Forecasting the Realized Variance in the Presence of Intraday Periodicity

Ana-Maria H. Dumitru*[a], Rodrigo Hizmeri†[b], Marwan Izzeldin‡[b]

[a]*Deutsche Bundesbank and University of Surrey*
[b]*Lancaster University*

March 5, 2019

Abstract

This paper examines the impact of intraday periodicity on forecasting realized volatility using a heterogeneous autoregressive model (HAR) framework. We show that periodicity inflates the variance of the realized volatility and biases jump estimators. This combined effect adversely affects forecasting. To account for this, we propose a periodicity-adjusted model, HARP, where predictors are built from the periodicity-filtered data. We demonstrate empirically (using 30 stocks from various business sectors and the SPY for the period 2000–2016) and via Monte Carlo simulations that the HARP models produce significantly better forecasts, especially at the 1-day and 5-days ahead horizons.

Keywords: realized volatility, heterogeneous autoregressive models, intraday periodicity, forecast.
JEL codes: C14, C22, C58, G17.

◇*The views expressed in this paper are those of the authors and do not represent the views of the Deutsche Bundesbank, the Eurosystem or its staff. The second author acknowledges the financial support from the Economic and Social Research Council (ESRC), Grant Number: ES/P000665/1.*

*Corresponding author: Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main, Germany. E-mail: a.dumitru@surrey.ac.uk.

†E-mail: r.hizmeri1@lancaster.ac.uk

‡E-mail: m.izzeldin@lancaster.ac.uk

1 Introduction

The last two decades following the seminal work of Andersen and Bollerslev (1998a) saw the emergence of realized volatility (RV) and related measures as proxies for the daily observed volatility of any financial security for which intraday price observations were available. The shift in volatility from latent to quasi-observable[1] meant forecasting could now rely on simple autoregressive models. Corsi (2009)'s heterogenous autoregressive model (HAR) emerged as the standard in forecasting univariate realized volatility.

In this paper, we show that the periodicity of intraday volatility impacts realized volatility forecasts based on autoregressive models through two channels. The first and most important channel is by distorting the variance of the realized volatility, which in turn contributes to biasing the coefficients of the forecasting models. The second channel is via the realized jumps regressors that appear in some predictive models and can also be biased in the presence of intraday periodicity.

To address the observed impact of periodicity, we propose a new class of forecasting models for the realized variance, HARP, where the predictors are based on data from which periodicity is filtered out. We compare the forecasting performance of the HARP models to several HAR models existing in the literature. To this end, we perform a simulation exercise, followed by an empirical application based on high frequency data for the SPDR S&P 500 ETF (SPY) and 30 S&P 500 constituents, observed over more than 4,000 days. Both simulation and empirical exercises attest the superiority of HARP models when forecasting 1-day to 5-days ahead. Specifically, for SPY, we observe improvements of over 10% for HARP models at the 1-day ahead horizon. For the average stock, depending on the model specification, filtering data reduces the forecast losses by approximately 2% to 4% at the 1-day horizon, and up to 5% at the 5-days horizon. At the 1-day horizon, the highest improvements are for models with realized jumps in their specifications, where data filtering leads to better proxies for the jump regressors.

Andersen et al. (2003) were the first to propose autoregressive models to forecast realized volatility. They document the presence of long memory in the time series of logarithmic realized volatilities and suggest a fractionally integrated autoregressive approach in modelling. Inspired by the heterogeneous autoregressive conditional heteroskedasticity (HARCH) model featured in Müller et al. (1997) and Dacorogna et al. (1997), Corsi (2009) proposes the HAR-RV model which regresses realized volatilities on past daily, weekly and monthly realized volatilities. This model can replicate the high levels of per-

[1]The use of the term quasi here is due to the fact that all realized measures are estimates.

sistence observed in the series of daily realized volatilities, without relying on fractional integration. Given its simple linear structure and ease in estimation, the HAR-RV has become the most popular option in forecasting realized volatilities.

When jumps are featured in the prices of financial securities, the realized volatility captures price variability due to both the diffusion and jump components of the price process. Numerous robust to jumps volatility estimators have been developed in Barndorff-Nielsen et al. (2006a), Barndorff-Nielsen and Shephard (2004, 2006b), Barndorff-Nielsen et al. (2006b) Mancini (2004, 2009), Christensen and Podolskij (2007), Andersen et al. (2012) and Corsi et al. (2010). In addition, there are abundant statistical procedures testing for the presence of jumps in the observed path of the price process: Aït-Sahalia and Jacod (2009), Andersen et al. (2007b), Barndorff-Nielsen and Shephard (2006a), Jiang and Oomen (2008), Lee and Mykland (2008), Podolskij and Ziggel (2010) among others.

As the daily integrated variance and jumps have different levels of persistence, they are likely to impact volatility forecasting in different ways. Andersen et al. (2007a) propose adding the lagged realized daily squared jump as an extra explanatory variable to the HAR-RV regression, leading to the HAR-RV-J model. They also propose the HAR-RV-CJ model, which uses as predictors daily, weekly and monthly estimates of the integrated variance and integrated squared jumps. They find that accounting for jumps generally leads to an increase in the explanatory power. This finding is also confirmed by Corsi et al. (2010), who perform a more exhaustive forecast exercise.

Corsi and Reno (2012) add negative returns to the previous HAR-RV specifications, in order to account for a potential leverage effect. They show improved accuracy in forecasting the S&P 500. Bollerslev et al. (2016) argue that all realized measures used in HAR models are bound to include measurement errors, which should be taken into account in modelling. The new model, abbreviated HARQ,[2] performs well in environments of high variability of the measurement error.

The impact of periodicity on the dynamic properties of high frequency returns was first examined by Andersen and Bollerslev (1997). They model intraday volatility as a product between two components: a deterministic periodic component and the actual volatility, i.e. a stochastic component reflecting variability in the fundamental value of the financial security. Such specification has become the literature standard and is also considered in our analysis. Andersen et al. (2001) and Bollerslev et al. (2000) employ similar specifications in modelling intraday volatility in the FX market and the US Treasury

[2] "Q" comes from the fact that the realized quarticity, as the estimated asymptotic variance of the realized variance, is included in the specification.

bond market.

While the periodicity component does not impact the realized variance, by integrating to 1 over the trading day, little is known on its impact on other realized measures. Dette et al. (2016) examine the effect of periodicity on the realized bi-power variation, its variance and covariance with the realized variance, as well as on the realized quarticity under a restrictive DGP, where volatility is constant during the trading day. A co-product of the present paper consists of some simulation evidence on the impact of periodicity on the realized volatility estimation error under a realistic DGP. In the high frequency literature, the only model-free theoretical approach accounting for intraday periodicity is in the work by Andersen et al. (2018), who propose a statistical test for time-varying intraday periodicity.

Intraday periodicity has also been shown to impact the jump detection ability of the intraday jump tests proposed by Andersen et al. (2007b) and Lee and Mykland (2008), where high levels of periodicity can increase the probability of type 1 error (Andersen et al., 2007b). This suggests that the effects of jumps and periodicity on the price and functions of the price can be confounding. Boudt et al. (2011) recommend applying intraday jump tests on returns from which periodicity is filtered out. As previous approaches to estimate intraday periodicity are not robust to jumps in prices and do not allow the actual volatility to vary over time, they propose non-parametric and parametric estimation methods for the periodicity component that allow for such data features. These properties make the Boudt et al. (2011) methods our choice for estimating periodicity.

There are a few other contributions in the literature that account for intraday periodicity when forecasting volatility. In most cases, the periodicity component is removed before modelling and forecasting the intraday returns. Then, the final intraday forecasts are obtained by adding back the estimated periodicity. Martens et al. (2002) forecast intraday volatility using various GARCH models. Deo et al. (2006) propose a long memory stochastic volatility model to forecast intraday returns which are further aggregated to obtain the forecast realized variance. Chortareas et al. (2011) compare daily aggregates of intraday volatility forecasts from a FI-GARCH model to the realized volatility forecasts from an ARFIMA. Frijns and Margaritis (2008) use the estimated periodicity function and the volatility level at the beginning of the trading day to forecast end-of-trading-day volatility. While these contributions model and forecast intraday data, our models apply to daily volatility.

The rest of the paper is structured as follows. Section 2 provides some theoretical background on defining and forecasting the realized variance and estimating the intraday periodicity component of the spot volatility. Section 3 discusses the impact of the intraday periodicity on modelling and forecasting the realized variance. Section 4 contains simulation and empirical results that compare the forecasts of the HARP and HAR models. Section 5 concludes the paper.

2 Theoretical Background

Let $p(t)$ denote a logarithmic asset price at time t belonging to a special class of semimartingales with jumps:

$$\mathrm{d}p(t) = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t) + \mathrm{d}L(t), \quad t \in [0, T] \tag{1}$$

where $\mu(t)$ is a continuous and locally bounded drift term, $\sigma(t)$ is the spot volatility which is adapted and càdlàg. $W(t)$ is one-dimensional standard Brownian motion, while $L(t)$ is a jump process. Without loss of generality, we assume T to be integer, representing the number of trading days over which we perform the analysis. Under this convention, all integers in $[0, T]$ mark the end of a trading day. The volatility at time t over the past day is given by the integrated variance, $IV_t = \int_{t-1}^{t} \sigma^2(u)\,\mathrm{d}u$.

Within each trading day, we are able to observe a number of M observations, equally spaced such that the time interval between any two consecutive observations is equal to $\Delta = \frac{1}{M}$.[3] Let $r_{t,i}$, $i = 1, ..., M$, be the i-th intraday return during the one-day interval $(t-1, t]$, such that $r_{t,i} = p(t-1+i\Delta) - p(t-1+(i-1)\Delta)$. In the absence of jumps, the integrated volatility is consistently estimated by the realized variance,(Andersen and Bollerslev, 1998a, Andersen et al., 2003) defined as $RV_t = \sum_{i=1}^{M} r_{t,i}^2$. If the price contains jumps, RV_t is no longer consistent for the integrated variance, converging to the quadratic variance of the price process, $\int_{t-1}^{t} [\sigma^2(u) + L^2(u)]\,\mathrm{d}u$. One needs to rely on a robust to jumps estimator of the integrated variance, such as the realized bipower variation of Barndorff-Nielsen and Shephard (2004) given by $BV_t = \frac{M}{M-1} 1.57 \sum_{i=2}^{M} |r_{t,i}||r_{t,i-1}|$.

We define the intraday volatility periodicity, $f(t)$, as a multiplicative component to the actual spot volatility, $s(t)$, as in Andersen and Bollerslev (1997, 1998b), Andersen et al. (2001), Boudt et al. (2011):

$$\sigma(t) = s(t) f(t), \tag{2a}$$

$$\text{such that} \int_{t-1}^{t} f^2(u)\,\mathrm{d}u = 1, \tag{2b}$$

[3]Note that defining realized volatility does not require equally spaced observations. We make this assumption here for simplicity.

so that intraday periodicity has no impact on the integrated variance, i.e. $\int_{t-1}^{t} \sigma^2(u)\, du = \int_{t-1}^{t} s^2(u)\, du$. In practice, as we observe a discrete number of observations, the condition in equation (2b) can be written using the following Riemann integral:

$$\Delta \sum_{i=1}^{M} f_i^2 = 1, \tag{3}$$

where f_i is the i-th value of the function $f(\cdot)$ observed during a trading day. Clearly, when Δ approaches 0, the Riemann sum converges to the integral in (2b).

The two components of spot volatility defined above in (2a) differ greatly. The periodic component is a deterministic function of intraday time and reflects intraday trading patterns. The actual spot volatility $s(t)$ is a stochastic process which varies in time reflecting the available information on the asset.

2.1 Models for Forecasting Realized Volatility

In the present analysis, we consider a wide range of HAR forecasting models. Let h be the forecasting horizon, measured in days.

HAR-RV

The original HAR-RV model proposed by Corsi (2009) is represented by the following auto-regression:

$$RV_{t,t+h-1} = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-5,t-1} + \beta_m RV_{t-22,t-1} + \epsilon_{t+h-1}, \tag{4}$$

where $RV_{t,t+h-1}$ is the forecasted realized variance over the next h days (starting from day t), RV_{t-1} the daily lag of the realized variance, $RV_{t-5,t-1}$ the average realized variance over the past week, $RV_{t-22,t-1}$ the average realized variance over the past month, and ϵ_{t+h-1} is the forecasting error. β_0 is the regression constant term, while β_d, β_w and β_m are the coefficients corresponding to the one-day, one-week and one-month lagged values of the realized variance.

HAR-RV-J

Andersen et al. (2007a) define the contribution of jumps to the daily quadratic variation of the price as $J_t = \max(RV_t - C_t, 0)$, for $t = 1, \cdots, T$, where C_t is a consistent estimator of the integrated variance. The daily lag of the jumps, J_{t-1}, is added to the HAR-RV forecasting regression:

$$RV_{t,t+h-1} = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-5,t-1} + \beta_m RV_{t-22,t-1} + \beta_J J_{t-1} + \epsilon_{t+h-1}. \tag{5}$$

HAR-RV-CJ

In this model, also by Andersen et al. (2007a), past lags of the estimated continuous and discontinuous components of the quadratic variation are considered in the forecasting regression, as follows:

$$RV_{t,t+h-1} = \beta_0 + \beta_{C_d} C_{t-1} + \beta_{C_w} C_{t-5,t-1} + \beta_{C_m} C_{t-22,t-1} + \beta_{J_d} J_{t-1} + \beta_{J_w} J_{t-5,t-1} + \beta_{J_m} J_{t-22,t-1} + \epsilon_{t+h-1}, \quad (6)$$

where C_{t-1}, $C_{t-5,t-1}$ and $C_{t-22,t-1}$ are the one-day, one-week and one-month lagged estimates of the integrated variance, and J_{t-1}, $J_{t-5,t-1}$ and $J_{t-22,t-1}$ are the one-day, one-week and one-month lagged estimates of the jumps' contribution to the quadratic variation. Following Andersen et al. (2007a), $C_t = RV_t \cdot \mathbb{I}_t(\text{no jumps}) + BV_t \cdot \mathbb{I}_t(\text{jumps})$, for $t = 1, \cdots, T$, where $\mathbb{I}_t(\cdot)$ is the indicator function for whether jumps were identified on day t or not.

HARQ

As Bollerslev et al. (2016) indicate, the variance of the realized volatility measurement error is a function of the integrated quarticity, $\int_{t-1}^{t} \sigma^4(u)\, du$, $t = 1, \cdots, T$. Their main forecasting model accounts for the error in measuring the one-day lagged realized variance,[4] as follows:

$$RV_{t,t+h-1} = \beta_0 + (\beta_d + \beta_{dQ} RQ_{t-1}^{1/2}) RV_{t-1} + \beta_w RV_{t-5,t-1} + \beta_m RV_{t-22,t-1} + \epsilon_{t+h-1}, \quad (7)$$

where $RQ_{t-1} = M \sum_{i=1}^{M} r_{t-1,i}^4$ is an estimator of the integrated quarticity.

2.2 Intraday Periodicity Estimation

We use the nonparametric methodology proposed by Boudt et al. (2011), which is robust to the presence of jumps in the price process. We define standardized intraday returns, $\bar{r}_{t,i}$, $i = 1, \ldots, M$, as the returns $r_{t,i}$ divided by the squared root of the realized bipower variation estimated on a local window. For a certain intraday time, i, we observe T standardized intraday returns.

[4]Authors explain that measurement errors for the one week and one month realized volatilities do not have a significant impact on forecasting.

The intraday periodicity estimator is defined as:

$$\hat{f}_i^{WSD} = \frac{WSD_i}{\sqrt{\Delta \sum_{j=1}^{M} WSD_j^2}} \tag{8}$$

$$WSD_i = \sqrt{1.081 \frac{\sum_{l=1}^{T} \chi_{l,i} \bar{r}_{l,i}^2}{\sum_{l=1}^{T} \chi_{l,i}}},$$

for all $i = 1, \ldots, M$, where WSD_i is the weighted standard deviation (WSD) and $\chi_{l,i}$, $l = 1, \ldots, T$ are weights computed using the shortest half periodicity estimator and defined in appendix B.1.

3 The Impact of Periodicity on the Forecasting Regression

To illustrate the impact of periodicity on the forecasting regression, we first consider a simple DGP and compare the coefficients of the forecasting regression in the presence of intraday periodicity to the coefficients obtained in the absence of periodicity. We further perform this comparison for a more complex and realistic DGP by relying on simulation evidence. Finally, we consider the impact of periodicity on forecasting via distortions in jump tests.

3.1 The Simple AR(1) Model

We assume the daily integrated variance evolves according to an AR(1) process.

$$IV_t = \Phi IV_{t-1} + \epsilon_t, \tag{9}$$

where $t \in \{1, 2, ..., T\}$, $|\Phi| < 1$, and ϵ_t is i.i.d. with $\text{Var}(\epsilon_t) = \sigma_\epsilon^2$. In addition, within each trading day, the actual spot volatility remains constant at a level equal to a fraction of the daily integrated variance, ΔIV_t. If we also account for intraday periodicity, the spot volatility for the i-th Δ-length window during a trading day equals $\Delta IV_t f_i^2$. Assuming no drift, the i-th return is $r_{t,i} = \sqrt{\Delta IV_t} f_i w_i$, where w_i is i.i.d. $\mathcal{N}(0, 1)$ and independent of present and past values of $s(\cdot)$. Suppose one attempts to forecast volatility using the following AR(1) model for the realized variance:

$$RV_t = \phi RV_{t-1} + u_t, \tag{10}$$

with ϕ equal to the well-known formula:

$$\phi = \frac{\text{cov}\,(RV_t, RV_{t-1})}{\text{Var}\,(RV_t)}. \tag{11}$$

In the above equations, as RV_t is only a proxy for the integrated variance, it is subject to measurement error. Bollerslev et al. (2016) show that when the true DGP for the integrated variance is an AR(1), as in (9), the presence of measurement error implies ARMA(1,1) dynamics for the realized volatility and not the AR(1) model specified in (10). Estimating this latter misspecified model leads to the presence of an attenuation bias in the estimator $\hat{\phi}$. Since this bias affects the errors of the model in (10) and not the autoregressive parameter in (11), our analysis is based on the AR(1) model due to its simplicity. Moreover, as shown in Bollerslev et al. (2016), bias correction is easy to implement, implying relatively simple extensions to our results, without modifying the conclusions in terms of the implications of the intraday periodicity.

To assess the impact of periodicity on the value of ϕ, we compute the numerator and denominator in equation (11) in the presence/ absence of periodicity. The required derivations are enclosed in appendix A. While the auto-covariance remains unaffected by periodicity, we obtain the following variance formulae for the case in which periodicity is present (equation (12a) below), compared to the case when it is absent (equation (12b)):

$$\text{Var}(RV_t) = \Delta^2 \frac{\sigma_\epsilon^2}{1-\Phi^2}\left(2\sum_{i=1}^{M} f_i^4 + \frac{1}{\Delta^2}\right) \tag{12a}$$

$$\text{Var}(RV_t)^{\overset{no}{periodicity}} = \Delta^2 \frac{\sigma_\epsilon^2}{1-\Phi^2}\left(\frac{2}{\Delta} + \frac{1}{\Delta^2}\right). \tag{12b}$$

The main difference between the above formulae resides in the term $\sum_{i=1}^{M} f_i^4$. From equation (3), we know that $\sum_{i=1}^{M} f_i^2 = \Delta^{-1} > 1$. Then, $\sum_{i=1}^{M} f_i^4 > \sum_{i=1}^{M} f_i^2 = \Delta^{-1}$, so that $\text{Var}(RV_t) > \text{Var}(RV_t)^{\overset{no}{periodicity}}$. This implies that ϕ, as given in equation (11), is lower –in absolute value– than the corresponding coefficient for the case of no periodicity. Thus, ϕ understates the true correlation coefficient, Φ, for two reasons. First, the presence of measurement error leads to the variance distortion in (12b), pushing ϕ downwards from Φ. Second, as shown in equation (12a), the presence of periodicity generates a further increase in the variance of realized volatility, further reducing ϕ.

3.2 Simulation Evidence

The analytic results in the previous section were possible because we relied on a simple DGP. Extending such results to a more complex DGP can be achieved via Monte Carlo simulation, whereby log-prices follow a one-factor stochastic volatility model with jumps:

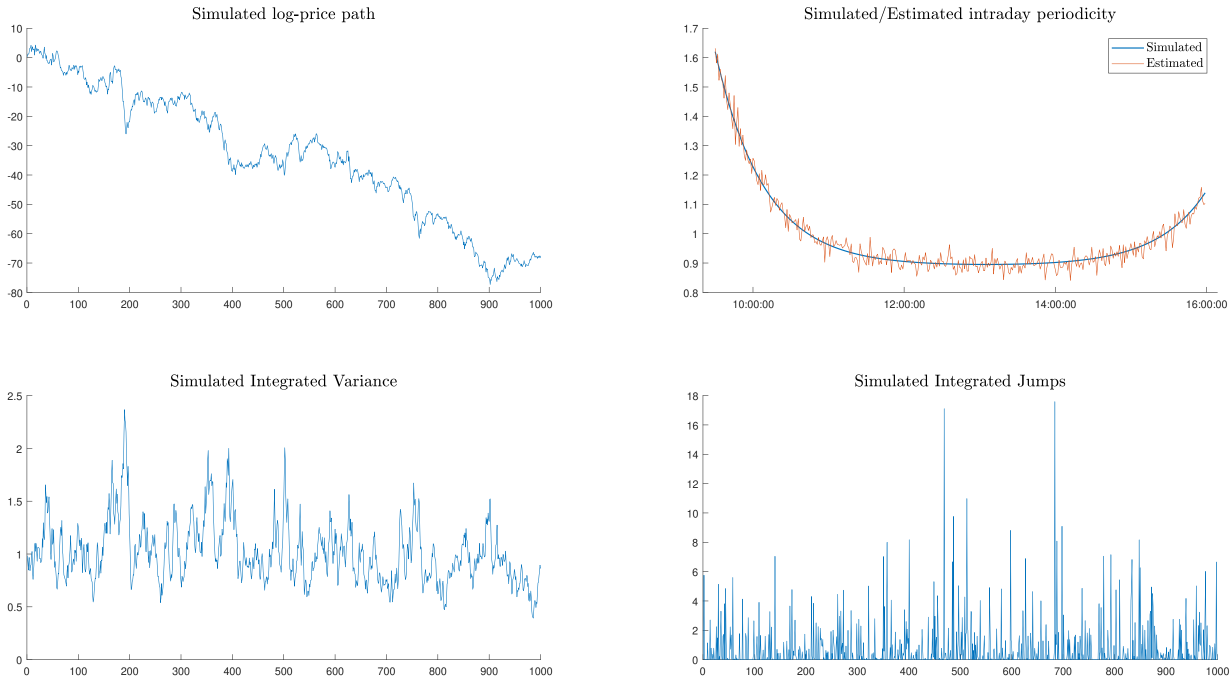
$$\begin{aligned} \mathrm{d}p(t) &= \eta\,\mathrm{d}t + f(t)\nu(t)\,\mathrm{d}W_p(t) + \mathrm{d}L(t) \\ \nu^2(t) &= \exp\{\gamma_0 + \gamma_1\nu_1^2(t)\} \\ \mathrm{d}\nu_1^2(t) &= \alpha_\nu(t)\nu_1(t)\,\mathrm{d}t + \mathrm{d}W_\nu(t) \\ f(t) &= C + Ae^{-at} + Be^{-b(1-t)} \end{aligned} \tag{13}$$

where $W's$ are correlated standard Brownian motions, $\nu(t)$ is a stochastic volatility factor, and $L(t)$ is a compound Poisson process. This model has been used by Huang and Tauchen (2005) among others. The price drift is set $\eta = 0.03$, while for the volatility factor, we have $\alpha_\nu = -0.1$, $\gamma_0 = 0$ and $\gamma_1 = 0.125$. Leverage $\rho = \mathrm{corr}(dW_p, dW_\nu)$ is set equal to -0.62, while jumps arrive with constant intensity $\lambda = 0.4$ and have sizes of up to 30% below or above the spot volatility.

The intraday volatility function $f(t)$ is defined as the sum of two exponentials, which provide the well-known U-shape. We follow Andersen et al. (2012) and Hasbrouck (1999) and set $A = 0.75$, $B = 0.25$, $C \approxeq 0.89$, and $a = b = 10$. Simulations are generated using an Euler scheme based on 23,400 initial data points (corresponding to seconds), which we further aggregate to lower sampling frequencies. We simulate a total of 150,000 sample paths of length 1,000 days.

Figure 1 shows the simulated path of the log price for 1,000 trading days in the upper left corner, along with corresponding paths for the integrated volatility and squared jumps in the lower half of the figure. The upper right corner plots the simulated and estimated values of the intraday periodicity function. The estimates are obtained by applying the formulas in (8) and follow closely the U-shaped periodicity curve.

Figure 1: Simulated log-price and its components



Note: This figure depicts the simulated log-price, simulated and estimated intraday periodicity, integrated volatility and jumps over 1000 days using the simulation outlined in equation (13).

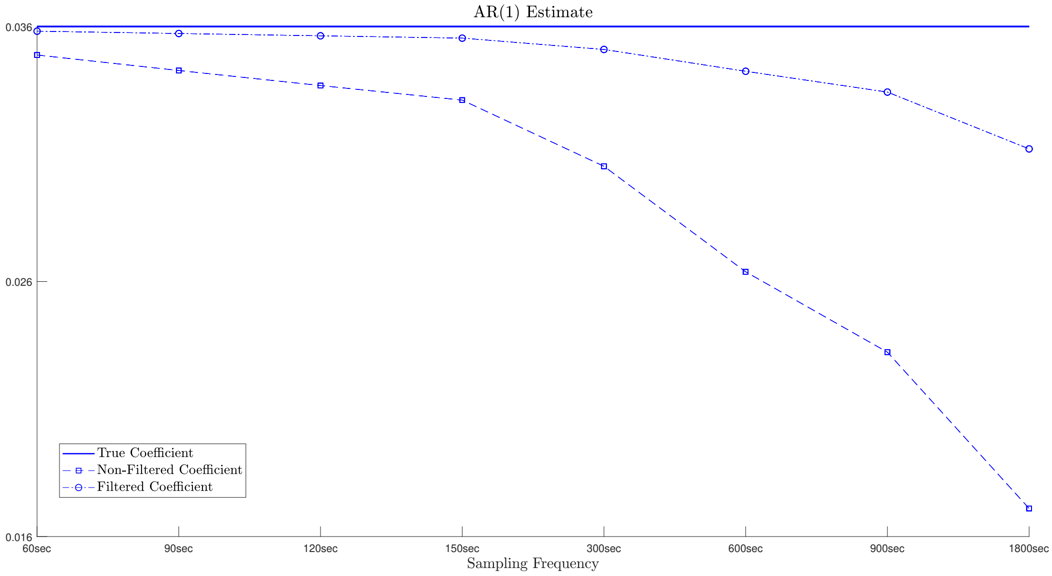
We can filter out periodicity by dividing each intraday return $r_{t,i}$ by the corresponding periodicity estimate, $\hat{f}_i^{WSD}$. This enables us to compare forecasts based on the original (unfiltered) returns to forecasts where predictors are obtained from filtered returns. We consider two forecasting models for the realized volatility. The first model is a simple regression with only one predictor: the lagged RV based on either unfiltered or filtered data. The second model uses as predictors the daily, weekly and monthly lags of the RV based on either unfiltered or filtered data. In the case of unfiltered data, the first model is equivalent to an AR(1) model for the realized variance, while the second corresponds to the HAR-RV model specified in (4).

Figure 2 shows the estimated coefficients for the simple model obtained by relying on filtered and unfiltered return data for different sampling frequencies. The straight line at the top of the figure represents the "true" coefficient, defined as the AR(1) coefficient estimated on the time series of daily quadratic variance.[5] The coefficient obtained when RV relies on filtered returns is much closer to the true coefficient across all sampling frequencies. Moreover, the gap between this coefficient and the one based on the unfiltered

[5] Note that the term "true" in this case is unrelated to the true DGP of the quadratic variance, given in (13).

returns increases as we decrease the sampling frequency.

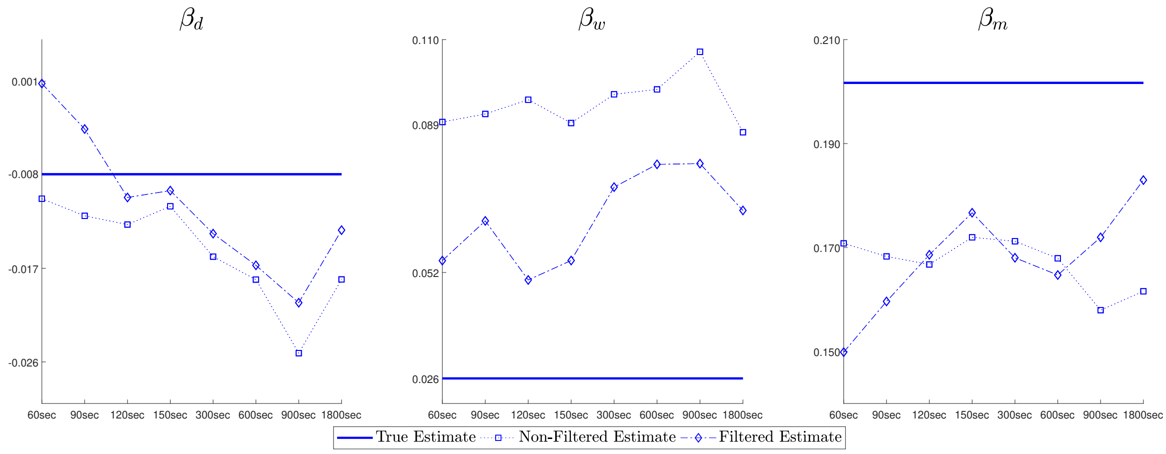
Figure 2: Simulated filtered vs. unfiltered AR(1) coefficients



Note: This figure plots the true coefficient of an AR(1) model versus the estimated coefficient using unfiltered data (squared marker) and filtered data (circled marker) across different sampling frequencies.

Figure 3 plots the three coefficients of the second model across decreasing sampling frequencies estimated on filtered and unfiltered data. From left to right, the first panel of the figure corresponds to β_d from (4), the middle panel to β_w, while the last panel to β_m. The straight line in each panel represents the corresponding estimates on the daily quadratic variance. These are referred to as "true" coefficients, as in the case of the AR(1) model above.

Figure 3: Simulated filtered vs. unfiltered HAR-RV coefficients



Note: This figure compared the estimates of the HAR-RV using filtered (circled marker) and unfiltered (squared marker) against the true estimates across different sampling frequencies.

Given that HAR-RV is a multiple regression model, coefficients include RV variance type terms in both the numerator and denominator. As a result, for this model, it is difficult to understand the direction of the impact of periodicity on coefficients. Simulation results in figure 2 show that, even with a complex DGP as in (13), both daily and weekly coefficients are closer to their true values once periodicity is filtered out, for almost all sampling frequencies. Unlike the case of the simple model, coefficients are higher –in absolute value– for unfiltered models than for filtered models. The high divergence of the daily coefficient for the filtered model estimated on 60 seconds data can be attributed to distortions from jumps. As jumps do not scale with the sampling frequency, their impact on periodicity estimates is greater at higher sampling frequencies. Moreover, in general, the filtered model coefficients show more variability across sampling frequencies due to the measurement error from estimating periodicity. Comparisons of the monthly coefficients are inconclusive for ranking filtered and unfiltered models. For this coefficient, the regressor is an average over 22 days of data, smoothing out the effect of periodicity.

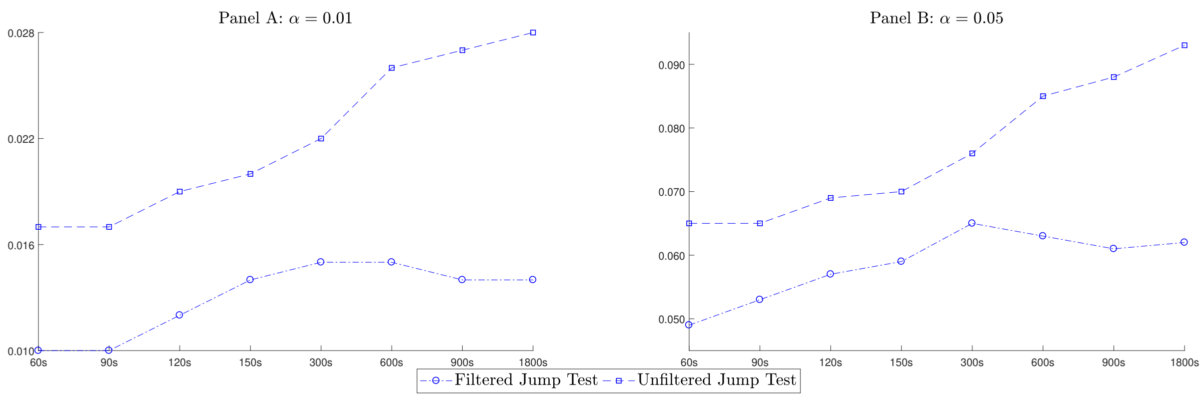
3.3 Impact on Detected Jumps

Two of the most popular HAR models, HAR-RV-J (equation (5)) and HAR-RV-CJ (equation (6)), use the estimated daily squared jumps as predictors. These estimates depend on the outcome of jump tests that decide whether jumps have occurred during a particular trading day. The most familiar test for jumps is the one proposed by Barndorff-Nielsen and Shephard (2006a), relying on a comparison between RV and the jump robust realized bipower variation, BV_t, defined in section 2. The test statistics for this test and for an alternative jump test are presented in appendix B.2.

While realized variance remains unchanged under periodicity, this does not hold for other realized-type estimators. Dette et al. (2016) derive the joint distribution of the realized volatility and the realized bipower variation in the presence of intraday periodicity for a simple DGP with constant intraday volatility. They show that not accounting for periodicity in this simple context can lead to spurious jump detection.

We use simulations from the DGP in (13) to explore the impact of intraday periodicity on spurious jump detection. We apply the Barndorff-Nielsen and Shephard (2006a) jump test to 30,000 replications of simulated data, aggregated at sampling frequencies ranging from 60 to 1,800 seconds. Results for the original return data and periodicity filtered data are shown in figure 4. The left hand panel of the figure shows results for a significance level $\alpha = 1\%$ and the panel on the right for a significance level of $\alpha = 5\%$.

Figure 4: Proportion of spurious jumps by sampling frequency for filtered and unfiltered data



Note: This plot graphs the proportion of spurious detection of jumps across sampling frequencies using the BNS test, Barndorff-Nielsen and Shephard (2006a), evaluated at the 1% and 5% significance level.

As the figure shows, the number of spurious jumps detected is higher for unfiltered returns, result that remains valid across all sampling frequencies. Moreover, this finding is confirmed in appendix C.1 for the case of the jump test proposed by Andersen et al. (2012).

4 HARP Models and their Forecasting Performance

Section 3 shows that the presence of intraday periodicity can distort the results for forecasting the realized variance based on autoregressive models. This can happen in a direct manner, with periodicity generating biases in the coefficients of the forecasting regressions, or indirectly by increasing the measurement error in the jump regressors.

In this section, we introduce a new class of models to forecast the realized variance, HARP, where the predictors rely on data from which periodicity is filtered out (the "P" in HARP stands for periodicity-filtered). Naturally, all HAR models illustrated in section 2.1 can be transformed into HARP models by simply using filtered data to compute all regressors. Unlike most HAR models,[6] HARP models are not autoregressions.[7] Equation (14) below introduces the HARP-RV model, i.e. the HARP version of the HAR-RV model given in equation (4). Similar equations can be written for the HARP versions of the rest of the HAR models in section 4.

$$RV_{t,t+h-1} = \beta_0 + \beta_d RV_{t-1}^{p-filtered} + \beta_w RV_{t-5,t-1}^{p-filtered} + \beta_m RV_{t-22,t-1}^{p-filtered} + \epsilon_{t+h-1}, \tag{14}$$

where the superscript ${}^{p-filtered}$ shows that the RV proxies rely on data from which periodicity was pre-filtered out.

We further compare the forecasting performance of the HARP models to that of the HAR models, using both simulation-based and empirical evidence. We demonstrate that considerable forecasting gains can be attained by filtering data, especially at short and medium horizons.

To evaluate the forecasting performance of the two classes of models, we use two separate loss functions, the mean squared error (MSE) and the quasi-likelihood (QLIKE) loss, defined in equation (15) below:

$$\begin{aligned} MSE(RV_t, F_t) &= (RV_t - F_t)^2 \\ QLIKE(RV_t, F_t) &= \frac{RV_t}{F_t} - \log\frac{RV_t}{F_t} - 1, \end{aligned} \tag{15}$$

where F_t is the out-of-sample forecast of the realized variance.

For forecast horizons beyond 1-day, both HARP and HAR models are adapted to the new time scale by replacing the daily RVs on the left-hand-side with the weekly and monthly RVs. Thus, separate models are fitted for each forecasting horizon. In the analysis based on real data, we perform both in-sample and out-of-sample forecasts, while the results for simulated data involve only the latter.[8]

4.1 Simulation Results

We rely on data simulated from the DGP in (13) to estimate all HAR models enumerated in section 2.1, as well as their HARP counterparts. For each 1000-day long

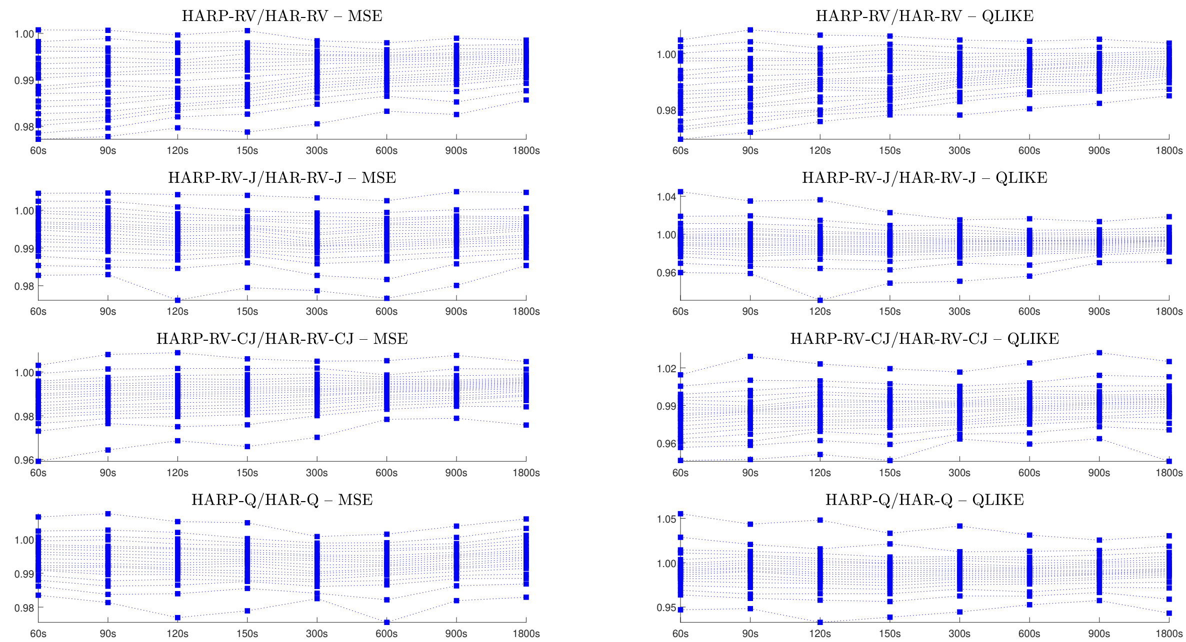
[6]The HAR-RV-CJ model is also not an autoregression.

[7]The HARP models resemble autoregressions where the dependent variable is measured with errors. As the dependent variable includes periodicity, it will have a conditional distribution with the same mean as the the periodicity-free RV, but a different variance, which is similar to having measurement errors in the dependent variable.

[8]When forecasting out-of-sample, we re-fit the models each day.

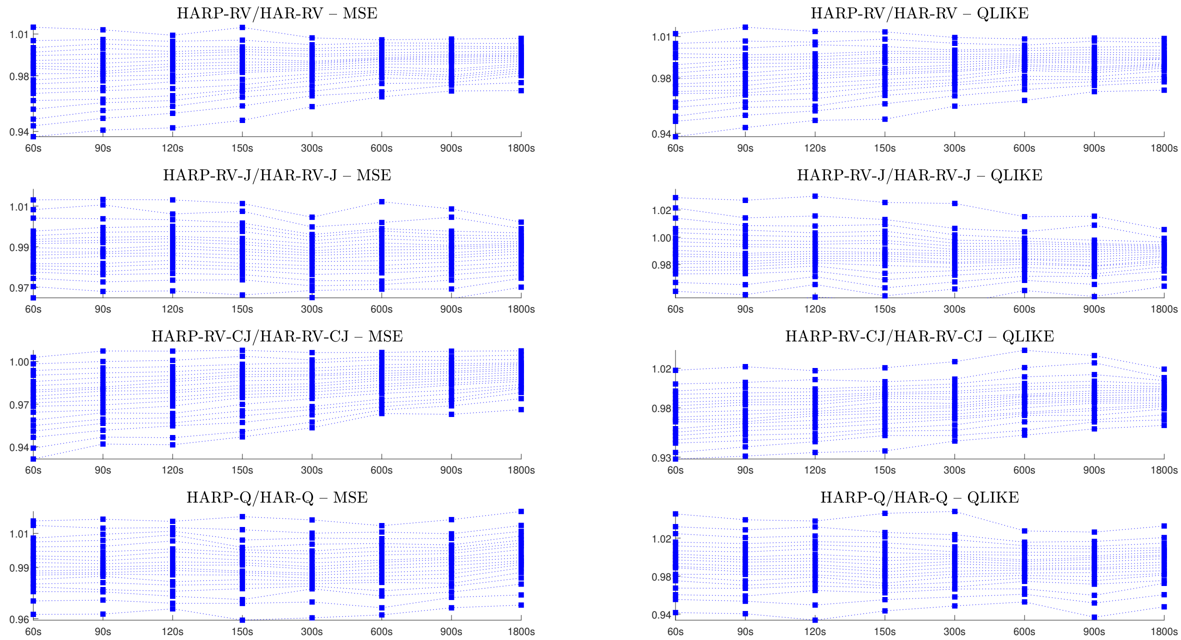
simulated path, we re-estimate the models each day on a rolling window of 350 days. For each forecasting model in 2.1 and forecast horizon, we compute the ratio of forecast losses for the HARP version of the model versus the HAR model. A ratio below one signals the superiority of the model based on filtered returns. The quantiles of such ratios for the 1-day and 1-week forecasting horizons are plotted in figures 5 and 6 against the sampling frequency.

Figure 5: Loss ratio for simulated data – $h = 1$



Note: The figure plots the quantiles ranging from 0.05 to 0.95 in increments of 0.05 for the MSE and QLIKE loss ratios, for HARP versus HAR models.

Figure 6: Loss ratio for simulated data – $h = 5$



Note: The figure plots the quantiles ranging from 0.05 to 0.95 in increments of 0.05 for the MSE and QLIKE loss ratios, for HARP versus HAR models.

Generally, the quantiles of the loss ratios are more dispersed at higher sampling frequencies. At lower frequencies, regressants and predictors in the forecasting equations are affected by measurement error. The reduced dispersion of loss ratios at low frequencies suggests that measurement error has a similar bearing on forecasts regardless on whether based on filtered or unfiltered returns.

For the MSE, in the case of both forecasting horizons, upper quantiles take values of around 1, while lower quantiles are below this value, indicating that the distributions of these ratios are skewed to the left of 1. This effect is generally more pronounced for the 5-day forecast, where the lower quantiles at 60 seconds are below 0.94 for the HARP-RV/ HAR-RV and HARP-RV-CJ/ HAR-RV-CJ loss ratios. The asymmetries in the distributions of HARP-RV-J/ HAR-RV-J and HARPQ/ HARQ loss ratios are more tempered, but maintained through most sampling frequencies. For the QLIKE loss, we observe a similar distributional asymmetry, more pronounced for the HARP-RV/ HAR-RV and HARP-RV-CJ/ HAR-RV-CJ ratios.

While results in figures 5 and 6 suggest the superiority of methods relying on filtered data for the 1-day and 1-week ahead forecasts, results for 1-month ahead are unclear (see appendix C.2). For this horizon, the distributions of the loss ratios do not show clear signs of asymmetry with respect to the value 1, making it difficult to rank predictions based on filtered versus unfiltered data. At longer horizons, the forecast error has a much bigger impact on both types of predictions, rendering data filtering less effective.

4.2 Empirical Analysis

4.2.1 Data

We use intraday price data from the TickData database for the SPDR S&P 500 ETF (SPY) and 30 individual stocks in the S&P 500 basket. We observe a total of 4,277 trading days between 2000 and 2016. Data is aggregated down from tick level using previous tick interpolation and is furthered sampled every 300 seconds, leaving us with 78 observations per day. The choice of this sampling frequency is standard in the literature, motivated by the trade-off between bias and variance (for more details, see Aït-Sahalia et al., 2005, Hansen and Lunde, 2006).

The intraday periodicity function is assumed not to vary from one day to another and is estimated as described in section 2.2. Figure 7 plots the estimated periodicity for SPY and the average estimated periodicity for the 30 S&P 500 stocks considered. Both plots reveal the characteristic U-shape for the estimated curve. The SPY estimated periodicity shows greater variation in comparison to the average stock periodicity, where averaging leads to a smoother curve.

Figure 7: Intraday estimated periodicity for SPY (left) and average periodicity for all stocks (right).

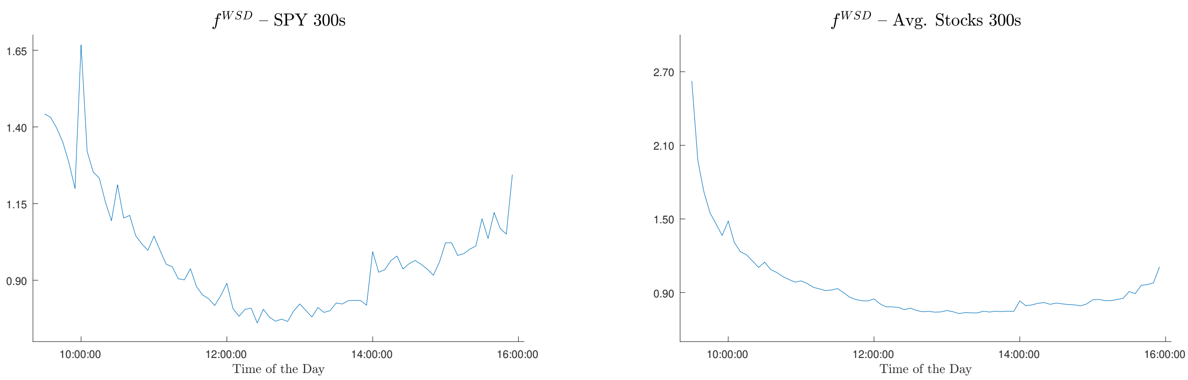


Table 1 below reports, for each ticker in our sample, the minimum, maximum and median values of the realized variance, the number of jumps detected and the estimated proportion of the continuous component relative to the total RV. The left hand side of the table reports these statistics for the unfiltered return data, and the right hand side for periodicity-filtered returns.

Table 1: Realized variance minimum, maximum and median, realized number of jumps and the estimated proportion of integrated variance in the quadratic variation for SPY and 30 stocks.

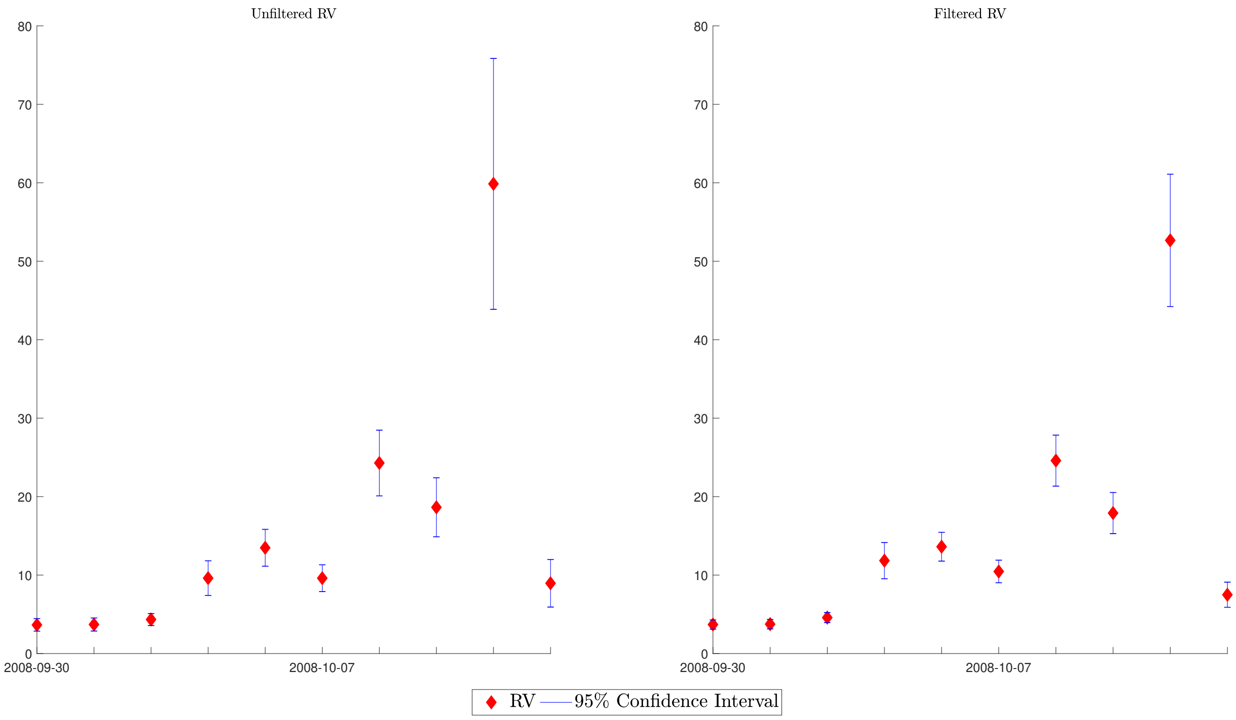
				Unfiltered					Filtered		
Stock	Ticker	Min RV	Max RV	Median RV	# Jumps	%QV	Min RV	Max RV	Median RV	# Jumps	%QV
SPDR ETF	SPY	0.013	59.863	0.485	353	98.153	0.012	52.663	0.486	281	98.580
3M	MMM	0.082	91.955	1.008	518	95.776	0.083	86.331	0.995	234	98.529
AK Steel	AKS	0.872	559.611	10.585	952	91.433	0.874	417.332	10.446	808	92.756
Arconic Inc.	ARNC	0.339	291.089	3.070	460	96.811	0.275	205.826	2.996	318	98.406
Brown-Forman	BFB	0.074	240.414	1.152	963	87.573	0.101	39.181	1.144	799	93.046
BT Group	BT	0.100	59.568	1.162	1386	81.165	0.109	44.555	1.170	1466	79.768
China Mobile	CHL	0.082	65.965	1.063	1040	89.395	0.086	66.370	1.055	907	91.155
Citigroup	C	0.137	975.858	2.110	449	96.535	0.117	967.488	2.087	206	98.210
Coca-Cola	KO	0.046	58.808	0.836	591	94.356	0.063	62.583	0.811	271	97.908
DUKE Energy	DUK	0.051	189.935	1.182	668	95.262	0.056	197.569	1.130	401	97.892
eBay	EBAY	0.202	236.419	2.782	469	97.037	0.210	352.784	2.729	200	98.806
General Dynamics	GD	0.081	63.282	1.281	582	94.066	0.064	60.334	1.255	331	96.701
General Electric	GE	0.108	180.389	1.303	465	96.230	0.099	139.389	1.299	257	98.033
Halliburton	HAL	0.229	265.432	3.579	429	95.756	0.207	374.087	3.545	198	98.101
Home Depot	HD	0.156	103.477	1.573	449	96.361	0.155	96.538	1.565	208	98.606
Honeywell	HON	0.104	268.331	1.609	506	95.204	0.103	158.574	1.581	235	97.447
Humana	HUM	0.240	157.529	2.609	750	88.447	0.300	302.448	2.509	468	90.376
Intel	INTC	0.154	89.885	2.038	489	97.437	0.154	91.724	2.002	223	98.739
LVLT	LVLT	0.242	1159.384	10.917	1010	91.358	0.258	1368.049	11.047	781	93.825
McDonald's	MCD	0.087	161.156	1.090	557	93.287	0.086	103.808	1.068	238	97.868
Microsoft	MSFT	0.083	62.386	1.416	490	96.654	0.054	61.070	1.377	224	98.767
ONEOK	OKE	0.160	411.055	1.668	957	86.341	0.161	147.493	1.622	800	89.533
Pfizer	PFE	0.150	62.697	1.382	520	94.356	0.140	61.198	1.370	252	97.889
Procter & Gamble	PG	0.101	79.549	0.766	538	94.281	0.090	125.615	0.771	176	97.195
Southern Co.	SO	0.092	97.041	0.937	633	93.667	0.109	82.402	0.914	336	96.547
Travelers Companies Inc	TRV	0.102	263.929	1.186	675	92.856	0.105	224.417	1.185	371	96.849
United Health	UNH	0.129	225.956	1.745	548	94.628	0.144	135.120	1.753	266	97.645
UPS	UPS	0.081	216.939	0.851	571	94.446	0.060	117.387	0.856	299	97.545
Verizon	VZ	0.122	102.221	1.162	544	94.670	0.130	100.391	1.144	254	97.897
Vodafone	VOD	0.110	70.936	0.926	391	97.419	0.076	79.999	0.922	355	97.671
Xerox	XRX	0.299	276.588	2.864	772	92.101	0.345	346.125	2.814	594	95.305
	Avg. Stocks	0.160	236.260	2.195	646	93.497	0.160	220.540	2.172	416	95.967

Note: This table reports the descriptive statistics for RV of the 30 individual stocks and the SPY estimated at the 300 second frequency. The %QV is estimated as $\%QV = \frac{\sum_{t=1}^{T} C_t}{\sum_{t=1}^{T}(C_t+J_t)}$, while # Jumps indicates the total number of days with jumps estimated at the 1% significance level using the Barndorff-Nielsen and Shephard (2006a) procedure.

For SPY, we detect 353 jumps for unfiltered data, meaning that we identify jumps on 8.25% of days. When data is filtered, the number of jumps drops to 281, suggesting that 6.57% days had jumps. Results for individual stocks show high variability in the number of jumps identified for both filtered and unfiltered data. On average, we observe 646 jumps for the unfiltered data, which decreases substantially after filtering to 416. As shown in section 3.3, the presence of intraday periodicity can lead to spurious jump detection. Removing the periodicity component from data can thus lead to fewer realized jumps. Additional results on jump identification based on the test by Andersen et al. (2012) are available in appendix C.3.

The analytic results for the simple AR(1) DGP in section 3.1 showed that periodicity impacted the autoregressive coefficient via the variance of the realized volatility. For a generic DGP, the asymptotic variance of the realized volatility is the integrated quarticity, which can be estimated via the realized quarticity. Both theoretical and realized quarticities are defined above in section 2.1. Figure 8 plots the SPY realized variance and its 95% confidence bands across time for a 10-days window, starting on 2008-09-30, differentiating again between unfiltered- and filtered -based estimates.

Figure 8: 95% confidence bands for the SPY realized variance based on unfiltered versus filtered data as a function of time, plotted for a 10-days window, starting on 2008-09-30



Confidence intervals obtained on the unfiltered returns are generally wider than confidence intervals obtained for filtered data. Thus, while the realized variance is unaffected by periodicity, its higher moments are impacted by it. This empirical finding backs the analytical and simulation results in section 3 in suggesting that periodicity alters the distribution of the realized variance, and consequently, its forecasts.

4.2.2 Forecasting Results

In-sample Forecasts

Tables 2, 3, 4 and 5 report the regression results for all HAR and HARP models, estimated on the entire sample, for SPY and a stock average. Estimated standard errors are robust to heteroscedasticity and autocorrelation, as we allow for serial correlation of up to orders 5, 10 and 44 for the 1-day, 5-days and 22-days models, respectively. We compute both in-sample and out-of-sample R-squared coefficients, reported as R^2_{is} and R^2_{oos}, where the computation of R^2_{oos} is based on Campbell and Thompson (2007) and uses over 3,000 observations. The bold R-squared coefficients are the ones bigger in the comparison between HARP and HAR models.

In general, all tables show that for SPY, R^2_{is} and R^2_{oos} from HARP models are higher in the majority of cases, irrespective of the forecasting horizon. In addition, the coefficients' standard errors for 1-day ahead models are generally lower following filtering, across all model specifications considered, suggesting that at short horizons, HARP models tend to be better specified than HAR models. Finally, averaging results across stocks leads to similar findings, with HARP models consistently outperforming HAR models.

All tables report for SPY $\widehat{\beta_d}+\widehat{\beta_w}+\widehat{\beta_m}$ ($\widehat{\beta_{C_d}}+\widehat{\beta_{C_w}}+\widehat{\beta_{C_m}}$ for the HAR-RV-CJ model), which represents the level of persistence when the models are autoregressions (all HAR except HAR-RV-CJ). For HARP models, this number gives some indication on the level of persistence in an autoregressive model where the dependent variable is measured with error, due to the presence of periodicity. As HARP models are not nested in the HAR class, comparisons of persistence between the HARP and HAR models should be interpreted with caution. While all models show a very high degree of persistence, we observe lower levels of persistence for all HARP models over all horizons. At the same time, the levels of persistence in the residuals of the estimated HARP models are much lower than for the HAR models. This confirms that HARP models are generally better specified and explains why these models outperform HAR models in forecasting.

Table 2: Estimated 1-, 5-, and 22- day ahead HAR(P)-RV models for SPY and a stock average.

	HAR			HARP		
	$h = 1$	$h = 5$	$h = 22$	$h = 1$	$h = 5$	$h = 22$
$\widehat{\beta_0}$	0.095*	0.148**	0.288***	0.097*	0.148***	0.289***
s.e.	(0.054)	(0.059)	(0.058)	(0.053)	(0.058)	(0.060)
$\widehat{\beta_d}$	0.246**	0.184***	0.103***	0.217**	0.147***	0.089***
s.e.	(0.099)	(0.052)	(0.021)	(0.089)	(0.054)	(0.026)
$\widehat{\beta_w}$	0.422***	0.347***	0.322***	0.435***	0.392***	0.343***
s.e.	(0.142)	(0.102)	(0.108)	(0.142)	(0.110)	(0.124)
$\widehat{\beta_m}$	0.238**	0.323***	0.290***	0.239**	0.302***	0.273***
s.e.	(0.097)	(0.097)	(0.085)	(0.102)	(0.103)	(0.095)
R^2_{is}	**0.512**	0.629	0.562	0.511	**0.635**	**0.56**
R^2_{oos}	0.426	0.590	0.496	**0.484**	**0.617**	**0.523**
$\widehat{\beta_d} + \widehat{\beta_w} + \widehat{\beta_m}$	0.906	0.854	0.716	0.891	0.842	0.705
Average Stocks						
$\overline{R}^2_{is}$	0.455	0.595	0.582	**0.471**	**0.608**	**0.587**
$\overline{R}^2_{oos}$	0.344	0.548	**0.499**	**0.351**	**0.563**	0.497
$\widehat{\beta_d} + \widehat{\beta_w} + \widehat{\beta_m}$	0.891	0.845	0.740	0.861	0.815	0.710

Note: This table reports the regression coefficients, standard errors in parentheses, and in- and out-of-sample R-squared for the HAR-RV and HARP-RV models based on various horizons, estimated on SPY data. The standard errors are estimated using the Newey-West HAC estimator. The bottom panel shows the stock average in- and out-of-sample R-squared obtained for HAR-RV and HARP-RV models of various horizons. Bold numbers indicate the R-squared coefficients that are higher for filtered as opposed to unfiltered models. *, ** and *** denote significance at 10%, 5% and 1% respectively.

For SPY, the in-sample R-squared has similar values for the HARP-RV and HAR-RV 1-day ahead models, while for the 5- and 22-days ahead models, this coefficient is higher when the forecast is based on filtered data. In terms of out-of-sample R-squared, the HARP-RV model outperforms the HAR-RV model uniformly across all horizons. In addition, for the 1-day ahead model, filtering data leads to a decrease in the standard errors of coefficients $\widehat{\beta_0}$ and $\widehat{\beta_d}$, while the standard error of $\widehat{\beta_w}$ remains at the same level. The in-sample R-squared averaged across stocks is higher for the HARP-RV models for all forecasting horizons, while in terms of average out-of-sample R-squared, we have higher values for the 1-day and 5-days ahead HARP-RV models.

Table 3: Estimated 1-, 5-, and 22- day ahead HAR(P)-RV-J models for SPY and a stock average.

	HAR			HARP		
	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$\widehat{\beta}_0$	0.096*	0.149**	0.288***	0.098*	0.149***	0.289***
s.e.	(0.054)	(0.058)	(0.058)	(0.053)	(0.057)	(0.061)
$\widehat{\beta}_d$	0.250**	0.189***	0.107***	0.223**	0.153***	0.093***
s.e.	(0.104)	(0.053)	(0.023)	(0.091)	(0.056)	(0.027)
$\widehat{\beta}_w$	0.421***	0.345***	0.321***	0.431***	0.388***	0.340***
s.e.	(0.145)	(0.103)	(0.107)	(0.142)	(0.109)	(0.122)
$\widehat{\beta}_m$	0.239**	0.325***	0.291***	0.242**	0.304***	0.275***
s.e.	(0.097)	(0.097)	(0.084)	(0.102)	(0.103)	(0.094)
$\widehat{\beta}_{J_d}$	−0.243	−0.274	−0.181	−0.313	−0.308	−0.211
s.e.	(0.270)	(0.195)	(0.176)	(0.227)	(0.191)	(0.167)
R^2_{is}	**0.512**	0.630	0.551	0.511	**0.635**	**0.568**
R^2_{oos}	0.417	0.598	0.508	**0.480**	**0.614**	**0.514**
$\widehat{\beta}_d+\widehat{\beta}_w+\widehat{\beta}_m$	0.910	0.858	0.719	0.895	0.845	0.708
Average Stocks						
$\overline{R}^2_{is}$	0.460	0.599	0.584	**0.473**	**0.610**	**0.588**
$\overline{R}^2_{oos}$	0.348	0.552	**0.498**	**0.354**	**0.562**	0.496
$\widehat{\beta}_d+\widehat{\beta}_w+\widehat{\beta}_m$	0.904	0.855	0.742	0.865	0.818	0.708

Note: This table reports the regression coefficients, standard errors in parentheses, and in- and out-of-sample R-squared for the HAR-RV-J and HARP-RV-J models based on various horizons, estimated on SPY data. The standard errors are estimated using the Newey-West HAC estimator. The bottom panel shows the stock average in- and out-of-sample R-squared obtained for the HAR-RV-J and HARP-RV-J models of various horizons. Bold numbers indicate the R-squared coefficients that are higher for filtered as opposed to unfiltered models. *, ** and *** denote significance at 10%, 5% and 1% respectively.

In the case of the HAR-RV-J and HARP-RV-J models, $\widehat{\beta_{J_d}}$ is always negative, in line with the existing literature (Andersen et al., 2007a, see). For the HARP-RV-J, $\widehat{\beta_{J_d}}$ is larger in absolute value and has smaller standard errors compared to the HAR-RV-J model. As filtering out periodicity reduces the number of detected spurious jumps (see section 3.3), the jump predictor for the HARP-RV-J model is less affected by measurement error and, as a result, it is more informative. In addition, in the case of the 1-day ahead model, we observe lower standard errors after filtering for almost all other coefficients, i.e. $\widehat{\beta}_0$, $\widehat{\beta}_d$ and $\widehat{\beta}_w$.

For SPY, the in-sample R-squared coefficients are higher for the HARP-RV-J models in the case of the 5- and 22-days horizons, while the out-of-sample R-squared is higher for these models across all horizons. A particularly high change following filtering is observed for the 1-day ahead model, where the out-of-sample R-squared increases from 0.417 for

HAR-RV-J to 0.480 for HARP-RV-J. Average results for stocks show higher average in-sample R-squared coefficients for HARP-RV-J models across all horizons. In the case of the average out-of-sample R-squared, the superiority of filtered data is preserved for the 1-day and 5-days ahead models.

Table 4: Estimated 1-, 5-, and 22-days ahead HAR(P)-RV-CJ models for SPY and a stock average.

	HAR			HARP		
	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$\widehat{\beta_0}$	0.098*	0.149***	0.284***	0.101*	0.152***	0.290***
s.e.	(0.054)	(0.053)	(0.055)	(0.052)	(0.053)	(0.058)
$\widehat{\beta_{C_d}}$	0.241**	0.172***	0.100***	0.218**	0.142***	0.088***
s.e.	(0.105)	(0.051)	(0.022)	(0.092)	(0.053)	(0.026)
$\widehat{\beta_{C_w}}$	0.448***	0.399***	0.350***	0.440***	0.412***	0.350***
s.e.	(0.149)	(0.098)	(0.117)	(0.142)	(0.102)	(0.123)
$\widehat{\beta_{C_m}}$	0.235***	0.306***	0.262***	0.247**	0.307***	0.275***
s.e.	(0.087)	(0.091)	(0.089)	(0.099)	(0.100)	(0.096)
$\widehat{\beta_{J_d}}$	0.255	0.379	0.100	0.107	0.280	0.051
s.e.	(0.246)	(0.329)	(0.142)	(0.201)	(0.326)	(0.119)
$\widehat{\beta_{J_w}}$	−0.961	−2.368**	−0.941	−0.542	−1.917*	−0.570
s.e.	(0.603)	(1.179)	(0.790)	(0.493)	(1.134)	(0.549)
$\widehat{\beta_{J_m}}$	0.522	1.520	2.010	0.163	0.873	0.567
s.e.	(1.735)	(2.054)	(2.051)	(1.460)	(1.806)	(1.708)
R^2_{is}	**0.514**	0.641	0.554	0.512	**0.643**	**0.570**
R^2_{oos}	0.407	0.589	0.485	**0.468**	**0.615**	**0.526**
$\widehat{\beta_{C_d}}+\widehat{\beta_{C_w}}+\widehat{\beta_{C_m}}$	0.924	0.878	0.712	0.905	0.862	0.714
Average Stocks						
$\overline{R}^2_{is}$	0.462	0.604	0.592	**0.475**	**0.613**	**0.595**
$\overline{R}^2_{oos}$	0.312	0.554	**0.508**	**0.338**	**0.565**	0.507
$\widehat{\beta_{C_d}}+\widehat{\beta_{C_w}}+\widehat{\beta_{C_m}}$	0.897	0.839	0.725	0.857	0.802	0.693

Note: This table reports the regression coefficients, standard errors in parentheses, and in- and out-of-sample R-squared for the HAR-RV-CJ and HARP-RV-CJ models based on various horizons, estimated on SPY data. The standard errors are estimated using the Newey-West HAC estimator. The bottom panel shows the stock average in- and out-of-sample R-squared obtained for the HAR-RV-CJ and HARP-RV-CJ models of various horizons. Bold numbers indicate the R-squared coefficients that are higher for filtered as opposed to unfiltered models. *, ** and *** denote significance at 10%, 5% and 1% respectively.

In line with our findings for the HARP-RV-J model, for the HARP-RV-CJ model, we notice an important reduction in the standard errors for the coefficients of all realized jumps repressors across all forecasting horizons in comparison to the unfiltered model. For the 1-day ahead model, the standard errors of $\widehat{\beta_0}$, $\widehat{\beta_{C_d}}$ and $\widehat{\beta_{C_w}}$ also show a substantial decrease after filtering.

The HARP-RV-CJ models generally show higher in-sample and out-of-sample R-squared for SPY and on average across all stocks. In particular, for SPY, the out-of-sample R-squared shows the highest increase after data filtering for this class of models compared to all other HARP models. The most extreme change occurs for the 1-day ahead model, where the out-of-sample R-squared increases from 0.407 for HAR-RV-CJ to 0.468 for HARP-RV-CJ.

Table 5: Estimated 1-, 5-, and 22-days ahead HAR(P)Q models for SPY and a stock average.

	HAR			HARP		
	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$\widehat{\beta_0}$	-0.029	0.072	0.218***	-0.002	0.077	0.228***
s.e.	(0.055)	(0.064)	(0.061)	(0.052)	(0.064)	(0.062)
$\widehat{\beta_d}$	0.658***	0.436***	0.334***	0.578***	0.407***	0.311***
s.e.	(0.097)	(0.116)	(0.101)	(0.083)	(0.105)	(0.077)
$\widehat{\beta_w}$	0.306**	0.275***	0.257***	0.305**	0.299***	0.264***
s.e.	(0.127)	(0.101)	(0.091)	(0.137)	(0.112)	(0.115)
$\widehat{\beta_m}$	0.129	0.257**	0.230**	0.159	0.244**	0.224**
s.e.	(0.116)	(0.112)	(0.101)	(0.102)	(0.109)	(0.101)
$\widehat{\beta_Q}$	-0.010***	-0.006***	-0.005***	-0.008***	-0.006***	-0.005***
s.e.	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
R^2_{is}	**0.539**	0.644	0.578	0.535	**0.653**	**0.585**
R^2_{oos}	0.544	0.628	**0.466**	**0.561**	**0.665**	0.463
$\widehat{\beta_d}+\widehat{\beta_w}+\widehat{\beta_m}$	1.093	0.968	0.821	1.042	0.950	0.798
Average Stocks						
$\overline{R}^2_{is}$	0.472	0.610	0.596	**0.485**	**0.622**	**0.603**
$\overline{R}^2_{oos}$	0.375	0.551	**0.495**	**0.399**	**0.567**	0.489
$\widehat{\beta_d}+\widehat{\beta_w}+\widehat{\beta_m}$	0.977	0.910	0.793	0.937	0.887	0.774

Note: This table reports the regression coefficients, standard errors in parentheses, and in- and out-of-sample R-squared for the HARQ and HARPQ models based on various horizons, estimated on SPY data. The standard errors are estimated using the Newey-West HAC estimator. The bottom panel shows the stock average in- and out-of-sample R-squared obtained for the HARQ and HARPQ models of various horizons. Bold numbers indicate the R-squared coefficients that are higher for filtered as opposed to unfiltered models. *, ** and *** denote significance at 10%, 5% and 1% respectively.

Standard errors for $\widehat{\beta_Q}$, the estimated coefficient for $RQ_t^{1/2}RV_t$, as illustrated in equation (7), are lower for the HARPQ model than for the HARQ model. As seen in section 3, periodicity impacts the estimated asymptotic RV variance, RQ_t, leading to additional distortions to results for this model. Pre-filtering data diminishes the periodicity-related bias in RQ_t and thus leads to more precise estimates of β_Q. For the 1-day ahead model, we also observe lower standard errors of $\widehat{\beta_0}$, $\widehat{\beta_d}$ and $\widehat{\beta_m}$ for the HARPQ model. Moreover, filtered models outperform unfiltered models in 5 out of 6 cases, for SPY, as well as on average across stocks.

Appendix D reports, as a robustness check, the results obtained for the HAR(P)-RV-J and HAR(P)-RV-CJ models where the Andersen et al. (2012) test was used to identify jumps.

Out-of-sample Forecasts

We compute out-of-sample forecast losses for all HAR and HARP models. The ratios of the losses from HARP versus HAR models for the 1-day, 1-week and 1-month horizons are reported in tables 6, 7, and 8. In each table, the top panel shows results for the SPY and the average across all stocks. Bold numbers indicate that filtered models outperform their counterpart. We further compare the performance of HARP models to that of HAR models by applying the Diebold and Mariano (1995) test. Let ϵ_t^u be the errors from one of the HAR models in equations (4), (5),(6) and (7) and ϵ_t^f errors from these models' HARP counterparts. Further, let $L(\cdot)$ denote one of the loss functions in (15) and $d_t^{u/f} = L(\epsilon_t^u) - L(\epsilon_t^f)$. Then, the Diebold and Mariano (1995) test statistic is defined as:

$$DM^{u/f} = \frac{\frac{1}{T}\sum_{t=1}^{T} d_t^{u/f}}{\sqrt{\widehat{\text{Var}}\left(\frac{1}{T}\sum_{t=1}^{T} d_t^{u/f}\right)}} \rightarrow \mathcal{N}(0,1), \tag{16}$$

where $\widehat{\text{Var}}\left(\frac{1}{T}\sum_{t=1}^{T} d_t^{u/f}\right)$ is a consistent estimator for the variance of the $d_t^{u/f}$ sample mean. We run a right tailed test, where rejection means that the average loss from HARP models is lower than the average loss from HAR models. In the tables, starred numbers indicate significance at 5% significance level.

The bottom panel in all tables states the number of stocks for which HARP models outperform HAR models based on the Diebold and Mariano (1995) test applied at a 5% significance level. For the models including jumps, we present results relying on both tests for jumps considered in this paper, the classic Barndorff-Nielsen and Shephard (2006a) test, based on the realized bipower variation (BV columns), as well as the Andersen et al. (2012) test, relying on the median realized variance (MedRV columns).

Table 6: Out-of-sample forecast losses – $h = 1$

		HARP-RV/ HAR-RV	HARPQ/ HARQ	BV HARP-RV-J/ HAR-RV-J	BV HARP-RV-CJ/ HAR-RV-CJ	MedRV HARP-RV-J/ HAR-RV-J	MedRV HARP-RV-CJ/ HAR-RV-CJ
SPY	MSE	**0.898****	**0.962**	**0.891****	**0.897****	**0.894****	**0.896****
	QLIKE	**0.998**	1.317	**0.995**	**0.988****	**0.985****	**0.985****
Avg. Stocks	MSE	**0.983**	**0.963**	**0.987**	**0.972**	**0.978**	**0.973**
	QLIKE	**0.964**	**0.956**	**0.975**	**0.978**	**0.974**	**0.973**
Diebold & Mariano Test – Individual Stocks							
	MSE	11	9	11	12	8	8
	QLIKE	20	20	18	10	17	15

Note: This table reports the ratio of the losses from HARP versus HAR models for various forecasting horizons. Bold numbers indicate that HARP models outperform their HAR counterparts. Doubly starred highlights the HARP models whose losses are significantly lower than their HAR model counterpart based on the Diebold and Mariano test at the 5% level. The entries in the bottom panel are the number of stocks for which the HARP model is significantly better than its HAR model counterpart at the 5% significance level.

In the case of the 1-day models for SPY, all except one loss ratios take values below 1, with the MSE ratios ranging just above 0.89. For both loss functions, the lowest ratios are observed for the models with realized jumps in their specifications (last four columns). This is in line with the in-sample results for SPY in paragraph 4.2.2, where we observed lower standard errors at 1-day ahead and higher R-squared coefficients for the HARP-RV-J and HARP-RV-CJ models in comparison to their HAR counterparts. As shown in section 3, periodicity impacts HAR models with jumps via two channels: distortions in the higher moments of the integrated variance estimators and measurement error in the jump regressor. Filtering out periodicity addresses distortions via both channels and leads to better specified models and improved forecasts at short horizons.

When using the MSE loss criterion, the Diebold and Mariano (1995) test indicates a significant gain (at 5% significance level) from forecasting the SPY RV based on filtered data for all but one models. When the QLIKE loss criterion is used, we find significance for all but one models with jumps in their specification.

The average loss ratios for all considered stocks and all models are below 1, indicating that filtering periodicity helps to improve forecasting for the majority of stocks. The second line in the bottom panel shows that the MSE from HARP models is significantly lower than the MSE from HAR models for a number of 8 to 12 stocks, and an average of 10 stocks.[9] The last line of the bottom panel shows that the QLIKE loss is lower for HARP models for a number of stocks ranging between 10 and 20, with a high average of 17, which is over half the number of stocks considered.

[9] Please note that failure to reject in this case does not imply superiority of HAR models, but just that differences are not statistically significant.

Table 7: Out-of-sample forecast losses – $h = 5$

				BV		MedRV	
		HARP-RV/ HAR-RV	HARPQ/ HARQ	HARP-RV-J/ HAR-RV-J	HARP-RV-CJ/ HAR-RV-CJ	HARP-RV-J/ HAR-RV-J	HARP-RV-CJ/ HAR-RV-CJ
SPY	MSE	**0.934****	**0.904****	**0.961****	**0.937****	**0.936**	**0.921**
	QLIKE	**0.994**	**0.932****	**0.998**	**0.968****	**0.983**	**0.951****
Avg. Stocks	MSE	**0.964**	**0.961**	**0.980**	**0.976**	**0.969**	**0.953**
	QLIKE	**0.978**	**0.983**	**0.985**	**0.993**	**0.984**	**0.980**
Diebold & Mariano Test – Individual Stocks							
	MSE	14	11	10	8	10	11
	QLIKE	13	14	13	8	12	14

Note: This table reports the ratio of the losses from HARP versus HAR models for various forecasting horizons. Bold numbers indicate that HARP models outperform their HAR counterparts. Doubly starred highlights the HARP models whose losses are significantly lower than their HAR model counterpart based on the Diebold and Mariano test at the 5% level. The entries in the bottom panel are the number of stocks for which the HARP model is significantly better than its HAR model counterpart at the 5% significance level.

For the 5-days models, all loss ratios are below 1 in the case of SPY. Moreover, the Diebold and Mariano (1995) test shows that the first four MSE losses are significantly lower for HARP models in comparison to HAR models. For the QLIKE loss criterion, we observe significantly lower losses for the HARPQ model and both HARP-RV-CJ models (BV and MedRV) in comparison to the HAR counterpart models. In the case of models with jumps in their specifications (last four columns), while loss ratios are always below 1, they are not lower than the ratios for the other models, which was the case when forecasting 1-day ahead. This is likely due to the fact that the impact of periodicity on jump predictors is diluted when they aggregate data over several days.

Across all models, the average loss ratios for all considered stocks are below 1, indicating, just as for the 1-day ahead model, that filtering periodicity is beneficial for the majority of stocks. Finally, the bottom panel shows that the MSE is significantly lower in the case of HARP models for a number of stocks ranging between 8 and 14, with an average of 11 stocks. In addition, the HARP QLIKE loss is significantly lower for a number of stocks ranging between 8 and 14, with an average of 12.

Table 8: Out-of-sample forecast losses – $h = 22$

				BV		MedRV	
		HARP-RV/ HAR-RV	HARPQ/ HARQ	HARP-RV-J/ HAR-RV-J	HARP-RV-CJ/ HAR-RV-CJ	HARP-RV-J/ HAR-RV-J	HARP-RV-CJ/ HAR-RV-CJ
SPY	MSE	**0.947****	1.006	**0.988**	**0.921****	**0.971**	**0.967**
	QLIKE	**0.970****	**0.999**	**0.994**	**0.971****	**0.977****	**0.973****
Avg. Stocks	MSE	**0.995**	1.006	**0.995**	**0.998**	**0.999**	**0.999**
	QLIKE	1.002	1.014	1.002	1.017	1.004	**0.998**
Diebold & Mariano Test – Individual Stocks							
	MSE	7	4	5	6	5	6
	QLIKE	8	3	7	6	7	8

Note: This table reports the ratio of the losses from HARP versus HAR models for various forecasting horizons. Bold numbers indicate that HARP models outperform their HAR counterparts. Doubly starred highlights the HARP models whose losses are significantly lower than their HAR model counterpart based on the Diebold and Mariano test at the 5% level. The entries in the bottom panel are the number of stocks for which the HARP model is significantly better than its HAR model counterpart at the 5% significance level.

Table 8 presents the out-of-sample results for the 22- month ahead models. For SPY, all but one ratios is below 1. The exception occurs for the MSE loss ratio HARPQ/HARQ, but even in this case, the ratio is very close to 1. The Diebold and Mariano (1995) test indicates that the MSE is significantly lower for HARP-RV and HARP-RV-CJ based on the realized bipower variation when these models are compared to their HAR counterparts. The QLIKE loss is significantly lower for HARP-RV, HARP-RV-CJ based on both jump tests, and HARP-RV-J based on the Andersen et al. (2012) test.

The stock average MSE ratio is lower than 1 in all cases except for HARPQ/HARQ, while the stock average QLIKE loss ratio is lower than 1 in only one case. In the bottom panel, MSE is significantly lower in the case HARP models for a number of stocks ranging between 4 and 7, with an average of 6 stocks, while the HARP QLIKE loss is significantly lower for a number of stocks ranging between 3 and 8, with an average of 7.

Both our Monte Carlo and empirical findings suggest that HARP models display a better out-of-sample performance than HAR models at short and medium horizons. However, we find mixed evidence for longer horizons, indicating that distortions due to intraday periodicity are mostly negligible in this case and realized measures based on unfiltered data produce as accurate forecasts as the filtered measures.

5 Conclusion

The contribution of this paper is twofold. Firstly, we document the impact of volatility intraday periodicity on forecasting the realized variance using heterogenous autoregressive (HAR) models. While periodicity has no impact on the realized volatility itself, it distorts its variance, leading to biases in the coefficients of forecasting models. We derive the variance and the 1-lag auto-correlation coefficient for the realized variance in the case of a very simple DGP and show that periodicity artificially inflates the variance and has a decreasing impact on the auto-correlation. For a more complex DGP, we provide simulation evidence showing that the estimated coefficients of the forecasting regression are closer to their true values when predictors are built from pre-filtered returns. In addition, we provide simulation evidence that periodicity distorts the jump test results, using two of the most known tests for jumps, the Barndorff-Nielsen and Shephard (2006a) and the Andersen et al. (2012) tests. As the realized daily squared jumps enter some of the frequently applied forecasting models, their periodicity related bias can further impact forecasting.

Secondly, we introduce a new class of forecasting models for the realized variance, HARP, where predictors rely on data from which periodicity is filtered out. We provide a thorough set of in-sample and out-of-sample forecasting comparisons between the HARP and HAR models, relying on both simulated and real data. Our analysis encompasses the HARP versions of the most common HAR models in the literature, the HAR-RV model by Corsi (2009), the HAR-RV-J and HAR-RV-CJ models by (Andersen et al., 2007a), and the HARQ model by Bollerslev et al. (2016). Our dataset includes intraday observations for the SPDR ETF and 30 S&P500 constituents, covering a period of over 4,000 trading days. The simulation and empirical evidence indicates that pre-filtering the data for periodicity leads to forecasting gains for all model specifications when forecasting 1-day to 5-days ahead. At the 1-day ahead horizon, the HARP-RV-J and HAR-RV-CJ models show the greatest improvements following filtering, due to lower distortions in the jump predictors. Findings are mixed when forecasting 22-days ahead, for which the increased forecasting error is likely to dilute the impact of periodicity.

References

Aït-Sahalia, Y. and Jacod, J. (2009). Testing for jumps in a discretely observed process. *Annals of Statistics*, 37:184–222.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*, 18(2):351–416.

Andersen, T., Bollerslev, T., and Diebold, F. (2007a). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89(4):701–720.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2):115 – 158.

Andersen, T. G. and Bollerslev, T. (1998a). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Andersen, T. G. and Bollerslev, T. (1998b). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *The Journal of Finance*, 53(1):219–265.

Andersen, T. G., Bollerslev, T., and Das, A. (2001). Variance-ratio statistics and high-frequency data: Testing for changes in intraday volatility patterns. *The Journal of Finance*, 56(1):305–327.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andersen, T. G., Bollerslev, T., and Dobrev, D. (2007b). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics*, 138:125–180.

Andersen, T. G., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169(1):75–93.

Andersen, T. G., Thyrsgaard, M., and Todorov, V. (2018). Time-varying periodicity in intraday volatility. *Journal of the American Statistical Association*, 0(ja):1–39.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., and Shephard, N. (2006a). Limit theorems for bipower variation in financial econometrics. *Econometric Theory*, 22(04):677–719.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Barndorff-Nielsen, O. E. and Shephard, N. (2006a). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30.

Barndorff-Nielsen, O. E. and Shephard, N. (2006b). Multipower variation and stochastic volatility. pages 73–82. Springer.

Barndorff-Nielsen, O. E., Shephard, N., and Winkel, M. (2006b). Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and their Applications*, 116:796–806.

Bollerslev, T., Cai, J., and Song, F. M. (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the us treasury bond market. *Journal of Empirical Finance*, 7(1):37 – 55.

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

Boudt, K., Croux, C., and Laurent, S. (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance*, 18(2):353–367.

Campbell, J. Y. and Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Chortareas, G., Jiang, Y., and Nankervis, J. C. (2011). Forecasting exchange rate volatility using high-frequency data: Is the euro different? *International Journal of Forecasting*, 27(4):1089 – 1107.

Christensen, K. and Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141(2):323 – 349.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Pirino, D., and Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288.

Corsi, F. and Reno, R. (2012). Discretetime volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30(3):368–380.

Dacorogna, M., Müller, A., Pictet, V., and Olsen, R. (1997). Modelling short-term volatility with garch and harch models. *Nonlinear Modelling of High Frequency Financial Time Series.*

Deo, R., Hurvich, C., and Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics*, 131(1):29 – 58.

Dette, H., Golosnoy, V., and Kellermann, J. (2016). *The Effect of Intraday Periodicity on Realized Volatility Measures.* Discussion Paper / SFB823. Universitätsbibliothek Dortmund.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Dumitru, A.-M. and Urga, G. (2012). Identifying jumps in financial assets: a comparison between nonparametric jump tests. *Journal of Business & Economic Statistics*, 30(2):242–255.

Frijns, B. and Margaritis, D. (2008). Forecasting daily volatility with intraday data. *The European Journal of Finance*, 14(6):523–540.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Hasbrouck, J. (1999). The dynamics of discrete bid and ask quotes. *The Journal of Finance*, 54(6):2109–2142.

Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3(4):456–499.

Jiang, G. J. and Oomen, R. (2008). Testing for jumps when asset prices are observed with noise- a “swap variance” approach. *Journal of Econometrics*, 144:352–370.

Lee, S. S. and Mykland, P. A. (2008). Jumps in financial markets: a new nonparametric test and jump dynamics. *Review of Financial studies*, 21:2535–2563.

Mancini, C. (2004). Estimation of the characteristics of the jumps of a general poisson-diffusion model. *Scandinavian Actuarial Journal*, 1:42–52.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36:270–296.

Martens, M., Chang, Y.-C., and Taylor, S. J. (2002). A comparison of seasonal adjustment methods when forecasting intraday volatility. *Journal of Financial Research*, 25(2):283–299.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., and von Weizsäcker, J. E. (1997). Volatilities of different time resolutions - analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2):213–239.

Podolskij, M. and Ziggel, D. (2010). New tests for jumps in semimartingale models. *Statistical Inference for Stochastic Processes*, 13:15–41.

A Some Proofs for the Simple AR(1) Model

Under the assumptions of section 3.1, $RV_t = \sum_{i=1}^{M} r_{t,i}^2 = \Delta IV_t \sum_{i=1}^{M} f_i^2 w_i^2$.

$$\mathbb{E}(RV_t) = \Delta\, \mathbb{E}(IV_t) \sum_{i=1}^{M} f_i^2 = 0,$$

where we used the fact that $\mathbb{E}(w_i^2) = 1$ and $\mathbb{E}(IV_t) = 0$ given the DGP for IV_t in equation (9).

Proof of equation (12a).

$$\begin{aligned}
\text{Var}(RV_t) = \ \mathbb{E}\left(RV_t^2\right) &= \Delta^2\, \mathbb{E}(IV_t^2)\, \mathbb{E}\left(\sum_{i=1}^{M} f_i^4 w_i^4 + \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j\neq i}}^{M} f_i^2 w_i^2 f_j^2 w_j^2\right) \\
&= \Delta^2\, \mathbb{E}(IV_t^2)\left(3\sum_{i=1}^{M} f_i^4 + \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j\neq i}}^{M} f_i^2 f_j^2\right) \\
&= \Delta^2 \frac{\sigma_\epsilon^2}{1-\Phi^2}\left[3\sum_{i=1}^{M} f_i^4 + \sum_{i=1}^{M} f_i^2 \left(\frac{1}{\Delta} - f_i^2\right)\right] \\
&= \Delta^2 \frac{\sigma_\epsilon^2}{1-\Phi^2}\left(2\sum_{i=1}^{M} f_i^4 + \frac{1}{\Delta^2}\right). \qquad \square
\end{aligned}$$

For comparison purposes, we compute the same variance in the absence of periodicity:

Proof of equation (12b).

$$\begin{aligned}
\text{Var}(RV_t)^{\substack{no \\ periodicity}} &= \Delta^2\, \mathbb{E}(IV_t^2)\, \mathbb{E}\left(\sum_{i=1}^{M} w_i^4 + \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j\neq i}}^{M} w_i^2 w_j^2\right) \\
&= \Delta^2\, \mathbb{E}(IV_t^2)\left[3M + M(M-1)\right] = \Delta^2 \frac{\sigma_\epsilon^2}{1-\Phi^2}\left(\frac{2}{\Delta} + \frac{1}{\Delta^2}\right). \qquad \square
\end{aligned}$$

Let w_i, $i = 1, \dots, M$ be a sequence of i.i.d. standard normal variables entering the intraday returns on day t and w_i^*, $i = 1, \dots, M$ another sequence of i.i.d. standard normals, independent of w_i, entering returns on day $t-h$, $h \geq 1$. The auto-covariance of lag h is obtained below.

Auto-covariance derivation.

$$\begin{aligned}
\text{cov}\left(RV_t, RV_{t-h}\right) = \; & \mathbb{E}\left(RV_t RV_{t-h}\right) = \Delta^2\,\mathbb{E}\left[\mathbb{E}\left(IV_t IV_{t-h}\sum_{i=1}^{M} f_i^2 w_i^2 \sum_{j=1}^{M} f_j^2 w_j^{*2} | IV_{t-h}\right)\right] \\
& = \Delta^2\,\mathbb{E}\left\{\mathbb{E}\left[IV_t IV_{t-h}\left(\sum_{i=1}^{M} f_i^4 w_i^2 w_i^{*2} + \sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M} f_i^2 w_i^2 f_j^2 w_j^{*2}\right)|IV_{t-h}\right]\right\} \\
& = \Delta^2\,\mathbb{E}\left[IV_{t-h}\,\mathbb{E}\left(IV_t|IV_{t-h}\right)\right]\left(\sum_{i=1}^{M} f_i^4 + \sum_{\substack{j=1\\ j\neq i}}^{M} f_i^2 f_j^2\right) \\
& = \Delta^2\,\mathbb{E}\left[IV_{t-h}\,\mathbb{E}\left(\Phi^h IV_{t-h} + \epsilon_t + \Phi\epsilon_{t-1} + \cdots + \Phi^{h-1}\epsilon_{t-h+1}|IV_{t-h}\right)\right]\cdot \\
& \left[\sum_{i=1}^{M} f_i^4 + \sum_{i=1}^{M} f_i^2\left(\frac{1}{\Delta} - f_i^2\right)\right] = \Delta^2\,\mathbb{E}\left(\Phi^h s^4(t-h)\right)\frac{1}{\Delta^2} = \Phi^h\frac{\sigma_\epsilon^2}{1-\Phi^2}.
\end{aligned}$$

□

B More on Methodology

B.1 Weighted Standard Deviation

For each $i = 1,\ldots,M$, we observe T standardized returns, $\bar{r}_{t,i}$, which we sort in increasing order, as follows:

$$\bar{r}_{(1),i} \leq \bar{r}_{(2),i} \leq \cdots \leq \bar{r}_{(T),i}$$

Given the above ordered set, we define the sub-sets containing half ($\kappa = \lfloor T/2 \rfloor + 1$) contiguous observations: $\{\bar{r}_{(1),i},\ldots,\bar{r}_{(\kappa),i}\},\ldots,\{\bar{r}_{(T-\kappa+1),i},\ldots,\bar{r}_{(T),i}\}$. The shortest half scale estimator is the shortest length of these subsets:

$$ShortH_i = 0.741\min(\bar{r}_{(\kappa),i} - \bar{r}_{(1),i},\ldots,\bar{r}_{(T),i} - \bar{r}_{(T-\kappa+1),i})$$

The short-half periodicity estimator is given by:

$$\hat{f}_i^{ShortH} = \frac{ShortH_i}{\Delta\sum_{j=1}^{M} ShortH_j^2}, \qquad \Delta \equiv 1/M$$

The weights used to compute the weighted standard deviation in section 2.2 are defined, for all $l = 1,\ldots,T$ and all $i = 1,\ldots,M$, as:

$$\chi_{l,i} = \chi(\bar{r}_{l,i}/\hat{f}_i^{ShortH}),$$
$$\chi(z) = \begin{cases} 1 & \text{if } z^2 \leq 6.635 \\ 0 & \text{otherwise.} \end{cases}$$

B.2 Jump Tests

In this paper, we identify jumps relying mostly on the test proposed by Barndorff-Nielsen and Shephard (2006a) and further developed by Huang and Tauchen (2005). The test statistic, Z_t^{BV}, is given by:

$$Z_t^{BV} = \frac{1 - BV_t/RV_t}{\sqrt{0.61M^{-1}\max(1, TPQ_t/BV_t^2)}} \sim \mathcal{N}(0,1)$$

where TPQ_t is the realized tripower quarticity, that consistently estimates the integrated quarticity in the presence of jumps and is defined as:

$$TPQ_t = M1.74\frac{M}{M-2}\sum_{i=3}^{M}|r_{t,i}|^{4/3}|r_{t,i-1}|^{4/3}|r_{t,i-2}|^{4/3} \xrightarrow{p} \int_{t-1}^{t}\sigma^4(u)\,\mathrm{d}u.$$

The above test is widely used in empirical work due to its simplicity and reasonable size and power properties under various scenarios (Dumitru and Urga, 2012, see). As documented in section 1, there are several other tests for jumps in the literature. In this paper, we also employ the test proposed by Andersen et al. (2012) to make sure our results are robust to the choice of the jump test. This test relies on the median realized variance to estimate the integrated variation and is shown to have better finite sample properties than the original test by Barndorff-Nielsen and Shephard (2006a). The test statistic is given below:

$$Z_t^{MedRV} = \frac{1 - MedRV_t/RV_t}{\sqrt{0.96M^{-1}\max(1, MedRQ_t/MedRV_t^2)}} \sim \mathcal{N}(0,1),$$

with

$$MedRV_t = \frac{M}{M-2}1.42\sum_{i=2}^{M-1}\mathrm{med}(|r_{t,i-1}|,|r_{t,i}|,|r_{t,i+1}|)^2 \xrightarrow{p} \int_{t-1}^{t}\sigma^2(u)\,\mathrm{d}u$$

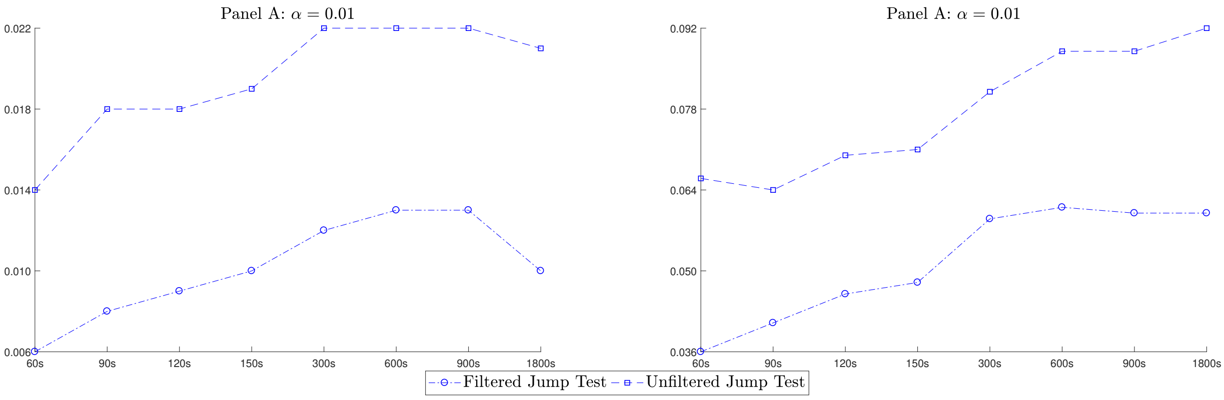
and

$$MedRQ_t = \frac{M}{M-2}0.92\sum_{i=2}^{M-1}\mathrm{med}(|r_{t,i-1}|,|r_{t,i}|,|r_{t,i+1}|)^4 \xrightarrow{p} \int_{t-1}^{t}\sigma^4(u)\,\mathrm{d}u.$$

C Additional Results

C.1 Spurious Detection of Jumps for the Jump Test by Andersen et al. (2012)

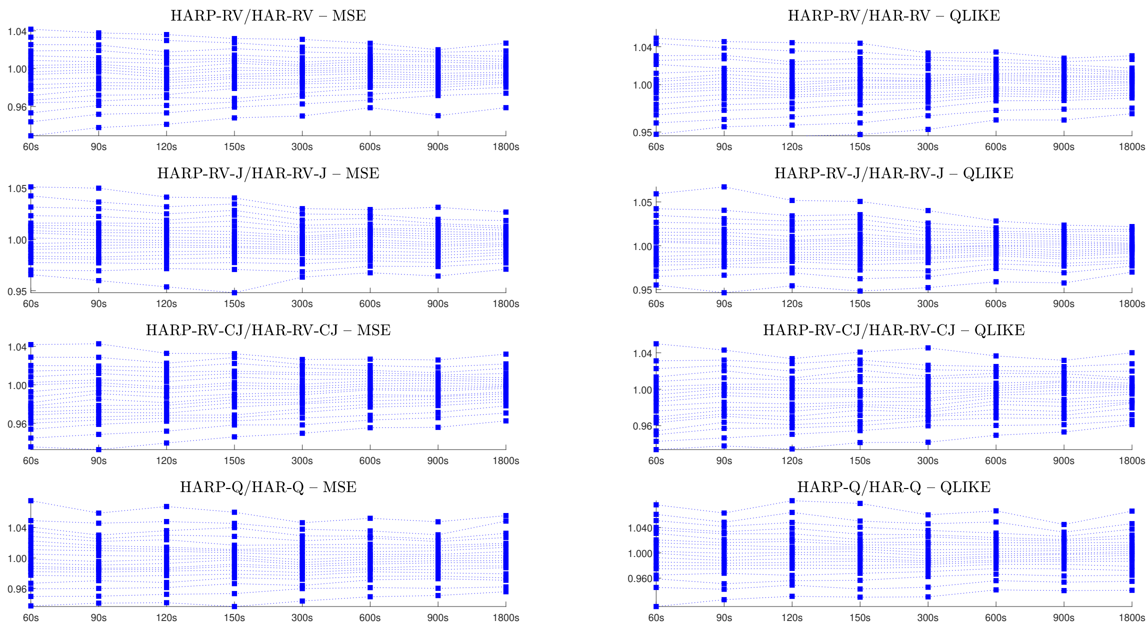
Figure 9: Proportion of spurious jumps by sampling frequency for filtered and unfiltered data



Note: This plot graphs the proportion of surious detection of jumps across sampling frequencies using the ADS test, Andersen et al. (2012), evaluated at the 1% and 5% significance level.

C.2 Simulated $h = 22$ Out-of-Sample Loss Ratio

Figure 10: Loss ratio for simulated data – $h = 22$



Note: This figure plots the quantiles ranging from 0.05 to 0.95 in increments of 0.05 for the MSE and QLIKE loss ratios for models relying on filtered data versus unfiltered data.

C.3 Results for the jump test by Andersen et al. (2012)

Table 9: Realized number of jumps and the estimated proportion of integrated variance in the quadratic variation.

Stock	Ticker	Unfiltered # Jumps	Unfiltered %QV	Filtered # Jumps	Filtered %QV
SPDR ETF	SPY	368	97.303	295	97.404
3M	MMM	536	94.738	218	97.916
AK Steel	AKS	491	94.478	312	96.945
Arconic Inc.	ARNC	342	96.660	171	98.882
Brown-Forman	BFB	712	89.189	464	95.309
BT Group	BT	561	90.807	583	90.271
China Mobile	CHL	604	92.477	401	95.365
Citigroup	C	380	96.424	148	98.576
Coca-Cola	KO	548	94.172	177	98.626
DUKE Energy	DUK	522	95.028	205	98.285
eBay	EBAY	464	96.433	172	98.756
General Dynamics	GD	538	93.426	250	97.177
General Electric	GE	389	95.934	177	98.187
Halliburton	HAL	389	95.362	144	98.050
Home Depot	HD	404	95.855	144	98.892
Honeywell	HON	487	94.442	200	97.534
Humana	HUM	526	93.730	226	97.521
Intel	INTC	399	97.071	167	98.651
LVLT	LVLT	528	93.722	254	97.179
McDonald's	MCD	474	93.268	150	98.631
Microsoft	MSFT	435	96.203	164	98.678
ONEOK	OKE	652	88.230	457	92.536
Pfizer	PFE	469	93.613	154	98.229
Procter & Gamble	PG	551	92.869	180	96.772
Southern Co.	SO	465	94.757	164	98.590
Travelers Companies Inc	TRV	599	91.497	241	96.669
United Health	UNH	502	93.415	192	98.288
UPS	UPS	561	93.810	203	97.980
Verizon	VZ	487	93.912	186	98.021
Vodafone	VOD	303	97.503	215	98.408
Xerox	XRX	488	93.787	268	97.793
	Avg. Stocks	494	94.094	233	97.424

Note: This table reports the total number of jump days using the ABD test at the 1% significance level and the %QV for filtered and unfiltered data. The %QV is estimated as $\%QV = \frac{\sum_{t=1}^{T} C_t}{\sum_{t=1}^{T}(C_t+J_t)}$.

D Additional In-Sample Results for Models with Jumps

Table 10: Estimated 1-, 5-, and 22- day ahead HAR(P)-RV-J models for SPY and a stock average, with jumps detected with the Andersen et al. (2012) test

	HAR			HARP		
	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$\widehat{\beta}_0$	0.095*	0.148***	0.288***	0.094*	0.146***	0.287***
s.e.	(0.054)	(0.057)	(0.058)	(0.051)	(0.055)	(0.061)
$\widehat{\beta}_d$	0.279**	0.209***	0.120***	0.297***	0.211***	0.135***
s.e.	(0.112)	(0.054)	(0.026)	(0.099)	(0.058)	(0.037)
$\widehat{\beta_w}$	0.396***	0.326***	0.309***	0.379**	0.348***	0.311**
s.e.	(0.149)	(0.104)	(0.107)	(0.151)	(0.117)	(0.127)
$\widehat{\beta_m}$	0.242**	0.327***	0.293***	0.232**	0.296***	0.269***
s.e.	(0.095)	(0.096)	(0.085)	(0.093)	(0.097)	(0.092)
$\widehat{\beta_{J_d}}$	−0.405**	−0.311**	−0.204	−0.523***	−0.411***	−0.303***
s.e.	(0.177)	(0.135)	(0.126)	(0.140)	(0.104)	(0.094)
R^2_{is}	0.515	0.632	0.553	**0.521**	**0.644**	**0.575**
R^2_{oos}	0.419	0.588	0.495	**0.481**	**0.614**	**0.510**
$\widehat{\beta}_d+\widehat{\beta_w}+\widehat{\beta_m}$	0.917	0.862	0.722	0.908	0.855	0.715
Average Stocks						
$\overline{R}^2_{is}$	0.461	0.599	0.584	**0.474**	**0.610**	**0.589**
$\overline{R}^2_{oos}$	0.341	0.552	**0.497**	**0.352**	**0.564**	0.494
$\widehat{\beta}_d+\widehat{\beta_w}+\widehat{\beta_m}$	0.904	0.854	0.743	0.865	0.817	0.710

Note: This table reports regression coefficients, standard errors in parentheses, and in- and out-of-sample R-squared for the HAR-RV-J and HARP-RV-J models based on various horizons, estimated on SPY data. The standard errors are estimated using the Newey-West HAC estimator. The bottom panel shows the stock average in- and out-of-sample R-squared obtained for the HAR-RV-J and HARP-RV-J models of various horizons. Bold numbers indicate the R-squared coefficients that are higher for filtered as opposed to unfiltered models. *, ** and *** denote significance at 10%, 5% and 1% respectively.

Table 11: Estimated 1-, 5-, and 22- day ahead HAR(P)-RV-CJ models for SPY and a stock average, with jumps detected with the Andersen et al. (2012) test

	HAR			HARP		
	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$\widehat{\beta_0}$	0.098*	0.150***	0.290***	0.096*	0.147***	0.289***
s.e.	(0.053)	(0.053)	(0.058)	(0.050)	(0.052)	(0.061)
$\widehat{\beta_{C_d}}$	0.268**	0.189***	0.110***	0.270***	0.171***	0.109***
s.e.	(0.113)	(0.052)	(0.023)	(0.102)	(0.053)	(0.026)
$\widehat{\beta_{C_w}}$	0.421***	0.378***	0.332***	0.432***	0.433***	0.364***
s.e.	(0.154)	(0.096)	(0.112)	(0.162)	(0.110)	(0.130)
$\widehat{\beta_{C_m}}$	0.245***	0.321***	0.296***	0.230**	0.281***	0.265***
s.e.	(0.095)	(0.099)	(0.094)	(0.098)	(0.102)	(0.099)
$\widehat{\beta_{J_d}}$	0.003	0.140	0.037	−0.078	0.015	−0.023
s.e.	(0.139)	(0.188)	(0.080)	(0.090)	(0.098)	(0.041)
$\widehat{\beta_{J_w}}$	−0.213	−0.889	−0.263	−0.230	−0.604**	−0.296**
s.e.	(0.238)	(0.572)	(0.256)	(0.147)	(0.304)	(0.130)
$\widehat{\beta_{J_m}}$	0.003	0.297	0.075	−0.276	−0.136	−0.174
s.e.	(0.780)	(0.950)	(0.575)	(0.300)	(0.378)	(0.298)
R^2_{is}	0.517	0.640	0.556	**0.526**	**0.657**	**0.582**
R^2_{oos}	0.401	0.581	0.494	**0.464**	**0.614**	**0.511**
$\widehat{\beta_d}+\widehat{\beta_w}+\widehat{\beta_m}$	0.934	0.888	0.738	0.932	0.885	0.738
Average Stocks						
$\overline{R}^2_{is}$	0.464	0.605	0.590	**0.476**	**0.615**	**0.595**
$\overline{R}^2_{oos}$	0.332	0.548	0.504	**0.347**	**0.565**	**0.505**
$\widehat{\beta_d}+\widehat{\beta_w}+\widehat{\beta_m}$	0.901	0.843	0.732	0.854	0.799	0.689

Note: This table reports the regression coefficients, standard errors in parentheses, and in- and out-of-sample R-squared for the HAR-RV-CJ and HARP-RV-CJ models based on various horizons, estimated on SPY data. The standard errors are estimated using the Newey-West HAC estimator. The bottom panel shows the stock average in- and out-of-sample R-squared obtained for the HAR-RV-CJ and HARP-RV-CJ models of various horizons. Bold numbers indicate the R-squared coefficients that are higher for filtered as opposed to unfiltered models. *, ** and *** denote significance at 10%, 5% and 1% respectively.