CARDIFF BUSINESS SCHOOL
WORKING PAPER SERIES

The role of jumps in anticipating volatility

Kefu Liao [1], Kevin P. Evans, and Dudley Gilder

December 2022

Cardiff Business School
Cardiff University
Colum Drive Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
www.cardiff.ac.uk/carbs

[1] Kefu Liao (liaok2@cardiff.ac.uk), Kevin Evans (evansk1@cardiff.ac.uk), and Dudley Gilder (gilderd@cardiff.ac.uk) at Cardiff Business School, Cardiff University. Kefu Liao is the contact author: LiaoK2@cardiff.ac.uk.

The role of jumps in anticipating volatility

Abstract

This paper investigates whether jumps predict future volatility. The results for S&P 500 index show that dividing volatility into jumps and continuous variation improves volatility forecasting. And this improvement becomes more prominent when the volatility decomposition is associated with more reliable jump tests, longer volatility lags, or jump sign disentanglement. Besides, our results show that although jumps only have limited in-sample, the models including jumps lead to significantly better out-of-sample. Our results are robust when accounting for drift bias, intraday volatility pattern, and ultra-high frequency sampling.

Keywords: high frequency; volatility forecasting; drift, jumps

JEL: G12, E44, E32

1 Introduction

Forecasting price volatility allows us to know how future prices vary, which is key to the financial industry for asset pricing (Black and Scholes, 1973), derivative pricing (Duffie et al., 2000), asset allocation (Merton, 1969), and risk management (Christoffersen and Diebold, 2000).

Literature has extensively studied the predictive usefulness of dividing past volatility into jumps and continuous variation, motivated by the fact that the return volatility consists of a continuous volatility component, which accommodates the smooth price dynamic, and of a jump component, which takes care of the rough price movements. However, whether this decomposition helps predict future volatility is still under debate. Some researchers show that dividing volatility into jumps and continuous variation does not contribute to explaining future volatility or forecasting precision (Andersen et al., 2007a, Santos and Ziegelmann, 2014, Sévi, 2014, Prokopczuk et al., 2016, Caporin, 2022). Whereas others argue that this decomposition explains future volatility or improves volatility forecasting accuracy (Corsi et al., 2010, Patton and Sheppard, 2015, Duong and Swanson, 2015, Clements and Liao, 2017, Ma et al., 2018). To provide the literature controversy with more evidence, we revisit the impact of volatility decomposition on future volatility. Our results for the S&P 500 index show that the volatility decomposition leads to significantly better out-of-sample forecasts. Besides, our out-of-sample analysis indicates that including jumps can significantly increase models' forecasting precision, although we find only limited in-sample evidence of jumps.

1.1 The jump tests

The nonparametric class of tests is widely applied as the methods for disentangling jumps in the above jump literature as its statistic inference relaxes the distribution assumption and more fits into the empirical distribution of the sample. The above literature centres on using the nonparametric method by Barndorff-Nielsen and Shephard (2006) (BNS) [2]. However, recent simulation studies show that some alternative jump tests may be more reliable in detecting jumps. The simulation analysis by Maneesoonthorn et al. (2020) shows that for different sampling frequencies, the Med and Min tests by Andersen et al. (2012) have an empirical size that is closer to the nominal value than the BNS test. Maneesoonthorn et al. (2020) also show that the test by Jiang and Oomen (2008) (JO) has better size and power performance than the BNS test. Further, Dumitru and Urga (2012) find that the test by Andersen et al. (2007b) and Lee and Mykland (2008) (ABD-LM) has a more accurate size and stronger power in detecting jumps, especially in the presence of microstructure noise, which is an indispensable fact in high-frequency prices. Moreover, Dumitru and Urga (2012) show that the power of the ASJ test by Aït-Sahalia and Jacod (2009) (ASJ)[3] is less sensitive to sample frequencies than the BNS test.

Motivated by the superiority of these alternative jump tests in detecting jumps, we apply them for decomposing volatility into jumps and continuous variation, in addition to the BNS test. Our results of the S&P 500 index show that the predictive superiority of decomposing volatility is mainly attributed to the use of these alternative jump tests. In contrast, decomposing volatility by the BNS test fails to provide significantly better forecasts.

[2] Caporin (2022) uses the alternative jump test by Andersen et al. (2007b). But their results are largely based on a group of different modified BNS test. Besides, the test by Corsi et al. (2010) is very similar to the BNS test, with the downward bias corrected. But this test is subject to a pre-specified threshold, which may be delicate when (latent) volatility is time varying. We consider the later Med and Min test as they are free from a pre-specified threshold.

[3] We use the power variation version of the ASJ test as opposed to the threshold version, since the threshold ASJ test has a very limited power to detect jumps (see simulation results in Maneesoonthorn et al., 2020 and Dumitru and Urga, 2012).

1.2 The longer lagged jumps

A large volume of papers decomposes daily lagged volatility only thus ignoring the information on more historical (or longer lagged) jumps (Corsi et al., 2010, Sévi, 2014, Patton and Sheppard, 2015, Clements and Liao, 2017, Ma et al., 2018, Prokopczuk et al., 2016). Are longer-lagged jumps important to volatility forecasting? The following paragraph shows our view on this question.

Daily lagged jumps give a timely measure of agents' perceptions of jump risk. However, it is noteworthy that the daily lagged jumps can also be very noisy since jumps are rare events. As discussed in Tauchen and Zhou (2011), jumps at longer lags may be less noisy. This is because longer lagged jumps estimate the average jumps within a longer previous window. Of course, longer-lagged jumps are less timely. To balance the trade-off between noisiness and timeliness, we suggest not omitting the weekly and monthly lagged jumps for forecasting future volatility. The empirical results for the S&P 500 index support our rationale: the decomposition of weekly and monthly-lagged volatility into jumps and continuous variations provides significantly better out-of-sample forecasts.

1.3 Separating jump signs

There seems a debate on the forecasting value of separating jump signs. Some researchers show that separating jump signs benefits volatility forecasting (Patton and Sheppard, 2015, Duong and Swanson, 2015). While others find that separating jump signs only provides limited forecasting value (Sévi, 2014, Caporin, 2022, Prokopczuk et al., 2016). We contribute to the literature by providing more evidence on this debate. Our results show that separating jump signs benefits predicting volatility, especially for the long-term horizons. Forecasting long-term volatility has some important implications, for example, it benefits option pricing (Ederington and Guan, 2010).

1.4 The jump test bias

Literature shows that the jump tests may be largely biased due to the microstructure noise contamination in the high-frequency sampling. For alleviating this bias, we use the noise-robust version of the jump tests. Moreover, we also consider other versions which immune to the distortion by drifts, intraday volatility pattern, and volatility bursts.

Adjusting the jump tests for alleviating the drift bias is important. For the finite sample, the drift component may harm the jump tests' power (Laurent and Shi, 2020). And the loss of power of the jump tests implies a large underestimation of the jump component (or overestimation of the continuous component). To the best of our knowledge, no paper has associated drift bias with volatility forecasting. We find that accounting for drift bias does not influence our above main findings.

Modifying the jump tests for the intraday volatility pattern is essential, as the jump tests may generate spurious jumps because of the U-shape pattern of intraday volatility. While linking this pattern with volatility forecasting remains a new research question in the literature. Only very recently, Caporin (2022) recently find modifying the jump tests for the intraday volatility pattern has a minor influence on their findings. Our findings are consistent with theirs: accounting for the intraday volatility pattern does not qualitatively change our results.

Our main results are based on the 5-minute frequency sample. However, Christensen et al. (2014) argue that the jump tests generally overestimate the jumps for the 5-minute sampling due to the volatility burst distortions. To alleviate this bias, they suggest testing jumps at ultra-high frequency. We check our results by increasing the frequency to 10-second, based on the volatility estimates and the jump tests adjusted to this frequency. The results for the ultra-high frequency are in line with what we find for the 5-minute frequency.

The remainder of the paper is organized as follows. Section 2 describes various jump tests for dividing volatility into jumps and continuous variation. Section 3 describes the data. Section 4 reports the empirical results for both in-sample and out-of-sample. Section 4 is the robustness checks. Section 5 concludes.

2 The jump tests for decomposing volatility

The volatility decomposition builds on the general theory that the logarithmic prices process follows a jump& drift-diffusion process defined by:

$$dp_t = \mu_t d_t + \sigma_t dW_t + \kappa_t dN_t, 1 \le t \le T,$$

where μ_t is the drift process, σ_t is the càdlàg stochastic (continuous) volatility process, W_t is a standard Brownian motion, κ_t represents the random jump size at time t, and the counting process N_t denotes the jumps amounts up to time t.

Given by the above difference equation, the price variation (V) (squared volatility) of log prices p_t over one trading day t can be decomposed into drift, continuous, and jump variation:

$$V_t = D_t + C_t + J_t,$$

where the $D_t = \int_0^t a_s^2 ds$ denotes the drift variation (D), $C_t = \int_0^t \sigma_{t_s}^2$ denotes the continuous variation (C) and $J_t = \sum_{1<s\le t} \kappa_s^2$stands for the jump variation (J). The drift variation is generally ignorable in the literature thus price variation equals the sum of jump and continuous variations, $V_t = C_t + J_t$. As in Shephard et al. (2008), jump variation can be further decomposed into positive jumps ($J_t^+ = \sum_{1<s\le t} \kappa_s^2 \, I(\kappa_s > 0)$) and negative jumps ($J_t^- = \sum_{1<s\le t} \kappa_s^2 \, I(\kappa_s < 0)$), with signed jumps defined as their difference ($J_t^\Delta = J_t^+ - J_t^-$).

In practice, all components (including V, C, and different Js) need to estimate. For the full variation V, we use a natural Realized Variance (RV) estimator by Andersen et al. (2000). This RV is defined as follows. Assume log prices p_{t_i} are observed for $i = 1, \dots, M$ at day t, with returns $r_{t_i} = p_{t_i} - p_{t_{i-1}}$, $RV_t = \sum_{i=1}^M r_{t_i}^2$. For the observed prices, we sample the prices at the conventional 5-minute frequency.

For estimating the continuous variation and (signed) jumps, we use the BNS test and a range of alternative jump tests. These tests have a variety of forms and transformations. We only choose the one with the best performance or the most widely applied in the literature. we use the ratio version of the BNS test. This version is identical to the one used by Andersen et al. (2007a) and has been shown most powerful based on the simulation analysis by Huang and Tauchen (2005). Besides, we also consider a sequential version of the BNS test by Andersen et al. (2010) (we term this test the s-BNS test), motivated by the fact the s-BNS test locates jumps more precisely (e.g., at the intraday level) than other versions (normally at daily level).

As the Med and Min tests have a similar argument as the BNS test, we again apply the ratio test version, which has been widely applied in the literature (Andersen et al., 2012, Maneesoonthorn et al., 2020, Dumitru and Urga, 2012). We also use the ratio form for the JO test as the simulation study in Jiang and Oomen (2008) indicates the ratio form is superior to other forms, in terms of detection power. Following Dumitru and Urga (2012), we use the ABD and LM tests jointly (ABD-LM) due to their similar structure and performance improvement in size and power.

The market microstructure noise may bias the reliability of these jump tests, for the high-frequency sampling. We use the BNS test that incorporates the staggered returns to alleviate the bias due to microstructure noise contamination. This is the same noise-robust BNS test used in Andersen et al. (2007a) and has been proven to be noise robust replying to the simulation study of Huang and Tauchen (2005). For the noise reduction, we also apply the staggered returns to the Bi-power type spot volatility estimator in the ABD-LM test.

The use of the truncation technique in the Med and Min tests makes it trivial to gain robustness to microstructure noise from the staggered approach (Andersen et al., 2012), especially for 5-minute frequency. Therefore, we discard using the staggered return method for these two tests. Also, we do not use the noise-adjusted version of the ASJ and JO tests, since the noise-robust version of these two tests is specially designed for the ultra-high frequency sampling and thus is subject to power loss for the 5-minute frequency (the loss of power is evidenced by the simulation results in Jiang and Oomen (2008) and Maneesoonthorn et al. (2020)). However, in the robustness section, we will apply these noise robust tests for the robustness checks.

The details of the above jump tests are listed in the appendix, where we also include the procedures for using these tests for separating volatility into (signed) jumps and continuous variation. As in Andersen et al. (2007a) and Dumitru and Urga (2012), we use 0.1% significance level for these jump tests. This lower significance level helps reduce a high number of spurious jumps

3 Data Description

We use the sample of SPRD S&P500 ETF (SPY) from Tick Data Inc. SPY is actively traded and tracks the S&P 500 index. The data is two decades of tick-by-tick records from January 2, 1997, to December 31, 2020 ($T = 6222$ days). SPY Tick data is cleaned as following procedures, following Barndorff-Nielsen et al. (2009), and Patton and Sheppard (2015).

1. Transactions outside of 9:30:00 and 16:00:00 were removed
2. Transactions with a 0 price or volume were removed
3. Retain entries originating from the most active exchange of each day and delete other entries.
4. Only trades with sale conditions ‘E’, ‘F’ or blank (blank, ‘*’, and ‘@’ in tick data) were retained.
5. If multiple transactions have the same timestamp, use the median price.
6. Delete entries with corrected trades

In addition to the realized measures, we obtain an alternative option-based S&P 500 index volatility measure (labelled SV) by Todorov (2019). The SV can be downloaded on www.tailindex.com website, ranging from January 2008 until December 2020. The SV is originally the percentage of annualized volatility, and we transform it to daily level variance to be consistent with the realized measures. As discussed by Andersen et al. (2021), the SV is constructed exclusively from option prices, and thus is void of the specific form for noise structure present in the high-frequency asset prices.

Table 1 reports the results for the statistical description. The upper panel shows the proportion of days in which jumps are not zero, across the jump tests. The empirical ranks are consistent with previous research (Dumitru and Urga, 2012, Maneesoonthorn et al., 2020, Bajgrowicz et al., 2016): a) the BNS test ranks higher than the Med and Min test; a) the Med test ranks higher than the Min test; b) ABD and LM tests rank higher than other tests; d) the JO test ranks higher than the BNS test.

Table 1. The proportion of jump days and the average proportion of jumps relative to RV

The proportion of jump days				
	$T^{-1}\sum_1^T I(J_t^\Delta \neq 0)$	$T^{-1}\sum_1^T I(J_t^+ \neq 0)$	$T^{-1}\sum_1^T I(J_t^- \neq 0)$	
BNS	0.042	0.022	0.020	
Med	0.031	0.016	0.016	
Min	0.010	0.006	0.005	
ASJ	0.022	0.011	0.011	
JO	0.048	0.026	0.022	
ABD-LM	0.073	0.030	0.043	
s-BNS	0.042	0.021	0.021	
The average proportion of jumps relative to RV				
	$\frac{1}{T}\sum_1^T(J_t/RV_t)$	$\frac{1}{T}\sum_1^T(J_t^\Delta/RV_t)$	$\frac{1}{T}\sum_1^T(J_t^+/RV_t)$	$\frac{1}{T}\sum_1^T(J_t^-/RV_t)$
BNS	1.502	0.260	0.788	-0.528
Med	1.364	0.178	0.584	-0.406
Min	0.568	0.080	0.241	-0.161
ASJ	0.331	0.021	0.245	-0.224
JO	1.265	0.378	1.295	-0.917
ABD-LM	2.267	-0.277	0.929	-1.207
s-BNS	1.228	0.037	0.537	-0.499

The lower panel of Table 1 reports the averaged ratio for jump variation relative to the RV. The results for the jump proportion across the jump tests (first column) are consistent with Dumitru and Urga (2012): there is a higher jump size for the ABD- LM test than those of the BNS, JO, Med, and Min tests.

4 Empirical analysis

4.1 Volatility forecasting setup and models

For investigating the predictive value of jumps, we use the popular HAR model by Corsi (2009):

$$V_{t+h-1|t} = \beta_0 + \beta_d\, RV_{t-1} + \beta_w\, RV_{t-1|t-5} + \beta_m\, RV_{t-1|t-22} + \varepsilon_t, (1)$$

while $V_{t+h-1|t}$ is the ex-post realized value of the volatility proxy we seek to forecast over [t+1, t+h]. Often, $V_{t+h-1|t} = RV_{t+h-1|t}$, but we also forecast the nonparametric option-based volatility by Todorov (2019)[4], as in Andersen et al. (2021). This volatility measure, unlike RV, is robust to jumps and is also less sensitive to microstructure noise.

The basic HAR model has subsequently been modified in numerous ways. In particular, Patton and Sheppard (2015) split the daily lagged RV of the model into continuous and jump components,

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,j}\, J_{t-1} + \beta_w\, RV_{t-1|t-5} + \beta_m\, RV_{t-1|t-22} + \varepsilon_t, (2)$$

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,J^\Delta} J^{\Delta}_{t-1} + \beta_w\, RV_{t-1|t-5} + \beta_m\, RV_{t-1|t-22} + \varepsilon_t, (3)$$

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,J^+}\, J^{+}_{t-1} + \beta_{d,J^-}\, J^{-}_{t-1} + \beta_w\, RV_{t-1|t-5} + \beta_m\, RV_{t-1|t-22} + \varepsilon_t. (4)$$

And we term the above three daily decomposed models the $\mathrm{C_d J_d}$, $\mathrm{C_d J_d^{\Delta}}$, and $\mathrm{C_d J_d^{\pm}}$ model, where the subscript d indicates the daily lag. Andersen et al. (2007) and Corsi and Renò (2012) suggest also including the weekly and monthly lagged jumps,

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,j}\, J_{t-1} + \beta_w\, C_{t-1|t-5} + \beta_{w,j}\, J_{t-1|t-5} + \beta_m\, C_{t-1|t-22} + \beta_{m,j}\, J_{t-1|t-22} + \varepsilon_t, (5)$$

[4] Following Andersen et al. (2021), we scaled SV so that its average value coincides with that of the ex-post TV measure:
$TV_t = \frac{1}{2}\sum_{i=1}^{M}(r_i)^2\, I\left(|r_i| \le 3\hat{\sigma}_t^{med}/\sqrt{M}\right)$, with $\left(\hat{\sigma}_t^{med}\right)^2 = \frac{\pi}{\pi+6-4\sqrt{3}}\left(\frac{M}{M-2}\right)\sum_{i=3}^{M}\mathrm{med}\left(|r_{t_{i-2}}|, |r_{t_{i-1}}|, |r_{t,i}|\right)^2$.

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,J^{\Delta}} J^{\Delta}_{t-1} + \beta_w\, C_{t-1|t-5} + \beta_{w,J^{\Delta}} J^{\Delta}_{t-1|t-5} + \beta_m\, C_{t-1|t-22} + +\beta_{m,J^{\Delta}} J^{\Delta}_{t-1|t-22} + \varepsilon_t, (6)$$

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,J^+}\, J^+_{t-1} + \beta_{d,J^-}\, J^-_{t-1} + \beta_w\, C_{t-1|t-5} + \beta_{w,J^+}\, J^+_{t-1|t-5} + \beta_{w,J^-}\, J^-_{t-1|t-5} + \beta_m\, C_{t-1|t-22} + \beta_{m,J^+}\, J^+_{t-1|t-22} + \beta_{m,J^-}\, J^-_{t-1|t-22} + \varepsilon_t, (7)$$

We term the above three fully decomposed models the $\mathrm{C_{d-m}J_{d-m}}$, $\mathrm{C_{d-m}J^{\Delta}_{d-m}}$, and $\mathrm{C_{d-m}J^{\pm}_{d-m}}$ model, where the $\mathrm{d-m}$ subscripts indicate the daily, weekly and monthly lags.

We use the Ordinary Least Square (OLS) method for estimating the above models. The coefficient statistical inference depends on the Heteroskedasticity- and autocorrelation-consistent (HAC) robust t-statistics by Newey and West (1987), with the bandwidth HAC equal to 2(h +1) as in Corsi and Renò (2012), where h is the lead length of the left-hand-side variable.

For the pseudo-out-of-sample, we forecast one-day, one-week, and one-month ahead volatility as in Corsi (2009), and we also consider three-month ahead horizons (h=66) as in Patton and Sheppard (2015). In addition, we make forecasts based on a 1000-rolling window (RW) and an expanding window (IW) based on the initial 1000 observations. To avoid abnormal forecasts (e.g., negative volatility forecasts), we apply the "insanity filter" suggested by Bollerslev et al. (2016) and Patton and Sheppard (2015). In detail, the "insanity filter" algorithm replaces any forecast falling outside the range of values of the target variable observed during the estimation period by the unconditional mean of the variable over that period.

The comparison of forecasting accuracy also, of course, requires a metric for measuring accuracy. Patton (2011a) shows that QLIKE is an unbiased loss function.

And this loss function is widely applied in key research (Andersen et al., 2021, Bollerslev et al., 2016). Thus, we measure forecast accuracy by the QLIKE loss function [5], defined as,

$$\text{QLIKE}(V_t, F_t) = \frac{V_t}{F_t} - \ln\frac{V_t}{F_t} - 1,$$

The statistical significance difference in forecasting precision is evaluated via the Model Confidence set (MCS) of Hansen et al. (2011) [6]. The MCS procedure consists of a sequence of tests which permits the construction of a set of "superior" models, where the null hypothesis of Equal Predictive Ability (EPA) is not rejected at a certain confidence level. The model in the superior model set, $\widehat{\mathcal{M}}^*$ significantly outperforms the models outside this set $\widehat{\mathcal{M}}^0$. Since the MCS should be used with caution when forecasts are based on estimated parameters and models are nested (Hansen et al., 2011), we complement a 90% confidence level MCS. This indicates that a model is in the superior set only if its p-value is larger than 10%.

For ease of comparison for a large group of paired models, the statistical significance is evaluated via the Diebold–Mariano-West (DMW) statistic [7] developed by Diebold and Mariano (1995) and West (1996), with adjustment to the Newey–West Heteroskedasticity and Autocorrelation Corrected (HAC) standard errors.

5 Simulation based evidence by Patton and Sheppard (2009) suggests the use of QLIKE rather than MSE due to the superior power of QLIKE in Diebold and Mariano (1995) and West (1996) type tests for equal predictive accuracy (EPA). But we confirm that the MSE results are qualitatively consistent. To save space, we do not report these results.
6 The MCS results presented here were obtained using the mcs function from the Oxford MFE Toolbox developed by Kevin Sheppard, https://www.kevinsheppard. com/code/matlab/mfe-toolbox/. We implement MCS by the SQ approach with 1000 stationary bootstrap replications and the average block size equal to 10). The results are consistent for an alternative R approach.
7 The DMW results in this paper were obtained using the robust_loss_1 function from Andrew Patton's Matlab code page, http://public.econ.duke.edu/~ap172/

4.2 Alternative jump tests

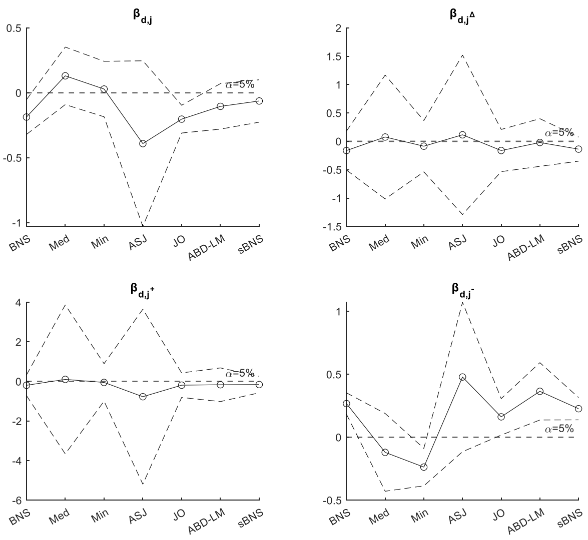
The purpose of this section is to explore whether the above more reliable alternative jump tests also provide better volatility forecasts than the BNS test. We begin the analysis by the in-sample estimation. Figure 1 depicts the coefficient estimates of (signed) jumps in the $\mathrm{C_dJ_d}$, $\mathrm{C_dJ_d^{\Delta}}$, and $\mathrm{C_dJ_d^{\pm}}$ models, across different jump tests, for the one-day-ahead forecast ($h = 1$). The upper panel shows the results for the RV forecast. As the panel shows, the impact of (signed) jumps is generally not highly significant, across the BNS and other BNS-alternative tests. Moreover, the lower panel for SV forecast again show limited in-sample evidence of (signed) jumps across these jump tests. In addition, unreported results show that the coefficient of the continuous variation is overwhelmingly significant for all the above cases. Overall, the limited (strong) in-sample evidence of jumps (continuous variation) is consistent with Andersen et al. (2007a), and the BNS-alternative jump tests appear to not make differences in the impact of jumps.

We then investigate the out-of-sample forecast. Table 2 compares the HAR model with the daily decomposed models by various jump tests, for the RV forecast. The top, middle, and bottom panel reports the results for the $\mathrm{C_dJ_d}$, $\mathrm{C_dJ_d^{\Delta}}$, and $\mathrm{C_dJ_d^{\pm}}$ models, respectively. The numbers in the table indicate the loss ratios relative to the HAR model. For each row, the lowest ratio is indicated in bold while the models included in the superior set $\widehat{\mathcal{M}}^*$ is indicated in asterisks.

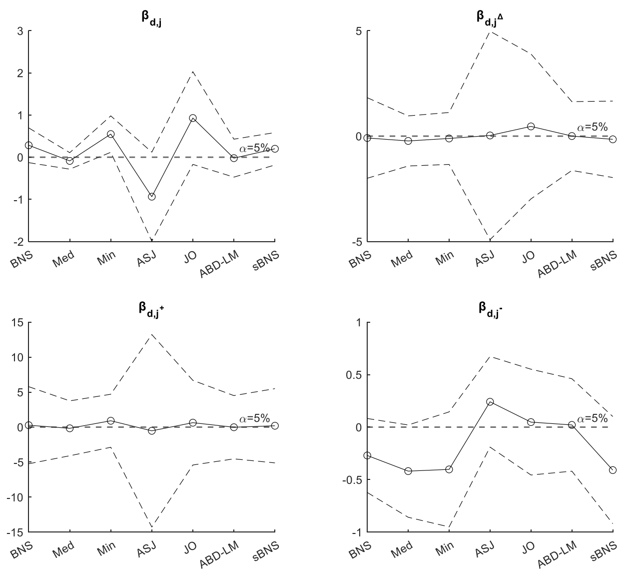
As the table shows, the HAR model is substantially excluded from the superior set $\widehat{\mathcal{M}}^*$ thus is significantly outperformed by some $\mathrm{C_dJ_d}$, $\mathrm{C_dJ_d^{\Delta}}$, and $\mathrm{C_dJ_d^{\pm}}$ models. Comparing the models across different jump tests, we find the models associated with the BNS-alternative tests generally perform better. For example, the lowest loss is always provided by the model with the BNS-alternative tests. Further, the model associated with the ABD-LM test has the lowest loss and significantly outperforms the model with the BNS test, for the daily forecast (h=1), for the $C_dJ_d^{\Delta}$, and $C_dJ_d^{\pm}$ models, and for both RW and IW forecasts. Moreover, for these different cases, the model with the s-BNS test systematically has lower loss and sometimes provides significantly better forecasts against the BNS test. The superiority of the s-BNS test indicates the importance of locating intraday jumps for volatility forecasting.

Figure 1. The coefficient estimates of jumps in the C_dJ_d, $C_dJ_d^{\Delta}$, and $C_dJ_d^{\pm}$ models, for the daily volatility forecast.

Panel A: RV forecast



Panel B: SV forecast



Note: the solid lines denote the coefficient estimates while the dashed lines indicate the 95% confidence intervals.

Table 2. Out-of-sample loss ratios for the daily decomposed models for RV forecasts

			BNS	Med	Min	ASJ	JO	ABD-LM	sBNS
		HAR				C_dJ_d			
h=1	*RW*	1.000*	1.004*	0.995	1.001*	0.955*	**0.869***	0.873*	0.985*
	IW	1.000	0.962	0.998	1.002	1.003	0.969	**0.947***	0.960
h=5	*RW*	1.000	0.997	1.001	1.000	**0.972***	0.983*	1.016	0.992*
	IW	1.000	0.976	0.998	1.001	1.001	0.982	**0.965***	0.973
h=22	*RW*	1.000*	1.004*	0.995*	1.000*	0.999*	0.999*	**0.982***	1.024*
	IW	1.000	0.990*	1.000	1.000	1.000	0.992	**0.987***	0.989*
h=66	*RW*	1.000*	1.004*	1.004*	1.003*	**0.997***	1.002*	1.006*	1.001*
	IW	1.000	0.997	1.000	1.000	1.000	0.998	**0.996***	0.996*
		HAR				$C_dJ_d^{\Delta}$			
h=1	*RW*	1.000	0.989	1.003	0.999	0.947*	0.885*	**0.870***	0.987
	IW	1.000	0.960	0.997	1.002	1.002	0.965	**0.944***	0.957
h=5	*RW*	1.000	0.996	0.997	1.001	**0.963***	1.007	1.013	0.996
	IW	1.000	0.975	0.996	1.000	1.000	0.977	**0.965***	0.971
h=22	*RW*	1.000*	1.002*	**0.994***	0.999*	0.998*	1.002*	0.994*	0.999*
	IW	1.000	0.988*	0.999	0.999	0.999	0.990	0.990*	**0.986***
h=66	*RW*	1.000	0.998*	**0.989***	1.001	0.994*	0.999*	0.996*	0.998*
	IW	1.000	0.997	1.000	1.000	1.000	0.998	0.997	**0.996***
		HAR				$C_dJ_d^{\pm}$			
h=1	*RW*	1.000	1.002	0.997	1.002	0.941*	0.881*	**0.868***	0.985
	IW	1.000	1.028	0.997	1.002	1.002	0.968	**0.944***	0.959
h=5	*RW*	1.000	0.994	0.997	1.001	**0.966***	1.000	1.026	0.993
	IW	1.000	0.977	0.996	1.000	1.003	0.978	**0.969***	0.973*
h=22	*RW*	1.000*	1.005*	**0.993***	1.001*	1.001*	1.001*	1.000*	1.013*
	IW	1.000	0.988*	1.000	0.999	0.999	0.990*	0.991*	**0.987***
h=66	*RW*	1.000*	1.003*	**0.995***	1.002*	0.999*	0.998*	0.999*	1.003*
	IW	1.000	0.997	1.001	1.001	1.001	0.998	0.998	**0.996***

Notes: The top, middle, and bottom panel reports the results for the C_dJ_d, $C_dJ_d^{\Delta}$, and $C_dJ_d^{\pm}$ models, respectively. The numbers in the table indicate the loss ratios relative to the benchmark HAR model, with the smallest ratio at each row in bold. Asterisks indicate the models included in the superior set $\widehat{\mathcal{M}}^*$.

Table 3. Out-of-sample loss ratios for the daily decomposed models for SV forecasts

			BNS	Med	Min	ASJ	JO	ABD-LM	sBNS
		HAR				C_dJ_d			
h=1	*RW*	1.000	1.006	1.002	1.002	**0.947***	1.009	0.979*	1.008
	IW	1.000	1.001	0.996	0.997	0.999	**0.989***	0.990*	1.001
h=5	*RW*	1.000	1.003	0.999	0.999	**0.964***	0.997	1.001	1.003
	IW	1.000	0.999	**0.993***	0.999	0.999	0.996	0.994*	1.000
h=22	*RW*	1.000	1.011	1.000	0.999	0.982*	0.999	**0.970***	1.009
	IW	1.000	0.998*	**0.998***	0.999*	1.001*	0.997*	1.001	0.999*
h=66	*RW*	1.000	1.007	1.007	1.002	**0.993***	1.016	0.995*	1.004
	IW	1.000*	1.000*	1.002*	1.000*	1.000*	**0.995***	1.000*	1.000*
		HAR				$C_dJ_d^{\Delta}$			
h=1	*RW*	1.000	1.005	1.004	1.008	**0.943***	1.014	**0.906***	1.010
	IW	1.000	1.002	0.994	1.003	1.000	1.011	**0.985***	1.002
h=5	*RW*	1.000	1.003	0.998	1.001	**0.960***	0.998	0.981*	1.003
	IW	1.000	1.002	**0.991***	1.001	0.999	1.002	**0.990***	1.000
h=22	*RW*	1.000	1.005	1.000	0.999	0.982*	0.997	**0.969***	1.008
	IW	1.000	1.003	**0.996***	1.001	0.998*	1.001	1.000	1.000*
h=66	*RW*	1.000	0.996*	0.987*	1.002	**0.993***	1.002*	1.005	0.999*
	IW	1.000*	1.003*	1.004*	1.002*	**0.998***	0.999*	1.002*	1.003*
		HAR				$C_dJ_d^{\pm}$			
h=1	*RW*	1.000	1.007	0.999	1.005	**0.933***	1.016	0.990*	1.012
	IW	1.000	1.001	0.993	0.996	1.000	0.990*	**0.986***	1.000
h=5	*RW*	1.000	1.003	0.999	1.000	**0.958***	0.998	1.008	1.002
	IW	1.000	1.000	**0.991***	0.999	0.998	0.993*	0.992*	0.999
h=22	*RW*	1.000	1.009	0.998	0.999	0.985*	1.003	**0.973***	1.011
	IW	1.000	1.001	0.996*	0.999	0.999	**0.994***	1.010	0.998*
h=66	*RW*	1.000*	0.998*	**0.996***	1.002*	0.999*	0.997*	1.001*	1.003*
	IW	1.000	1.000	1.003	1.001	0.999*	**0.992***	0.999*	1.001

Notes: The top, middle, and bottom panel reports the results for the C_dJ_d, $C_dJ_d^{\Delta}$, and $C_dJ_d^{\pm}$ models, respectively. The numbers in the table indicate the loss ratios relative to the benchmark HAR model, with the smallest ratio at each row in bold. Asterisks indicate the models included in the superior set $\widehat{\mathcal{M}}^*$.

In addition, we also check the out-of-sample for the SV forecast, reported in Table 3. The results reveal even stronger evidence [8]. Specifically, the HAR model, with only very few exceptions for h=66, is overwhelmingly cut off from the superior set $\widehat{\mathcal{M}}^*$, thus is clearly outperformed by the C_dJ_d, $C_dJ_d^{\Delta}$, and $C_dJ_d^{\pm}$ models. In addition, the superiority

[8] The discrepancies between Table 2 and Table 3 also reflect the different sample periods, as the SV forecasts are initiated only during the great financial crisis of 2008-2009. However, we confirm that the difference remains substantial, even if we generate the forecasts over the identical time period.

of these decomposed models is almost achieved by the BNS alternatives. Specifically, the model with the BNS test generally fails to beat the HAR model, with most loss ratios greater than one. In contrast, the model associated with the alternative jump tests (e.g., Med, ASJ, JO, or LM) always provides the lowest ratio and is substantially included in the superior set $\widehat{\mathcal{M}}^*$, across all of these scenarios. However, the model with the sequential BNS test somewhat makes no obvious difference. This indicates that this sequential version is not able to increase the predictive value of the BNS test for the SV forecast.

Overall, we find the BNS-alternative jump tests are more important than the BNS test, for both SV and RV forecasts. For the remaining analysis in this paper, we only report the results for the SV forecast in the interest of brevity, with the results for the RV forecast (qualitatively similar but with a clear-cut for short horizons) reported in the appendix.

4.3 The longer lagged jumps

This section aims to study whether the decomposition of longer-lagged RV is important for volatility forecasting. To explore this, we begin by presenting the in-sample estimation results for the fully decomposed models, including $\mathrm{C_{d-m}J_{d-m}}$, $\mathrm{C_{d-m}J^{\Delta}_{d-m}}$, and $\mathrm{C_{d-m}J^{\pm}_{d-m}}$ models. Figure 2 shows the jump coefficients for these three models for SV forecasts. As the figure shows, across all different jumps, tests, and lags, the impact of jumps tends to be statistically insignificant. These results are again in line with Andersen et al. (2007a). Table 4 compares the above fully decomposed models (lower panel) with the daily decomposed models (upper panel), in terms of the in-sample goodness fits, for the SV forecast. The evaluation criteria rely on the R^2 ratio for the competing models relative to the HAR model. Comparing these two panels, the improvement of R^2 in the upper panel is generally minor (only around 0.1%) while this improvement in the lower panel is much more prominent (e.g., for the Med test R^2 increase by 3.1%). This finding indicates that the decomposing weekly and monthly lagged RV leads to the much better in-sample performance of the models. However, whether the in-sample improvement is of practical importance remains an empirical question to be answered in the out-of-sample forecasting exercise below. Besides, the results for both panels also imply that the in-sample advantage of these alternative tests is robust to both daily and fully decomposed model types. Specifically, for both panels, the models with the alternative jump tests (except for the sequential test) systematically have a greater R^2 than that with the BNS test.

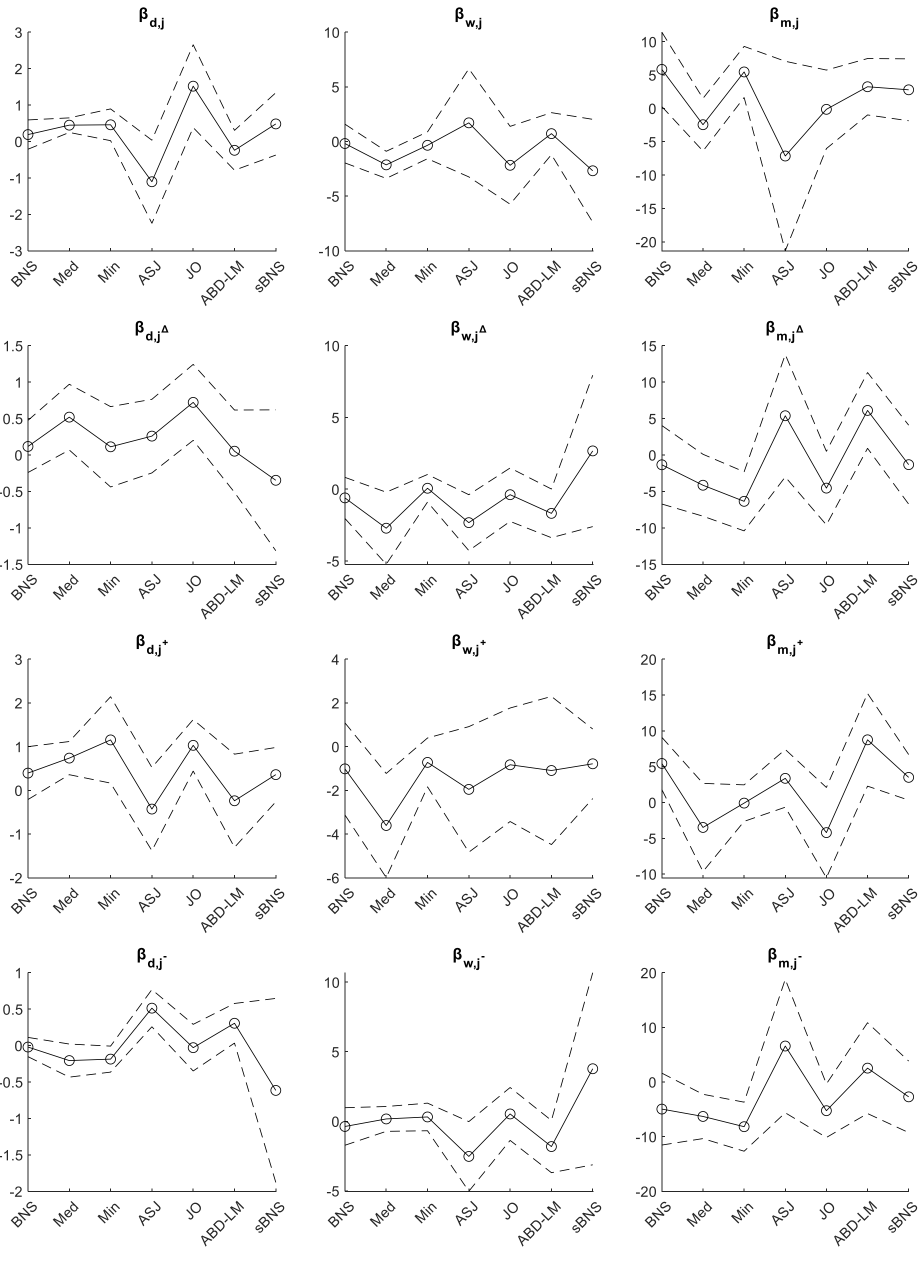
Table 4. Full in-sample R^2 for the models associated with different jump tests for SV forecast.

	BNS	Med	Min	ASJ	JO	ABD-LM	sBNS
			Daily decomposed models				
$C_d J_d$	1.0000	1.0013	1.0003	**1.0025**	1.0016	1.0008	1.0000
$C_d J_d^{\Delta}$	0.9999	1.0016	0.9996	1.0008	**1.0024**	1.0008	1.0000
$C_d J_d^{\pm}$	1.0001	1.0017	1.0003	1.0020	**1.0034**	1.0008	1.0001
			Fully decomposed models				
$C_{d-m} J_{d-m}$	1.0018	**1.0295**	1.0027	1.0074	1.0052	1.0098	1.0007
$C_{d-m} J_{d-m}^{\Delta}$	1.0000	**1.0289**	1.0027	1.0140	1.0134	1.0200	1.0001
$C_{d-m} J_{d-m}^{\pm}$	1.0029	**1.0317**	1.0045	1.0167	1.0164	1.0288	1.0016

Notes: This table reports the R^2 ratio for the competing models relative to the HAR model. The greatest ratio in each row is indicated in **bold**.

We next investigate the out-of-sample results. For the out-of-sample analysis, we compare the fully decomposed models with the daily decomposed models, as reported in Table 5. The results are for the daily forecast, but we confirm that the results are similar for longer horizons. The left panel compares the $C_d J_d$ model with the $C_{d-m} J_{d-m}$ model, via their loss ratios relative to that of the HAR model, across different jump tests and both rolling and increasing windows. And the asterisks indicate the models included in the superior set $\widehat{\mathcal{M}}^*$ among these three models.

Figure 2. The jump coefficient estimates for the $C_{d-m}J_{d-m}$, $C_{d-m}J^{\Delta}_{d-m}$, and $C_{d-m}J^{\pm}_{d-m}$ models for daily SV forecast.



Note: the solid lines denote the coefficient estimates while the dashed lines indicate the 95% confidence intervals.

As the left panel shows, the $C_{d-m}J_{d-m}$ model generally win the lowest ratio and are substantially included in the superior set $\widehat{\mathcal{M}}^*$, across these different cases. The middle panel for comparing the $C_dJ_d^{\Delta}$ model with the $C_{d-m}J_{d-m}^{\Delta}$ model is somewhat mixed: the $C_{d-m}J_{d-m}^{\Delta}$ model with the BNS, Med, and s-BNS tests, fail to beat the HAR and the $C_dJ_d^{\Delta}$ model with the equivalent tests while the results of the $C_{d-m}J_{d-m}^{\Delta}$ model with other tests is still consistent. The right panel for comparing the $C_dJ_d^{\pm}$ model with the $C_{d-m}J_{d-m}^{\pm}$ the model reveals a cleaner picture: the $C_{d-m}J_{d-m}^{\pm}$ model exhibits the lowest loss for almost all of these different cases. The above results are for the SV forecast, but we confirm that the (unreported) results for RV are qualitatively very similar, with a clear-cut for short-term forecasts.

In addition, we also check whether the out-of-sample advantage of alternative jump tests is robust in the case of the fully decomposed models. Table 6 compares the HAR model with the $C_{d-m}J_{d-m}$, $C_{d-m}J_{d-m}^{\Delta}$, and $C_{d-m}J_{d-m}^{\pm}$ models by various jump tests, in terms of the out-of-sample performance. The results again reveal that the models with the alternative jump tests perform better: the models with the BNS struggle to beat the HAR model and are substantially excluded from the superior model set $\widehat{\mathcal{M}}^*$, while the model with the alternative tests always has the lowest loss and is included in the set $\widehat{\mathcal{M}}^*$. These results imply that the predictive advantage of alternative jump tests found in Table 3 is robust when the models are fully decomposed into jumps and continuous variation.

Table 5. Comparing the daily decomposed models with the fully decomposed models for daily SV forecast

		C_dJ_d - $C_{d-m}J_{d-m}$			$C_dJ_d^{\Delta}$- $C_{d-m}J_{d-m}^{\Delta}$			$C_dJ_d^{\pm}$- $C_{d-m}J_{d-m}^{\Delta\pm}$		
		HAR	C_dJ_d	$C_{d-m}J_{d-m}$	HAR	$C_dJ_d^{\Delta}$-	$C_{d-m}J_{d-m}^{\Delta}$	HAR	$C_dJ_d^{\pm}$-	$C_{d-m}J_{d-m}^{\Delta\pm}$
BNS	*RW*	**1.000***	1.006	1.044	**1.000***	1.005*	1.086*	**1.000***	1.007	1.081
	IW	1.000	1.001	**0.965***	**1.000***	1.002	1.019	1.000	1.001	**0.952***
Med	*RW*	1.000	1.002	**0.967***	1.000*	1.004*	**0.993***	1.000	0.999	**0.956***
	IW	1.000	0.996	**0.938***	1.000	0.994*	**0.958***	1.000	0.993	**0.935***
Min	*RW*	1.000	1.002	**0.983***	**1.000***	1.008	1.032	1.000*	1.005*	**0.996***
	IW	1.000	0.997	**0.947***	**1.000***	1.003	1.016	1.000	0.996	**0.968***
ASJ	*RW*	1.000	0.947	**0.881***	1.000	0.943	**0.910***	1.000	0.933*	**0.924***
	IW	1.000*	**0.999***	1.001*	1.000	1.000	**0.987***	1.000	1.000	**0.982***
JO	*RW*	**1.000***	1.009*	1.028	1.000	1.014	**0.961***	1.000*	1.016	**0.966***
	IW	1.000	0.989	**0.966***	1.000	1.011	**0.962***	1.000	0.990	**0.885***
ABD-LM	*RW*	1.000*	**0.979***	0.992*	1.000*	**0.906***	1.211*	1.000*	**0.990***	1.320*
	IW	1.000	**0.990***	1.065	1.000	0.985	**0.864***	1.000	0.986*	**0.904***
sBNS	*RW*	**1.000***	1.008	1.056	**1.000***	1.010	1.098	**1.000***	1.012	1.087
	IW	1.000*	1.001	**0.978***	**1.000***	1.002	1.017	1.000*	1.000*	**0.978***

Table 6. Out-of-sample results for the fully decomposed model for SV forecast.

			BNS	Med	Min	ASJ	JO	LM	sBNS
		HAR				CJ_{d-m}			
h=1	*RW*	1.000	1.044	0.967	0.983	**0.881***	1.028	1.003	1.056
	IW	1.000	0.965*	**0.938***	0.947*	1.001	0.966*	0.940*	0.978
h=5	*RW*	1.000	1.037	0.990	0.990	**0.928***	1.009	0.983*	1.050
	IW	1.000	0.969	0.951*	0.954*	0.989	0.953*	**0.916***	0.980
h=22	*RW*	1.000*	1.047	1.030*	0.996*	**0.992***	1.096	1.068*	1.042
	IW	1.000	0.997	0.982*	0.986*	0.991*	0.974*	**0.958***	1.000
h=66	*RW*	1.000*	1.015*	0.974*	0.994*	0.981*	0.965*	**0.955***	1.005*
	IW	1.000	0.989	1.029*	0.996	0.999	**0.920***	0.952*	0.995
		HAR				CJ^{Δ}_{d-m}			
h=1	*RW*	1.000	1.086	0.993	1.032	**0.910***	0.961	1.450	1.098
	IW	1.000	1.019	**0.958***	1.016	0.987	0.962*	1.006	1.017
h=5	*RW*	1.000	1.062	0.992	1.027	**0.930***	0.979	1.295	1.056
	IW	1.000	1.022	**0.939***	1.012	0.979	0.982	1.043	1.020
h=22	*RW*	1.000	0.998*	0.995*	1.010	**0.973***	1.034	1.166	1.004
	IW	1.000	1.021	**0.969***	1.001	0.991	1.027	1.027	1.018
h=66	*RW*	1.000*	0.848*	**0.734***	0.948*	0.979*	0.845*	0.956*	0.901*
	IW	1.000	1.033	1.095	1.026	0.989*	**0.974***	0.979*	1.022
		HAR				$CJ^{\pm}_{d-m}$			
h=1	*RW*	1.000	1.081	0.956*	0.996	**0.924***	0.966*	1.114	1.087
	IW	1.000	0.952	0.935*	0.968	0.982	**0.885***	0.916*	0.978
h=5	*RW*	1.000	1.107	0.981	0.995	**0.943***	0.969*	1.198	1.065
	IW	1.000	0.959	0.938*	0.973	0.968	**0.895***	0.905*	0.983
h=22	*RW*	1.000*	1.036	1.011*	1.000*	**0.990***	1.033*	1.155	1.063
	IW	1.000	0.994*	0.969*	0.987*	0.988*	**0.958***	0.960*	1.004
h=66	*RW*	1.000*	0.940*	0.842*	0.922*	0.980*	**0.838***	0.882*	0.986*
	IW	1.000	0.974	1.091	1.007	0.992	**0.922***	0.924*	0.993

Notes: The competing model is CJ_{d-m}, CJ^{Δ}_{d-m}, and $CJ^{\pm}_{d-m}$ model while the benchmark model is the HAR model.

4.4 Separating jump signs

The aim of this section is to investigate whether separating jump signs benefit volatility forecasting. For this investigation, we compare the CJ model with the CJ^{Δ} and $CJ^{\pm}$ models. The (unreported) daily SV forecast results are mixed across these different scenarios: only for the cases of the Med test, the CJ^{Δ} and $CJ^{\pm}$ models outperform the CJ model, while for the cases of other tests, there is no clear evidence of one model type dominating another.

Table 7. Comparing the CJ model with the CJ^{Δ} and $CJ^{\pm}$ models for seasonal SV forecast (h=66).

			HAR	CJ	CJ^{Δ}	$CJ^{\pm}$
BNS	*Daily*	*RW*	1.000*	1.007	**0.996***	0.998*
		IW	**1.000***	1.000*	1.003*	1.000*
	Fully	*RW*	1.000*	1.015*	**0.848***	0.940*
		IW	1.000*	0.989*	1.033*	**0.974***
Med	*Daily*	*RW*	1.000*	1.007*	**0.987***	0.996*
		IW	**1.000***	1.002*	1.004*	1.003*
	Fully	*RW*	1.000*	0.974*	**0.734***	0.842*
		IW	**1.000***	1.029*	1.095*	1.091*
Min	*Daily*	*RW*	**1.000***	1.002*	1.002*	1.002*
		IW	**1.000***	1.000*	1.002*	1.001*
	Fully	*RW*	1.000	0.994	0.948	**0.922***
		IW	1.000*	**0.996***	1.026*	1.007*
ASJ	*Daily*	*RW*	1.000	0.993*	**0.993***	0.999*
		IW	1.000	1.000	**0.998***	0.999*
	Fully	*RW*	1.000	0.981*	**0.979***	0.980*
		IW	1.000	0.999	**0.989***	0.992*
JO	*Daily*	*RW*	1.000*	1.016*	1.002*	**0.997***
		IW	1.000	0.995*	0.999	**0.992***
	Fully	*RW*	1.000*	0.965*	0.845*	**0.838***
		IW	1.000	**0.920***	0.974	0.922*
ABD-LM	*Daily*	*RW*	**1.000***	0.995*	1.005*	1.001*
		IW	1.000*	1.000*	1.002*	**0.999***
	Fully	*RW*	1.000	0.940*	0.947*	**0.884***
		IW	1.000	0.997	0.930*	**0.876***
sBNS	*Daily*	*RW*	1.000*	1.004	**0.999***	1.003*
		IW	**1.000***	1.000*	1.003*	1.001*
	Fully	*RW*	1.000*	1.005*	**0.901***	0.986*
		IW	1.000*	0.995*	1.022*	**0.993***

Notes: This table reports the loss ratios of CJ, CJ^{Δ} and $CJ^{\pm}$ models relative to the HAR model, across various jump tests, both daily and fully decomposed model versions, and both forecasting windows. "Daily" indicates that the competing models are based on the daily decomposed model specification while "Fully" indicates that the competing models are based on the fully decomposed model specification.

However, the long-term SV forecast (h=66), reported in Table 7, exhibits a much cleaner picture: the CJ^{Δ} and $CJ^{\pm}$ models substantially outperform the CJ model. Moreover, unreported results show that the predictive superiority of jump signs holds for long-term RV forecasts. The finding that separating jump signs is important for long-term volatility forecasting is consistent with Patton and Sheppard (2015).

5 Robustness analysis

The robustness checks involve alternative benchmark models, drift bias and intraday volatility pattern, and ultra-high frequency sampling.

5.1 Exploring the contribution of jumps via alternative benchmark models

The limited in-sample evidence of jumps reported in section 4 indicates that jumps weakly affect future volatility, which motivates that the jump component should be excluded from the model, as in Bollerslev et al. (2016) and Andersen et al. (2021). However, the null in-sample evidence does not necessarily imply that jumps have no contribution to the out-of-sample, which is the primary interest of this paper. Based on alternative benchmark models, this section studies whether jumps should be excluded or not for the out-of-sample. To explore the contribution of daily lagged jumps, we compare the competing $\mathrm{C_d J_d}$, $\mathrm{C_d J_d^{\Delta}}$, and $\mathrm{C_d J_d^{\pm}}$ models with the benchmark $\mathrm{C_d}$ model by Patton and Sheppard (2015),

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_w\, RV_{t-1|t-5} + \beta_m\, RV_{t-1|t-22} + \varepsilon_t. \quad (8)$$

Setting the $\mathrm{C_d}$ as the benchmark model facilitates investigating whether jumps provide an incremental value, as the $\mathrm{C_d}$ model only excludes jumps from the three competing models.

To investigate the importance of longer lagged jumps, we compare the competing $\mathrm{C_{d-m} J_{d-m}}$, $\mathrm{C_{d-m} J_{d-m}^{\Delta}}$, and $\mathrm{C_{d-m} J_{d-m}^{\pm}}$ models with the $\mathrm{C_{d-m} J_d}$, $\mathrm{C_{d-m} J_d^{\Delta}}$, and $\mathrm{C_{d-m} J_d^{\pm}}$ models, respectively. The latter three models are the benchmark models and are defined by Corsi et al. (2010),

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,j}\, J_{t-1} + \beta_w\, C_{t-1|t-5} + \beta_m\, C_{t-1|t-22} + \varepsilon_t, \quad (9)$$

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,J^{\Delta}} J_{t-1}^{\Delta} + \beta_w\, C_{t-1|t-5} + \beta_m\, C_{t-1|t-22} + \varepsilon_t, \quad (10)$$

$$V_{t+h-1|t} = \beta_0 + \beta_d\, C_{t-1} + \beta_{d,J^+}\, J_{t-1}^{+} + \beta_{d,J^-}\, J_{t-1}^{-} + \beta_w\, C_{t-1|t-5} + \beta_m\, C_{t-1|t-22} + \varepsilon_t. \quad (11)$$

The only difference between these two groups of models is the inclusion of weekly and monthly jump lags, which facilitates studying the value of longer lagged jumps.

Table 8 reports the daily out-of-sample forecasting results for the comparisons of the above models. Since the competing models and the benchmark models are compared in a large number of pairs, we report the DMW statistics for presentation purposes. The left panel shows the DMW statistics of the C_dJ_d, $C_dJ_d^{\Delta}$, and $C_dJ_d^{\pm}$ models, relative to the benchmark C_d model, across different tests and both forecasting windows. The results show that for all jump tests, the competing models can significantly outperform the benchmark model. But this result is sensitive to different forecasting windows.

The right panel exhibits the DMW statistics of the $C_{d-m}J_{d-m}$, $C_{d-m}J_{d-m}^{\Delta}$, and $C_{d-m}J_{d-m}^{\pm}$ models, relative to their benchmark models. As the results show, for all jump tests, these competing models again can significantly outperform the benchmark model. Moreover, the superiority of the competing models (except the models by the BNS and s-BNS test) is robust to both rolling and increasing windows. This indicates the out-of-sample importance of including longer lagged jumps in the model.

Table 8. The out-of-sample contribution of jumps

		Daily decomposed			*Fully decomposed*		
		C_dJ_d	$C_dJ_d^{\Delta}$	$C_dJ_d^{\pm}$	$C_{d-m}J_{d-m}$	$C_{d-m}J_{d-m}^{\Delta}$	$C_{d-m}J_{d-m}^{\pm}$
BNS	*RW*	-1.51	-1.66	-1.76	-3.21	-1.36	-2.12
	IW	1.75	-0.30	**2.32**	**2.69**	-1.86	**4.49**
Med	*RW*	0.69	-1.31	1.12	**3.71**	**2.38**	**4.49**
	IW	-6.01	**2.06**	**3.04**	**3.02**	0.57	1.63
Min	*RW*	0.92	-1.63	-0.88	**2.27**	-1.93	1.24
	IW	**4.40**	-0.70	**3.47**	**4.75**	-0.52	**7.70**
ASJ	*RW*	**3.24**	**5.91**	**5.75**	**5.29**	**2.52**	0.12
	IW	-0.28	-3.85	-2.02	-0.60	**9.38**	**3.35**
JO	*RW*	-1.10	-1.64	-1.48	-1.46	**3.43**	**2.57**
	IW	**4.53**	-3.79	**5.18**	**2.32**	**6.37**	**7.38**
ABD-LM	*RW*	0.41	0.96	0.14	-1.06	-1.27	-1.19
	IW	1.34	**5.25**	**4.57**	-6.53	**8.50**	1.43
sBNS	*RW*	-1.69	-1.69	-2.31	-3.68	-2.00	-2.64
	IW	1.18	-0.22	1.35	**2.03**	-2.16	**2.07**

Notes: The left panel reports the DMW statistics for the competing C_dJ_d, $C_dJ_d^{\Delta}$, and $C_dJ_d^{\pm}$ models relative to the benchmark C_d model while the right panel reports the DMW statistics for the competing $C_{d-m}J_{d-m}$, $C_{d-m}J_{d-m}^{\Delta}$, and $C_{d-m}J_{d-m}^{\pm}$ models relative to the $C_{d-m}J_d$, $C_{d-m}J_d^{\Delta}$, and $C_{d-m}J_d^{\pm}$ benchmark models. The positive and significant DMW statistics are indicated in **bold**.

5.2 The drift bias

Reducing drift bias for the jump tests is essential. Laurent and Shi (2020) show that although the drift component is asymptotically small, it may harm the jump test power for the 5-minute sample. As a result, the loss of power in the jump tests leads to underestimated jumps. To reduce drift bias, we modify the jump tests based on the method by Laurent and Shi (2020). In detail, we substitute the intraday return r_{t_i} with the drift-robust returns, which are defined by centring r_{t_i} with its median [9],

$$\bar{r}_{t_i} = r_{t_i} - \text{median}(r_{t_i}).$$

For comparison, we only modify the jump test statistics [10] and we do not make any changes to the volatility estimates. For the BNS test, for example, we only modify the test statistic $T_{BNS,t}$ by using the centred return $\bar{r}_{t_i}$, but we do not alter any volatility measures such as RV_t, RV_t^+, RV_t^-, BV_t.

We find that the jump tests corrected for drift bias generally detect more jumps, and this is consistent with Laurent and Shi (2020), who find that the drift bias correction helps improve the testing power. The forecasting results based on the drift-bias adjusted jump tests are quite similar to those based on unadjusted jump tests, which confirms the drift-bias robustness of our out-of-sample results. While the forecasting results from these robustness checks are not reported here, they are available upon request.

[9] The drift-robust returns can also be obtained by mean. But mean is sensitive to jumps thus the more jump-robust median is preferred, as argued in Laurent and Shi (2020).

[10] All test statistics are modified by simply using centred returns, expect for the LM test. For the LM test, we follow the procedure by Laurent and Shi (2020). We first define $m^*(r_{t,i}) = \text{median}(r_{t,i-k+1} \dots r_{t,i})$, and then modify the test as $T_{LM,t,i} = |r_{t,i} - m^*(r_{t,i})|/\hat{\sigma}_{t,i}$,where $\hat{\sigma}_{t,i} = \left(\frac{\pi}{2}\frac{1}{k-2}\sum_{i=j-k+2}^{j-1}|r_{t,i} - m^*(r_{t,i})||r_{t,i-1} - m^*(r_{t,i})|\right)^{1/2}$. The modified LM test is only for identifying the location of intraday jumps but not for generating realized measures. Instead, the realized measures are still obtained by returns centred within one trading day, $\tilde{r}_{t,i} = r_{t,i} - \text{median}(r_{t,i}, \dots, r_{t,M})$, consistent with those of other jump tests.

5.3 The intraday volatility pattern

The jump tests may erroneously identify many spurious jumps because of the well-known U-shape pattern of intraday volatility. Therefore, the spurious jumps may largely contribute to our results. As a robustness check of our results, we modify the jump tests for reducing the intraday volatility pattern bias. Specifically, we modify these jump tests by using the returns divided by the WSD corrector ($r_{t_i}/\hat{f}_{WSD_i}$) of Boudt et al. (2011). The method to obtain the corrector $\hat{f}_{WSD_i}$ is as follows.

First, define,

$$\bar{r}_{t,i} = \frac{r_{t_i}}{\sqrt{M^{-1}BV_t}},$$

Then,

$$ShortH_i = 0.741 \cdot \min\{\bar{r}_{(h_i),i} - \bar{r}_{(1),i}, \dots, \bar{r}_{(T_i),i} - \bar{r}_{(T_i-h_i+1),i}\},$$

where T_i is the total number of observations of intraday i, say, the number of observation days, $\bar{r}_{(j),i}$ are the order statistics of $\bar{r}_{t,i}$, and $h_i = \lfloor T_i/2 \rfloor + 1$.

Then,

$$\hat{s}^2_{ShortH_i} = \frac{M \cdot ShortH_i^2}{\sum_{i=1}^{M} ShortH_i^2}.$$

The WSD estimator is then given by,

$$WSD_i^2 = 1.081 \frac{\sum_{t=1}^{T} w_{t_i} \bar{r}_{t_i}^2}{\sum_{t=1}^{T} w_{t_i}},$$

where $w_{t_i} = w\left(\frac{\bar{r}_{t_i}}{\hat{s}_{ShortH_i}}\right)$ and $w(z) = 1$ if $z^2 \leq 6.635$ and 0 otherwise.

Finally, the estimator for the intraday volatility corrector f_i is obtained by,

$$\hat{f}_{WSD_i} = \sqrt{\frac{M \cdot WSD_i^2}{\sum_{i=1}^{M} WSD_i^2}}.$$

As in section 5.1. we only modify the jump test statistics while we do not make any changes to the volatility estimates. Again, the WSD corrector is only for finite sample performance and does not influence the asymptotic property of the jump tests. Therefore, there is no need to make further adjustments. Unreported results show that the WSD-corrected jump tests generally identify fewer jumps. This is consistent with Boudt et al. (2011), who find the WSD corrector reduces the possibility of detecting the spurious jumps caused by the intraday volatility pattern.

Unreported results for the jump tests corrected for the intraday volatility pattern are very similar. We again find the out-of-sample worthiness of dividing volatility, BNS-alternative jump tests, longer-lagged jumps, and jump signs. Thus, our results are robust when the jump test is modified for the intraday volatility pattern.

5.4 Ultra-high frequency

Christensen et al. (2014) argue that the jump tests generally overestimate the jumps for 5-minute sampling due to the volatility burst distortions. And they suggest that testing jumps for the ultra-high frequency sample can alleviate this bias. This section studies whether our results hold at the ultra-high frequency setting. For the ultra-high frequency, Christensen et al. (2014) suggest modifying the RV by using the pre-averaged noisy returns,

$$RV_t^* = \frac{M}{M-H+2}\frac{1}{H\psi_H}\sum_{i=1}^{M-H+2}\left|r_{t_i}^*\right|^2 - \frac{\widehat{\omega}_t^2}{\theta^2\psi_H},$$

with $H = \lceil\theta\sqrt{M}\rceil$, $\psi_H = (1+2H^{-2})/12$, $r_{t_i}^* = \sum_{j=0}^{H-2} g(j+1/H) r_{t_{i+j}}$, $g(x) = \min\big(x,(1-x)\big)$ and $\widehat{\omega}_t^2 = (1/2(\mathrm{M}-1))\sum_{i=2}^{M}\left|r_{t_i}\right|\left|r_{t_{i-1}}\right|$.

For the ultra-high frequency, all jump tests are based on their noise robust version. Literature has shown that in some noise-robust jump tests: the BNS* and ABD-LM* tests by Christensen et al. (2014), the ASJ* test by Aït-Sahalia et al. (2012), and the JO* test by Jiang and Oomen (2008). For notation purposes, we term the modified jump tests using an asterisk. In the spirit of Christensen et al. (2014), we also modify the Med and Min, tests (termed by Med* and Min*) for the ultra-high frequency. We use these robust jump tests for decomposing the previously defined RV^{*} into jumps and continuous variation, with the significance level still 0.1%. The details of these tests are provided in appendix.

Table 9. Out-of-sample loss ratios for daily decomposed models for ultra-high frequency

			BNS*	Med*	Min*	ASJ*	JO*	ABD-LM*
		HAR^{*}	CJ^{*}					
h=1	*RW*	1.000*	1.004*	1.001*	1.001*	1.002*	**0.998***	1.225*
	IW	**1.000***	1.001	1.001	1.000	1.001	1.000*	1.006
h=5	*RW*	1.000*	1.002*	1.001*	1.001*	**0.999***	1.000*	1.005*
	IW	1.000*	1.001	1.001	1.000*	1.003	**0.999***	1.004
h=22	*RW*	**1.000***	1.002*	1.017*	1.009*	1.071*	1.005*	1.016*
	IW	1.000*	1.001	1.000	1.000*	0.997*	1.001	**0.993***
h=66	*RW*	1.000*	**0.973***	1.000*	1.000*	1.000*	0.978*	0.980*
	IW	1.000*	1.001	1.000*	1.000*	**0.996***	0.998*	0.998*
		HAR^{*}	$C_dJ_d^{\Delta*}$					
h=1	*RW*	**1.000***	1.004*	1.000*	1.001*	1.003*	1.017*	1.224*
	IW	1.000	1.002	1.001	1.001	1.012	**0.999***	1.003
h=5	*RW*	1.000*	1.002*	1.000*	1.001*	**0.999***	1.000*	1.006*
	IW	1.000	1.001	1.001	1.001	1.003	**0.995***	1.001
h=22	*RW*	**1.000***	1.004*	1.002*	1.002*	1.027*	1.004*	1.044*
	IW	1.000	1.000	1.001	1.001	1.002	**0.987***	0.992
h=66	*RW*	1.000*	1.001*	1.000*	1.000*	**0.970***	0.999*	0.977*
	IW	1.000*	1.000*	1.000*	1.000*	**0.998***	0.998*	0.998*
		HAR^{*}	$C_dJ_d^{\pm*}$					
h=1	*RW*	1.000*	1.006*	1.001*	1.001*	**0.998***	1.005*	1.232*
	IW	1.000	1.002	1.002	1.002	**0.983***	0.999	1.005
h=5	*RW*	1.000	1.004	1.002	1.002	**0.994***	1.002	1.008
	IW	1.000	1.002	1.002	1.001	**0.988***	0.995	1.006
h=22	*RW*	1.000*	1.005*	1.000*	**0.999***	1.026*	1.006*	1.015*
	IW	1.000	1.001	1.001	1.001	**0.981***	0.987	0.994
h=66	*RW*	1.000*	**0.973***	0.998*	0.999*	0.976*	1.004*	0.980*
	IW	1.000*	1.000*	1.000*	1.000*	**0.993***	0.998*	0.998*

Following Bajgrowicz et al. (2016), we sample the SPY prices at 10-second for the ultra-high frequency setting. Unreported results show that the (signed) jumps detected in 10-second frequency overwhelmingly have a smaller size and occurrence than in 5-minute frequency. This evidence corroborates that testing jumps for ultra-high frequency alleviates the upward bias due to the volatility bursts, and is in line with Christensen et al. (2014) and Bajgrowicz et al. (2016), who also find decreased jumps size and occurrences in ultra-high frequency asset prices.

We check the robustness of the out-of-sample results. The results again evidence the out-of-sample importance of dividing volatility, BNS-alternative jump tests, longer lagged jumps, and jump signs. For brevity, we only report the results for the former two findings. Table 9 compares the HAR* model with the $C_dJ_d^*$, $C_dJ_d^{\Delta *}$, and $C_dJ_d^{\pm *}$ models [11] by various jump tests, for the out-of-sample. The results are qualitatively in line with those in Table 2. The results show that the daily decomposed model is able to beat the HAR* model, but mostly due to the use of alternative tests such as ASJ* and JO*.

[11] As the benchmark model and different competing models are all based on the volatility estimates under 10-second frequency, we term these modified models their original names with an asterisk,

6 Conclusion

This paper investigates whether jumps predict future volatility. The results for S&P 500 index show that dividing volatility into jumps and continuous variation can significantly improve volatility out-of-sample forecast. But this improvement is mainly attributed to the use of the jump tests which are more powerful and accurate than the BNS test. In addition, we find decomposing longer volatility lags is essential as this makes significant contributions to the out-of-sample.

Although our results show that the jumps have only limited in-sample, we still find that including jumps to the model significantly enhances out-of-sample. Moreover, we find separating the signs of jumps is also important for the out-of-sample, especially for the long-term horizons. Besides, our results are robust when the jump tests are adjusted for drift bias, intraday volatility pattern, and ultra-high frequency sampling.

7 Bibliography

AÏT-SAHALIA, Y. & JACOD, J. 2009. Testing for jumps in a discretely observed process. *The Annals of Statistics,* 37**,** 184-222.

AÏT-SAHALIA, Y., JACOD, J. & LI, J. 2012. Testing for jumps in noisy high frequency data. *Journal of Econometrics,* 168**,** 207-222.

ANDERSEN, T. G., BOLLERSLEV, T. & DIEBOLD, F. X. 2007a. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics,* 89**,** 701-720.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. & LABYS, P. 2000. Great realizations. *Risk,* 13**,** 105-108.

ANDERSEN, T. G., BOLLERSLEV, T. & DOBREV, D. 2007b. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics,* 138**,** 125-180.

ANDERSEN, T. G., BOLLERSLEV, T., FREDERIKSEN, P. & ØRREGAARD NIELSEN, M. 2010. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics,* 25**,** 233-261.

ANDERSEN, T. G., DOBREV, D. & SCHAUMBURG, E. 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics,* 169**,** 75-93.

ANDERSEN, T. G., LI, Y., TODOROV, V. & ZHOU, B. 2021. Volatility measurement with pockets of extreme return persistence. *Journal of Econometrics*.

BAJGROWICZ, P., SCAILLET, O. & TRECCANI, A. 2016. Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science,* 62**,** 2198-2217.

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., HANSEN, P. R., LUNDE, A. & SHEPHARD, N. 2009. Realized kernels in practice: Trades and quotes. Oxford University Press Oxford, UK.

BLACK, F. & SCHOLES, M. 1973. The pricing of options and corporate liabilities. *Journal of political economy,* 81**,** 637-654.

BOLLERSLEV, T., PATTON, A. J. & QUAEDVLIEG, R. 2016. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics,* 192**,** 1-18.

BOUDT, K., CROUX, C. & LAURENT, S. 2011. Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance,* 18**,** 353-367.

CAPORIN, M. 2022. The Role of Jumps in Realized Volatility Modeling and Forecasting. *Journal of Financial Econometrics*.

CHRISTENSEN, K., OOMEN, R. C. & PODOLSKIJ, M. 2014. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics,* 114**,** 576-599.

CHRISTOFFERSEN, P. F. & DIEBOLD, F. X. 2000. How relevant is volatility forecasting for financial risk management? *Review of Economics and Statistics,* 82**,** 12-22.

CLEMENTS, A. & LIAO, Y. 2017. Forecasting the variance of stock index returns using jumps and cojumps. *International Journal of Forecasting,* 33**,** 729-742.

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. & RENÒ, R. 2012. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics,* 30**,** 368-380.

DIEBOLD, F. X. & MARIANO, R. S. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics,* 13.

DUFFIE, D., PAN, J. & SINGLETON, K. 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica,* 68**,** 1343-1376.

DUMITRU, A.-M. & URGA, G. 2012. Identifying jumps in financial assets: a comparison between nonparametric jump tests. *Journal of Business & Economic Statistics,* 30**,** 242-255.

DUONG, D. & SWANSON, N. R. 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics,* 187**,** 606-621.

EDERINGTON, L. H. & GUAN, W. 2010. Longer-term time-series volatility forecasts. *Journal of Financial and Quantitative Analysis,* 45**,** 1055-1076.

HANSEN, P. R., LUNDE, A. & NASON, J. M. 2011. The model confidence set. *Econometrica,* 79**,** 453-497.

HUANG, X. & TAUCHEN, G. 2005. The relative contribution of jumps to total price variance. *Journal of financial econometrics,* 3**,** 456-499.

JIANG, G. J. & OOMEN, R. C. 2008. Testing for jumps when asset prices are observed with noise–a “swap variance” approach. *Journal of Econometrics,* 144**,** 352-370.

LAURENT, S. & SHI, S. 2020. Volatility estimation and jump detection for drift–diffusion processes. *Journal of Econometrics,* 217**,** 259-290.

LEE, S. S. & MYKLAND, P. A. 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies,* 21**,** 2535-2563.

MA, F., WEI, Y., LIU, L. & HUANG, D. 2018. Forecasting realized volatility of oil futures market: A new insight. *Journal of Forecasting,* 37**,** 419-436.

MANEESOONTHORN, W., MARTIN, G. M. & FORBES, C. S. 2020. High-frequency jump tests: Which test should we use? *Journal of Econometrics,* 219**,** 478-487.

MERTON, R. C. 1969. Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics***,** 247-257.

NEWEY, W. & WEST, K. 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.

PROKOPCZUK, M., SYMEONIDIS, L. & WESE SIMEN, C. 2016. Do jumps matter for volatility forecasting? Evidence from energy markets. *Journal of Futures Markets,* 36**,** 758-792.

SANTOS, D. G. & ZIEGELMANN, F. A. 2014. Volatility forecasting via MIDAS, HAR and their combination: An empirical comparative study for IBOVESPA. *Journal of Forecasting,* 33**,** 284-299.

SÉVI, B. 2014. Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research,* 235**,** 643-659.

SHEPHARD, N., KINNEBROCK, S. & BARNDORFF-NEILSEN, O. E. 2008. Measuring downside risk-realised semivariance. *Department of Economics, University of Oxford*.

TAUCHEN, G. & ZHOU, H. 2011. Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics,* 160**,** 102-118.

TODOROV, V. 2019. Nonparametric spot volatility from options. *The Annals of Applied Probability,* 29**,** 3590-3636.