1/2}} \right)^2 \\
 \text{MAE} &= \frac{1}{N} \sum_{t=1}^N |RV_{t:t+h}^{1/2} - F_{t:t+h}^{1/2}| & \text{MAPE} &= \frac{1}{N} \sum_{t=1}^N \left| \frac{RV_{t:t+h}^{1/2} - F_{t:t+h}^{1/2}}{F_{t:t+h}^{1/2}} \right| \\
 \text{LL} &= \frac{1}{N} \sum_{t=1}^N \left[\log(RV_{t:t+h}^{1/2}) - \log(F_{t:t+h}^{1/2}) \right] & \text{QLIKE} &= \frac{1}{N} \sum_{t=1}^N \left[\log(F_{t:t+h}^{1/2}) + \frac{RV_{t:t+h}^{1/2}}{F_{t:t+h}^{1/2}} \right]
 \end{aligned}$$

where N is the number of out-of-sample forecasts, $RV_{t:t+h}^{1/2}$ is the ex-post realized volatility, and $F_{t:t+h}^{1/2}$ is the volatility forecast from each of the five forecasting models.

Table VII presents the forecasting errors. Each panel focuses on a specific loss function. While each row corresponds to a specific market, each column represents a specific forecasting model. We present the results for each forecasting horizon.

We observe that the forecast errors of the more complex models are of the same order of magnitude as those of the baseline HAR-RV, indicating that modeling jumps does not noticeably improve forecast accuracy. For instance, the MSEs of natural gas (monthly horizon) vary within a tight range from 0.837 (HAR-ARJ) to 0.847 (HAR-C-J). Clearly, there is very little to distinguish between all competing models. This example also reveals that the most

¹²Section 4 considers other window sizes such as 400, 800, and 1,000. Our main conclusions are robust to the width of the rolling window.

TABLE VII
Forecasting Errors

	1-Day Horizon						5-Day Horizon						22-Day Horizon					
	HAR-RV			HAR-RJ			HAR-RV			HAR-RJ			HAR-RV			HAR-RJ		
	HAR-RV	HAR-RJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-C-J
A. MSE																		
Crude Oil	0.535	0.533	0.530	0.530	0.533	0.538	0.376	0.375	0.373	0.374	0.381	0.401	0.401	0.400	0.403	0.419		
Heating Oil	0.399	0.400	0.396	0.400	0.400	0.400	0.247	0.247	0.247	0.247	0.248	0.244	0.244	0.247	0.245	0.249		
Natural Gas	1.494	1.487	1.491	1.493	1.483	1.483	0.842	0.836	0.849	0.841	0.848	0.841	0.841	0.838	0.838	0.837		
Gasoline	0.484	0.483	0.484	0.483	0.483	0.499	0.297	0.299	0.299	0.299	0.314	0.289	0.289	0.291	0.293	0.292		
	HAR-RV	HAR-RJ	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J		
B. MSPE																		
Crude Oil	7.339	7.342	7.275	7.351	7.502	7.511	3.540	3.543	3.480	3.504	3.718	4.320	4.320	4.348	4.303	4.354		
Heating Oil	7.488	7.513	7.422	7.502	7.502	7.501	3.525	3.532	3.511	3.529	3.502	4.054	4.054	4.060	4.072	4.063		
Natural Gas	8.250	8.209	8.241	8.253	8.204	8.204	3.079	3.044	3.104	3.087	3.065	3.128	3.128	3.111	3.130	3.128		
Gasoline	7.740	7.733	7.766	7.744	7.950	7.950	3.663	3.668	3.682	3.677	3.830	4.197	4.197	4.217	4.227	4.226		
	HAR-RV	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J	HAR-C-J	HAR-RV	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J		
C. MAE																		
Crude Oil	5.150	5.165	5.144	5.169	5.169	5.214	4.206	4.205	4.196	4.199	4.303	5.012	5.012	5.010	5.005	5.011		
Heating Oil	4.505	4.511	4.491	4.511	4.511	4.510	3.649	3.650	3.653	3.644	3.650	4.102	4.102	4.103	4.117	4.100		
Natural Gas	8.721	8.715	8.739	8.731	8.721	8.721	6.229	6.212	6.290	6.224	6.287	6.632	6.632	6.629	6.631	6.616		
Gasoline	5.142	5.148	5.134	5.132	5.233	5.233	4.003	4.015	4.011	4.004	4.136	4.360	4.360	4.376	4.390	4.379		
	HAR-RV	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J	HAR-C-J	HAR-RV	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J	HAR-RV	HAR-RJ	HAR-RJ	HAR-ARJ	HAR-C-J		
D. MAPE																		
Crude Oil	18.872	18.921	18.839	18.935	18.935	19.157	14.226	14.220	14.161	14.179	14.676	17.047	17.047	17.050	17.011	17.039		
Heating Oil	18.976	18.991	18.906	18.988	18.988	19.001	14.474	14.472	14.458	14.448	14.432	16.410	16.410	16.412	16.441	16.386		
Natural Gas	21.150	21.136	21.205	21.179	21.179	21.141	13.646	13.603	13.789	13.641	13.718	14.174	14.174	14.162	14.181	14.147		
Gasoline	19.832	19.847	19.812	19.797	19.797	20.134	14.616	14.632	14.615	14.591	15.004	16.239	16.239	16.273	16.321	16.290		

Continue

TABLE VII
Continued

	1-Day Horizon						5-Day Horizon						22-Day Horizon					
	HAR-RV		HAR-RJ		HAR-ARJ		HAR-C-J		HAR-RV		HAR-RJ		HAR-ARJ		HAR-C-J		HAR-RV	
	HAR-RV	HAR-RJ	HAR-RV	HAR-RJ	HAR-RV	HAR-RJ	HAR-RV	HAR-RJ	HAR-RV	HAR-RJ	HAR-RV	HAR-RJ	HAR-RV	HAR-RJ	HAR-RV	HAR-RJ	HAR-RV	HAR-RJ
E. LL																		
Crude Oil	5.633	5.634	5.597	5.645	5.708	5.593	3.577	3.549	3.551	3.676	4.238	4.239	4.228	4.242	4.456	4.456	4.456	4.456
Heating Oil	5.610	5.615	5.561	5.614	5.626	3.321	3.323	3.318	3.321	3.315	3.539	3.541	3.560	3.546	3.544	3.544	3.544	3.544
Natural Gas	6.615	6.597	6.621	6.629	6.562	3.104	3.079	3.132	3.102	3.080	3.250	3.237	3.244	3.241	3.239	3.239	3.239	3.239
Gasoline	5.914	5.901	5.915	5.909	6.072	3.459	3.475	3.482	3.476	3.641	3.658	3.680	3.696	3.691	3.851	3.851	3.851	3.851
F. QLIKE																		
Crude Oil	-25.534	-25.536	-25.552	-25.530	-25.507	-23.233	-23.245	-23.255	-23.258	-23.209	-21.189	-21.192	-21.193	-21.191	-21.103	-21.103	-21.103	-21.103
Heating Oil	-37.243	-37.242	-37.266	-37.242	-37.236	-35.058	-35.058	-35.059	-35.059	-35.059	-33.582	-33.581	-33.570	-33.577	-33.580	-33.580	-33.580	-33.580
Natural Gas	14.556	14.548	14.557	14.563	14.523	16.996	16.984	17.010	16.994	16.980	18.177	18.170	18.172	18.171	18.167	18.167	18.167	18.167
Gasoline	-29.255	-29.263	-29.255	-29.258	-29.171	-27.032	-27.021	-27.019	-27.022	-26.931	-25.457	-25.444	-25.435	-25.438	-25.355	-25.355	-25.355	-25.355

Note. This table presents out-of-sample forecasting errors for the five volatility models considered. Each panel focuses on a specific loss function. *MSE* is the mean squared error, *MSPE* is the mean squared percentage error, *MAE* is the mean absolute error, *MAPE* is the mean absolute percentage error, *LL* is the logarithmic loss, and *QLIKE* is the quasi-likelihood loss function. We consider three forecast horizons, namely 1, 5, and 22 days. Out-of-sample forecasts are obtained using a rolling window of 600 observations. In order to facilitate the presentation of our results, we multiply each loss function by 100.

elaborated model, i.e., HAR–C–J, often produces the worst forecast, thus strengthening our main conclusion.

Up to this point, we only analyze the magnitudes of the loss functions and do not formally investigate whether the economically small differences are statistically significant. We rigorously address this question by implementing the statistical test of Giacomini and White (2006), which accounts for parameter uncertainty and allows for comparison of nested models.

The GW test is based on the expected difference in forecast errors between two competing models. Let h and $\Delta L_{i,j}$ denote the forecast horizon and the vector of the loss differences between models i and j , respectively. The null hypothesis of the GW test is as follows:

$$H_0 : E [\Delta L_{i,j}] = 0 \quad (25)$$

The test follows a chi-squared distribution with one degree of freedom and the null is evaluated on the basis of the following test statistic:

$$GW = P \left(P^{-1} \sum_{t=1}^{T-h} \Delta L_{t+h,i,j} \right)' \hat{V}_h^{-1} \left(P^{-1} \sum_{t=1}^{T-h} \Delta L_{t+h,i,j} \right) \sim \chi_1^2 \quad (26)$$

where P is the total number of out-of-sample forecasts, $\Delta L_{t+h,i,j}$ is the loss difference at time $t + h$, and \hat{V}_h is a heteroskedasticity and autocorrelation consistent (HAC) estimator of the asymptotic variance of $P^{-1} \sum_t \Delta L_{t+h}$. Following Giacomini and White (2006), we employ the Newey-West (1987) estimator with $h-1$ lags to account for the serial dependence in multistep-ahead forecasts. Using a significance level α , the null of equal predictive ability is rejected if $|GW| > \chi_{1,1-\alpha}^2$, where $\chi_{1,1-\alpha}^2$ is the critical value from a chi-squared distribution with one degree of freedom.

Tables VIII–XI summarize our results. The test statistics presented in the table are based on the mean difference between the model [name in row] and the model [name in column]. Hence, a negative test statistics means that the model [name in row] yields more accurate forecasts than the model [name in column]. We highlight in bold statistically significant test statistics at the 5% significance level.

Comparing our baseline model (HAR–RV) to its more sophisticated rivals, we find very little evidence to suggest that explicitly modeling jumps significantly improves the accuracy of volatility forecasts. To quickly see this, notice that very few entries in the column headed “HAR–RV” are boldfaced, suggesting that the more elaborated models yield forecasts that are not statistically distinguishable from those of the simple and parsimonious HAR–RV. This is true, irrespective of the forecasting horizon, the market, and the loss function. Moreover, the most complex model, i.e., HAR–C–J, significantly underperforms all other models (including the benchmark HAR–RV). This is particularly noticeable in the crude oil and gasoline markets, where significantly positive entries are often reported in the last row. This result echoes our core finding: the simpler the model, the better.

In sum, our out-of-sample analysis reveals that models that explicitly seek to capture the dynamics of jumps do not significantly improve the accuracy of volatility forecasts: there is virtually no gain in modeling the dynamics of jumps in energy markets.

TABLE VIII
Out-of-Sample Forecast Comparisons for Crude Oil

	1-Day Horizon				5-Day Horizon				22-Day Horizon			
	HAR-RV		HAR-RJ		HAR-RV		HAR-RJ		HAR-RV		HAR-RJ	
	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ
A. SE												
HAR-RV	–				HAR-RV	–			HAR-RV	–		
HAR-J	–0.60	–			HAR-J	–0.31	–		HAR-J	0.06		
HAR-RJ	–2.75	–2.25	–		HAR-RJ	–1.61	–0.70	–	HAR-RJ	–0.17	–0.23	–
HAR-ARJ	–0.35	0.53	4.98	–	HAR-ARJ	–1.11	–0.90	0.10	HAR-ARJ	0.40	0.59	0.57
HAR-C-J	0.38	4.30	6.30	3.04	HAR-C-J	0.58	1.05	1.67	HAR-C-J	1.01	1.06	1.09
	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ
B. SPE												
HAR-RV	–				HAR-RV	–			HAR-RV	–		
HAR-J	0.00	–			HAR-J	0.01	–		HAR-J	1.98		
HAR-RJ	–1.35	–3.36	–		HAR-RJ	–1.98	–2.54	–	HAR-RJ	–0.61	–1.90	–
HAR-ARJ	0.05	0.24	7.43	–	HAR-ARJ	–0.79	–2.90	0.80	HAR-ARJ	1.05	0.05	1.62
HAR-C-J	4.33	6.73	8.96	5.53	HAR-C-J	3.79	3.76	6.73	HAR-C-J	2.76	2.54	3.00
	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ
C. AE												
HAR-RV	–				HAR-RV	–			HAR-RV	–		
HAR-J	0.53				HAR-J	0.00	–		HAR-J	–0.03		
HAR-RJ	–0.12	–2.97	–		HAR-RJ	–0.46	–0.45	–	HAR-RJ	–0.34	–0.09	–
HAR-ARJ	0.79	0.71	5.33		HAR-ARJ	–0.21	–0.35	0.06	HAR-ARJ	–0.01	0.02	0.11
HAR-C-J	5.00	6.44	8.97	5.17	HAR-C-J	3.76	4.37	4.74	HAR-C-J	0.57	0.65	0.64
								4.70				0.63

Continue

TABLE VIII
Continued

1-Day Horizon					5-Day Horizon					22-Day Horizon				
HAR-RV	HAR-J	HAR-RJ	HAR-ARJ		HAR-RV	HAR-J	HAR-RJ	HAR-ARJ		HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	
D. APE														
HAR-RV	–				HAR-RV	–				HAR-RV	–			
HAR-J	0.34	–			HAR-J	–0.01	–			HAR-J	0.00	–		
HAR-RJ	–0.26	–3.05	–		HAR-RJ	–1.04	–1.10	–		HAR-RJ	–0.75	–0.49	–	
HAR-ARJ	0.55	0.65	4.93	–	HAR-ARJ	–0.62	–1.11	0.14	–	HAR-ARJ	–0.02	–0.12	0.25	
HAR-C-J	6.42	10.22	13.10	8.49	HAR-C-J	5.90	6.80	8.00	7.67	HAR-C-J	1.33	1.41	1.50	
HAR-RV	HAR-J	HAR-RJ	HAR-ARJ		HAR-RV	HAR-J	HAR-RJ	HAR-ARJ		HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	
E. LL														
HAR-RV	–				HAR-RV	–				HAR-RV	–			
HAR-J	0.00	–			HAR-J	–0.57	–			HAR-J	0.00	–		
HAR-RJ	–1.36	–2.42	–		HAR-RJ	–2.54	–1.42	–		HAR-RJ	–0.39	–0.26	–	
HAR-ARJ	0.08	1.45	4.58	–	HAR-ARJ	–2.39	–3.26	0.01	–	HAR-ARJ	0.03	0.06	0.34	
HAR-C-J	2.15	5.81	7.99	3.70	HAR-C-J	1.48	2.35	3.59	3.64	HAR-C-J	1.44	1.56	1.57	
HAR-RV	HAR-J	HAR-RJ	HAR-ARJ		HAR-RV	HAR-J	HAR-RJ	HAR-ARJ		HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	
F. QLIKE														
HAR-RV	–				HAR-RV	–				HAR-RV	–			
HAR-J	–0.01	–			HAR-J	–1.33	–			HAR-J	–0.16	–		
HAR-RJ	–1.34	–1.72	–		HAR-RJ	–2.86	–0.84	–		HAR-RJ	–0.28	–0.01	–	
HAR-ARJ	0.03	1.79	3.31	–	HAR-ARJ	–3.31	–3.12	–0.07	–	HAR-ARJ	–0.03	0.08	0.05	
HAR-C-J	1.07	3.78	5.37	2.09	HAR-C-J	0.48	1.29	1.88	2.22	HAR-C-J	0.95	1.14	1.05	

Note. This table presents test statistics from pairwise comparisons of equal predictive accuracy of forecasting models for crude oil volatility. Each day, we use a trailing window of 600 observations to estimate the parameters of the HAR models. Equipped with these estimates, we then make out-of-sample forecasts of volatility. We consider three forecasting horizons: daily, weekly, and monthly. We report the test statistics from comparing the mean difference between the forecast errors of model [name in row] and those of the model [name in column]. The Giacomini and White (2006) test statistic is distributed as a chi-squared random variable with 1 degree of freedom. We highlight in bold all the significant test statistics based on the 95% confidence level.

TABLE IX
Out-of-Sample Forecast Comparisons for Heating Oil

	1-Day Horizon				5-Day Horizon				22-Day Horizon			
	HAR-RV		HAR-RJ		HAR-RV		HAR-RJ		HAR-RV		HAR-RJ	
	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ
A. SE	HAR-RV	-			HAR-RV	-			HAR-RV	-		
	HAR-J	0.11	-		HAR-J	0.44	-		HAR-J	0.84	-	
	HAR-RJ	-2.23	-2.65	-	HAR-RJ	0.23	0.01	-	HAR-RJ	2.02	1.07	-
	HAR-ARJ	0.08	0.00	2.73	HAR-ARJ	0.19	-0.01	-0.02	HAR-ARJ	0.34	0.06	-1.06
	HAR-C-J	0.34	0.20	2.61	HAR-C-J	0.29	0.08	0.04	HAR-C-J	0.61	0.51	0.16
	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ
B. SPE	HAR-RV	-			HAR-RV	-			HAR-RV	-		
	HAR-J	0.27	-		HAR-J	0.13	-		HAR-J	0.32	-	
	HAR-RJ	-1.50	-2.94	-	HAR-RJ	-0.26	-0.54	-	HAR-RJ	0.77	0.30	-
	HAR-ARJ	0.07	-0.33	2.66	HAR-ARJ	0.03	-0.05	0.33	HAR-ARJ	0.18	0.03	-0.24
	HAR-C-J	0.03	-0.04	0.93	HAR-C-J	-0.29	-0.49	-0.04	HAR-C-J	0.03	0.01	0.00
	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ
C. AE	HAR-RV	-			HAR-RV	-			HAR-RV	-		
	HAR-J	0.14	-		HAR-J	0.02	-		HAR-J	0.10	-	
	HAR-RJ	-0.87	-1.55	-	HAR-RJ	0.09	0.04	-	HAR-RJ	1.20	0.92	-
	HAR-ARJ	0.12	-0.01	1.54	HAR-ARJ	-0.25	-0.77	-0.45	HAR-ARJ	-0.02	-0.10	-2.50
	HAR-C-J	0.07	-0.01	0.99	HAR-C-J	0.00	0.00	-0.02	HAR-C-J	0.65	0.63	0.31
	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ	HAR-RV	HAR-J	HAR-RJ	HAR-ARJ

Continue

TABLE IX
Continued

22-Day Horizon											
5-Day Horizon											
1-Day Horizon											
HAR-RV HAR-J HAR-RJ HAR-ARJ				HAR-RV HAR-J HAR-RJ HAR-ARJ				HAR-RV HAR-J HAR-RJ HAR-ARJ			
D. APE											
HAR-RV	-			HAR-RV	-			HAR-RV	-		
HAR-J	0.04	-		HAR-J	0.00	-		HAR-J	0.01	-	
HAR-RJ	-1.06	-1.56	-	HAR-RJ	-0.08	-0.07	-	HAR-RJ	0.34	0.29	-
HAR-ARJ	0.02	-0.04	1.57	HAR-ARJ	-0.40	-0.81	-0.05	HAR-ARJ	-0.25	-0.37	-2.07
HAR-C-J	0.07	0.04	1.34	HAR-C-J	-0.19	-0.19	-0.07	HAR-C-J	0.21	0.21	0.09
HAR-RV HAR-J HAR-RJ HAR-ARJ				HAR-RV HAR-J HAR-RJ HAR-ARJ				HAR-RV HAR-J HAR-RJ HAR-ARJ			
E. LL											
HAR-RV	-			HAR-RV	-			HAR-RV	-		
HAR-J	0.03	-		HAR-J	0.01	-		HAR-J	0.06	-	
HAR-RJ	-2.63	-3.37	-	HAR-RJ	-0.03	-0.06	-	HAR-RJ	0.96	0.66	-
HAR-ARJ	0.02	-0.01	3.53	HAR-ARJ	0.00	-0.06	0.03	HAR-ARJ	0.24	0.15	-0.76
HAR-C-J	0.18	0.25	3.09	HAR-C-J	-0.04	-0.07	-0.01	HAR-C-J	0.01	0.00	-0.07
HAR-RV HAR-J HAR-RJ HAR-ARJ				HAR-RV HAR-J HAR-RJ HAR-ARJ				HAR-RV HAR-J HAR-RJ HAR-ARJ			
F. QLIKE											
HAR-RV	-			HAR-RV	-			HAR-RV	-		
HAR-J	0.00	-		HAR-J	0.00	-		HAR-J	0.02	-	
HAR-RJ	-2.93	-3.26	-	HAR-RJ	0.00	0.00	-	HAR-RJ	0.96	0.72	-
HAR-ARJ	0.01	0.01	3.53	HAR-ARJ	-0.01	-0.02	0.00	HAR-ARJ	0.35	0.26	-0.77
HAR-C-J	0.16	0.31	2.95	HAR-C-J	0.00	0.00	0.00	HAR-C-J	0.00	0.00	-0.11
											-0.01

Note. This table presents test statistics from pairwise comparisons of equal predictive accuracy of forecasting models for heating oil volatility. Each day, we use a trailing window of 600 observations to estimate the parameters of the HAR models. Equipped with these estimates, we then make out-of-sample forecasts of volatility. We consider three forecasting horizons: daily, weekly, and monthly. We report the test statistics from comparing the mean difference between the forecast errors of model [name in row] and those of the model [name in column]. The Giacomini and White (2006) test statistic is distributed as a chi-squared random variable with 1 degree of freedom. We highlight in bold all the significant test statistics based on the 95% confidence level.

TABLE X
Out-of-Sample Forecast Comparisons for Natural Gas

1-Day Horizon				5-Day Horizon				22-Day Horizon						
HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ
A. SE														
HAR-RV	-				HAR-RV	-				HAR-RV	-			
HAR-J	-1.19	-			HAR-J	-1.57	-			HAR-J	-0.47	-		
HAR-RJ	-0.08	0.66	-		HAR-RJ	1.01	7.76	-		HAR-RJ	-0.33	-0.03	-	
HAR-ARJ	-0.03	3.55	0.18	-	HAR-ARJ	-0.03	2.45	-2.09	-	HAR-ARJ	-0.57	-0.10	-0.15	-
HAR-C-J	-1.11	-0.23	-0.79	-1.39	HAR-C-J	0.13	0.56	0.00	0.18	HAR-C-J	0.16	0.33	0.33	0.37
HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ
B. SPE														
HAR-RV	-				HAR-RV	-				HAR-RV	-			
HAR-J	-0.39	-			HAR-J	-1.47	-			HAR-J	-0.52	-		
HAR-RJ	-0.01	0.37	-		HAR-RJ	0.50	10.53	-		HAR-RJ	0.01	0.45	-	
HAR-ARJ	0.00	1.36	0.06	-	HAR-ARJ	0.19	1.79	-0.23	-	HAR-ARJ	0.00	0.23	-0.05	-
HAR-C-J	-0.31	-0.01	-0.32	-0.69	HAR-C-J	-0.05	0.23	-0.78	-0.13	HAR-C-J	0.01	0.10	0.01	0.01
HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ
C. AE														
HAR-RV	-				HAR-RV	-				HAR-RV	-			
HAR-J	-0.07	-			HAR-J	-1.14	-			HAR-J	-0.03	-		
HAR-RJ	0.26	1.53	-		HAR-RJ	3.11	9.21	-		HAR-RJ	0.00	0.02	-	
HAR-ARJ	0.16	1.41	-0.23	-	HAR-ARJ	-0.09	1.10	-6.01	-	HAR-ARJ	-0.47	-0.51	-1.80	-
HAR-C-J	0.00	0.06	-0.36	-0.14	HAR-C-J	1.26	2.86	-0.01	1.75	HAR-C-J	-0.21	-0.19	-0.21	-0.09

TABLE XI
Continued

1-Day Horizon					5-Day Horizon					22-Day Horizon				
HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ
D. APE														
HAR-RV	–				HAR-RV	–				HAR-RV	–			
HAR-J	0.05	–			HAR-J	0.13	–			HAR-J	0.55	–		
HAR-RJ	–0.11	–0.58	–		HAR-RJ	0.00	–0.07	–		HAR-RJ	1.60	1.11	–	
HAR-ARJ	–0.25	–1.14	–0.39	–	HAR-ARJ	–0.09	–0.34	–1.61	–	HAR-ARJ	0.77	0.15	–2.62	–
HAR-C-J	5.40	6.69	7.57	8.45	HAR-C-J	6.06	6.62	6.46	6.88	HAR-C-J	3.18	3.40	2.99	3.26
HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ
E. LL														
HAR-RV	–				HAR-RV	–				HAR-RV	–			
HAR-J	–0.29	–			HAR-J	0.88	–			HAR-J	2.40	–		
HAR-RJ	0.00	0.64	–		HAR-RJ	0.77	0.09	–		HAR-RJ	3.47	1.44	–	
HAR-ARJ	–0.04	0.20	–0.54	–	HAR-ARJ	0.47	0.00	–0.56	–	HAR-ARJ	3.14	0.77	–0.75	–
HAR-C-J	8.45	12.33	9.73	10.90	HAR-C-J	11.78	13.47	10.33	10.20	HAR-C-J	3.79	3.62	3.18	3.31
HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ	HAR-RV		HAR-J	HAR-RJ	HAR-ARJ
F. QLIKE														
HAR-RV	–				HAR-RV	–				HAR-RV	–			
HAR-J	–0.56	–			HAR-J	1.34	–			HAR-J	2.77	–		
HAR-RJ	0.00	0.86	–		HAR-RJ	1.02	0.03	–		HAR-RJ	3.89	1.91	–	
HAR-ARJ	–0.07	0.35	–0.53	–	HAR-ARJ	0.67	–0.01	–0.72	–	HAR-ARJ	3.55	0.96	–1.17	–
HAR-C-J	9.60	13.84	11.12	12.11	HAR-C-J	12.40	14.14	11.40	11.49	HAR-C-J	3.88	3.61	3.20	3.35

Note. This table presents test statistics from pairwise comparisons of equal predictive accuracy of forecasting models for gasoline volatility. Each day, we use a trailing window of 600 observations to estimate the parameters of the HAR models. Equipped with these estimates, we then make out-of-sample forecasts of volatility. We consider three forecasting horizons: daily, weekly, and monthly. We report the test statistics from comparing the mean difference between the forecast errors of model [name in row] and those of the model [name in column]. The Giacomini and White (2006) test statistic is distributed as a chi-squared random variable with 1 degree of freedom. We highlight in bold all the significant test statistics based on the 5% significance level.

4. ROBUSTNESS CHECKS

In this section, we conduct several additional tests to investigate the robustness of our findings. We begin by analyzing whether our main findings hold if we consider the task of predicting variance and log variance, rather than volatility. We then explore the robustness of our results with respect to the jump detection procedure by using the nearest neighbor estimator of Andersen et al. (2012). Additionally, we show that our results are robust to the width of the window used to obtain rolling forecasts. Finally, we consider a WLS (rather than OLS) estimation to establish that our findings are not affected by the method of estimation.

4.1. Variance and Log Variance Forecasts

Up to this point, our analysis focuses on the task of forecasting volatility. As previously discussed, we focus on volatility instead of variance because of the key role it plays in modern finance. For instance, volatility (not variance) is a key input in option pricing and modern portfolio theories. Nonetheless, one may argue that the jump detection tests identify jumps in variance not in volatility, and this subtle difference may matter for our analysis.

Tables B.6–B.10 of the complementary appendix investigate whether modeling jumps can improve the accuracy of *variance* forecasts. Similarly, Tables C.11–C.15 of the appendix focuses on the task of forecasting *log variance*. Consistent with our main findings, these tables establish that more sophisticated models do not generally outperform the baseline specification. The upshot of this is that our results are the same, irrespective of whether we look at volatility, variance, or log variance forecasting.

4.2. Alternative Jump-Robust Estimators

Andersen et al. (2012) point out that the standard multipower variations may be biased in finite samples. The authors then propose jump-robust volatility estimators that use the nearest neighbor truncation. They forcefully show that the “median realized variance estimator” (MedRV_t) is more efficient and robust to jumps than its main rivals. As a robustness check, we repeat our analysis replacing BPV with the MedRV variation estimator. This estimator, using staggered (skip–1) returns, is defined as follows:

$$\text{MedRV}_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{M}{M - 2(k+1)} \right) \sum_{j=2k+3}^M \text{med} \left(|r_{t_{j-2(k+1)}}|, |r_{t_{j-(k+1)}}|, |r_{t_j}| \right)^2 \quad (27)$$

where $\text{med}(\cdot)$ stands for the median operator. As in our main analysis, we set $k=1$ (skip–1 return). The corresponding jump test statistic is as follows:

$$z_{\text{Med},t} = \Delta^{-1/2} \frac{(\text{RV}_t - \text{MedRV}_t) / \text{RV}_t}{\sqrt{0.96 \max(1, \frac{\text{MedRQ}_t}{\text{MedRV}_t^2})}} \quad (28)$$

The number 0.96 comes from the asymptotic distribution of the MedRV estimator.¹³ Notice also that the tripower quarticity in the test statistic of Equation (10) is replaced with the

¹³For further details, we refer the interested reader to Propositions 1–3 in Andersen et al. (2012).

median-realized quarticity given by the following:

$$\text{MedRQ}_t = \frac{3\pi N}{9\pi + 72 - 52\sqrt{3}} \left(\frac{M}{M - 2(k+1)} \right) \cdot \sum_{j=2k+3}^M \text{med} \left(|r_{t_{j-2(k+1)}}|, |r_{t_{j-(k+1)}}|, |r_{t_j}| \right)^4 \quad (29)$$

Finally, the decomposition of realized variance into its continuous and jump components is done exactly as in Equations (12) and (13) replacing, BPV with MedRV. Tables D.16–D.20 of the online appendix confirm our main findings: specifically accounting for jumps in volatility forecasting does not significantly improve forecasting accuracy.

4.3. Alternative Estimation Periods

Our out-of-sample analysis rests on a rolling window of 600 observations. One may argue that this choice is somewhat arbitrary and wonder what effect, if any, it may have on our results. To investigate this point, we consider windows of 400, 800, and 1,000 observations. Tables E.21–E.35 of the supplementary appendix clearly show that changing the width of the rolling window has very little impact on our main conclusions.

4.4. Alternative Estimation Methods

Patton and Sheppard (2015) argue that because the dependent variable in the models is volatility, the OLS estimation may put too much weight on highly volatile periods. To address the concern that this may be the driving force behind our results, we estimate each model with WLS (rather than OLS). To be more specific, we first estimate each model using OLS and then employ the inverse of the fitted values as weights for the WLS estimations. Equipped with the parameter estimates, we repeat our main analyses (both in- and out-of-sample) and obtain very similar conclusions (See Tables F.36–F.40 of the appendix).

5. CONCLUSIONS

This paper uses high-frequency data on four deep and liquid commodity futures markets, namely crude oil, heating oil, natural gas, and gasoline, to identify jumps and analyze their impact on future volatility.

Our analysis establishes that jumps are rare events and their intensity substantially varies over time. We then investigate the importance of jumps for forecasts of realized volatility over horizons ranging from 1 to 22 days. To this end, we estimate and empirically analyze several extensions of the HAR–RV model that explicitly seek to capture the dynamics of jumps. We employ six distinct loss functions and the GW test to carefully assess the predictive ability of these models. Analyzing the magnitude of the error metrics, we find very little to distinguish between the benchmark model and its more complex competitors. Moreover, our rigorous econometric analysis establishes that the differences in forecast errors are not only economically small but also statistically insignificant. Collectively, our results suggest that explicitly modeling jumps does not significantly improve the accuracy of volatility forecasts in energy markets.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Table A1: Volatility Forecasting Errors.

Table A2: Comparisons of Out-of-Sample Volatility Forecasts for Crude Oil.

Table A3: Comparisons of Out-of-Sample Volatility Forecasts for Heating Oil.

Table A4: Comparisons of Out-of-Sample Volatility Forecasts for Natural Gas.

Table A5: Comparisons of Out-of-Sample Volatility Forecasts for Gasoline.

Table B6: Variance Forecasting Errors.

Table B7: Comparisons of Out-of-Sample Variance Forecasts for Crude Oil.

Table B8: Comparisons of Out-of-Sample Variance Forecasts for Heating Oil.

Table B9: Comparisons of Out-of-Sample Variance Forecasts for Natural Gas.

Table B10: Comparisons of Out-of-Sample Variance Forecasts for Gasoline.

Table C11: Forecasting Errors: Logarithmic Variance.

Table C12: Comparisons of Out-of-Sample Logarithmic Variance Forecasts for Crude Oil.

Table C13: Comparisons of Out-of-Sample Logarithmic Variance Forecasts for Heating Oil.

Table C14: Comparisons of Out-of-Sample Logarithmic Variance Forecasts for Natural Gas.

Table C15: Comparisons of Out-of-Sample Logarithmic Variance Forecasts for Gasoline.

Table D16: Volatility Forecasting Errors (MedRV Estimator).

Table D17: Comparisons of Out-of-Sample Volatility Forecasts for Crude Oil (MedRV Estimator).

Table D18: Comparisons of Out-of-Sample Volatility Forecasts for Heating Oil (MedRV Estimator).

Table D19: Comparisons of Out-of-Sample Volatility Forecasts for Natural Gas (MedRV Estimator).

Table D20: Comparisons of Out-of-Sample Volatility Forecasts for Gasoline (MedRV Estimator).

Table E21: Volatility Forecasting Errors (Rolling Window of 400 Observations).

Table E22: Out-of-Sample Volatility Forecast Comparisons for Crude Oil (Rolling Window of 400 Observations).

Table E23: Out-of-Sample Volatility Forecast Comparisons for Heating Oil (Rolling Window of 400 Observations).

Table E24: Out-of-Sample Volatility Forecast Comparisons for Natural Gas (Rolling Window of 400 Observations).

Table E25: Out-of-Sample Forecast Comparisons for Gasoline (Rolling Window of 400 Observations).

Table E26: Volatility Forecasting Errors (Rolling Window of 800 Observations).

Table E27: Out-of-Sample Volatility Forecast Comparisons for Crude Oil (Rolling Window of 800 Observations).

Table E28: Out-of-Sample Volatility Forecast Comparisons for Heating Oil (Rolling Window of 800 Observations).

Table E29: Out-of-Sample Volatility Forecast Comparisons for Natural Gas (Rolling Window of 800 Observations).

Table E30: Out-of-Sample Volatility Forecast Comparisons for Gasoline (Rolling Window of 800 Observations).

Table E31: Volatility Forecasting Errors (Rolling Window of 1,000 Observations).

Table E32: Out-of-Sample Volatility Forecast Comparisons for Crude Oil (Rolling Window of 1,000 Observations).

Table E33: Out-of-Sample Volatility Forecast Comparisons for Heating Oil (Rolling Window of 1,000 Observations).

Table E34: Out-of-Sample Volatility Forecast Comparisons for Natural Gas (Rolling Window of 1,000 Observations).

Table E35: Out-of-Sample Volatility Forecast Comparisons for Gasoline (Rolling Window of 1,000 Observations).

Table F36: Volatility Forecasting Errors (Weighted Least Squares Estimation).

Table F37: Out-of-Sample Volatility Forecast Comparisons for Crude Oil (Weighted Least Squares Estimation).

Table F38: Out-of-Sample Volatility Forecast Comparisons for Heating Oil (Weighted Least Squares Estimation).

Table F39: Out-of-Sample Volatility Forecast Comparisons for Natural Gas (Weighted Least Squares Estimation).

Table F40: Out-of-Sample Volatility Forecast Comparisons for Gasoline (Weighted Least Squares Estimation).

REFERENCES

- Aït-Sahalia, Y., Laeven, R. J., & Pelizzon, L. (2014). Mutual excitation in eurozone sovereign CDS. *Journal of Econometrics*, 183, 151–167.
- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39, 885–905.
- Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89, 701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (1999). Realized volatility and correlation. Working paper, University of Pennsylvania, US.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96, 42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 579–625.
- Andersen, T. G., Bollerslev, T., & Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics*, 160, 176–189.
- Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169, 75–93.
- Areal, N. M., & Taylor, S. J. (2002). The realized volatility of FTSE-100 futures prices. *Journal of Futures Markets*, 22, 627–648.
- Bajgrowicz, P., Scaillet, O., & Treccani, A. (2014). Jumps in high-frequency data: Spurious detections, dynamics, and news. Forthcoming in *Management Science*.
- Barndorff-Nielsen, O. E., Graversen, S. E., & Shephard, N. (2004). Power variation and stochastic volatility: A review and some new results. *Journal of Applied Probability*, 41, 133–143.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64, 253–280.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17, 457–477.
- Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2, 1–37.
- Barndorff-Nielsen, O. E., & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4, 1–30.
- Barndorff-Nielsen, O. E., Shephard, N., & Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and their Applications*, 116, 796–806.

- Bradley, D., Clarke, J., Lee, S., & Ornathanalai, C. (2014). Are analysts' recommendations informative? Intraday evidence on the impact of time stamp delays. *Journal of Finance*, 69, 645–673.
- Busch, T., Christensen, B. J., & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160, 48–57.
- Chung, S.-L., Tsai, W.-C., Wang, Y.-H., & Weng, P.-S. (2011). The information content of the s&p 500 index and vix options on the dynamics of the S&P 500 index. *Journal of Futures Markets*, 31, 1170–1201.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7, 174–196.
- Corsi, F., Pirino, D., & Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159, 276–288.
- Corsi, F., & Reno, R. (2009). HAR volatility modelling with heterogeneous leverage and jumps. Working paper, SSRN.
- Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance*, 59, 1367–1404.
- Eraker, B., Johannes, M., & Polson, N. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, 58, 1269–1300.
- Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74, 1545–1578.
- Giot, P., & Laurent, S. (2007). The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Futures Markets*, 27, 337–359.
- Huang, X., & Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3, 456–499.
- Lee, S. S., & Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21, 2535–2563.
- Maneesoonthorn, W., Forbes, C. S., & Martin, G. M. (2014). Inference on self-exciting jumps in prices and volatility using high frequency measures. Working paper, Cornell University Library.
- Martens, M., & Zein, J. (2004). Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets*, 24, 1005–1028.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 703–708.
- 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.
- Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, 235, 643–659.
- Tauchen, G., & Zhou, H. (2011). Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics*, 160, 102–118.
- Thomakos, D. D., & Wang, T. (2003). Realized volatility in the futures markets. *Journal of Empirical Finance*, 10, 321–353.
- Wang, T., Wu, J., & Yang, J. (2008). Realized volatility and correlation in energy futures markets. *Journal of Futures Markets*, 28, 993–1011.