Essays on High Frequency Financial Econometrics: (Co)Jumps, Aggregation, Asymmetry and Measurement Error

by

Rodrigo Hizmeri

A Thesis Submitted to
the Economics Department at Lancaster University
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in Economics

May 28, 2021

Everything has been thought before, but the problem is to think of it again. Knowing is not enough; we must apply. Willing is not enough; we must do.
-Goethe

Declaration

I hereby declare that this thesis is my own work and has not been submitted for the award of a higher degree elsewhere. This thesis contains no material previously published or written by any other person except where references have been made in the thesis.

Rodrigo Hizmeri

June 15, 2021

Acknowledgements

Over the last four years I have greatly benefited from the support of many people. First and foremost, I want to thank my supervisors, Professor Marwan Izzeldin and Professor Mike Tsionas, for the great discussions, motivation and continuous support. I especially want to thank Marwan for: i) introducing me to high-frequency finance while I was a postgraduate student; ii) the persuasive conversations about pursuing PhD studies and following an academic path; iii) the very enjoyable discussions about research, life, football and more. I also enjoyed our football games every Friday night, and let me add they will be missed.[1]

I am grateful to Ingmar Nolte, Anthony Murphy and Ana-Maria Dumitru for the great conversations and support on many topics, including those that were beyond the scope of my research. A special tribute goes to Gerry Steele, whose editorial input has been instrumental, and for the many discussions and feedback about writing.

I also want to thank my PhD colleagues for the company, discussions and great time during our studies, I wish them all the very best and hope to keep in contact with each one of you. I am thankful to the academic and professional staff of the Economics Department at Lancaster University. I particularly want to thank Caren, Craig, Emma and Vicki for their administrative support throughout these years.

My gratitude also goes to the Economics Department at Lancaster University, the Gulf One Lab for Computational and Economic Research (GOLCER) and the Economic and Social Research Council (ESRC), whose financial support helped me in attending and presenting my work at very prestigious conferences around the world. I am particularly

[1] I am also thankful to all the football players that attended every rainy Friday night, for their friendship and for making this path even more joyful. I am not going to mention all of you, but you know who you are.

indebted to GOLCER and its executive training program for inviting me to teach modules in high-frequency finance, programming, time series econometrics, applied econometrics and LaTeX. Here is when I thank Teresa for helping us in organizing, advertising, handling, and many more things related to these events.

I am also grateful to the members of the dissertation committee, for their time and reading of this thesis.

Last but not least, I am supremely thankful to my wife Alejandra and my son Rod for, first of all, accepting the challenge of leaving our home country almost 10 years ago to pursue our dreams. With your support, help and understanding everything becomes much easier and enjoyable. Let me add that it has been a hell of a ride! I am also grateful to my parents and siblings, who always encouraged me to find and pursue my intellectual goals. I am deeply grateful for the many years of their unconditional support: Thank you!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Over the last twenty years, the use of statistical and econometric methods for analyzing high-frequency data has increased substantially. This growth has been driven by an increase in the availability of intraday data and technological advancements. High-frequency data provides a finer characterization of the elements comprised in asset prices; for instance, it makes it possible to discriminate jumps from the diffusive component. However, the most notable contribution of high-frequency data is the Realized Volatility (RV), which is estimated as the sum of all squared intraday returns. The RV is a consistent estimator (as $\Delta_n \to 0$) of the true latent volatility process, and as such it enables to treat volatility as "quasi"-observable. Whereas in the absence of jumps, the RV converges to the integrated variance, in the presence of jumps it converges to the quadratic variance, i.e. the sum of the integrated variance and integrated jumps.

However, since RV is only a proxy for the true latent volatility process, this measure is, of course, subject to estimation error. There are many potential sources yielding RV to be an imperfect measure. Nevertheless, the most relevant one is that we work with limited samples, which makes RV a less efficient estimator. While this issue can be mitigated by increasing the sample size, there are well-known high-frequency features that spoil this alternative. The most notable cases are the presence of microstructure noise and intraday periodicity.[1] Therefore, the parameter estimates of the econometric

[1] Whereas intraday periodicity does not impact the realized variance, as it integrates to 1, it does impact other realized measures that are essential for estimating jumps, and higher-order moments.

models based on realized measures are subject to the error-in-variables problem.

Accurate estimates and forecasts of both univariate and multivariate volatility play a central role in many financial economic applications. Examples include the comparison of total risk of two portfolios measured by their volatility and, of course, the estimation of portfolio weights. Besides that poor out-of-sample forecasts of the volatility leads to poor financial decisions, inaccurate forecasts of the covariance matrix generally lead to extreme positions that increase both transaction costs and the risk of the portfolio. Thus, an investor may end up with a riskier portfolio with smaller expected return.

Another important issue related to modelling and forecasting asset price volatility resides in understanding the underlying component of high-frequency data. Jumps are the main culprit of the extreme variations and the fat-tails observed in asset prices. The current evidence suggests that jumps are unpredictable and have different sizes and signs. Therefore, it is imperative to underscore the information content of these different types of jumps, and also evaluate whether assets with distinct levels of liquidity share similar underlying components.

This dissertation focuses on the aforementioned issues, and therefore we split the contribution into two main parts. The first part examines the impact of estimation error in the modelling and forecasting of both univariate and multivariate volatility, and their impact on portfolio choice. The second part evaluates the underlying components of high-frequency data, and explores the predictive information content of different types of jumps and the role of systematic jumps to, respectively, modelling and forecasting realized variances and covariances.

A summary of each chapter is as follows. Chapter 2 examines the impact of intraday periodicity on forecasting realized variance using a heterogeneous autoregressive (HAR) model framework. We show that intraday periodicity inflates both the unconditional and conditional variance of the realized variance, and therefore biases the autoregressive parameter estimates and jump estimators. This combined effect adversely affects forecasting. To overcome this issue, we propose a periodicity-adjusted model, HARP, where predictors are built from periodicity-filtered data. We demonstrate empirically –using 30

stocks and the SPDR S&P 500 ETF– and via Monte Carlo simulations that the HARP models produce significantly better forecasts. We also show that our results are robust to various sources of intraday periodicity estimation error and to a 'possibly" time-varying feature of the intraday periodicity.

Chapter 3 proposes a dilution bias correction approach to deal with the error-in-variables problem observed in realized volatility (RV) measures. Given that the weekly and monthly measures of the RV are less prone to measurement error, we show that the absolute difference between the daily and monthly RV is proportional to the relative magnitude of the estimation error. Therefore, in implementing the latter metric, and in allowing the daily autoregressive parameter to vary as a function of the error term, the result is more responsive forecasts with greater persistence (faster mean-reversion) when the measurement error is low (high). Empirical results indicate that our models outperform some of the most popular HAR and GARCH models across various forecasting horizons.

In chapter 4, we model and forecast realized (co)variances using a factor-structure, which suggests that (co)variances are formed by the sum of systematic and idiosyncratic components. First, we show that idiosyncratic volatility is the main driver of total realized volatility. Given the evidence that assets with a high level of idiosyncratic volatility suffer from low predictability, (co)variance forecasts of these assets are likely to have higher forecasting errors. To take this issue into account we incorporate the market factor information, and show significant improvements in the in- and out-of-sample performance of the models. We evaluate these forecasting gains using statistical loss functions and global minimum variance portfolios. We create 100 random portfolios of 5 and 10 assets, and show that the proposed models not only improve statistically upon their benchmark models, but also economically, in that a risk-averse investor is willing to sacrifice up to 157 annual basis points to obtain greater forecasting accuracy that translates in more informed financial decisions.

Chapter 5 examines the underlying components of high-frequency data using novel theoretical tests that the presence of: a) Brownian motion; b) jumps; c) finite vs. infinite

activity jumps. Given that the asymptotic distribution of most of these procedures has been derived under the assumption of noiseless prices, we first evaluate the finite sample properties under different types of microstructure noise such as Gaussian, t-distributed and Gaussian-T mixture noise. The Monte Carlo results show that 1-min data provide a good trade-off between bias and enough statistical power. Using 100 stocks and SPY, we find that both a Brownian and a jump component characterize the 1-min data, and jumps should allow for both finite and infinite activity. We also find evidence of time-varying rejection rates, such that more jump days are usually associated with an increase of infinite jumps vis-à-vis finite jumps.

Chapter 6 proposes a novel approach for disentangling realized jumps measures by activity (infinite/finite) and by sign (positive/negative). It also provides noise-robust versions of the ABD jump test (Andersen et al., 2007b) and realized semivariance measures for use at high-frequency sampling intervals. The volatility forecasting exercise involves the use of different types of jumps, forecast horizons, sampling frequencies, calendar and transaction time-based sampling schemes, as well as standard and noise-robust volatility measures. We find that infinite (finite) jumps improve the forecasts at shorter (longer) horizons; but the contribution of signed jumps is limited. Noise-robust estimators, that identify jumps in the presence of microstructure noise, deliver substantial forecast improvements at higher sampling frequencies. However, standard volatility measures at the 300-second frequency generate the smallest MSPEs. Since no single model dominates across sampling frequency and forecasting horizon, we show that model-averaged volatility forecasts –using time-varying weights and models from the model confidence set– generally outperform forecasts from both the benchmark and single best extended HAR model.

Finally, Chapter 7 proposes a robust framework for disentangling undiversifiable common jumps within the realized covariance matrix. Simultaneous jumps detected in our empirical study are strongly related to major financial and economic news, and their occurrence raises correlation and persistence among assets. Our application shows that common jumps and directional common jumps substantially improve the in- and out-

of-sample forecasts of the realized variance at the day-, week- and month-horizon. This finding is corroborated via Monte Carlo simulations. Applying these new specifications to minimum variance portfolios results in superior positions from reduced turnover. Thus, investors willingly sacrifice up to 100 annual basis points in switching to those strategies.

Part I

The Impact of Estimation Error for Modelling and Forecasting Univariate and Multivariate Volatility

Chapter 2

The Impact of Intraday Periodicity in the Autoregressive Regression Estimates and the Identification of Jumps

2.1 Introduction

Leading from the seminal work of Andersen and Bollerslev (1998a), realized volatility (RV) and related measures were developed as proxies for the daily observed volatility of all financial securities 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

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

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 models for forecasting 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 the period 2000-2016.

Our analysis attests the superiority of HARP models across all forecasting horizons, with greater gains for the 1-day to 1-week ahead forecasts. 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 1-week horizon. At the 1-day horizon, the highest improvements are for models with realized jumps in their specifications, where data filtering yields better proxies for the jump regressors. These results are confirmed when using a time-varying window to estimate periodicity. Finally, we explore how our results are impacted by the error in estimating periodicity, showing that the presence of jumps widens the distribution of forecast losses at high frequencies.

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 model which regresses realized volatilities on past daily, weekly and monthly realized volatilities. This model can replicate the high levels of persistence observed in the series of daily realized volatilities, without relying on fractional integration. Given its simple linear structure and ease in estimation, the HAR has become the

most popular option in forecasting realized volatilities.

The daily quadratic variation includes a continuous component and a jump part,[2] with the former component featuring a high level of persistence, while the jump component shows little or no persistence. To account for the different levels of persistence in the two components, Andersen et al. (2007a) propose adding the lagged realized daily squared jump as an extra explanatory variable to the HAR regression, leading to the HAR-J model. They also propose the HAR-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 Renò (2012) add negative returns to the previous HAR 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 HAR-Q,[3] 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. (2001a) and Bollerslev et al. (2000) employ similar specifications in modelling intraday volatility in the FX market and the US Treasury bond market.

While the periodicity component does not impact the realized variance, by integrating

[2]See, for instance, Barndorff-Nielsen et al. (2006a); Barndorff-Nielsen and Shephard (2004b, 2006); Barndorff-Nielsen et al. (2006b) Mancini (2004, 2009), Christensen and Podolskij (2007), Andersen et al. (2012) and Corsi et al. (2010).

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

to 1 over the trading day, little is known on its impact on other realized measures. Andersen et al. (2018) propose a statistical test for time-varying intraday periodicity in high frequency data and associated realized measures. Christensen et al. (2018) develop a test for the hypothesis that time-variation in intraday volatility is caused solely by intraday periodicity. 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 constant volatility data generating process (DGP hereafter).

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 highlights the confounding impact of jumps and periodicity on the price process and related functions. Boudt et al. (2012, 2011) recommend applying intraday jump tests on returns from which periodicity is filtered out. They propose non-parametric and parametric methods to estimate periodicity that are robust to jumps in prices and time-varying volatility.

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 2.2 provides the theoretical background on defining and forecasting the realized variance and estimating the intra-

day periodicity component of the spot volatility. Section 2.3 presents the empirical and simulated datasets used to generate results throughout the paper. Section 2.4 discusses the impact of the intraday periodicity on modelling and forecasting the realized variance. Section 2.5 presents the simulation and empirical results that compare the forecasts of the HARP and HAR models. Section 2.6 concludes the paper.

2.2 Theoretical Background

Let X_t denote a logarithmic asset price at time t belonging to a special class of semimartingales with jumps:

$$\mathrm{d}X_t = \mu_t \,\mathrm{d}t + \sigma_t \,\mathrm{d}W_t + \,\mathrm{d}L_t, \quad t \in [0, T] \tag{2.1}$$

where μ_t is a continuous and locally bounded drift term, σ_t is the spot volatility which is adapted and càdlàg. W_t is a 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. 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, $IV_t = \int_{t-1}^{t} \sigma_u^2 \,\mathrm{d}u$.

Within each trading day, there are n observations, equally spaced such that the time interval between any two consecutive observations is equal to $\Delta_n = \frac{1}{n}$.[4] Let $\Delta_i^n X$, $i = 1, ..., n$, be the i-th intraday return during the one-day interval $(t-1, t]$, such that $\Delta_i^n X = X_{i\Delta} - X_{(i-1)\Delta}$. 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:

$$RV_t = \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^2.$$

If the price contains jumps, RV_t is no longer consistent for the integrated variance, converging to the quadratic variance of the price process, $\int_{t-1}^{t} [\sigma_u^2 + L_u^2] \,\mathrm{d}u$. To estimate

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

the integrated variance, one needs to rely on a robust to jumps estimator, such as the realized bipower variation of Barndorff-Nielsen and Shephard (2004b) given by:

$$BV_t = \frac{n}{n-1}\mu_1^{-2} \sum_{i=2}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X||\Delta_{i-1}^n|, \tag{2.2}$$

where $\mu_p \equiv \mathbb{E}[|Z|^p]$, $Z \sim \mathcal{N}(0,1)$, $p > 0$. 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. (2001a); Boudt et al. (2011):

$$\sigma_t = s_t f_t, \tag{2.3a}$$

$$\text{such that } \int_{t-1}^{t} f_u^2 \, \mathrm{d}u = 1, \tag{2.3b}$$

so that intraday periodicity has no impact on the integrated variance, i.e. $\int_{t-1}^{t} \sigma_u^2 \, \mathrm{d}u = \int_{t-1}^{t} s_u^2 \, \mathrm{d}u$. In practice, as we observe a discrete number of observations, the condition in equation (2.3b) can be written using the following Riemann integral:

$$\Delta_n \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^2 = 1, \tag{2.4}$$

where f_i is the i-th value of the function $f(\cdot)$ observed during a trading day. Clearly, when Δ_n approaches 0, the Riemann sum converges to the integral in (2.3b).

The two components of spot volatility defined above in (2.3a) 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 over time reflecting the available information on the asset.

To estimate the intraday periodicity, we use the non-parametric approach proposed by Boudt et al. (2011), which is robust to the presence of jumps in the price process. Let $\Delta_i^n \overline{X} = \frac{\Delta_i^n X}{\sqrt{\Delta_n BV_t}}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$ be the standardized intraday returns, with BV_t given in equation (2.2). For a certain intraday time, i, we observe T standardized

intraday returns, which are sorted in increasing order, as follows:

$$\Delta^n_{(1),i}\overline{X} \leq \Delta^n_{(2),i}\overline{X} \leq \cdots \leq \Delta^n_{(T),i}\overline{X}.$$

Given the above ordered set, we define the subsets containing half ($\kappa \equiv \lfloor T/2 \rfloor + 1$) contiguous observations: $\left\{\Delta^n_{(1),i}\overline{X}, \ldots, \Delta^n_{(\kappa),i}\overline{X}\right\}, \ldots, \left\{\Delta^n_{(T-\kappa+1),i}\overline{X}, \ldots, \Delta^n_{(T),i}\overline{X}\right\}$. The shortest half scale estimator is the shortest length of these subsets:

$$ShortH_i = 0.741 \min\left(\Delta^n_{(\kappa),i}\overline{X} - \Delta^n_{(1),i}\overline{X}, \ldots, \Delta^n_{(T),i} - \Delta^n_{(T-\kappa+1),i}\overline{XX}\right).$$

The shortest half scale periodicity estimator is given by:

$$\hat{f}_i^{ShortH} = \frac{ShortH_i}{\Delta_n \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} ShortH_j^2}, \quad \Delta_n \equiv 1/n.$$

Finally, the weighted standard deviation intraday periodicity estimator is defined as:[5]

$$\hat{f}_i^{WSD} = \frac{WSD_i}{\sqrt{\Delta_n \sum_{j=1}^{n} WSD_j^2}} \tag{2.5}$$

$$WSD_i = \sqrt{1.081 \frac{\sum_{l=1}^{T} \chi_{l,i} \left(\Delta^n_{l,i}\overline{X}\right)^2}{\sum_{l=1}^{T} \chi_{l,i}}},$$

for all $i = 1, \ldots, n$, where WSD_i is the weighted standard deviation (WSD) and $\chi_{l,i}$, $l = 1, \ldots, T$ are weights computed using the shortest half scale periodicity estimator which are defined as:

$$\chi_{l,i} = \chi\left(\Delta^n_{l,i}X / \hat{f}_i^{ShortH}\right),$$

$$\chi(z) = \begin{cases} 1, & if \;\; z^2 \leq 6.635 \\ 0, & \text{otherwise.} \end{cases}$$

[5]The $ShortH$ is highly robust to jumps, but it has only 37% efficiency under normality of returns. For this reason Boudt et al. (2011) propose the use of a weighted standard deviation, where the threshold corresponds to the 99% quantile of the χ^2_1. The WSD has a 69% efficiency under normality.

The periodicity-filtered returns are defined as:

$$\Delta_i^n X^f = \frac{\Delta_i^n X}{\hat{f}_i^{WSD}}. \tag{2.6}$$

We can further define periodicity-filtered realized measures, such as the filtered realized variance $RV_t^f = \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X^f|^2$, the filtered realized bipower variation $BV_t^f = \frac{n}{n-1}\mu_1^{-2}$ $\sum_{i=2}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X^f||\Delta_{i-1}^n X^f|$ and so on.

In the present analysis, we rely on four models to forecast realized volatility: the HAR model proposed by Corsi (2009), the HAR-J and HAR-CJ models by Andersen et al. (2007a), and the HAR-Q model by Bollerslev et al. (2016). In addition, we introduce a new class of models, HARP, where the predictors rely on data from which periodicity is filtered out (hence the "P" in HARP stands for periodicity-filtered). Naturally, all HAR models can be transformed into HARP models by simply using filtered data, $\Delta_i^n X^f$, to compute all regressors. Note that unlike most HAR models,[6] HARP models are not autoregressions.[7]

Let h be the forecasting horizon, measured in days. Then, $RV_{t,t+h-1}$ is the forecasted realized variance over the next h days (starting from day t). Below, we present the forecasting regressions for the four HAR models and their HARP counterparts.

HAR and HARP

Let RV_{t-1} be the first lag of the (daily) realized variance, $RV_{t-5,t-1}$ the average realized variance over the past week and $RV_{t-22,t-1}$ the average realized variance over the past month. In a similar way, we define RV_{t-1}^f, $RV_{t-5,t-1}^f$ and $RV_{t-22,t-1}^f$ as the periodicity-filtered one-day, one-week and one-month lagged realized variances. Let ϵ_{t+h-1} be the

[6] The HAR-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.

forecasting error for the HAR, while ϵ^f_{t+h-1} is the forecasting error for the HARP model.

$$\text{HAR} \quad RV_{t,t+h-1} = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-5,t-1} + \beta_m RV_{t-22,t-1} + \epsilon_{t+h-1}, \tag{2.7}$$

$$\text{HARP} \quad RV_{t,t+h-1} = \beta_0 + \beta_d RV^f_{t-1} + \beta_w RV^f_{t-5,t-1} + \beta_m RV^f_{t-22,t-1} + \epsilon^f_{t+h-1}, \tag{2.8}$$

where 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-J and HARP-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 $t = 1, \ldots, T$, where C_t is a consistent estimator of the integrated variance. Similarly, we can define a jump regressor based on periodicity-filtered returns, J^f_t. The HAR-J and HARP-J models are obtained by including J_{t-1} and J^f_{t-1}, respectively, in the forecasting regression:

$$\text{HAR-J} \quad RV_{t,t+h-1} = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-5,t-1} + \beta_m RV_{t-22,t-1} + \beta_J J_{t-1} + \epsilon_{t+h-1}, \tag{2.9}$$

$$\text{HARP-J} \quad RV_{t,t+h-1} = \beta_0 + \beta_d RV^f_{t-1} + \beta_w RV^f_{t-5,t-1} + \beta_m RV^f_{t-22,t-1} + \beta_J J^f_{t-1} + \epsilon^f_{t+h-1}. \tag{2.10}$$

HAR-CJ and HARP-CJ

In this model, past lags of the estimated continuous and discontinuous components of the quadratic variation are considered in the forecasting regression, as follows:

$$\begin{aligned}\text{HAR-CJ} \quad RV_{t,t+h-1} \quad &= \beta_0 + \beta_{C_d} C_{t-1} + \beta_{C_w} C_{t-5,t-1} + \beta_{C_m} C_{t-22,t-1} + \beta_{J_d} J_{t-1} + \\ &\quad \beta_{J_w} J_{t-5,t-1} + \beta_{J_m} J_{t-22,t-1} + \epsilon_{t+h-1},\end{aligned} \tag{2.11}$$

$$\begin{aligned}\text{HARP-CJ} \quad RV_{t,t+h-1} \quad &= \beta_0 + \beta_{C_d} C^f_{t-1} + \beta_{C_w} C^f_{t-5,t-1} + \beta_{C_m} C^f_{t-22,t-1} + \beta_{J_d} J^f_{t-1} + \\ &\quad \beta_{J_w} J^f_{t-5,t-1} + \beta_{J_m} J^f_{t-22,t-1} + \epsilon^f_{t+h-1},\end{aligned} \tag{2.12}$$

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. In equation (2.12), all these regressors are computed on periodicity-filtered returns, hence the ^{f} superscript. In computing C_t and C_t^f, we employ the method in Andersen et al. (2007a): $C_t = RV_t \cdot \mathbb{I}_t(\text{no jumps}) + BV_t \cdot \mathbb{I}_t(\text{jumps})$, for $t = 1, \ldots, T$, where $\mathbb{I}_t(\cdot)$ is the indicator function for whether jumps were identified on day t or not.

HAR-Q and HARP-Q

As Bollerslev et al. (2016) indicate, the variance of the realized volatility measurement error is a function of the integrated quarticity, $\int_{t-1}^{t} \sigma_u^4 \, \mathrm{d}u$, $t = 1, \ldots, T$. Their main forecasting model accounts for the error in measuring the one-day lagged realized variance,[8] as follows:

$$\text{HAR-Q} \quad RV_{t,t+h-1} = \beta_0 + (\beta_d + \beta_{dQ} RQ_{t-1}^{1/2}) RV_{t-1} + \beta_w RV_{t-5,t-1} + \beta_m RV_{t-22,t-1} + \epsilon_{t+h-1}, \tag{2.13}$$

$$\text{HARP-Q} \quad RV_{t,t+h-1} = \beta_0 + (\beta_d + \beta_{dQ} (RQ_{t-1}^f)^{1/2}) RV_{t-1}^f + \beta_w RV_{t-5,t-1}^f + \beta_m RV_{t-22,t-1}^f + \epsilon_{t+h-1}^f, \tag{2.14}$$

where $RQ_t = \frac{n}{3} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^4$ estimates the integrated quarticity using unfiltered data, while RQ_t^f its counterpart estimate is based on periodicity-filtered data.

2.3 Data

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

2.3.1 Empirical 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 further sampled every 5 minutes. 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 generally assumed not to vary from one day to another and is estimated as described in section 2.2.[9] Figure 2.1 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.

Table 2.1 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 (right) panel of the table reports these statistics for the unfiltered (filtered) return data.

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 2.4.3 below, the presence of intraday periodicity can lead to spurious jump detection.[10]

2.3.2 Simulated Data

Here, we introduce the stochastic volatility processes from which we simulate data throughout the paper. We start from the one- and two-factor stochastic volatility models

[9] For SPY, in section 2.5.2 below, we also show results where periodicity is estimated on a past rolling window of varying size.

[10] Additional results on jump identification based on the test by Andersen et al. (2012) are reported in Table 2.8 of Section 2.C.2.

previously analyzed by Huang and Tauchen (2005) and given in equations (2.15a) and (2.15b) below.[11] For the intraday periodicity function, $f(t)$, we employ the specification in Andersen et al. (2012) and Hasbrouck (1999).

One volatility factor (SV1F) (2.15a)

$$\begin{aligned}
\mathrm{d}p(t) &= 0.03\,\mathrm{d}t + f(t)\nu(t)\left(-0.62\,\mathrm{d}W_{\nu_1}(t) + \sqrt{0.6156}\,\mathrm{d}W_p(t)\right),\\
\nu^2(t) &= \exp\{0.125\nu_1^2(t)\},\\
\mathrm{d}\nu_1^2(t) &= -0.1\nu_1^2(t)\,\mathrm{d}t + \mathrm{d}W_{\nu_1}(t).
\end{aligned}$$

Two volatility factors (SV2F) (2.15b)

$$\begin{aligned}
\mathrm{d}p(t) &= 0.03\,\mathrm{d}t + f(t)\nu(t)\left(-0.3\,\mathrm{d}W_{\nu_1}(t) - 0.3\,\mathrm{d}W_{\nu_2}(t) + \sqrt{0.82}\,\mathrm{d}W_p(t)\right),\\
\nu^2(t) &= \text{s-exp}\{-1.2 + 0.04\nu_1^2(t) + 1.5\nu_2^2(t)\},\\
\mathrm{d}\nu_1^2(t) &= -0.00137\nu_1^2(t)\,\mathrm{d}t + \mathrm{d}W_{\nu_1}(t),\\
\mathrm{d}\nu_2^2(t) &= -1.386\nu_2^2(t)\,\mathrm{d}t + \left(1 + 0.25\nu_2^2(t)\right)\mathrm{d}W_{\nu_2}(t).
\end{aligned}$$

$$f(t) = 0.88929198 + 0.75e^{-10t} + 0.25e^{-10(1-t)}, \tag{2.15c}$$

where W's are correlated standard Brownian motions, and s-exp denotes the exponential function with a polynomial spline at high values to avoid explosive behavior.

For the SV1F model, volatility is predictable and moderate and does not lead to a large number of extreme returns. In fact, in this set-up, the only way to generate a reasonable number of extreme returns is by adding jumps to the price process. Jumps arrive with constant intensity $\lambda = 0.4$ and have sizes drawn from $\mathcal{N}(0, \iota^2)$, where $\iota^2 = 1.284$, accounting for approximately 30% of the quadratic variation in the SV1F model.[12] In the SV1F model without with jumps, dependence in the second order moment of the returns is unrelated to the occurrence of extreme returns. By contrast, the SV2F model generates jointly extreme volatility and extreme returns. This is possible because this model allows for larger levels of dependence in volatility, as well as dependence in the volatility of volatility.[13]

[11] Abbreviations for the two models are as in Huang and Tauchen (2005).

[12] ι^2 is set to equal $\exp(0.125)^2$, where the value 0.125 is the coefficient multiplying the volatility factor in equation (2.15a) for the SV1F model.

[13] Figure 2.12, in section 2.C.1, plots returns simulated from both stochastic volatility models.

The SV1F model corresponds to periods of tranquility, when the occasional new information on the traded security is rapidly incorporated into the price via an added jump, followed by a return to the status quo. The SV2F model corresponds to turbulent periods when extreme returns and volatility are likely followed by more extreme returns and volatility.

Simulations are generated using an Euler scheme based on 23,400 initial data points (corresponding to seconds). We further aggregate data up to the following lower sampling frequencies: 5 seconds (4680 observations), 30 seconds (780 observations), 1 minute (390 observations), 1.5 minutes (260 observations), 2 minutes (195 observations), 2.5 minutes (156 observations), 5 minutes (78 observations), 10 minutes (39 observations), 15 minutes (26 observations) and 30 minutes (13 observations). We simulate a total of 1,000 sample paths of length 1,000 days.

Figure 2.2 plots $\nu(t)$, $f(t)$ and $f(t)\nu(t)$ simulated during the course of a trading day for the two stochastic volatility models specified in (2.15a) and (2.15b). At the start and end of the day, periodicity is higher than 1, leading to values for the spot volatility, $f(t)\nu(t)$ higher than in the absence of periodicity ($\nu(t)$). At the same time, in the middle of the day, as periodicity is lower than 1, spot volatility is lower when periodicity is present in the data. This effect is less obvious for the SV2F model, which features larger variability in $\nu(t)$, making variations due to periodicity relatively smaller.

2.4 Periodicity and 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 more complex DGPs using both simulations and real data. Finally, we consider the impact of periodicity on jump-type predictors.

2.4.1 The Simple AR(1) Model

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

$$IV_t = \Theta + \Phi IV_{t-1} + \epsilon_t, \tag{2.16}$$

where $t \in \{1, 2, ..., T\}$, $\Theta > 0$, $|\Phi| < 1$, 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, $\Delta_n IV_t$. If we also account for intraday periodicity, the spot volatility for the i-th Δ_n-length window during a trading day equals $\Delta_n IV_t f_i^2$. Assuming no drift, the i-th return is $\Delta_i^n X = \sqrt{\Delta_n IV_t} f_i w_i$, 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:

$$RV_t = \theta + \phi RV_{t-1} + u_t, \tag{2.17}$$

with ϕ equal to the well-known formula:

$$\phi = \frac{\text{cov}\left(RV_t, RV_{t-1}\right)}{\text{Var}\left(RV_t\right)}. \tag{2.18}$$

In the above equations, as RV_t is only a proxy for the integrated variance, it is subject to measurement error, leading to an attenuation bias in the estimate of ϕ (Bollerslev et al., 2016). Below, we show that periodicity further increases this bias, resulting in a further reduction - in absolute value - in the ϕ estimate.

To assess the impact of periodicity on the value of ϕ, we compute the numerator and denominator in equation (2.18) in the presence/absence of periodicity. The required derivations are enclosed in section 2.A of the appendix. While the auto-covariance remains unaffected by periodicity, we obtain the following variance formulae for the case in which periodicity is present (equation (2.19a) below), compared to the case when it is absent

(equation (2.19b)):[14]

$$\text{Var}(RV_t) = \frac{\sigma_\epsilon^2}{1-\Phi^2} + 2\Delta_n^2 \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4 \left[\frac{\sigma_\epsilon^2}{1-\Phi^2} + \left(\frac{\Theta}{1-\Phi} \right)^2 \right], \tag{2.19a}$$

$$\text{Var}(RV_t)^{NP} = \frac{\sigma_\epsilon^2}{1-\Phi^2} + 2\Delta_n^2 \frac{1}{\Delta_n} \left[\frac{\sigma_\epsilon^2}{1-\Phi^2} + \left(\frac{\Theta}{1-\Phi} \right)^2 \right], \tag{2.19b}$$

Proof. See Appendix 2.A. □

The main difference between the above formulae resides in the term $\sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4$. From equation (2.4), we know that $\sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^2 = \Delta^{-1} > 1$. Then, $\sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4 > \sum_{i=1}^{M} f_i^2 = \Delta^{-1}$, so that $\text{Var}(RV_t) > \text{Var}(RV_t)^{NP}$. As we sample at lower frequencies, $\sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4$ becomes smaller, as it sums up fewer positive terms than at higher frequencies. This effect is counterbalanced by a bigger Δ_n, making $\Delta_n^2 \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4$ in equation (2.19a) bigger and bigger in comparison to its counterpart term in equation (2.19b), $\Delta_n^2 \frac{1}{\Delta_n} = \Delta_n$. As a result, the gap between $\text{Var}(RV_t)$ and $\text{Var}(RV_t)^{NP}$ will be bigger at lower sampling frequencies than at higher frequencies, as periodicity inflates the bias due to measurement error.

The established inequality $\text{Var}(RV_t) > \text{Var}(RV_t)^{NP}$ implies that ϕ, as given in equation (2.18), 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 (2.19b), pushing ϕ downwards from Φ. Second, as shown in equation (2.19a), the presence of periodicity generates a further increase in the variance of realized volatility, further reducing ϕ.

2.4.2 Simulation and empirical evidence

The analytic results in the previous section were facilitated by the simple DGP considered. Extending such results to a more complex DGP can be achieved via Monte Carlo simulation. In this section, we rely on data simulated from the two-factor stochastic volatility model, given in equation (2.15b) above.

[14] The superscript NP above stands for "no periodicity".

We start by illustrating the impact of periodicity on the variance of the realized volatility. Let $Var(RV_t^{unf})$, $Var(RV_t^{ft})$ and $Var(RV_t^{WSD})$ be the variances of the realized volatility estimators based on, respectively, unfiltered returns, returns filtered by the true periodicity, and returns filtered with the weighted standard deviation method as shown in section 2.2. Figure 2.3 shows the histograms of the ratios $Var(RV_t^{unf})/Var(RV_t^{ft})$ and $Var(RV_t^{WSD})/Var(RV_t^{ft})$ computed on simulated returns. We consider two sampling frequencies: 1-second, the frequency at which data is generated, and 5-minute, the standard sampling frequency used in applications.

Both plots on the left show that when periodicity is present, the distribution of $Var(RV_t^{unf})/Var(RV_t^{ft})$ is almost entirely shifted to the right of 1, suggesting that the realized volatility variance increases substantially when periodicity is present. The plots on the right show that filtering out periodicity using the weighted standard deviation method is on average beneficial, as the distributions of $Var(RV_t^{WSD})/Var(RV_t^{ft})$ for both sampling frequencies are centered around 1.

We further explore the impact of periodicity on the estimated coefficients of the more complex HAR forecasting model, defined in equation (2.7). Figure 2.4 plots, for SV2F, the three coefficients of the model applied on unfiltered (HAR) and filtered (HARP) returns across decreasing sampling frequencies.[15] From left to right, the first panel of the figure corresponds to β_d from (2.7), 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.

Results in Figure 2.4 show that all coefficients are closer to their true values once periodicity is filtered out, for all sampling frequencies. At the same time, the unfiltered coefficients are closer to 0 in comparison to filtered coefficients. This finding is in line with general textbook results on errors in variables in OLS regression, where coefficients are biased towards 0.

We further rely on real return data to capture the impact of periodicity on the variance of the realized volatility. For a generic DGP, the asymptotic variance of the realized

[15]Section 2.C.1 shows results for the case when a simple AR(1) model is used in forecasting, as well as results for HAR(P)-J models.

volatility can be estimated via the realized quarticity, defined in section 2.2. Figure 2.5 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. Confidence intervals obtained on the unfiltered returns are generally wider than confidence intervals obtained for filtered data.

2.4.3 Impact on Detected Jumps

Two of the most popular HAR models, HAR-J (equation (2.9)) and HAR-CJ (equation (2.11)), 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 that proposed by Barndorff-Nielsen and Shephard (2006), which relies on a comparison between RV and the jump robust realized bipower variation, BV_t, defined in section 2.2. Throughout the paper, most results involving jump identification are based on this test.[16]

To examine the impact of intraday periodicity on spurious jump detection, we rely on data generated from the one-factor stochastic volatility model (SV1F) plus jumps, as described in section 2.3.2. The moderate levels of volatility for this model make jumps easier to identify. Figure 2.6 illustrates the proportions of spurious jumps detected with the Barndorff-Nielsen and Shephard (2006) jump test on both the original return data and periodicity-filtered data. 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 $\alpha = 5\%$.

The figure shows that the number of spurious jumps detected is higher for unfiltered returns, result that remains valid across all sampling frequencies.[17] This suggests that jump regressors in models HAR-J and HAR-CJ are likely to be affected by estimation error, which can further impact the forecast of the realized variance.

[16] The test statistics for this test and for an alternative jump test are presented in section 2.A.1 of the appendix.

[17] This result is verified when applying the jump test proposed by Andersen et al. (2012). See Figure 2.15 in Section 2.C.1.

2.5 HARP Forecasting Performance

In this section, we compare the forecasting performance of the HARP models to that of the HAR models, using both simulated and empirical data. We demonstrate that considerable forecasting gains can be attained by using periodicity-filtered data, especially at short and medium horizons.

To evaluate the forecasting performance of the two classes of models, we use two distinct loss functions, the mean squared error (MSE) and the quasi-likelihood (QLIKE) loss, defined in equation (2.20) below:

$$\begin{aligned} MSE(RV_t, F_t) &= (RV_t - F_t)^2 \\ QLIKE(RV_t, F_t) &= \frac{RV_t}{F_t} - \log \frac{RV_t}{F_t} - 1, \end{aligned} \tag{2.20}$$

where F_t denotes 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. For the analysis based on empirical data, we perform both in-sample and out-of-sample forecasts, while the results for simulated data involve only the latter.[18]

2.5.1 Simulation Results

We estimate HAR and HARP models on data simulated from models SV2F (2.15b) and SV1F (2.15a) with jumps, both defined in section 2.3.2. For each 1000-day simulated path, we re-estimate the models each day on a rolling window of 350 days.[19] For each forecasting model introduced in section 2.2 and each 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.

In the case of the SV2F model, Figure 2.7 plots the median and the 5% and 95% quantiles of the ratios of the forecast losses from the HARP model versus the HAR

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

[19] The rolling window length is equivalent to 35% of the length of the dataset.

model against the sampling frequency. All three forecasting horizons (1-day, 1-week and 1-month) are included.[20]

The distribution of the loss ratio is similar across sampling frequency, forecasting horizons and loss functions under consideration. The median is always below 1 and distributions are skewed to the left of 1 in all cases and uniformly across sampling frequency. At frequencies below 5 minutes (78 observations), we observe slightly wider distributions. For such frequencies, forecasts become more uncertain because they rely on less precisely estimated realized variances. Overall, for this DGP, where volatility is high and persistent, filtering for periodicity always improves forecasting results.

Figure 2.8 below plots, for the SV1F model with jumps, the median and the 5% and 95% quantiles of the one-day ahead forecast loss ratios against the sampling frequency.[21]

Generally, the distributions of the loss ratios are more dispersed at higher sampling frequencies. This mainly reflects the impact of periodicity estimation error, which is relatively larger at higher frequencies due to the presence of jumps. For models including jumps in their specification, this effect is less visible. Medians are always below 1, indicating that the distributions are shifted to the left of 1. Section 2.5.3 below discusses in more detail the impact of jump-related estimation error on our analysis.

2.5.2 Empirical Results

In-Sample Forecasting Results

Tables 2.2, 2.3, 2.4 and 2.5 report the regression results for all HAR and HARP models, estimated on the entire sample, for SPY and a stock average.[22] 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

[20] Figure 2.16, in Section 2.C.1, reports similar results for the HARP-Q and HAR-Q models.

[21] Section 2.C.1 reports results for the 1-week and 1-month forecasting horizons.

[22] Table 2.9 in Section 2.C.2 also reports results for the Rivers and Vuong (2002) test, comparing the in-sample performance of HARP and HAR models for SPY. Results indicate HARP models generally outperform their HAR counterparts.

uses over 3,000 observations.[23]

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-CJ model), which represents the level of persistence when the models are autoregressions (all HAR except HAR-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 levels 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.

For SPY, the in-sample R-squared has similar values for the HARP and HAR 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 model outperforms the HAR 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 unchanged. The in-sample R-squared averaged across stocks is higher for the HARP models for all forecasting horizons, while in terms of average out-of-sample R-squared, we observe higher values for the 1-day and 5-days ahead HARP models.

In the case of the HAR-J and HARP-J models, $\widehat{\beta_{J_d}}$ is always negative, in line with the

[23]Section 2.C.2 reports, as a robustness check, the results obtained for the HAR(P)-J and HAR(P)-CJ models where the Andersen et al. (2012) test is used to identify jumps.

existing literature (see Andersen et al., 2007a). For the HARP-J, $\widehat{\beta_{J_d}}$ is larger in absolute value and has smaller standard errors compared to the HAR-J model. As filtering out periodicity reduces the number of detected spurious jumps (see section 2.4.3), the jump predictor for the HARP-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-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. In particular, for SPY, the out-of-sample R-squared features the highest increase post 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 0.417 for HAR-J to 0.480 for HARP-J. Average results for stocks show higher average in-sample R-squared coefficients for HARP-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.

In line with our findings for the HARP-J model, for the HARP-CJ model, we notice an important reduction in the standard errors for the coefficients of all realized jumps regressors 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 feature a substantial decrease after filtering.

The HARP-CJ models generally show higher in-sample and out-of-sample R-squared for SPY and on average across all stocks. A particularly high change following filtering is observed for the 1-day ahead model for SPY, where the out-of-sample R-squared increases from 0.407 for HAR-CJ to 0.468 for HARP-CJ.

For the HARP-Q model, standard errors for $\widehat{\beta_Q}$, the estimated coefficient for $RQ_t^{1/2}RV_t$, as illustrated in equation (2.13), are lower than for the HAR-Q model. As seen in section 2.4, periodicity impacts the RV variance and thus, its 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 HARP-Q model. Moreover, filtered models outperform unfiltered models in 4 out of 6 cases, for SPY, and 5 out of 6 cases on average across stocks.

Out-of-Sample Forecasting Results

We re-estimate all HAR and HARP models on rolling windows of 1000 days and compute out-of-sample forecast losses.[24] The ratios of the losses from HARP versus HAR models for the 1-day, 1-week and 1-month horizons are reported in Table 2.6. For each forecasting horizon, the top panel shows results for the SPY and the average across all stocks. We apply the Diebold and Mariano (1995) test to assess the significance of the forecasting gains attained for the HARP models relative to the HAR models. Let ϵ_{t+h-1} be the errors from one of the HAR models in equations (2.7), (2.9),(2.11) and (2.13) and ϵ^f_{t+h-1} the errors from these models' HARP counterparts. Further, let $L(\cdot)$ denote one of the loss functions in (2.20) and $d_t = L(\epsilon^f_{t+h-1}) - L(\epsilon_{t+h-1})$. Then, the Diebold and Mariano (1995) test statistic is defined as:

$$DM = \frac{\frac{1}{T}\sum_{t=1}^{T} d_t}{\sqrt{\widehat{\text{Var}}\left(\frac{1}{T}\sum_{t=1}^{T} d_t\right)}} \to \mathcal{N}(0,1), \tag{2.21}$$

where $\widehat{\text{Var}}\left(\frac{1}{T}\sum_{t=1}^{T} d_t\right)$ is a consistent estimator for the variance of the d_t sample mean. We run a two-tailed test, where rejection when $DM < 0$ means that the average loss from HARP models is lower than the average loss from HAR models. In Table 2.6, we use starred numbers to indicate this at a 5% significance level, while numbers with a diamond superscript indicate that HAR models significantly outperform HARP models.

For each forecasting horizon, the last three rows show the results of the Diebold and Mariano (1995) test for the 30 stocks in our sample. For each entry, the first value indicates the number of stocks for which HARP models outperform HAR models at 5% significance level, while the second value shows the number of stocks for which the opposite is true. For the models including jumps, we present results relying on both tests

[24] The rolling window length is equivalent to approximately 30% of the length of the dataset.

for jumps considered in this paper, the classic Barndorff-Nielsen and Shephard (2006) 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).

In the case of the 1-day ahead forecasts ($h = 1$), all except one loss ratios take values below 1 for SPY, 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 section 2.5.2, where we observed lower standard errors at 1-day ahead and higher R-squared coefficients for the HARP-J and HARP-CJ models in comparison to their HAR counterparts. As shown in section 2.4, 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.

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 using the QLIKE loss criterion, we find a significant gain 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. Between 8 and 11 stocks feature a significantly lower HARP MSE, and between 10 and 20 a significantly lower HARP QLIKE loss. By comparison, the number of stocks for which HAR losses are significantly lower than HARP losses ranges between 0 and 2, equalling 0 most of the times.

For the 1-week ahead forecasts ($h = 5$), 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. For the QLIKE loss criterion, we observe significantly lower losses for the HARP-Q model and both HARP-CJ models (BV and MedRV) relative to the benchmark HAR models.

Across all models, the average loss ratios for all stocks under consideration are below 1, indicating, just as for the 1-day ahead forecasts, that filtering periodicity is beneficial

for the majority of stocks. Between 7 and 10 stocks feature a significantly lower HARP MSE, and between 8 and 13 a significantly lower HARP QLIKE loss. By comparison, HAR models outperform HARP models for a number of stocks between 0 and 3.

For the 1-month ahead forecasts ($h = 22$), all but one ratios are below 1 for SPY. The exception occurs for the MSE loss ratio HARP-Q/HAR-Q, but even in this case, the ratio remains close to 1. The Diebold and Mariano (1995) test indicates that the MSE is significantly lower for HARP and HARP-CJ based on the realized bipower variation when these models are compared to their HAR counterparts. The QLIKE loss is significantly lower for HARP, HARP-CJ based on both jump tests, and HARP-J based on the Andersen et al. (2012) test.

The stock average MSE ratio is lower than 1 in all cases except for HARP-Q/HAR-Q, while the stock average QLIKE loss ratio is lower than 1 in only one case. Furthermore, MSE is significantly lower in the case of HARP models for a number of stocks ranging between 3 and 6, while the HARP QLIKE loss is significantly lower for a number of stocks ranging between 3 and 8. While these numbers are lower than the numbers reported for the 1-day and 1-week ahead forecasts, we generally observe more stocks with significantly lower HARP losses than with significantly lower HAR losses.

To account for the fact that the periodicity function might be time-varying, we perform the out-of-sample analysis for SPY using periodicity estimates obtained on time windows of varying length. For each day, we use W past days to compute the periodicity function, where W varies from 22 trading days, i.e. 1 month, to 1008 days, equivalent to a 4-year long window. Results are shown in Table 2.7.

Irrespective of the forecasting model and the window length for estimating periodicity, the vast majority of loss ratios are below 1, indicating that HARP models dominate in performance. This is further confirmed by the results for the Diebold and Mariano (1995) test, based on which HARP models significantly (at 5%) outperform HAR models in most of the cases, while the opposite is true for just a few cases when using the QLIKE criterion for the HAR-Q model when $h = 1$. All results in Table 2.7 are dispersed around the values obtained when estimating periodicity on the full sample (see Table 2.6), with

some estimation window sizes outperforming, while others under-performing the general case.

In terms of changes in results with the varying of the length of the periodicity estimation window, we find no clear pattern pointing towards an optimal window length. In general, the 1-month window seems too short, as we tend to attain better performance for longer windows. The results remain qualitatively similar to the general case using the full sample size for filtering.

In summary, evidence from both Monte Carlo simulations and empirical data supports that filtering periodicity leads to superior forecasting performance at the 1- and 5- days forecasting horizons. For the longer horizons, we argue that distortions due to intraday periodicity are mostly negligible and that HAR and HARP models forecasting performance is similar.

2.5.3 Sensitivity and validity analysis

This section acts as a robustness check for our results. It explores how various sources of estimation error in the intraday periodicity estimates can impact our findings. We ultimately show that our results hold even in the presence of such errors. The estimation error in disentangling periodicity has two main sources: the number of days used to estimate periodicity is too short and the jumps in the price interfere with periodicity estimation, especially at higher frequencies.

A time-varying periodicity function (see Andersen et al., 2018) calls for estimating periodicity over shorter windows of time. As shorter estimation windows can lead to less reliable periodicity estimates, Figure 2.9 explores the sensitivity of our results on forecasting RV to the length of the periodicity estimation window. We plot the distribution of the HARP/HAR loss ratios obtained at the highest sampling frequency for the SV2F model against the length of the periodicity estimation window. At very high sampling frequencies and in the absence of jumps, the impact of measurement error emanating from any other source than the length of the estimation window is insignificant.

The distribution of the loss ratios does not change much with the length of the esti-

mation window for periodicity. The median is always below 1, confirming that filtering improves the forecasting performance. The distribution of the QLIKE ratios is slightly more dispersed than the distribution of the MSE ratios for shorter estimation windows.

We further consider the impact of the jump-related periodicity estimation error on our analysis. To this end, we compare the HARP forecast loss for the filtered SV1F process with jumps to the forecast loss for the filtered SV1F model to which we add jumps only after applying the periodicity filter. Specifically, for the latter forecast loss, we apply the periodicity filter at different sampling frequencies before adding the jumps also sampled correspondingly. The first forecast loss is impacted by jump-related periodicity estimation error, while the second loss is not. The distributions of the ratios of the two losses for different forecasting horizons are plotted against the sampling frequencies in Figure 2.10. The "$HARP^{fj}$" notation indicates the HARP model where filtering ("f") occurs on data with jumps ("j"), while the "$HARP^{fnj}$" denotes the forecasting model where filtering ("f") occurs on data with no jumps ("nj").

At high frequencies, the distribution of the loss ratio shifts above 1 and is more dispersed than at lower frequencies. This shift is mostly visible for the 1-day and 1-week ahead forecasts, where all three quantiles are located above 1 for sampling frequencies higher than 30 seconds. For the one-month ahead forecasts, the median and the 95% quantile at high frequency are above 1, indicating an upwards shift, but the distribution is a lot more dispersed, with the 5% quantile well below 1. In this case, aggregation of data over long horizons makes the impact of jump-related estimation error less clear in terms of direction, but still very much visible in terms of dispersion. For all forecasting horizons, the impact of jump-related estimation error gradually decreases with the sampling frequency.

Finally, we examine whether excessive or insufficient filtering impacts our results. We rely on the SV2F 5-minute data and employ the forecast loss definition used in presenting the Diebold and Mariano (1995) test in section 2.5.2. Let $L(\epsilon_{t+1}^{WSD})$ and $L(\epsilon_{t+1}^{unf})$ be, respectively, the HARP and HAR model forecast losses computed using the MSE loss function. Let RV_t^{ft} and RV_t^{WSD} be the realized variance estimators based on, respectively,

returns filtered by the true periodicity, and returns filtered with the weighted standard deviation method outlined in section 2.2. We define excessive filtering the situation for which $RV_t^{WSD} < RV_t^{ft}$ and insufficient filtering when $RV_t^{WSD} > RV_t^{ft}$. In Figure 2.11, we plot the loss differential $L(\epsilon_{t+1}^{WSD}) - L(\epsilon_{t+1}^{unf})$ against $RV_t^{WSD} - RV_t^{ft}$. The surface of the plot is split in four quadrants based on the criteria: $RV_t^{WSD} - RV_t^{ft} \lessgtr 0$ and $L(\epsilon_{t+1}^{WSD}) - L(\epsilon_{t+1}^{unf}) \lessgtr 0$. In each quadrant, we also report the average loss difference per quadrant, $\overline{\Delta L}_{t+1}$, as well as the percentage of points.

Overall, filtering leads to forecast gains, as more than 75% of the points in the scatter plot are situated below the line $L(\epsilon_{t+1}^{WSD}) = L(\epsilon_{t+1}^{unf})$, where loss differentials are also higher in absolute value. Over 67% of the points are to the right of the $RV_t^{WSD} = RV_t^{ft}$ line, out of which 51% are in the fourth quadrant, where $L(\epsilon_{t+1}^{WSD}) - L(\epsilon_{t+1}^{unf}) < 0$ and $RV_t^{WSD} - RV_t^{ft} > 0$. This quadrant also features the most extreme points of the scatter plot and the highest loss difference in absolute value. The second quadrant, defined by $L(\epsilon_{t+1}^{WSD}) - L(\epsilon_{t+1}^{unf}) > 0$ and $RV_t^{WSD} - RV_t^{ft} < 0$, also contains some extreme points, showing that excessive filtering can have adverse effects. This is not worrisome though, as this quadrant has the lowest percentage of points and a relatively low loss differential in absolute value.

2.6 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 the 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 autocorrelation. 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 periodicity-filtered returns.

In addition, we also document that periodicity leads to spurious jumps detection.

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 empirical data. Our analysis encompasses the HARP versions of the most common HAR models in the literature, the HAR model by Corsi (2009), the HAR-J and HAR-CJ models by Andersen et al. (2007a), and the HAR-Q model by Bollerslev et al. (2016). Our dataset includes intraday observations for the SPDR ETF and 30 S&P500 constituents for the period 2000 to 2016. 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 1-week ahead. At the 1-day ahead horizon, the HARP-J and HAR-CJ models show the greatest improvements following filtering, owing to lower distortions in the jump predictors. At the 1-month horizon, results show little to no gains from filtering periodicity, as the increase in the forecasting error at this horizon is likely to dilute the impact of periodicity. Finally, for robustness, we examine the impact of time varying periodicity using SPY data and the impact of measurement error using simulated data. Our results are robust to the various lengths of the periodicity estimation window and the different sources of error in estimating periodicity.

Appendix 2.A Some Proofs for the Simple AR(1) Model

Under the assumptions of section 2.4.1, $RV_t = \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^2 = \Delta_n IV_t \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^2 w_i^2$.

$$\mathbb{E}(RV_t) = \Delta_n \, \mathbb{E}(IV_t) \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^2 = \frac{\Theta}{1-\Phi},$$

where we used the fact that $\mathbb{E}(w_i^2) = 1$ and $\mathbb{E}(IV_t) = \frac{\Theta}{1-\Phi}$ given the DGP for IV_t in equation (2.16).

Proof of equation (2.19a).

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

$$\begin{aligned}
\mathrm{Var}(RV_t) = \mathbb{E}\left(RV_t^2\right) - \left[\mathbb{E}\left(RV_t\right) \right]^2 &= \Delta_n^2 \left[\frac{\sigma_\epsilon^2}{1-\Phi^2} + \left(\frac{\Theta}{1-\Phi} \right)^2 \right] \left(2 \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4 + \frac{1}{\Delta_n^2} \right) - \left(\frac{\Theta}{1-\Phi} \right)^2 \\
&= \Delta_n^2 \frac{\sigma_\epsilon^2}{1-\Phi^2} \left(2 \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4 + \frac{1}{\Delta_n^2} \right) + 2\Delta_n^2 \left(\frac{\Theta}{1-\Phi} \right)^2 \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4 \\
&= \frac{\sigma_\epsilon^2}{1-\Phi^2} + 2\Delta_n^2 \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} f_i^4 \left[\frac{\sigma_\epsilon^2}{1-\Phi^2} + \left(\frac{\Theta}{1-\Phi} \right)^2 \right]. \qquad \square
\end{aligned}$$

For comparison purposes, we compute the same variance in the absence of periodicity, where the superscript NP below stands for "no periodicity".:

Proof of equation (2.19b).

$$\begin{aligned}
\mathrm{Var}(RV_t)^{NP} &= \Delta_n^2\,\mathbb{E}(IV_t^2)\,\mathbb{E}\left(\sum_{i=1}^{\lfloor 1/\Delta_n\rfloor} w_i^4 + \sum_{i=1}^{\lfloor 1/\Delta_n\rfloor}\sum_{\substack{j=1\\ j\neq i}}^{\lfloor 1/\Delta_n\rfloor} w_i^2 w_j^2\right) - \left(\frac{\Theta}{1-\Phi}\right)^2 \\
&= \Delta_n^2\left\{\mathrm{Var}(IV_t) + \left[\mathbb{E}(RV_t)\right]^2\right\}\left[3n + n(n-1)\right] - \left(\frac{\Theta}{1-\Phi}\right)^2 \\
&= \Delta_n^2\frac{\sigma_\epsilon^2}{1-\Phi^2}\left(\frac{2}{\Delta_n} + \frac{1}{\Delta_n^2}\right) + 2\Delta_n\left(\frac{\Theta}{1-\Phi}\right)^2 \\
&= \frac{\sigma_\epsilon^2}{1-\Phi^2} + 2\Delta_n\left[\frac{\sigma_\epsilon^2}{1-\Phi^2} + \left(\frac{\Theta}{1-\Phi}\right)^2\right].
\end{aligned}$$

□

Let w_i, $i = 1, \ldots, n$ be a sequence of i.i.d. standard normal variables entering the intraday returns on day t and w_i^*, $i = 1, \ldots, n$ 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.

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

2.A.1 Jump Tests

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

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

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

$$TPQ_t = n1.74\frac{n}{n-2}\sum_{i=3}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^{4/3}|\Delta_{i-1}^n X|^{4/3}|\Delta_{i-2}^n X|^{4/3} \xrightarrow{p} \int_{t-1}^{t} \sigma_u^4 \, \mathrm{d}u.$$

The above test is widely used in empirical work due to its simplicity and reasonable size and power properties under various scenarios (see Dumitru and Urga, 2012). As documented in the introduction to this paper, 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 (2006). The test statistic is given below:

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

with

$$MedRV_t = \frac{n}{n-2}1.42\sum_{i=2}^{\lfloor 1/\Delta_n \rfloor - 1} \mathrm{med}(|\Delta_{i-1}^n X|, |\Delta_i^n X|, |\Delta_{i+1}^n X|)^2 \xrightarrow{p} \int_{t-1}^{t} \sigma_u^2 \, \mathrm{d}u$$

and

$$MedRQ_t = \frac{n^2}{n-2}0.92\sum_{i=2}^{\lfloor 1/\Delta_n \rfloor - 1} \mathrm{med}(|\Delta_{i-1}^n X|, |\Delta_i^n X|, |\Delta_{i+1}^X|)^4 \xrightarrow{p} \int_{t-1}^{t} \sigma_u^4 \, \mathrm{d}u.$$

Appendix 2.B Tables and Figures

Table 2.1: 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 C. 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 the 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 (2006) procedure.

Table 2.2: Estimated 1-, 5-, and 22- day ahead HAR(P) models for SPY and stocks 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	0.568
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 and HARP 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 and HARP models of various horizons. *, ** and *** denote significance at 10%, 5% and 1% respectively.

Table 2.3: Estimated 1-, 5-, and 22- day ahead HAR(P)-J models for SPY and stocks average.

	HAR-J			HARP-J		
	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$\widehat{\beta_0}$	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-J and HARP-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-J and HARP-J models of various horizons. *, ** and *** denote significance at 10%, 5% and 1% respectively.

Table 2.4: Estimated 1-, 5-, and 22-days ahead HAR(P)-CJ models for SPY and stocks average.

	HAR-CJ			HARP-CJ		
	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$\widehat{\beta_0}$	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-CJ and HARP-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-CJ and HARP-CJ models of various horizons. *, ** and *** denote significance at 10%, 5% and 1% respectively.

Table 2.5: Estimated 1-, 5-, and 22-days ahead HAR(P)-Q models for SPY and stocks average.

	HAR-Q			HARP-Q		
	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$\widehat{\beta_0}$	-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 HAR-Q and HARP-Q 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-Q and HARP-Q models of various horizons. *, ** and *** denote significance at 10%, 5% and 1% respectively.

Table 2.6: Out-of-sample forecast losses

(a) $h = 1$							
				BV		MedRV	
		HARP/ HAR	HARP-Q/ HAR-Q	HARP-J/ HAR-J	HARP-CJ/ HAR-CJ	HARP-J/ HAR-J	HARP-CJ/ HAR-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	9 : 2	9 : 0	8 : 2	11 : 0	8 : 0	8 : 0
	QLIKE	20 : 0	18 : 0	18 : 0	10 : 0	17 : 1	15 : 1
(b) $h = 5$							
				BV		MedRV	
		HARP/ HAR	HARP-Q/ HAR-Q	HARP-J/ HAR-J	HARP-CJ/ HAR-CJ	HARP-J/ HAR-J	HARP-CJ/ HAR-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	10 : 0	10 : 0	10 : 1	7 : 1	10 : 1	11 : 1
	QLIKE	13 : 0	12 : 0	12 : 1	8 : 3	12 : 1	13 : 3
(c) $h = 22$							
				BV		MedRV	
		HARP/ HAR	HARP-Q/ HAR-Q	HARP-J/ HAR-J	HARP-CJ/ HAR-CJ	HARP-J/ HAR-J	HARP-CJ/ HAR-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	6 : 0	3 : 0	4 : 0	6 : 4	5 : 1	6 : 2
	QLIKE	8 : 6	3 : 3	7 : 5	6 : 5	7 : 3	8 : 4

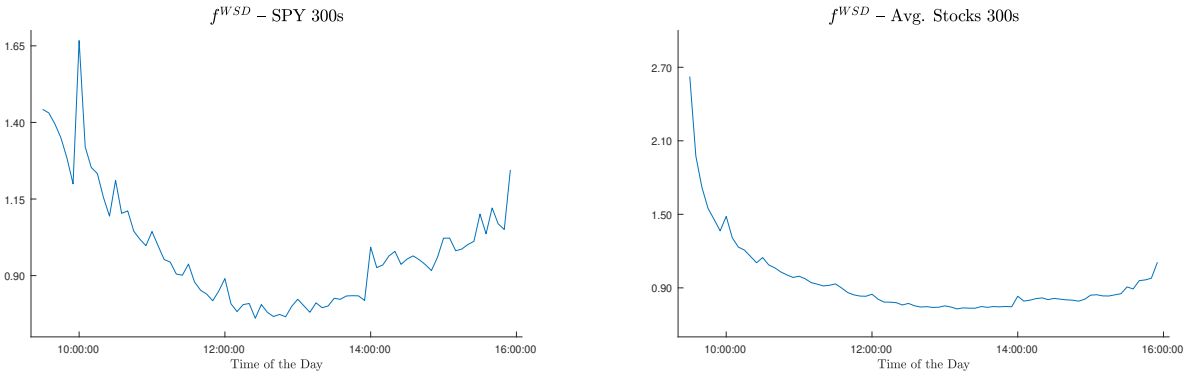
Note: This table reports the ratio of the losses from HARP versus HAR models for various forecasting horizons. * (◆) indicates that the losses of the HARP (HAR) models are significantly lower compared to the HAR (HARP) model at the 5% significance level based on the Diebold and Mariano test. The entries of type "xx : yy" summarize the results of the Diebold and Mariano test for the 30 stocks considered. The first number, "xx" shows the number of stocks for which the HARP model significantly outperforms the HAR model, while the second number, "yy" indicates the number of stocks for which the opposite is true.

Table 2.7: Results for SPY assuming Time-Varying Periodicity

		$W=22$	$W=44$	$W=126$	$W=252$	$W=504$	$W=1008$	$W=22$	$W=44$	$W=126$	$W=252$	$W=504$	$W=1008$
		HARP/HAR						HARP-Q/HAR-Q					
MSE	$h=1$	0.958*	0.915*	0.914*	0.916*	0.910*	0.917*	0.970*	1.066	0.973*	0.959*	0.964*	0.965*
QLIKE	$h=1$	0.991	0.972*	0.984*	0.991	0.992	0.997	1.519♦	1.521♦	1.341♦	1.247♦	1.310♦	1.358♦
MSE	$h=5$	0.979	0.983	0.995	0.942*	0.933*	0.935*	0.952*	0.945*	0.974	0.977	0.975	0.978
QLIKE	$h=5$	0.998	0.986	0.993	0.971	0.972	0.976	0.953*	0.962*	0.970	0.971	0.970	0.964*
MSE	$h=22$	0.951*	0.905*	0.898*	0.896*	0.939*	0.937*	0.964*	0.940*	0.940*	0.985	0.936*	0.968
QLIKE	$h=22$	0.973	0.949*	0.951*	0.950*	0.966	0.969	0.985	0.971	0.989	1.006	1.013	1.008
		HARP-J/HAR-J						HARP-CJ/HAR-CJ					
MSE	$h=1$	0.947*	0.906*	0.909*	0.910*	0.905*	0.911*	0.932*	0.925*	0.913*	0.917*	0.913*	0.921*
QLIKE	$h=1$	0.980*	0.958*	0.984*	0.988	0.990	0.994	0.977*	0.970*	0.977*	0.982*	0.981*	0.988
MSE	$h=5$	0.891*	0.846*	0.912*	0.866*	0.857*	0.859*	0.892*	0.853*	0.922*	0.934*	0.886*	0.890*
QLIKE	$h=5$	0.952*	0.942*	0.948*	0.928*	0.930*	0.934*	0.918*	0.899*	0.937*	0.927*	0.916*	0.930*
MSE	$h=22$	0.949*	0.969	0.905*	0.929*	0.958*	0.970	0.919*	0.948*	0.888*	0.939*	0.971	0.935*
QLIKE	$h=22$	0.946*	0.951*	0.934*	0.940*	0.946*	0.957*	0.964*	0.965	1.000	1.016	1.004	0.996

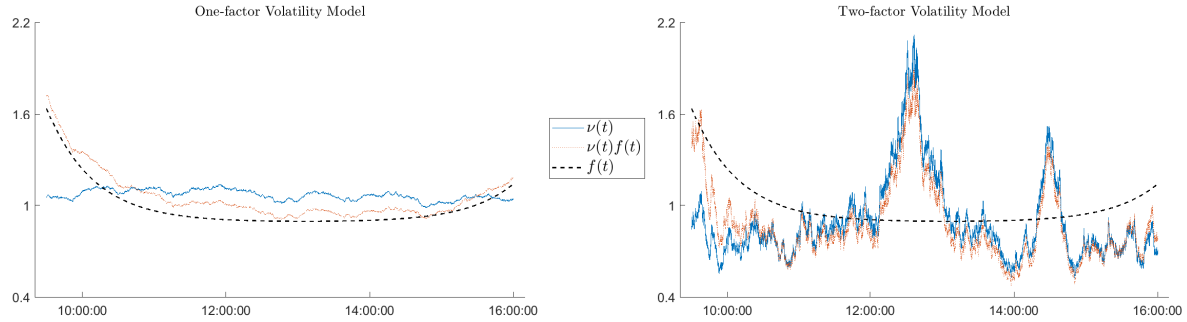
Notes: This table reports the loss ratio across different size windows used to estimate the intraday periodicity. * (♦) indicates that the losses of the HARP (HAR) models are significantly lower compared to the HAR (HARP) model at the 5% significance level. The first row shows the length of the window used to estimate the intraday periodicity, given in number of trading days.

Figure 2.1: Intraday estimated periodicity for SPY (left) and average periodicity for all stocks (right).



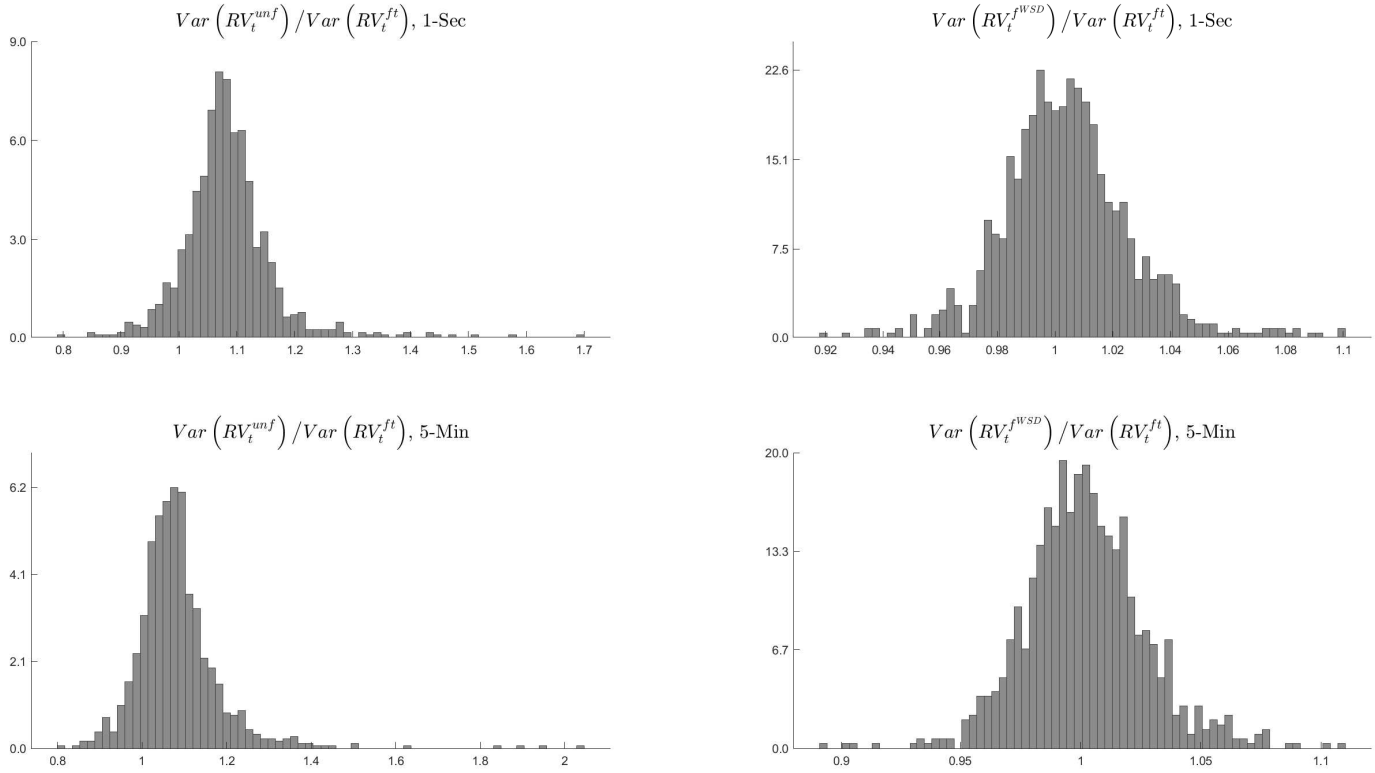
Note: The panel on the left shows the estimated periodicity for SPY, while the panel on the right shows the average estimated periodicity for the 30 S&P 500 stocks considered. Periodicity was estimated using all available data and a 5-minute sampling frequency.

Figure 2.2: The impact of periodicty on intraday volatility



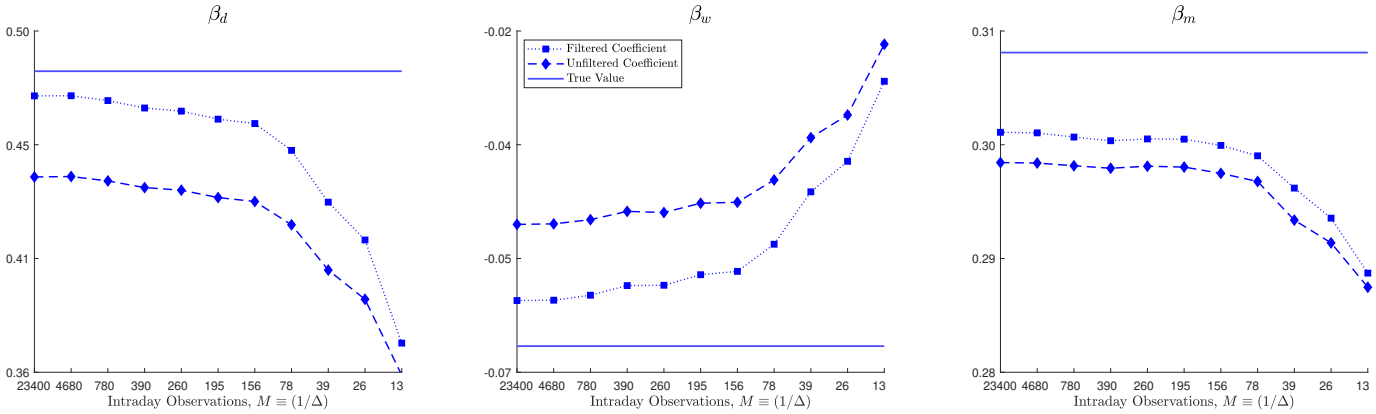
Note: The figure plots one trading day simulations of the periodicity function ($f(t)$), the actual spot volatility ($\nu(t)$) and the final spot volatility including periodicity ($f(t)\nu(t)$) for the SV1F and SV2F models. Data frequency is 1 second.

Figure 2.3: Impact of Periodicity on the Unconditional Variance of RV



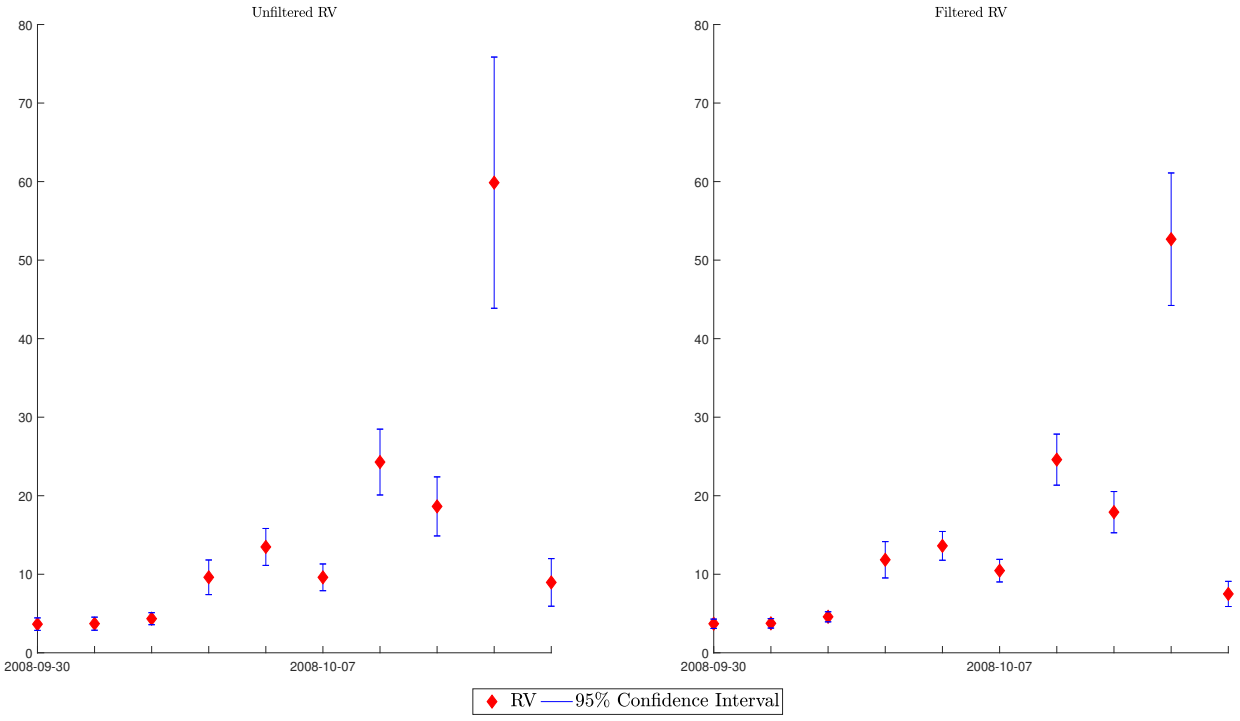
Note: This figure plots the distributions of the ratios $Var(RV_t^{unf})/Var(RV_t^{ft})$ and $Var(RV_t^{WSD})/Var(RV_t^{ft})$ for simulated returns sampled every second and every 5 minutes. $Var(RV_t^{unf})$, $Var(RV_t^{ft})$ and $Var(RV_t^{WSD})$ are the variances of the realized volatility estimators based on, respectively, unfiltered returns, returns filtered by the true periodicity, and returns filtered with the weighted standard deviation method as shown in section 2.2. Data is generated from the SV2F model.

Figure 2.4: HARP and HAR coefficients



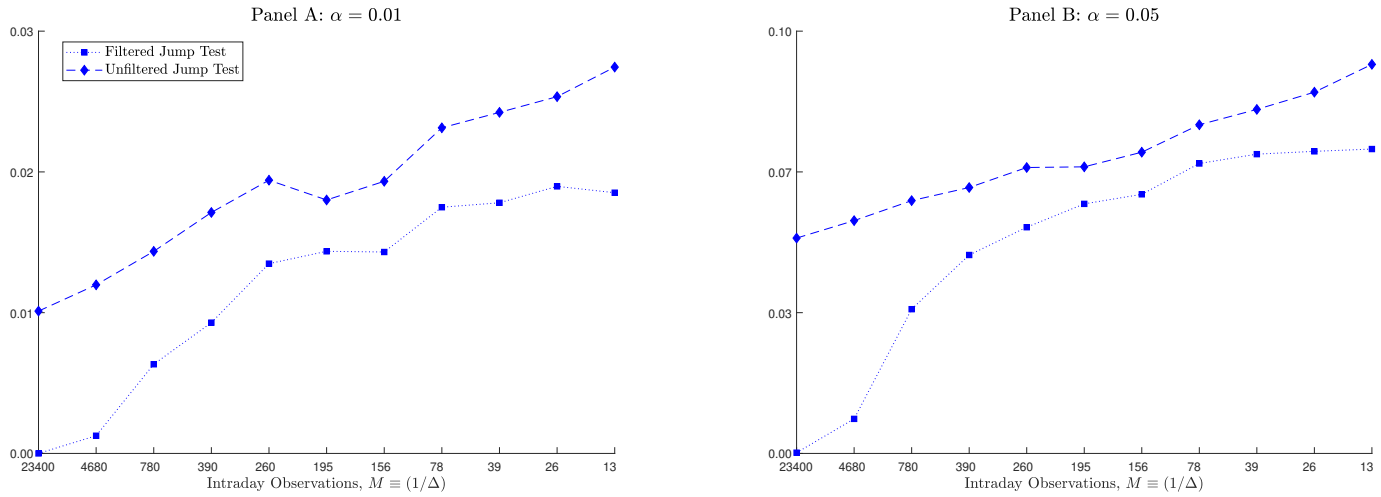
Note: This figure compares the estimates of the HARP (squared marker) and HAR (diamond marker) models against the true estimates across different sampling frequencies. Data is generated from the SV2F model. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.5: Unfiltered versus filtered confidence bands for the SPY realized variance



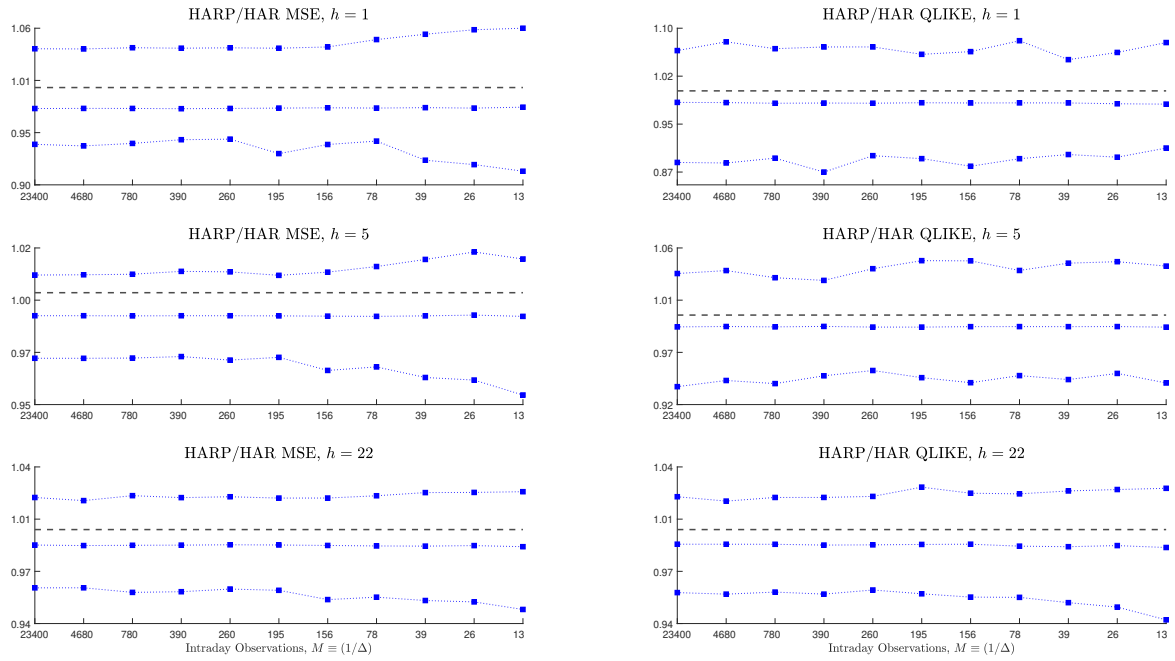
The figure plots the realized variance and its 95% confidence intervals for SPY based on unfiltered data (left plot) versus filtered data (right plot) for 10 consecutive days, starting with 2008-09-30.

Figure 2.6: Proportion of spurious jumps by sampling frequency for filtered and unfiltered data



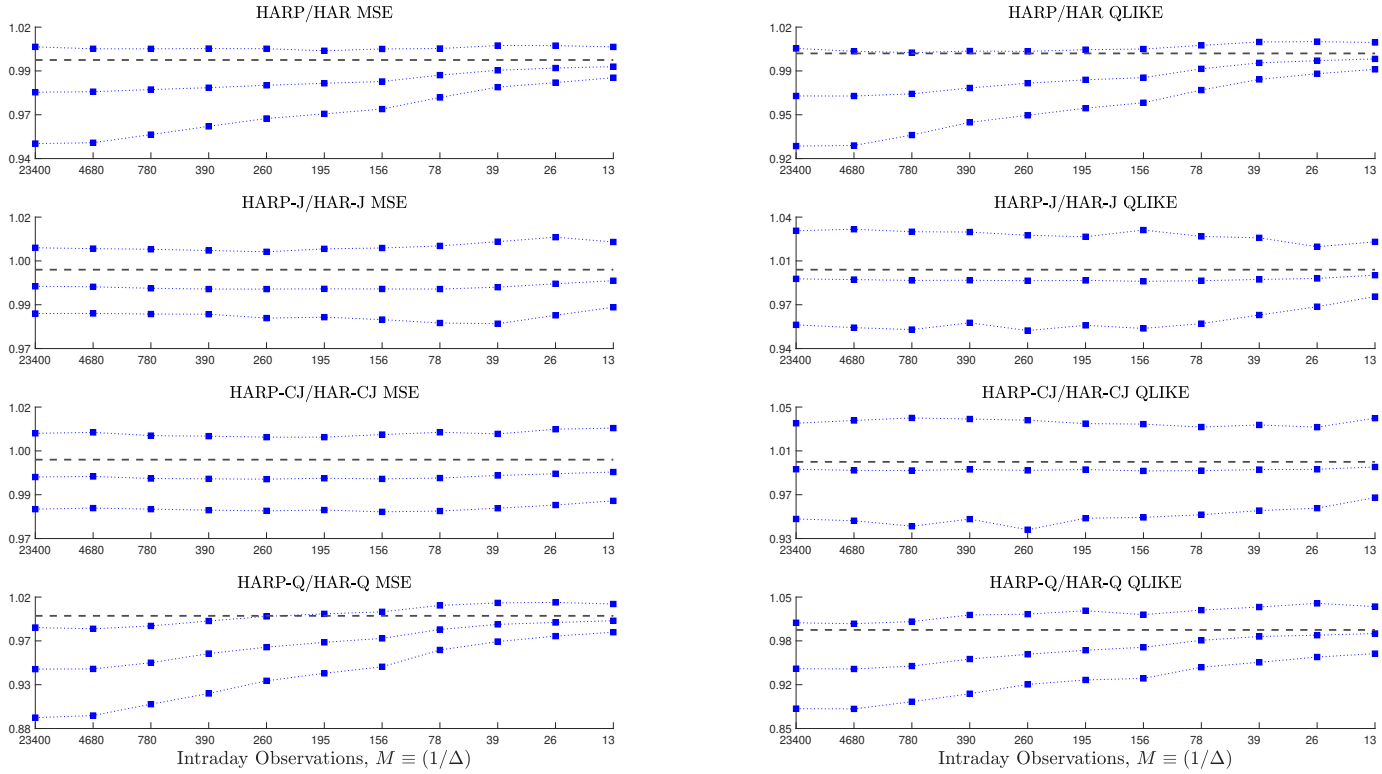
Note: This plot graphs the proportion of spurious jumps across sampling frequencies. Jumps were detected using the Barndorff-Nielsen and Shephard (2006) jump test evaluated at the 1% and 5% significance level. Data is generated from the SV1F model plus jumps. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.7: Loss ratio for the simulated SV2F model



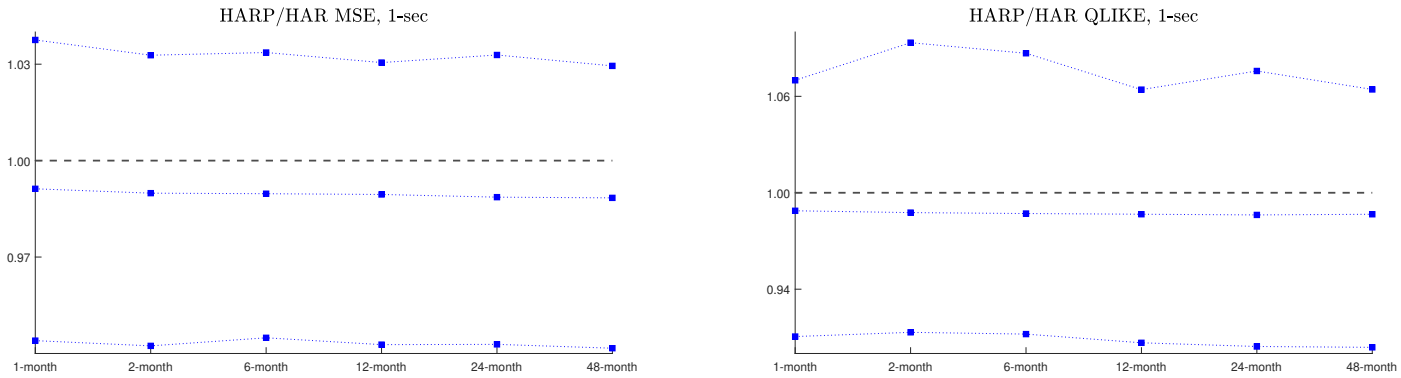
Note: The figure plots the median and the 5% and 95% quantiles for the MSE and QLIKE loss ratios, for the HARP versus the HAR model. All forecasting horizons are included: one-day ($h = 1$), one-week ($h = 5$) and one-month ($h = 22$). The dashed horizontal line corresponds to the value 1. Data is generated from the SV2F model. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.8: One-day ahead loss ratio for the simulated SV1F model with jumps



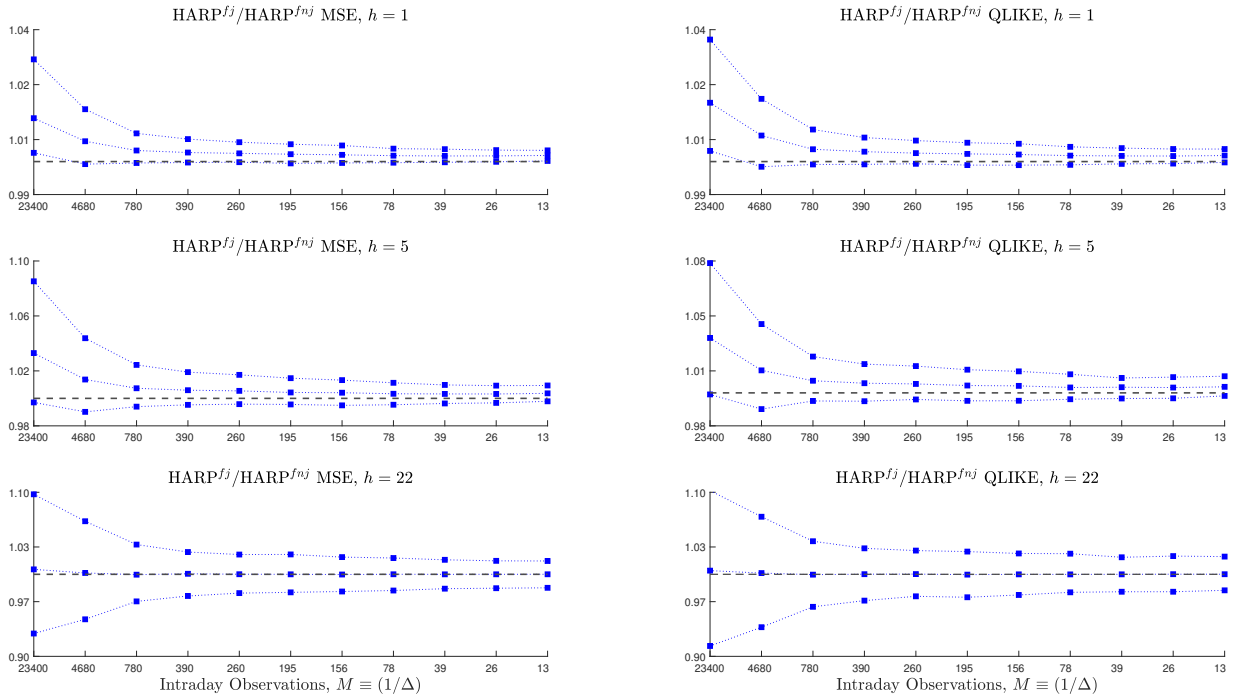
Note: The figure plots the median and the 5% and 95% quantiles for the MSE and QLIKE loss ratios, for HARP versus HAR models. The dashed horizontal line corresponds to the value 1. Data is generated from the SV1F model plus jumps. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.9: The impact of the length of the periodicity estimation window on the performance of HARP models



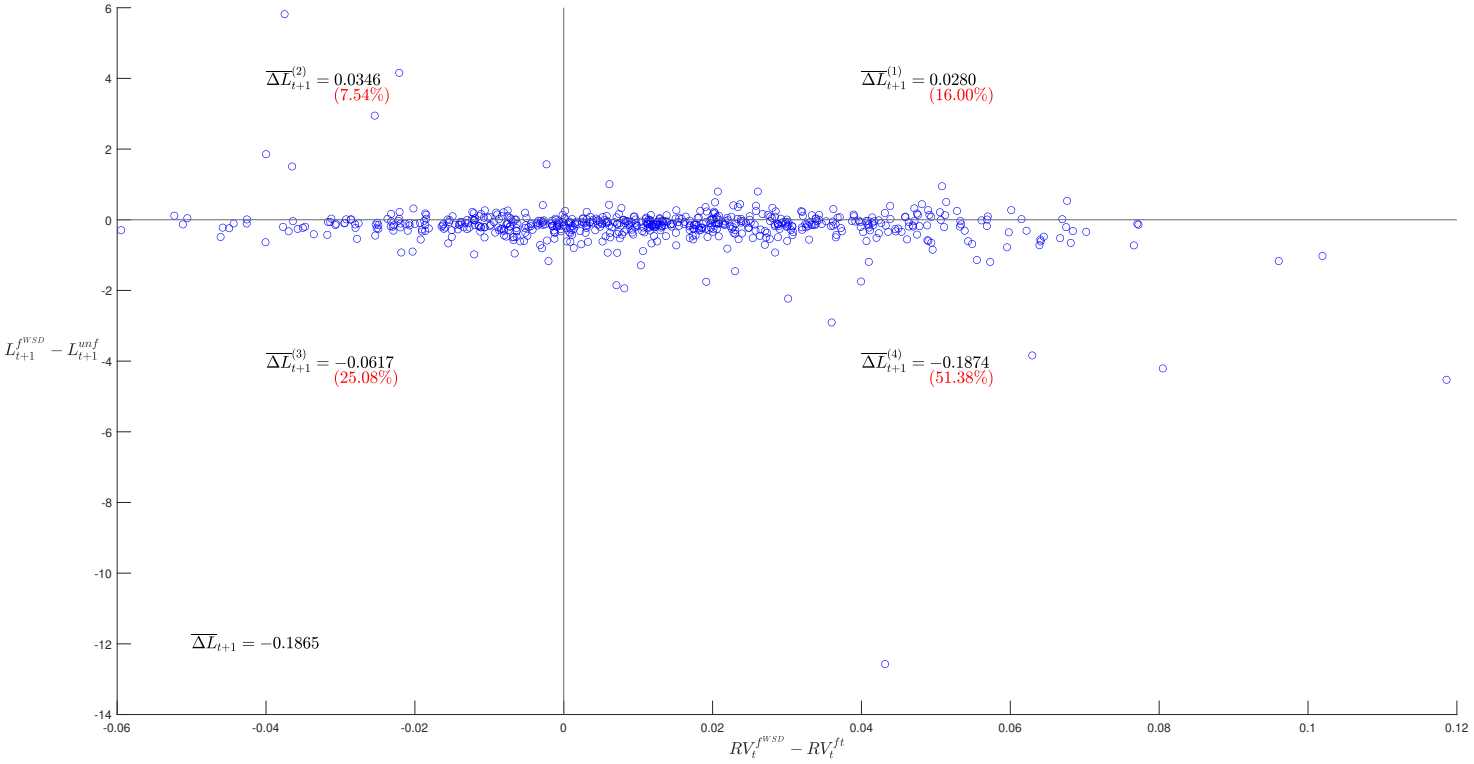
Note: The figure plots the median and the 5% and 95% quantiles for the MSE and QLIKE loss ratios, for the HARP versus the HAR model against the length of the time window over which periodicity was estimated. Simulations are based on the SV2F model and shown results are for a sampling frequency equal to 1 sec.

Figure 2.10: The impact of the jump-related periodicity estimation error on the performance of HARP models



Note: The figure plots the median and the 5% and 95% quantiles for the MSE and QLIKE loss ratios, for the $HARP^{fj}$ model, for which filtering is applied to data containing jumps, versus the $HARP^{fnj}$ model, for which filtering is performed before adding jumps to the data. Simulations are based on the SV1F model plus jumps. All forecasting horizons are included: one-day ($h = 1$), one-week ($h = 5$) and one-month ($h = 22$). The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.11: One-day ahead loss differential as a function of the amount of filtering

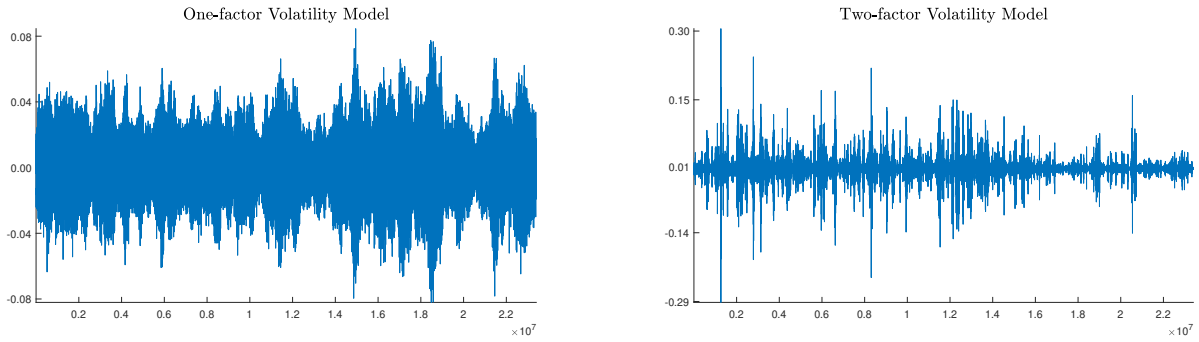


Note: The figure depicts the loss differential for HARP versus HAR models, $L(\epsilon^{WSD}_{t+1}) - L(\epsilon^{unf}_{t+1})$, against the previous day difference between the filtered and no periodicity realized volatilities, $RV^{WSD}_t - RV^{ft}_t$. The loss function considered is the MSE. Each quadrant reports the average loss difference, $\overline{\Delta L}_{t+1}$, as well as the percentage number of points in that quadrant (red ink). Data is generated from the SV2F model.

Appendix 2.C Additional Results on Simulated and Empirical Data

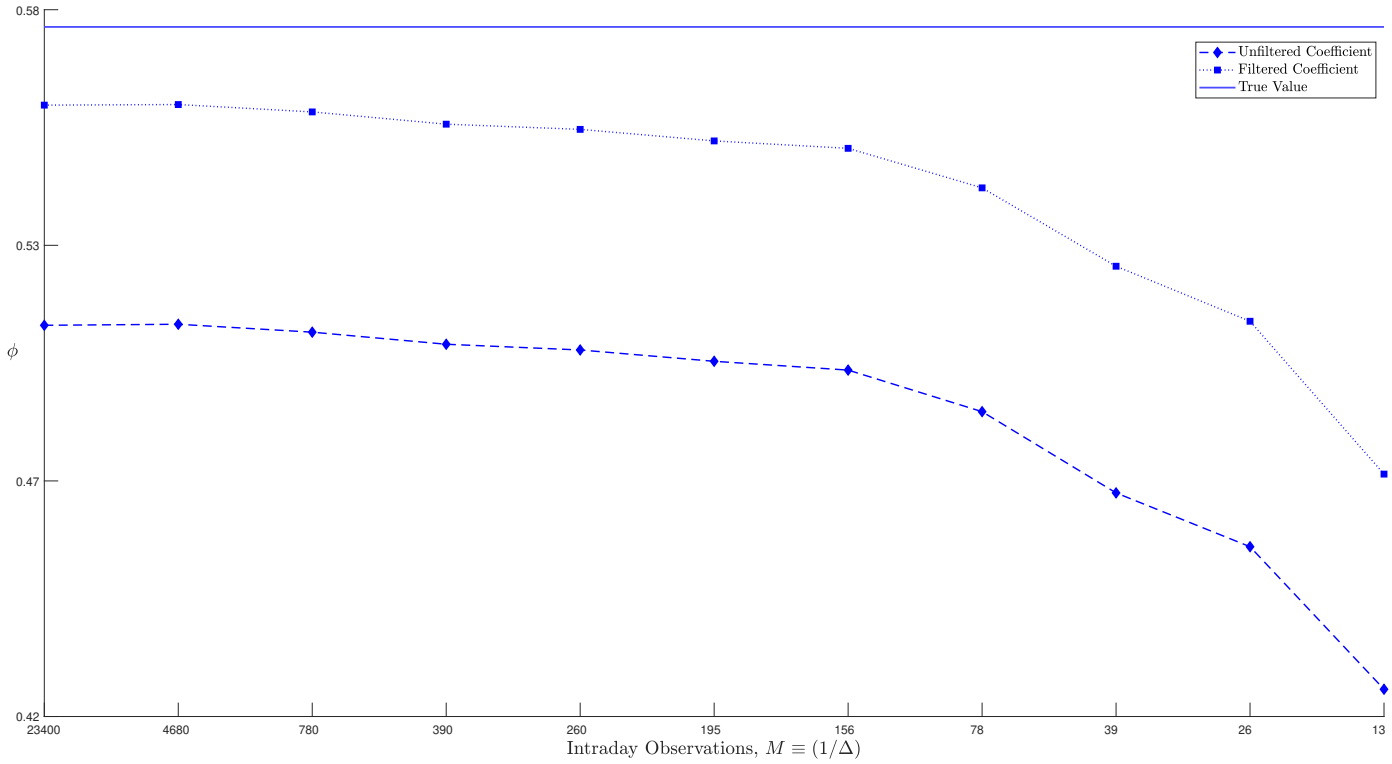
2.C.1 Additional Results on Simulated Data

Figure 2.12: Simulated returns for the two stochastic volatility models



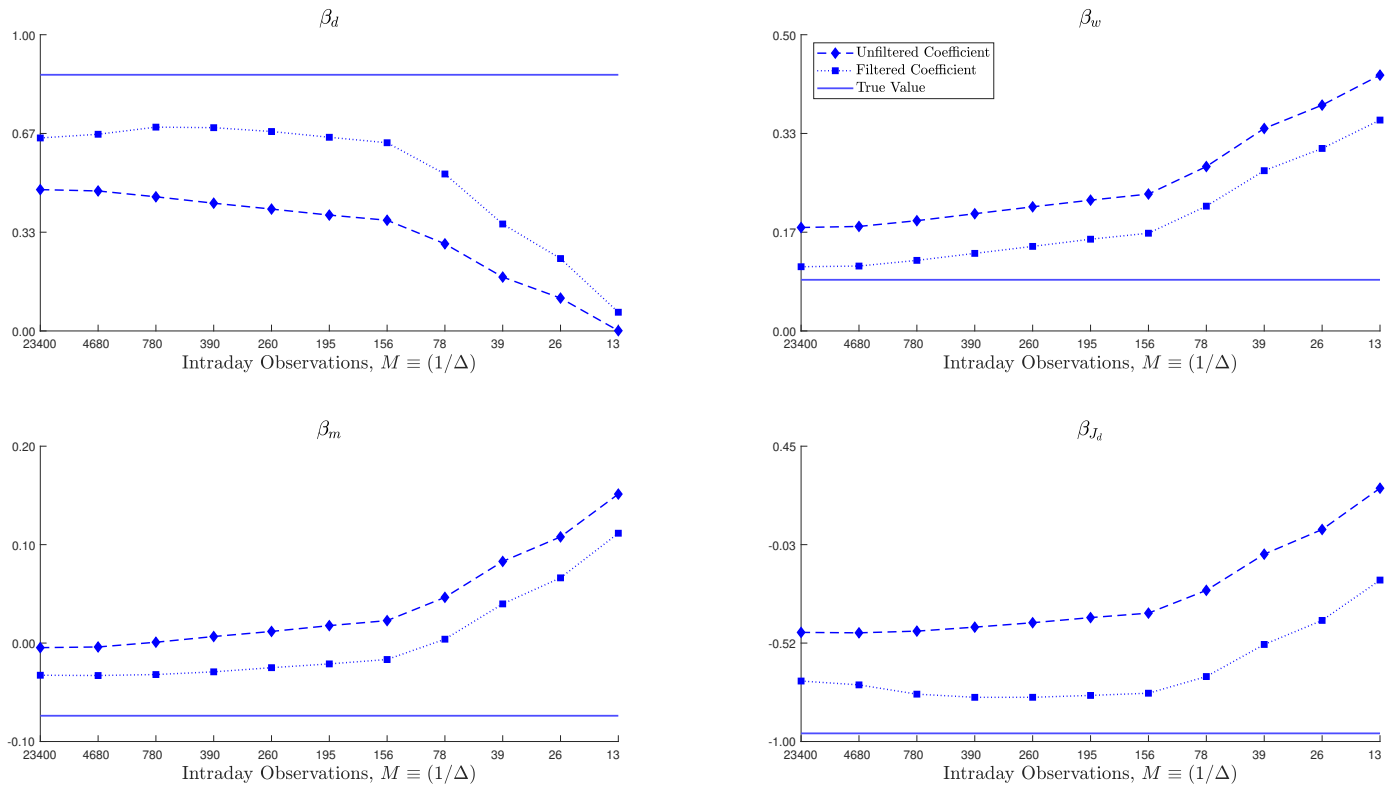
Note: The plot shows the dynamics of simulated 1-second returns over 1000 trading days for the one-factor stochastic volatility model (SV1F) and the two-factor stochastic volatility model (SV2F).

Figure 2.13: Filtered vs. unfiltered AR(1) coefficients for the SV2F model



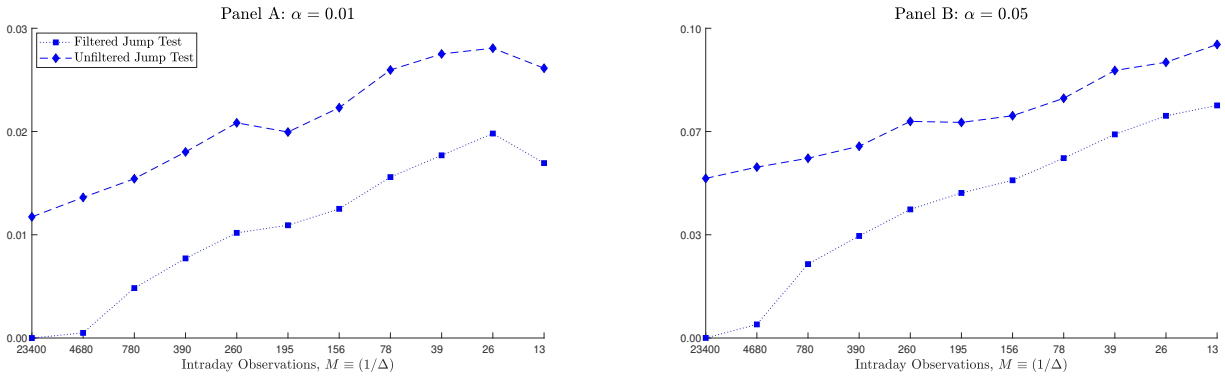
Note: This figure plots the true coefficient of an AR(1) model versus the estimated coefficient based on unfiltered (diamond marker) and filtered (squared marker) data across different sampling frequencies. Data is generated from the SV2F model, outlined in section 2.3.2 of the paper. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.14: HARP-J and HAR-J coefficients for the SV1F model with jumps



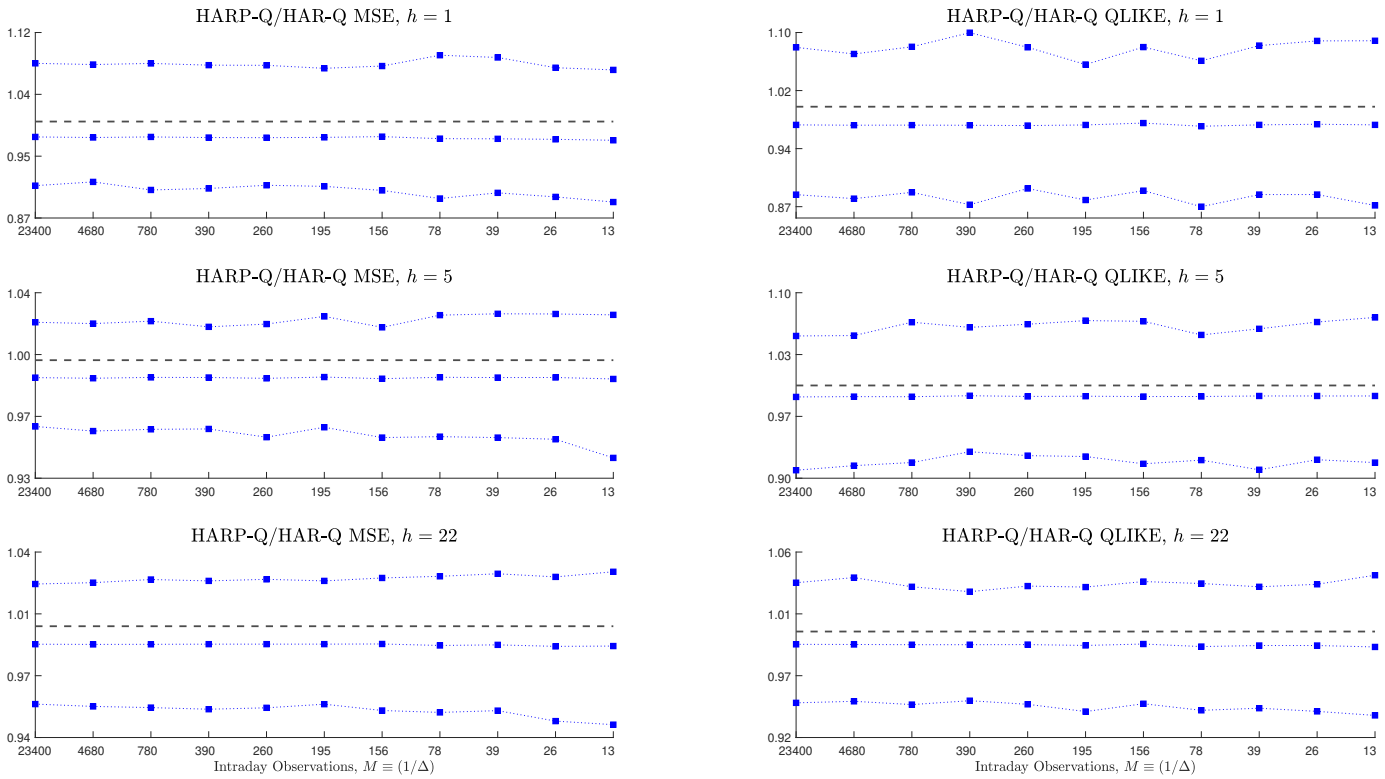
Note: This figure compares the estimates of the HARP-J (squared marker) and HAR-J (diamond marker) models against the true estimates across different sampling frequencies. Data is generated from the SV1F model plus jumps, as outlined in section 2.3.2 of the paper. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.15: Proportion of spurious jumps by sampling frequency for filtered and unfiltered data-Med-RV test



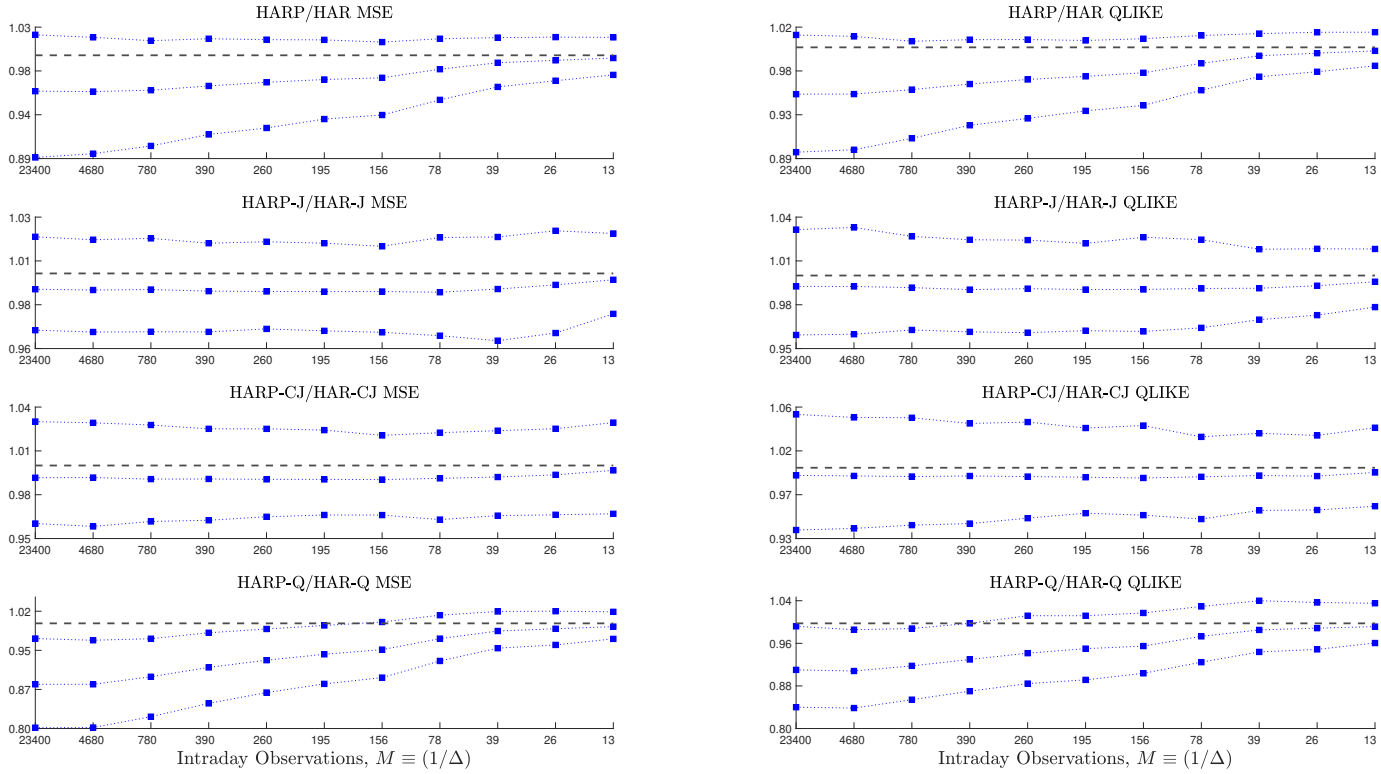
Note: This plot graphs the proportion of spuriously detected jumps across sampling frequencies using the jump test by jump test by Andersen et al. (2012), evaluated at the 1% and 5% significance level. Data is generated from the SV1F model plus jumps. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.16: HARP-Q/HAR-Q loss ratio



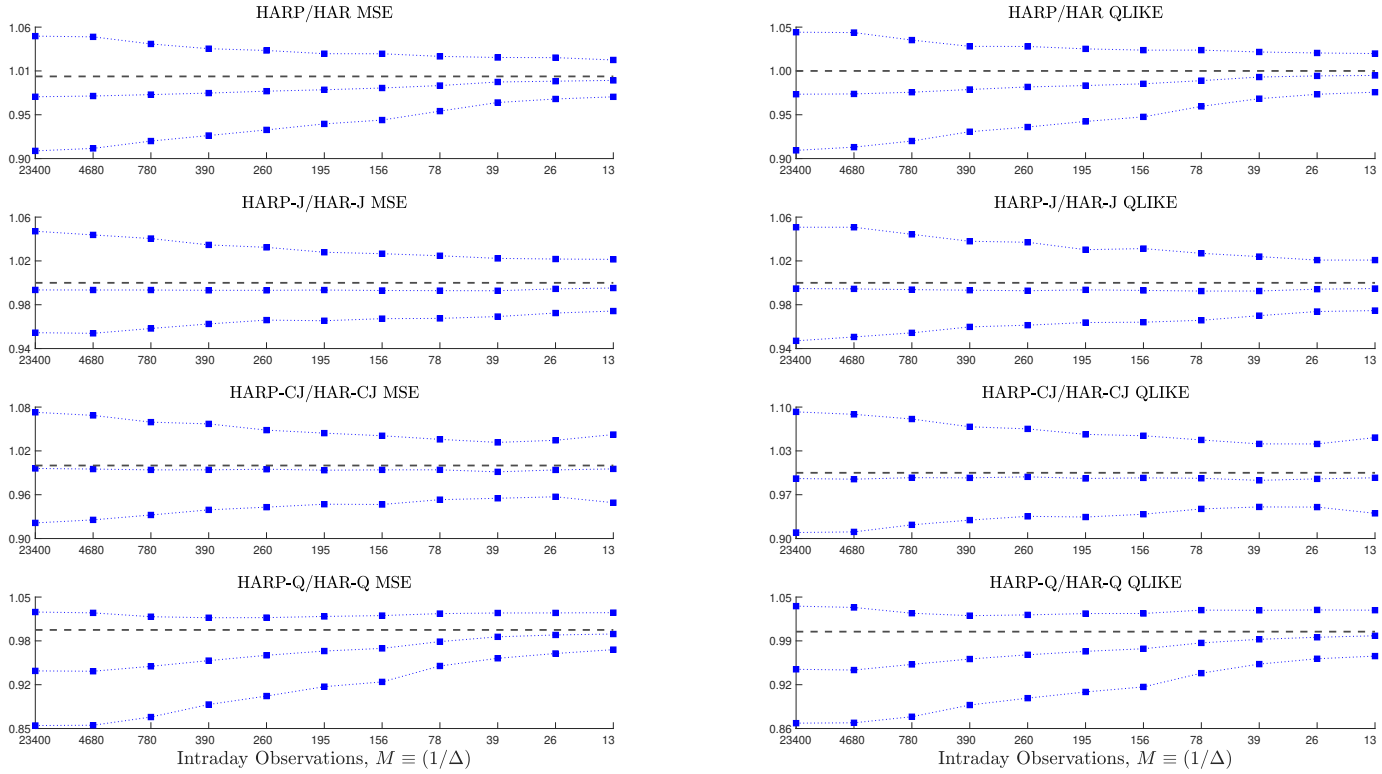
Note: The figure plots the median and the 5% and 95% quantiles for the MSE and QLIKE loss ratios, for the HARP versus the HAR model. All forecasting horizons are included: one-day ($h = 1$), one-week ($h = 5$) and one-month ($h = 22$). The dashed horizontal line corresponds to the value 1. Data is generated from the SV2F model. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.17: One-week ahead loss ratio for the simulated SV1F model with jumps



Note: The figure plots the median and the 5% and 95% quantiles for the MSE and QLIKE loss ratios, for HARP versus HAR models. The dashed horizontal line corresponds to the value 1. Data is generated from the SV1F model plus jumps. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

Figure 2.18: One-month ahead loss ratio for the simulated SV1F model with jumps



Note: The figure plots the median and the 5% and 95% quantiles for the MSE and QLIKE loss ratios, for HARP versus HAR models. The dashed horizontal line corresponds to the value 1. Data is generated from the SV1F model plus jumps. The number of observations on the x axis corresponds to the following sampling frequencies: 1 second (23400), 5 seconds (4680), 30 seconds (780), 1 minute (390), 1.5 minutes (260), 2 minutes (195), 2.5 minutes (156), 5 minutes (78), 10 minutes (39), 15 minutes (26) and 30 minutes (13).

2.C.2 Additional Results on Empirical Data

Table 2.8: Realized number of jumps and the estimated proportion of integrated variance in the quadratic variation using the jump test by Andersen et al. (2012)

		Unfiltered		Filtered	
Stock	Ticker	# Jumps	%QV	# Jumps	%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 $\%QV = \frac{\sum_{t=1}^{T} C_t}{\sum_{t=1}^{T}(C_t+J_t)}$.

Table 2.9: Results for the Rivers and Vuong (2002) test applied to SPY data

			BV		Med-RV	
	HARP/ HAR	HARPQ/ HARQ	HARP-J/ HAR-J	HARP-CJ/ HAR-CJ	HARP-J/ HAR-J	HARP-CJ/ HAR-CJ
$h = 1$	0.811	0.129	0.850	0.674	0.003*	0.000*
$h = 5$	0.039*	0.000*	0.474	0.439	0.000*	0.000*
$h = 22$	0.000*	0.000*	0.040*	0.026*	0.025*	0.006*

Note: The table reports the p-values for the two-sided Rivers and Vuong (2002) test. The test evaluates at the 5% significance level whether the in-sample performance of the HARP (HAR) models is significantly better relative to the performance of the HAR (HARP) models. Starred values indicate significance for the HARP models, while a diamond superscript indicates significance in favour of the HAR models.

Table 2.10: Estimated 1-, 5-, and 22- day ahead HAR(P)-J models for SPY and stocks 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-J and HARP-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-J and HARP-J models of various horizons. *, ** and *** denote significance at 10%, 5% and 1% respectively.

Table 2.11: Estimated 1-, 5-, and 22- day ahead HAR(P)-CJ models for SPY and stocks 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-CJ and HARP-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-CJ and HARP-CJ models of various horizons. *, ** and *** denote significance at 10%, 5% and 1% respectively.

Chapter 3

A Simple Model Correction for Modelling and Forecasting (Un)Reliable Realized Volatility

3.1 Introduction

Asset return volatility plays a crucial role in a number of practical financial management decisions, for which reason extensive efforts have been made to provide real-time estimates and forecasts of current and future volatility. Unlike returns, volatility is not directly observable, but rather inherently latent. This poses several challenges, where the most relevant is related to the errors-in-variable problem associated with the measurement of the realization of the forecasted variable. Constructed as the summation of finely sampled squared high-frequency returns, the use of so-called realized volatility (RV) does provide consistent estimates of the true latent volatility, but in finite samples the estimates are subject to measurement errors (e.g. Andersen et al., 2005; Andreou and Ghysels, 2002).[1] The additive errors-in-variables problem leads to a dilution bias in the HAR models, that automatically delivers insignificant coefficients with values decreasing towards zero. This is because the errors make the RV less persistent than the true la-

[1]The use of RV as a proxy for the true latent proxy is documented in the work of Andersen and Bollerslev (1998a); Andersen et al. (2001b, 2003); Barndorff-Nielsen and Shephard (2002a,b), inter alios.

tent process (e.g. Andersen et al., 2003; Bollerslev et al., 2016). This also produces a downward bias in the estimates of the R^2s (e.g. Andersen et al., 2011; Meddahi, 2003).

In light of the above, this paper proposes an alternative approach to correct the dilution bias observed in the HAR-type models when RV is unreliable.[2] We named our model DBC-HAR, where "DBC" stands for dilution bias correction. We note that on days where RV is unreliable, the daily RV is much larger than both the weekly and monthly RV, with the latter being less prone to errors. Therefore, we rely on the absolute difference between the daily and monthly RV as a metric of the relative magnitude of the measurement error. By implementing this metric in the DBC-HAR model, we allow the daily autoregressive parameter to dynamically load more (less) weight to the daily coefficient in days with low (high) levels of measurement error. Thereby, the DBC-HAR model displays stronger persistence during reliable days, and faster mean-reversion during unreliable days.

Although the literature regarding modelling RV and RV-based forecasting is extensive, the issue of measurement error has generally been ignored. Hansen and Lunde (2014) and Andersen et al. (2011) are among the few who propose a remedy for modelling unreliable RV. Their remedy is to use standard instrumental variables to deal with the errors-in-variables and an ARMA model to correct the errors in RV. However, they assume that the errors are homoskedastic. Bollerslev et al. (2016) were the first to successfully deal with the issue of heteroskedastic measurement error. They show that, with errors present, the RV is equal to the sum of the true latent volatility process and an error term. Thus, they propose the HAR-Q model that uses the asymptotic distribution of the RV, where they show that the square root of the Realized Quarticity (RQ) is proportional to the measurement error. More recently, Cipollini et al. (2020) argue that the square root of the RQ offers a numerical approximation to the RV, and therefore it is more appropriate to use the RV as the RQ is more sensitive to outliers.[3]

The specifications of Bollerslev et al. (2016) and Cipollini et al. (2020) address the issue of the measurement error using very similar approaches. The $RQ^{1/2} \approx \delta RV$, where

[2] We refer to (un)reliable RV to the case where RV is (less) more precisely estimated.
[3] Buccheri and Corsi (2019) use a Kalman filter to forecasting unreliable RV.

δ is a correction for numerical differences, as that RQ and RV rely respectively in fourth and second order returns. It is well-known (Andersen et al., 2010; Barndorff-Nielsen and Shephard, 2002a,b) that both the RQ and RV are very sensitive to the presence of jumps, noise and volatility bursts. This means that the level of the RQ and RV can increase substantially for reasons other than the degree of the measurement error, sending a wrong signal to the HAR specification. By contrast, the DBC-HAR model relies in the difference between the daily RV and monthly RV, as the latter averages 22 daily RVs, the measurement error issue is very small. Thereby, in the absence (presence) of measurement error the average difference of these two measures is expected to be (much greater) zero, while when jumps and volatility bursts are present, they impact both measures providing a more robust metric for the measurement error than just using the $RQ^{1/2}$ or RV.

We evaluate the forecasting improvements afforded by the DBC-HAR model relative to the baseline HAR and HAR-Q models. Our analysis begins with a Monte Carlo simulation, using the two-factor stochastic volatility model.[4] Here, the usual features of the data are incorporated, such as intraday volatility patterns and market microstructure noise. The empirical analysis comprises the SPDR S&P 500 ETF (SPY) and 12 individual Dow Jones constituents as traded continuously during the sample period, i.e. from January 3, 2000 to December 31, 2010. To avoid market microstructure effects, we use data sampled from 2 to 10 minutes.

The main findings are summarized as follows. Using Monte Carlo simulations and empirical analyses we show that the time-varying nature of our model provides significant in- and out-of-sample improvements relative to the HAR model. The improved performance of the DBC-HAR model is obtained by the dynamic allocation of weights, so delivering more persistent forecasts during periods where the measurement error is low, and quickly mean reversion forecasts on periods where RV is unreliable. This finding is confirmed by the greater level of persistence and smaller mean lag value of the DBC-HAR models, as compared with the HAR model. These forecasting gains are also obtained when using both longer forecasting horizons and different sampling frequencies. The DBC-HAR

[4]This model has been normally employed in the literature since it generates both extreme volatility and extreme returns (e.g. Bandi and Russell, 2008; Bollerslev et al., 2016; Huang and Tauchen, 2005).

model consistently outperforms the forecasts of the HAR model when the data are split by regimes (pre-crisis and crisis period). This highlights that our model performs equally well across periods of low and high measurement error.

The DBC-HAR model generally outperforms the forecasts of the HAR-Q models for SPY, whereas for the stock average the performance of both models is of similar magnitude. By comparison, the DBC-HAR model is always retained by the model confidence set of Hansen et al. (2011) for SPY and excluded up to two-times for the individual stocks. On the other hand, the HAR-Q model is always excluded at $h = 1$ for SPY and up to five-times for the stock average. This suggests that the DBC-HAR model provides on average more accurate out-of-sample forecasts than both the HAR and HAR-Q models.

The model correction afforded by the DBC-HAR specification can easily be incorporated into more sophisticated HAR models, such as the HAR-J and HAR-CJ of Andersen et al. (2007a),[5] as well as into the continuous HAR (CHAR) model outlined in Bollerslev et al. (2016). The results confirm the superiority of our proposed specification, as the extended DBC-HAR models generally produce significantly smaller QKILE losses than both the HAR and the standard extended HAR models.

We also study the sensitivity of our metric to the use of different functional forms, such as the logarithmic and square root functions, and compare the forecasts of the DBC-HAR model to those of more efficient noise-robust measures based on a HAR specification. First, using the square root function to estimate our metric, results in good out-of-sample performance, albeit the standard metric generally dominates. Second, the DBC-HAR model significantly outperforms the forecasts of the noise-robust measures across all forecasting horizons.

Finally, we show that our correction mechanism can be successfully applied to the GARCH family. Using the GARCH and GJR-GARCH models, we show that taking into account the measurement error in RV, not only improves the in-sample fitting of the models and reduces the standard errors, it also increases the out-of-sample forecast

[5]The HAR-J and HAR-CJ models have been extensively used in the RV-based forecasting literature (e.g. Busch et al., 2011; Corsi et al., 2010; Duong and Swanson, 2015; Giot and Laurent, 2007, and the references therein).

accuracy between 0.8–2 percentage points.

The remainder of the paper is structured as follows. Section 3.2 provides the modelling framework and outlines the forecasting models. Here, the proposed DBC-HAR model is described and a new metric for measuring the relative magnitude of the measurement error is proposed. The forecast evaluation criteria are also presented. Section 3.3 presents the simulation setup and results, together with the empirical study that reports the in- and out-of-sample forecasting results for the SPY and 12 individual stocks. Section 3.4 provides a set of robustness checks that: a) study the sensitivity of our measurement error metric to different functional forms; b) evaluate the out-of-sample performance of the DBC-HAR models to HAR models based on more efficient noise-robust measures; c) evaluate the performance of alternative DBC-HAR models; d) extend the DBC approach to the GARCH family, showing that GARCH models also improve their performance after accounting for the errors in RV. Section 3.5 concludes.

3.2 Modelling Framework

Let us assume that the data generating process X_t (log-price) is a real-valued process that can be included in a standard probability space, in the form of an Ito's semimartingale:

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t, \tag{3.1}$$

where W_t is a standard Brownian motion, μ_t is a predictable drift, σ_t is spot volatility which is both adapted and cádlág. $dJ_t = k_t dq_t$ is a jump process where q_t is a non-explosive Poisson process whose intensity is an adapted stochastic process λ_t, and k_t is the adapted random variable measuring the size of the jump at time t and satisfying, $\forall t \in [0, T]$, $\mathbb{P}[k_t = 0] = 0$.

In practice, say, for risk management purposes, we are interested in forecasting the Quadratic Variation (QV) of the process in equation (3.1), which takes the following

form:

$$QV_t = \int_0^t \sigma_s^2 ds + \sum_{0 \leq s \leq t} (\Delta X_s)^2, \tag{3.2}$$

where the time unit is one day. $\Delta X_s := X_s - X_{s-} \neq 0$, if and only if X jumps at time s. The QV is not directly observable, but can be proxied using the realized variance defined by the sum of intraday squared returns:

$$RV_t = \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^2, \tag{3.3}$$

where $\Delta_i^n = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ is the Δ_n-period intraday returns. The RV is a consistent estimator as $\Delta_n \to 0$ (see, e.g. Andersen and Bollerslev, 1998a; Andersen et al., 2003).

In the absence of jumps, Barndorff-Nielsen and Shephard (2002a) show that the RV converges in probability to the Integrated Variation (IV), thereby allowing the asymptotic distribution of the realized variance to be derived:[6]

$$(RV_t - IV_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\Delta_n IQ_t), \tag{3.4}$$

where the IQ_t, is defined as:

$$IQ_t = \int_0^t \sigma_s^4 ds. \tag{3.5}$$

Barndorff-Nielsen and Shephard (2002a) show that the IQ_t is consistently estimated by the realized quarticity:

$$RQ_t = \frac{1}{3\Delta_n} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^4. \tag{3.6}$$

As shown by Andersen et al. (2010), the RQ involves the estimation of fourth order return moments. This provides a very imprecise estimator, which is highly non-robust in many

[6] Jacod (2008) derives the asymptotic distribution of RV for Brownian semimartingale processes with jumps.

realistic scenarios, even in the absence of jumps. In the absence of market microstructure noise and jumps, the asymptotic distribution of the RQ is:

$$(RQ_t - IQ_t) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Delta_n \vartheta \int_0^t \sigma_s^8 ds\right), \tag{3.7}$$

where ϑ is a known constant and is approximately 10.66.[7]

3.2.1 Forecasting Models

The HAR Model

The standard HAR model introduced by Corsi (2009) has become the natural benchmark for modelling and forecasting realized volatility. As demonstrated by Corsi (2009), the HAR model captures volatility persistence in a simple and parsimonious way using daily, weekly, and monthly lags of the volatility series. The HAR model is defined as:

$$RV_{t,t+h} = \beta_0 + \beta_d RV_t + \beta_w RV_{t-5,t} + \beta_m RV_{t-22,t} + \epsilon_{t+h}, \tag{3.8}$$

where $RV_{t,t+h} = h^{-1}[RV_{t+1} + RV_{t+2} + \cdots + RV_{t+h}]$ aggregates information between $\{t+1, t+h\}$.

The HAR-Q Model

As it is only a proxy of the true latent volatility, RV is subject to measurement error. As shown by Bollerslev et al. (2016), the variance of the RV measurement error is a function of the IQ; the greater the variance of the measurement error, the less persistent the observed process. To cater for this, they allow the daily parameter to vary as a function of the square root of the RQ as follows:

$$RV_{t,t+h} = \beta_0 + \underbrace{\left(\beta_d + \beta_{dQ} RQ_t^{1/2}\right)}_{\beta_{d,t}} RV_t + \beta_w RV_{t-5,t} + \beta_m RV_{t-22,t} + \epsilon_{t+h}, \tag{3.9}$$

[7] Andersen et al. (2012) show that the minimum and median RQ are less efficient than the RQ_t. However, these estimators are robust to jumps.

where the interaction between the lagged RV and the lagged $\text{RQ}^{1/2}$ provides a remedy for the dilution bias, as the daily parameter receives less weight when the value of $\text{RQ}^{1/2}$ is high, provided that $\beta_{d_Q} < 0$.

The DBC-HAR Model

The proposed Dilution Bias Correction HAR (DBC-HAR) offers a simple model adjustment for the attenuation bias induced by the additive errors-in-variables problem. We show that the measurement error is proportional to the absolute difference between the daily and monthly RV. As the monthly RV is constructed using a moving average of 22 RVs, this estimator is less prone to measurement error. This observation follows a similar rationale to the sub-sampling estimator of Zhang et al. (2005) and the pre-averaging method of Jacod et al. (2009). To illustrate, let K be the number of lags to include, or sometimes called a bandwidth parameter, so if one averages K RVs, i.e. $(RV_t + RV_{t-1} + \cdots + RV_{t-K})/K$, the averaged $\overline{RV}$ is closer to the true latent process, and it turns out that the variance of the error term is reduced by a factor of $1/K$. This can be shown using a simple example, in which the error term is i.i.d. and the RV follows the process:

$$RV_t = IV_t + \eta_t, \quad \eta_t \sim i.i.d.\mathcal{N}(0, \omega^2), \tag{3.10}$$

$$\overline{RV}_t = \overline{IV}_t + \overline{\eta}_t, \quad \overline{\eta}_t \sim i.i.d.\mathcal{N}(0, \omega^2/K). \tag{3.11}$$

To provide a more realistic example, Figure 3.1 plots the empirical distribution of the measurement error estimated as $RV_t/IV_t - 1$ across both sampling frequencies and forecasting horizons using the simulation set by equation (3.15).[8] Figure 3.1 shows that daily RV always has greater measurement error, with a distribution that is much more dispersed and right skewed, than the weekly and monthly RV. The latter is the least affected measure of the three horizons, with most of its values centered on zero. By implication, with measurement error present, the RV generally overestimates the IV. Hence, the absolute

[8] As shown in Section 3.3.1, our structure accounts for the main stylized facts observed in asset prices. These are: leverage effect, diurnal effect, extreme returns, and microstructure noise. The variance of the microstructure noise is assumed to be constant throughout a day, but changes from day-to-day.

difference between the daily and monthly RV indicates the relative magnitude of the measurement error observed in the daily RV.

The HAR model normally assigns greater weights to weekly and monthly RVs, as these variables are less prone to errors (see, Figure 3.1). Thereby, the HAR model adjusts slowly to new market information. By contrast, the time-varying allocation allows the DBC-HAR model to react faster to new information. This is because in periods of low measurement error, the DBC-HAR model places greater weight on more recent RV, generating more persistent forecasts.

We define the DBC-HAR model as follows:

$$RV_{t,t+h} = \beta_0 + \underbrace{(\beta_d + \alpha|RV_t - RV_{t-22,t}|)}_{\theta_{d,t}} RV_t + \beta_w RV_{t-5,t} + \beta_m RV_{t-22,t} + \epsilon_{t+h}, \quad (3.12)$$

where α is expected to be negative. In other words, when the measurement error rises ($RV_t \gg RV_{t-22,t}$), the information contained in the current RV decreases, and therefore $\theta_{d,t}$ tends to zero. This effect is very similar to that produced by the HAR-Q model; when the daily RV is high (and hence a high RQ) less weight is allocated to the daily lagged RV. Along these lines, Figure 3.2 illustrates the time-varying dynamics of the DBC-HAR and HAR-Q models for SPY. These parameters indeed capture the heteroskedastic feature of the measurement error, where their weights are generally well above the constant daily estimate of the HAR model. Few differences can be drawn between $\theta_{d,t}$ and β_{d_Q}, where the most relevant is that β_{d_Q} becomes largely negative on days with high RQ.

Forecasting Evaluation

We evaluate the out-of-sample performance of the DBC-HAR model in relation to two benchmarks: the HAR and HAR-Q models. We consider horizons $h = 1, 5, 22$, corresponding to one day, one week, and one month. To estimate the models, we use a rolling window regression of size 1000 days, which is approximately 4 years.

We use the quasi-likelihood (QLIKE) loss function to evaluate the prediction of the models. The QLIKE has been shown to be robust to noise in the proxy for volatility in

Patton (2011b).

$$QLIKE(RV_t, F_t) = \frac{RV_t}{F_t} - \log \frac{RV_t}{F_t} - 1, \tag{3.13}$$

where F_t is the out-of-sample forecast of the realized variance. We also employ the Model Confidence Set (MCS) procedure of Hansen et al. (2011) to identify the (sub)set of models that significantly outperform the others. Let $\mathcal{M}$ denote the set of all the models under analysis. We define $d_{h,i,j} = L(RV_t, F_t^{(i)}) - L(RV_t, F_t^{(j)})$ as the difference in the loss of the model i and the model j, where $i \neq j$. We use the QLIKE as loss function L and define the following test statistics:

$$t_{i,j}^h = \frac{\bar{d}_{h,i,j}}{\sqrt{\text{Var}(\bar{d}_{h,i,j})}}, \ \forall i, j \in \mathcal{M}, \tag{3.14}$$

where $\bar{d}_{h,i,j}$ is the average loss difference. The MCS test statistics are given by $T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t_{i,j}^h|$ and the null hypothesis, H_0, is equal predictive ability. We implement the MCS using 10,000 bootstrap resamples and a block window of 20 days. Surviving models are then retained with a confidence level $z_\alpha = 0.1$.

3.3 Simulated and Empirical Results

3.3.1 Monte Carlo Evidence

A simulation study is undertaken to investigate further the performance of the DBC-HAR model. Simulations are based on the two-factor stochastic volatility (SV2F) model as commonly used to generate intraday returns (e.g. Barndorff-Nielsen et al., 2008; Boller-

slev et al., 2016; Huang and Tauchen, 2005):

$$
\begin{aligned}
dp_t &= \mu dt + \sigma_{ut}\nu_t \left(\rho_1 dW_t^{(\nu_1)} + \rho_2 dW_t^{(\nu_2)} + \sqrt{1-\rho_1^2-\rho_2^2} dW_t^{(p)} \right), \\
\nu_t^2 &= \text{s-exp}\left(\gamma_0 + \gamma_1 \nu_{1t}^2 + \gamma_2 \nu_{2t}^2\right), \\
d\nu_{1t}^2 &= \alpha_1 \nu_{1t}^2 dt + dW_t^{(\nu_1)}, \\
d\nu_{2t}^2 &= \alpha_2 \nu_{2t}^2 dt + (1+\phi\nu_{2t}^2) dW_t^{(\nu_2)}, \\
\sigma_{ut} &= C + Ae^{-at} + Be^{-b(1-t)},
\end{aligned}
\tag{3.15}
$$

where $W_t^{(\nu_1)}$, $W_t^{(\nu_2)}$, and $W_t^{(p)}$ are standard Brownian motions, and s-exp is the usual exponential function with a linear growth function splined in at high values of its argument to avoid an explosive behavior.[9] The process ν_{1t} is the persistent factor and the process ν_{2t} is the strongly mean-reverting factor. The persistent factor is initialized by drawing $\nu_{10} \sim \mathcal{N}(0, -1/2\alpha_1)$ and $p_0 = \log(25)$. Following Huang and Tauchen (2005), we set $\mu = 0.03$, $\gamma_0 = -1.2$, $\gamma_1 = 0.04$, $\gamma_2 = 1.5$, $\alpha_1 = -0.00137$, $\alpha_2 = -1.386$, $\phi = 0.25$, and $\rho_1 = \rho_2 = -0.3$, where the parameters are expressed in daily units. For the diurnal U-shape function, we follow Hasbrouck (1999) and Andersen et al. (2012), and set $A = 0.75$, $B = 0.25$, $C = 0.88929198$, and $a = b = 10$, respectively. One common feature of high-frequency data is the presence of market microstructure noise. To account for this feature, we follow Barndorff-Nielsen et al. (2008) generating i.i.d. noise from $u_{t,i} \sim \mathcal{N}(0, \omega_t^2)$, with $\omega_t^2 = \xi^2 \int_0^t \nu_s^2 ds$, and where ξ is the noise-to-signal ratio. The variance of the noise is constant during any given day, but changes from day to day. So, we observe a contaminated process of the form: $Y_{t,i} = X_{t,i} + u_{t,i}$.

The simulations are generated using an Euler scheme based on 23,400 intervals for each of the $T = 2000$ days in the sample. We then aggregate these prices sparsely sampled and construct 5 different sampling frequencies, 1-min, 2-min, 5-min, 10-min, and 15-min which correspond to 390, 195, 78, 39, and 26 intraday observations per day. The forecasting exercise uses a window size of 700 days,[10] updating the coefficient estimates

[9] s-exp$(x) = \begin{cases} \frac{\exp(x_0)}{\sqrt{x_0 - x_0^2 + x^2}}, & if\ x > x_0, \\ \exp(x), & otherwise. \end{cases}$ Where $x_0 \equiv \log(1.5)$.

[10] The window size mimics the proportion used in our empirical analysis, i.e. $\approx 35\%$ of the sample

at every iteration. We use 200 replications for every sampling frequency.

Table 3.1 summarizes the key in-sample and out-of-sample findings. In 4 panels, we report i) the adjusted R^2, ii) the out-of-sample QLIKE, which is standardized by the loss of the HAR model, iii) the mean lag, and iv) the persistence. The mean lag indicates the location of the lag weights. The lower the mean lag, the greater the weight on more recent RVs.

The DBC-HAR model always improves on the fit of the HAR model irrespective of the forecasting horizon or sampling frequency under analysis. Compared to the HAR-Q model, the DBC-HAR generally delivers a greater R^2_{adj}, where the difference narrows as the time-interval widens. The out-of-sample standardized losses show the DBC-HAR model to give more accurate volatility predictions, with gains as large as 3 percentage points observed at $h = 1$. Whereas the DBC-HAR model normally outperforms the HAR-Q model across sampling frequencies and forecasting horizons, the gains are narrower than those observed over the HAR model. This is expected as both models account for the dilution bias observed in the HAR structure, with the only difference being the metric employed in estimating the measurement error.

The mean lag (third panel) shows the DBC-HAR model placing greater weight on the daily RV than the HAR model. The smaller mean lag ensures higher daily estimates, which in turn provide both a faster reaction to new information, and more persistent forecasts in periods of low measurement error. The greater persistence is confirmed in the fourth panel of Table 3.1, where the DBC-HAR model is shown to be more persistent than the HAR model. With few differences between the DBC-HAR and HAR-Q models, we nevertheless observe slightly bigger persistence in the DBC-HAR model across all forecasting horizons when returns are sampled every 5 minutes. As shown in panels 3 and 4, the mean lag increases and persistence decreases for all of the models as the number of observations per day falls (i.e. from 1-min to 15-min). This implies that lower frequencies generally reduce forecasting performance, because finite sample problems increases the level of measurement error. However, the HAR model is most affected.

---

size.

3.3.2 Empirical Results

Data

Our empirical study uses the S&P 500 ETF (SPY) to represent an aggregate market index and 12 Down Jones constituents. These stocks were continuously traded during our sample period, i.e. from January 3, 2000 to December 31, 2010. The aggregate market index and all the individual stocks are obtained from TickData database, and the data for each individual stock and aggregate market are aggregated from tick level using the previous tick interpolation. We retrieve data sampled every 2-, 5-, and 10-minutes, as equivalent to 390, 78, and 39 intraday observations per day. These sampling frequencies provide a good variance-bias trade-off (see, e.g. Aït-Sahalia et al., 2005; Bandi and Russell, 2006; Hansen and Lunde, 2006), and are standard choices in the literature when market microstructure noise effects are not a subject of interest.

Table 3.2 reports summary statistics for the RV of the SPY and the 12 stocks under analysis. In addition to the usual summary measures for the RV, we report summary statistics for the absolute difference of the RV_t and $RV_{t-22,t}$ (Panel B), and for cases where this difference is positive $(RV_t - RV_{t-22,t})^+$ (Panel C), and negative $(RV_t - RV_{t-22,t})^-$ (Panel D). The observed variation demonstrated by the descriptive statistics in Panel A illustrates the heterogeneity of our data, which is desirable for the empirical analysis.

Panel B shows the absolute difference between the daily and monthly RVs. If no measurement error were present, one would expect the average of this difference to be zero. However, the mean and standard deviation range respectively from 0.669–2.757 and 1.862–7.660. The implied presence of errors in RV, is confirmed by the maximum value of this absolute difference. Across our sample, AXP shows the biggest RV with a max RV of 299.968, while the max value of AXP in Panel B is 250.039. This implies that during this day the RV is less reliable. Figure 3.3 reaffirms this finding by plotting the 95% confidence bands for the RV of AXP together with its daily and monthly RV. The wider confidence bands on 2008-10-10 confirms that the RV is less reliable. A similar conclusion is reached via our metric which also shows the largest difference between the daily and monthly RV to coincide with the 2008-10-10 date.

Panels C and D split the absolute difference by sign, so that Panel C reports positive values ($RV_t > RV_{t-22,t}$), and Panel D reports the negative difference ($RV_t < RV_{t-22,t}$). For almost two-thirds of the sample data, monthly RV exceeds daily RV, with the biggest difference occurring where $RV_t > RV_{t-22,t}$. On average the largest positive difference is 5 times bigger than its negative counterpart. These results are in line with the simulation result in Figure 3.1, which shows the distribution of the errors in daily RV to be generally shifted to the right.

In-sample Estimation Results

Table 3.3 reports in-sample coefficient estimates together with measures of fit, mean lag, and persistence values for the HAR, HAR-Q, and DBC-HAR models for SPY using three forecast horizons based on 5 minutes returns; i.e. $h = 1$ (day), $h = 5$ (week), and $h = 22$ (month).[11] Standard errors are estimated using a Newey-West HAC robust estimator. This allows for serial correlation of up to order 5 ($h = 1$), 10 ($h = 5$), and 44 ($h = 22$), since the random error term in the models is serially correlated at least up to order $h - 1$. Starred values indicate the in-sample fit of the models to be significantly better than that of the HAR model based on the Rivers and Vuong (2002) test at the 5% significance level. The bottom panel of Table 3.3 presents the mean lag, adjusted R^2 and persistence level as an average of all the individual stocks.

The estimates of the HAR model are generally significant and more weight is allocated to the weekly and monthly lags, which is in line with both previous findings in the literature (e.g. Andersen et al., 2007a; Corsi, 2009) and our Monte Carlo exercise. On the other hand, the DBC-HAR model significantly improves on the fit of the HAR model, and displays a daily coefficient almost 3 times higher than the HAR daily coefficient. These results hold true across forecasting horizons and sampling frequencies. The α estimate is consistently negative and strongly significant. This means that α reduces the weight allocated to the daily estimate when the RV is less reliable (greater measurement error) reverting quickly to its long term mean while, when the reverse occurs, the DBC-HAR

[11] In-sample estimates are not reported for other sampling frequencies since the results are qualitatively similar. They are available upon request.

model displays stronger persistence than the HAR model.[12] In addition, the DBC-HAR model responds better to changes in information signal since greater weight is assigned to more recent information, as opposed to the HAR model that gives greater weight to older information. This finding is also supported by the mean lags of the DBC-HAR model which are almost 2 and 1.2 times smaller than those of the HAR model at $h = 1$ and $h = 22$, respectively.

Although the DBC-HAR provides a better model specification for modelling and forecasting RV than the HAR model, the HAR-Q model of Bollerslev et al. (2016) is a natural benchmark in that it provides a time-varying structure which addresses the measurement error based on the level of the RQ. While both models significantly outperform the in-sample fit of the HAR model, based on the Rivers and Vuong test at the 5% level, few differences can be drawn from the in-sample fit of these two models. Nevertheless, we find that the DBC-HAR (HAR-Q) model improves on the fit of the HAR-Q (DBC-HAR) model for the stock average (SPY).

To further contrast the dynamics of the time-varying parameters of the DBC-HAR ($\theta_{d,t}$) and HAR-Q ($\beta_{d,t}$) models with the constant daily estimate of the HAR model, Figures 3.4 and 3.5 respectively plot the time-varying and constant daily estimates of the SPY and IBM. For ten consecutive trading days, both Figures plot the time-varying parameters (top panel), the daily and monthly RV (middle panel), and the 1-day ahead fitted values (bottom panel). The plots on the left (right) panel represent days with low (high) volatility. As shown in Figures 3.4 and 3.5, $\theta_{d,t}$ and $\beta_{d,t}$ are of similar magnitude and well above the daily constant estimate of the HAR model when the level of measurement error is low. By contrast, the right top panel shows that both time-varying estimates fall below the daily HAR coefficient when $RV_t \gg RV_{t-22,t}$. During less volatile periods, and therefore low levels of measurement error, the DBC-HAR model provides more accurate 1-day ahead forecasts than both the HAR and HAR-Q models. Conversely, when the level of measurement error is high, the three models rely heavily in information from

[12] The sum of the autoregressive coefficients for the DBC-HAR and HAR-Q models occasionally exceed unity at $h = 1$. However, this does not necessarily imply non-stationary, as the temporal variation in the realized variance measures may induce stationarity (see, for instance, Conley et al., 1997; Nielsen and Rahbek, 2014, for a discussion of volatility induced stationarity).

weekly and monthly lags, thereby resulting in similar out-of-sample performance across these days.

Out-of-sample Forecasts Results

Table 3.4 reports the out-of-sample results for the SPY and for the average of the individual stocks across all sampling frequencies (2-, 5-, and 10-minutes), and forecasting horizons (1-, 5-, and 22-day).[13] The reported QLIKE losses are standardized by the losses of the HAR models. This clarifies comparisons since values below 1.0 indicate that the models outperform the HAR model. We also report the p-values of the Model Confidence Set (MCS) of Hansen et al. (2011), which are estimated using 10,000 bootstrap resamples and a block window of 20 days length. The surviving models are retained using a 10% confidence level. The starred values highlight the retained models for the SPY while, for the average stocks, we report the number of times each model is retained by the MCS. The smallest and highest p-value are reported in brackets.

For SPY, unlike the DBC-HAR model, the HAR and HAR-Q models are excluded by the MCS. For instance, the HAR-Q model is excluded at $h = 1$ across all horizons, while the HAR model is never retained when returns are sampled every 5 minutes. On several occasions the DBC-HAR model renders significantly smaller losses than the HAR and HAR-Q models. We find only two cases in which the DBC-HAR model is outperformed: at $h = 5$ when sampling every 2 minutes, and at $h = 22$ when the sampling occurs every 10 minutes.

For average stocks, the DBC-HAR model is generally retained by the MCS. The one exception is when returns are sampled every 2- and 5-minutes, and there are two exceptions when returns are sampled every 10-minutes. In line with the results of the SPY, the HAR model is excluded 45% of the time across sampling frequency and forecasting horizon. In general, the DBC-HAR (HAR-Q) model consistently outperforms the HAR-Q (DBC-HAR) model at $h = 1$ ($h = 22$), whereas for $h = 5$ the results are mixed. These findings confirm that the specification of the DBC-HAR model is superior to that of the

[13] The models are estimated using a rolling window of 1000 days.

HAR model, and it performs equally well as the HAR-Q model.

Figures 3.6 and 3.7 plot the one-day and one-month ahead standardized losses of the DBC-HAR and HAR-Q models for each individual stock and the SPY against the standard deviation of our measurement error metric, i.e. std($|RV_t - RV_{t-22,t}|$). In the absence or under constant measurement error, the average difference between the daily and monthly RV would be expected to be zero, so reducing the DBC-HAR model to the HAR model. However, we observe that the biggest forecasting gains are attained for stocks that display larger variations in the level of the measurement error as estimated using $|RV_t - RV_{t-22,t}|$. Although this finding also applies for the HAR-Q model, Figure 3.6 shows that on several occasions the standardized losses of the HAR-Q model are above 1.0.

To further examine the prediction performance of the DBC-HAR model, we partition the sample data as two sub-samples. The first covers the period prior to the sub-prime crisis, i.e. January 3, 2000 to December 29, 2006; the second covers the crisis period, i.e. 3, 2007 to December 31, 2010. The two panels of Table 3.5 report the out-of-sample results for the SPY and the average of the individual stocks.[14]

In line with the full sample results, during the pre-crisis period (Panel A), the HAR model is significantly outperformed by both the DBC-HAR and HAR-Q models. This holds true for the SPY and stock average. For SPY, the MCS excludes the HAR model across all frequencies and forecasting horizons, except for $h = 1$ using returns sampled every 2- and 5-minutes. The statistics for the stock average confirm the poor out-of-sample performance of the HAR model, which is constantly excluded by the MCS irrespective of the sampling frequency or forecasting horizon under analysis. Whether we refer to the SPY or the stock average, we find that both the DBC-HAR and HAR-Q models are generally retained by the MCS, and on average across all forecasting horizons and sampling frequencies, the DBC-HAR model outperforms the HAR-Q model.

The crisis period (Panel B) shows more dispersed out-of-sample results, yet the performance of both the DBC-HAR and HAR-Q models dominate that of the HAR model.

[14]The models are also estimated using a rolling window of 615 and 350 days for the pre-crisis and crisis period, respectively. These values represent about 35% of each sub-sample.

These out-of-sample gains decrease as the forecasting horizon increases, although the HAR model seldom outperforms its counterparts. In part, this is because less reliable RVs are often found during periods of financial distress. Unreliable RVs make predictions of volatility less accurate, and the greater levels of measurement error dwarf $\theta_{d,t}$ and $\beta_{d,t}$ towards zero, so that the specifications quickly revert to their long-term mean. The latter behavior is also observed in the HAR model. However, unlike the DBC-HAR and HAR-Q models, the HAR model lacks the capacity to dynamically update the loading of the daily estimates, so explaining how the HAR model underperforms the DBC-HAR and HAR-Q models.

In summary, the out-of-sample results show the superiority of the DBC-HAR model over the HAR model across both sampling frequencies and forecasting horizons. The superior in performance is also attained when the sample data are sub-categorized by pre-crisis and crisis period. During tranquil periods the greater persistence obtained by allocating more weight to the daily estimate, results in bigger out-of-sample gains. By contrast, in periods of financial turmoil, the time-varying nature of the DBC-HAR model increases/decreases the weight of the daily estimate as the measurement error decreases/increases, so rendering more accurate forecasts.

3.4 Robustness Check

3.4.1 Noise-robust Realized Measures

The literature generally advocates the use of 5-minutes returns as a way of mitigating the impact of market microstructure noise (e.g. Aït-Sahalia et al., 2005; Hansen and Lunde, 2006). However, following statistical principles, more data are preferred to less, as more data points contain more information. Based on this principle, the literature has proposed alternative noise-robust estimators for RV, which are meant to be more efficient as they use many more observations relying on pre-averaging or sub-samples techniques to remove the market microstructure noise effects. In this section, we compare the forecasts of the DBC-HAR model to those of HAR models based on more efficient realized measures,

which are estimated using 1- and 2-min returns. We consider two of the most common noise-robust estimators: the Two-Time Scale Realized Variance (TSRV) of Zhang et al. (2005) and the Pre-averaging Realized Variance (PRV) of Jacod et al. (2009). The TSRV is outlined as follows:

$$TSRV_t = \frac{1}{n_k} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor - k + 1} |X_{k+i\Delta_n} - X_{i\Delta_n}|^2 - \frac{\overline{n}}{n} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^2, \tag{3.16}$$

where $\overline{n} = (n - k + 1)/k$, $n_k = \frac{N_k}{n/k} \approx k$, N_k represents the number of observations after using the scale k. $k = [cn^{2/3}]$, c is the bandwidth stated as in Zhang et al. (2005), and n is the number of intraday observations.

The PRV takes the following form:

$$PRV_t = \frac{n}{n - L + 2} \frac{1}{L\psi_2^L} \sum_{i=0}^{\lfloor 1/\Delta_n \rfloor - L + 1} |\Delta_i^n X^*|^2 - \frac{\psi_1^L \hat{\omega}_{AC}^2}{\theta^2 \psi_2^L}, \tag{3.17}$$

where $L = \theta\sqrt{n} + o(n^{-1/4})$, $n/(n - L + 2)$ is a small sample bias correction, while the right hand side of equation (3.17) aims to remove the residual effect of noise that is not eliminated by the pre-averaging estimator. The variance of the noise is estimated following Oomen (2006a), $\hat{\omega}_{AC}^2 = -\frac{1}{n-1} \sum_{i=2}^{\lfloor 1/\Delta_n \rfloor} \Delta_{i-1}^n X \Delta_i^n X$. Finally, the pre-averaging returns are estimated as follows:

$$\Delta_i^n X^* = \sum_{j=1}^{L-1} g\left(\frac{j}{L}\right) \Delta_{i+j}^n X, \tag{3.18}$$

where $g = (x \wedge 1 - x)$. The constants associated with g are defined as:

$$\psi_1^L = L \sum_{j=1}^{L} \left[g\left(\frac{j}{L}\right) - g\left(\frac{j-1}{L}\right) \right]^2, \quad \psi_2^L = \frac{1}{L} \sum_{j=1}^{L-1} g^2\left(\frac{j}{L}\right). \tag{3.19}$$

Table 3.6 reports the out-of-sample results for the alternative HAR models using the noise-robust measures and the DBC-HAR model using 5-min returns. The QLIKE losses of the models are standardized by the loss of the DBC-HAR model. To facilitate comparison, we also report the results for the HAR-RV and HAR-Q models based on 5-min

returns. Whether we refer to the SPY or stock average, the DBC-HAR model consistently outperforms the forecasts of the noise-robust adjusted HAR models across all sampling frequencies and forecasting horizons, but the HAR-PRV at $h = 22$. In fact, the noise-robust forecasts generally fail to outperform the HAR models using 5-min RV across all the forecasting horizons. This finding is not surprising, as shown by Liu et al. (2015), 5-min RV is very difficult to beat. However, we observe a decrease in the out-of-sample performance of the noise-robust based forecasts as the time-interval increases, which indicates that more data increases the efficiency of these estimators. Eventually, higher sampling frequencies might increase their out-of-sample accuracy.

3.4.2 Alternative Functional Forms for the Measurement Error Metric

An alternative estimate of the relative magnitude of the measurement error is to use a logarithmic and square root transformation of the daily and monthly RV. Our choice for using level daily and monthly RV is motivated by the fact that logarithmic and square root transformations alter the distribution of our variables, reducing their right skewness. That reduction in skewness decreases the spread between the two variables, which can make the measurement error look more homoskedastic. However, the square root has a weaker effect than the logarithmic; hence, we should expect the square root to perform closer to our original metric than its logarithmic counterpart.

The transformed DBC-HAR models are outlined below:

$$RV_{t,t+h} = \beta_0 + (\beta_d + \alpha|\log RV_t - \log RV_{t-22,t}|)RV_t + \beta_w RV_{t-5,t} + \beta_m RV_{t-22,t} + \epsilon_{t+h}, \tag{3.20}$$

$$RV_{t,t+h} = \beta_0 + (\beta_d + \alpha|\sqrt{RV_t} - \sqrt{RV_{t-22,t}}|)RV_t + \beta_w RV_{t-5,t} + \beta_m RV_{t-22,t} + \epsilon_{t+h}, \tag{3.21}$$

where the DBC-HAR-log and DBC-HAR-sqrt models use the logarithmic and square root transformation as shown in equations (3.20) and (3.21), respectively.

The out-of-sample results for these forecasts are reported in Table 3.7. The QLIKE losses are standardized by the losses of the DBC-HAR model. Bold numbers highlight the best DBC-HAR specification, and for comparison we also report the HAR and HAR-Q models. Although the logarithmic transformation generally fails to outperform the square root and level version of the metric, the DBC-HAR-log model normally outperforms the HAR model. The decrease in performance of the DBC-HAR-log model is expected as the reduction in the spread spuriously mitigates the impact of the measurement error. On the other hand, the DBC-HAR-sqrt model performs well throughout our analysis, albeit the DBC-HAR model generally dominates across forecasting horizon and sampling frequency.

3.4.3 Alternative Forecasting Models

Alternative DBC-HAR Models

Since the DBC-HAR model provides a simple correction based on the absolute different between short-term (daily) and long-term (monthly) volatility, its extension to more sophisticated models is straightforward. For instance, one can modify the HAR-J model proposed by Andersen et al. (2007a) and obtain the DBC-HAR-J model as follows:

$$RV_{t,t+h} = \beta_0 + (\beta_d + \alpha|RV_t - RV_{t-22,t}|)RV_t + \beta_w RV_{t-5,t} + \beta_m RV_{t-22,t} + \beta_{J_d} J_t + \epsilon_{t+h}. \tag{3.22}$$

Similarly, the HAR-CJ of Andersen et al. (2007a) can be adapted to the DBC-HAR structure. However the difference between the short- and long-term variance is based on an estimator of the integrated variation rather than the usual RV. The DBC-HAR-CJ model is outlined as follows:

$$\begin{aligned} RV_{t,t+h} = \beta_0 + (\beta_{C_d} + \alpha|C_t - C_{t-22,t}|)C_t + \beta_{C_w} C_{t-5,t} + \beta_{C_m} C_{t-22,t} + \\ \beta_{J_d} J_t + \beta_{J_w} J_{t-5,t} + \beta_{J_m} J_{t-22,t} + \epsilon_{t+h}. \end{aligned} \tag{3.23}$$

To identify jumps in the price process we use the jump test proposed by Barndorff-Nielsen and Shephard (2006), which takes the following form:

$$Z_t = \frac{1 - BV_t/RV_t}{\sqrt{0.61\Delta_n \max(1, TPQ_t/BV_t^2)}} \sim \mathcal{N}(0,1), \tag{3.24}$$

where BV_t and TPQ_t are the respective realized bipower variation and realized tripower quarticity, which estimate the integrated variation and integrated quarticity.

$$BV_t = \mu_1^{-2} \frac{n}{n-1} \sum_{i=2}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X||\Delta_{i-1}^n X| \xrightarrow{p} \int_0^t \sigma_s^2 ds, \tag{3.25}$$

$$TPQ_t = n\mu_{4/3}^{-3} \frac{n}{n-2} \sum_{i=3}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^{4/3} |\Delta_{i-1}^n X|^{4/3} |\Delta_{i-2}^n X|^{4/3} \xrightarrow{p} \int_0^t \sigma_s^4 ds, \tag{3.26}$$

where $\mu_p \equiv 2^{p/2}\Gamma((p+1)/2)/\Gamma(1/2)$, for $p > 0$. C_t and J_t are estimated as follows:

$$C_t = RV_t \cdot \mathbb{1}(Z_t = 0) + BV_t \cdot \mathbb{1}(Z_t = 1), \tag{3.27}$$

$$J_t = (RV_t - BV_t) \cdot \mathbb{1}(Z_t = 1), \tag{3.28}$$

with $\mathbb{1}(\cdot)$ being an indicator function. The Continuous HAR (CHAR) model extended to our time-varying set up, which we term DBC-CHAR model, is outlined as follows:

$$RV_{t,t+h} = \beta_0 + (\beta_{C_d} + \alpha|C_t - C_{t-22,t}|)C_t + \beta_{C_w} C_{t-5,t} + \beta_{C_m} C_{t-22,t} + \epsilon_{t+h}. \tag{3.29}$$

Table 3.8 reports the 1-day ahead in-sample fit for the alternative DBC-HAR models and their benchmark counterparts. We observe an increase in the daily estimate which is almost 3 times bigger than those of their benchmark models. This confirms that DBC-HAR models allocate more weight to the daily lag. In addition, the standard errors of the daily estimates are reduced significantly, implying that the daily coefficients of the DBC-HAR models are more informative. This dynamic allocation of weight offers a better model specification, leading to an increase of about 4 percentage points in the adjusted R^2s. Moreover, the Rivers and Vuong test rejects the null of equal predicting ability

in favor of all the alternative DBC-HAR models, highlighting the substantial in-sample increments for models accounting for the measurement error.

As reported in Table 3.9, out-of-sample results of the alternative HAR models corroborate the in-sample performance. This means that the DBC-HAR models significantly outperform their benchmark models. This finding is supported by the Giacomini and White (2006) test, which finds the losses of the alternative DBC-HAR models to be significantly smaller than those of their benchmark models. Similarly, the MCS usually retains all the extended DBC-HAR models, while excluding their baseline models. For instance, the HAR-CJ and CHAR models are always excluded across sampling frequencies, as opposed to the DBC-HAR-CJ and DBC-CHAR models that are always retained by the MCS.

Alternative GARCH Models

Our implementation is not only limited to autoregressive models. The simple correction afforded by our metric offers a straightforward extension to other families of volatility models such as the (G)ARCH family proposed by Engle (1982) and Bollerslev (1986). In this section, we show that GARCH models improve their in- and out-of-sample performance when they account for the measurement error observed in RV. Let $\hat{\sigma}_t^2$ be a realized measure of the variance such as RV and $\nu_t = \mathbb{E}_{t-1}[\hat{\sigma}_t^2]$ its conditional expectation at time $(t-1)$. To simplify notations and comparisons we only consider the $(1,1)$ specification. The GARCH model of Bollerslev (1986) is then outlined as follows:

$$\nu_{t+h} = \omega + \beta\nu_t + \pi RV_t. \tag{3.30}$$

Similarly the DBC-GARCH model takes the following form:

$$\begin{aligned} \nu_{t+h} &= \omega + \beta\nu_t + \underbrace{(\pi + \alpha\varsigma_t)}_{\phi_t} RV_t, \\ \varsigma_t &= |RV_t - RV_{t-22,t}|. \end{aligned} \tag{3.31}$$

The GJR-GARCH model of Glosten et al. (1993) is described as:[15]

$$\nu_{t+h} = \omega + \beta\nu_t + \pi RV_t + \gamma R_t^2 \mathbb{1}\left(R_t < 0\right). \tag{3.32}$$

The DBC-GJR-GARCH modifies the original model as follows:

$$\begin{aligned} \nu_{t+h} &= \omega + \beta\nu_t + \underbrace{\left(\pi + \alpha\varsigma_t\right)}_{\phi_t} RV_t + \gamma R_t^2 \mathbb{1}\left(R_t < 0\right), \\ \varsigma_t &= |RV_t - RV_{t-22,t}|. \end{aligned} \tag{3.33}$$

We refrain from using the GARCH-X structure motivated by Engle (2002b), who shows that, when a GARCH-X model is created by adding a realized measure, the ARCH parameter is close to zero, in consequence of the realized measure subsuming most of the information. The conditional variances are estimated using a Gaussian quasi log-likelihood function that is given by:

$$\mathscr{L} = -\sum_{t=1}^{T}\left[\log h_t + \frac{\hat{\sigma}_t^2}{h_t}\right] \tag{3.34}$$

Table 3.10 reports the in-sample estimates of the GARCH and alternative GARCH specifications for the SPY along with the log-likelihood, BIC, and 1-day ahead out-of-sample standardized QLIKE loss. The bottom panel reports the log-likelihood, BIC, and the standardized QLIKE loss as an average of all the individual stocks. The general pattern in the results is similar to that found in the DBC-HAR models, where the BIC indicates that our proposed DBC-GARCH and DBC-GJR improve on the fit of their standard counterparts. In addition, there is an increase in the persistence level of our models and a decrease in the standard errors of their parameters. The out-of-sample standardized QLIKE losses indicate that the better in-sample fitting directly translates in more accurate volatility forecasts, with improvements between 0.8 and 2 percentage points.

[15] R_t represents the daily returns, which are estimated as $R_t = \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} \Delta_i^n X$.

3.5 Conclusion

We have proposed a simple model correction that successfully accounts for the dilution bias induced by the errors-in-variable problem of RV-based forecasting models. We have shown that the absolute difference between the daily and monthly RV is proportional to the relative magnitude of these errors. Therefore, our proposed models improve upon the in- and out-of-sample performance of standard volatility forecasting models, by allowing the daily autoregressive coefficient to vary as a function of the measurement error. This results in more responsive forecasts with greater persistence on days where the RV is more precisely estimated, and faster mean-reversion forecasts on days where the RV is unreliable.

Implementing our model correction in the SPY and 12 individual stocks for an 11-year period, we were able to show substantial improvements in the fit of the models together with a decrease in the standard errors of the coefficients. The better fit of the models directly translates into more accurate out-of-sample forecasts, which significantly outperform standard HAR and GARCH specifications. These findings held true across various sampling frequencies, different forecasting horizons, and when sub-samples of the data were used to distinguish between pre-crisis and crisis periods.

Appendix 3.A Tables and Figures

Table 3.1: Simulation Results

	HAR	HAR-Q	DBC-HAR	HAR	HAR-Q	DBC-HAR	HAR	HAR-Q	DBC-HAR
	$h = 1$			$h = 5$			$h = 22$		
	In-Sample R^2_{adj}								
1-min	0.229	0.244	0.245	0.218	0.219	0.221	0.359	0.359	0.360
2-min	0.216	0.239	0.239	0.218	0.219	0.222	0.360	0.360	0.361
5-min	0.212	0.219	0.220	0.206	0.208	0.212	0.343	0.343	0.346
10-min	0.175	0.196	0.192	0.190	0.200	0.202	0.332	0.334	0.334
15-min	0.166	0.177	0.177	0.195	0.199	0.199	0.339	0.342	0.342
	Out-of-Sample QLIKE								
1-min	1.000	0.973	0.969	1.000	0.992	0.988	1.000	0.988	0.985
2-min	1.000	0.978	0.973	1.000	0.996	0.985	1.000	0.988	0.987
5-min	1.000	0.993	0.983	1.000	0.989	0.986	1.000	0.989	0.987
10-min	1.000	0.982	0.984	1.000	0.992	0.987	1.000	0.988	0.986
15-min	1.000	1.000	0.998	1.000	0.993	0.987	1.000	0.989	0.985
	Mean Lag								
1-min	6.375	4.810	4.573	9.164	8.672	8.273	10.663	10.632	10.210
2-min	6.499	4.695	4.436	9.322	8.766	8.366	10.672	10.553	10.194
5-min	6.368	5.436	5.107	9.131	8.644	8.068	10.660	10.435	10.031
10-min	6.593	4.930	4.910	9.159	8.037	7.885	10.761	10.311	10.189
15-min	6.850	5.562	5.495	8.971	8.224	8.164	10.765	10.222	10.238
	Persistence								
1-min	0.755	0.938	0.917	0.644	0.675	0.692	0.598	0.600	0.617
2-min	0.752	0.973	0.940	0.647	0.683	0.698	0.599	0.605	0.620
5-min	0.741	0.834	0.839	0.632	0.661	0.686	0.585	0.595	0.610
10-min	0.714	0.845	0.833	0.618	0.675	0.678	0.576	0.594	0.597
15-min	0.716	0.820	0.808	0.620	0.661	0.659	0.582	0.605	0.602

Note: The table reports the adjusted R^2 as a measure of the in-sample fit for the different models, along with the average out-of-sample QLIKE loss, which is standardized by the loss of the HAR model. The persistence of the models are estimated as $(\beta_d + \beta_w + \beta_m)$, while the mean lag is estimated as $\sum_{i=1}^{22} ib_i / \sum_{i=1}^{22} b_i$, and b_i represents the value at each lag; i.e. $b_1 = \beta_d + \beta_w/5 + \beta_m/22$, and $b_6 = \beta_m/22$.

Table 3.2: Summary Statistics

		Panel A: RV_t					Panel B: $\lvert RV_t - RV_{t-22,t}\rvert$					Panel C: $(RV_t - RV_{t-22,t})^+$					Panel D: $(RV_t - RV_{t-22,t})^-$				
Company	Ticker	Obs	Mean	St. Dev.	Min	Max	obs	Mean	St. Dev.	Min	Max	Obs	Mean	St. Dev.	Min	Max	Obs	Mean	St. Dev.	Min	Max
S&P 500 ETF	SPY	2767	1.325	2.678	0.041	59.863	2746	0.669	1.862	0.000	49.432	1066	0.853	2.649	0.000	49.432	1680	-0.553	1.087	-15.049	0.000
American Express	AXP	2767	5.497	11.125	0.088	299.968	2746	2.757	7.660	0.000	250.039	1046	3.601	11.087	0.000	250.039	1700	-2.238	4.299	-47.725	0.000
DuPont	DD	2767	3.495	4.779	0.100	83.487	2746	1.510	3.053	0.000	66.703	1054	1.936	4.323	0.000	66.703	1692	-1.245	1.820	-22.864	0.000
Disney	DIS	2767	3.907	5.920	0.221	145.706	2746	1.726	4.239	0.000	131.016	1020	2.296	6.376	0.000	131.016	1726	-1.390	2.067	-21.554	-0.001
Home Depot	HD	2767	4.129	5.644	0.201	103.477	2746	1.825	3.553	0.000	82.872	1060	2.334	4.950	0.004	82.872	1686	-1.504	2.214	-24.340	0.000
IBM	IBM	2767	2.597	4.200	0.102	71.293	2746	1.237	2.767	0.000	54.149	1078	1.547	3.874	0.002	54.149	1668	-1.036	1.676	-17.889	0.000
Coca-Cola	KO	2767	1.988	2.994	0.046	58.808	2746	0.856	2.022	0.000	50.959	1035	1.112	2.960	0.000	50.959	1711	-0.700	1.095	-12.800	-0.001
McDonald's	MCD	2767	2.917	5.164	0.087	161.156	2746	1.408	4.283	0.000	150.089	1017	1.864	6.709	0.000	150.089	1729	-1.140	1.576	-17.221	0.000
3M	MMM	2767	2.378	3.809	0.082	91.955	2746	1.101	2.654	0.000	79.198	1066	1.400	3.822	0.001	79.198	1680	-0.912	1.470	-20.425	0.000
Merck	MRK	2767	3.151	6.341	0.137	223.255	2746	1.714	5.311	0.000	208.576	978	2.385	8.432	0.000	208.576	1768	-1.343	2.030	-21.198	-0.001
Microsoft	MSFT	2767	3.336	4.478	0.083	62.386	2746	1.449	2.775	0.000	48.735	1111	1.756	3.713	0.000	48.735	1635	-1.240	1.861	-19.453	0.000
Procter & Gamble	PG	2767	1.862	2.997	0.105	70.015	2746	0.900	2.121	0.000	60.119	1032	1.139	3.084	0.000	60.119	1714	-0.756	1.196	-13.880	-0.001
Exxon Mobil	XOM	2767	2.452	4.754	0.155	141.130	2746	1.171	3.470	0.000	119.346	1071	1.483	4.928	0.004	119.346	1675	-0.972	2.029	-29.086	0.000

Note: The table reports the summary statistics for the RV of all the stocks and SPY, along with the summary statistics for the absolute difference between the daily and monthly RV, and the summary statistics for the positive, $(RV_t - RV_{t-22,t})^+$, and negative, $(RV_t - RV_{t-22,t})^-$, difference between the daily and monthly RV.

Table 3.3: In-sample Fit SPY and Stock Average

	HAR	HAR-Q	DBC-HAR	HAR	HAR-Q	DBC-HAR	HAR	HAR-Q	DBC-HAR
		$h=1$			$h=5$			$h=22$	
β_0	0.123	−0.015	0.006	0.191	0.107	0.140	0.383	0.320	0.343
s.e.	(0.073)	(0.077)	(0.072)	(0.078)	(0.079)	(0.083)	(0.084)	(0.082)	(0.084)
β_d	0.226	0.666	0.592	0.177	0.443	0.335	0.102	0.300	0.225
s.e.	(0.105)	(0.106)	(0.087)	(0.055)	(0.122)	(0.098)	(0.022)	(0.090)	(0.082)
β_w	0.452	0.306	0.334	0.372	0.284	0.321	0.353	0.287	0.314
s.e.	(0.155)	(0.124)	(0.128)	(0.112)	(0.109)	(0.109)	(0.119)	(0.098)	(0.099)
β_m	0.227	0.104	0.119	0.304	0.229	0.257	0.251	0.196	0.215
s.e.	(0.107)	(0.123)	(0.122)	(0.108)	(0.118)	(0.117)	(0.099)	(0.112)	(0.115)
β_{dQ}		−0.007			−0.004			−0.003	
s.e.		(0.002)			(0.001)			(0.001)	
α			−0.013			−0.006			−0.004
s.e.			(0.002)			(0.002)			(0.002)
Mean Lag	4.633	2.583	2.831	5.615	4.113	4.659	5.736	4.359	4.826
Persistence	0.905	1.075	1.044	0.853	0.956	0.913	0.706	0.783	0.753
R^2_{adj}	0.508	0.553⋆	0.544⋆	0.631	0.655⋆	0.641⋆	0.554	0.571⋆	0.561⋆
Average Stocks									
Mean Lag	5.983	4.525	4.421	6.117	4.741	4.685	6.391	5.051	5.042
Persistence	0.887	0.986	0.988	0.837	0.907	0.897	0.707	0.768	0.750
$\bar{R}^2_{Adj}$	0.465	0.492	0.494	0.596	0.620	0.622	0.545	0.562	0.565

Note: The table reports the in-sample parameter estimates and measures of fit for the HAR-RV, HAR-Q and DBC-HAR models at the $h = 1$ (day), $h = 5$ (week), and $h = 22$ (monthly) horizons using 5 minutes returns. The first panel reports parameter estimates for the SPY with robust standard errors in parentheses, along with the mean lag, persistence and adjusted R^2's. The bottom panel summarizes the measure of fit, mean lag and persistence level as an average across all of the individual stocks. Starred values indicate that the in-sample losses of the models are significantly smaller relative to the losses of the HAR model based on the Rivers and Vuong (2002) test at the 5% confidence level.

Table 3.4: Out-of-sample Relative QLIKE and MCS p-values

		HAR		HAR-Q		DBC-HAR	
		QLIKE	MCS	QLIKE	MCS	QLIKE	MCS
		Panel A: 2 minutes Sampling Frequency					
SPY	$h = 1$	1.000	0.309*	1.088	0.071	0.958	1.000*
	$h = 5$	1.000	0.951*	0.997	1.000*	1.016	0.875*
	$h = 22$	1.000	0.021	0.967	0.570*	0.957	1.000*
Avg. Stocks	$h = 1$	1.000	5 [0, 1]	0.962	8 [0, 1]	0.931	12 [0.3, 1]
	$h = 5$	1.000	6 [0, 0.7]	0.922	12 [0.5, 1]	0.924	12 [0.2, 1]
	$h = 22$	1.000	8 [0, 0.3]	0.956	10 [0, 1]	0.963	11 [0, 1]
		Panel B: 5 minutes Sampling Frequency					
SPY	$h = 1$	1.000	0.046	1.018	0.000	0.942	1.000*
	$h = 5$	1.000	0.084	0.956	0.114*	0.928	1.000*
	$h = 22$	1.000	0.000	0.959	0.415*	0.939	1.000*
Avg. Stocks	$h = 1$	1.000	7 [0, 1]	0.973	9 [0, 1]	0.954	12 [0.6, 1]
	$h = 5$	1.000	5 [0, 1]	0.948	12 [0.3, 1]	0.961	11 [0, 1]
	$h = 22$	1.000	9 [0, 0.7]	0.953	11 [0.1, 1]	0.962	12 [0.3, 1]
		Panel C: 10 minutes Sampling Frequency					
SPY	$h = 1$	1.000	0.363*	1.131	0.067	0.965	1.000*
	$h = 5$	1.000	0.006	0.974	0.255*	0.942	1.000*
	$h = 22$	1.000	0.037	0.948	1.000*	0.965	0.510*
Avg. Stocks	$h = 1$	1.000	7 [0, 1]	0.999	7 [0, 1]	0.971	12 [0.1, 1]
	$h = 5$	1.000	7 [0, 1]	0.964	12 [0.1, 1]	0.950	12 [0.3, 1]
	$h = 22$	1.000	5 [0, 0.5]	0.945	12 [0.4, 1]	0.969	10 [0, 1]

Note: This table reports the out-of-sample QLIKE losses and the p-values of the Model Confidence Set (MCS). The losses are standardized by the loss of the HAR model. Each panel is split in two sub-panels. The top sub-panel reports the relative loss and the MCS's p-value. The starred values indicates the model(s) that are retained by the MCS. the bottom sub-panel reports the average relative loss across all of the individual stocks and the number of stocks for which each model has been retained by the MCS (in squares brackets we report the lowest/highest pvalue). The significance level of the MCS is set to $z_\alpha = 10\%$ and use 10,000 bootstrap resamples (with a block length of 20 days).

Table 3.5: Out-of-sample Results – Regimes

		HAR		HAR-Q		DBC-HAR		HAR		HAR-Q		DBC-HAR	
		QLIKE	MCS	QLIKE	MCS	QLIKE	MCS	QLIKE	MCS	QLIKE	MCS	QLIKE	MCS
		Panel A: Pre-Crisis Period						Panel B: Crisis Period					
		Sub-panel A.1: 2 minutes Sampling Frequency						Sub-panel B.1: 2 minutes Sampling Frequency					
SPY	$h=1$	1.000	0.230⋆	0.969	0.966⋆	0.969	1.000⋆	1.000	0.603⋆	1.119	0.603⋆	0.902	1.000⋆
	$h=5$	1.000	0.007	0.960	0.014	0.883	1.000⋆	1.000	0.006	0.928	1.000⋆	0.943	0.682⋆
	$h=22$	1.000	0.002	0.930	0.112⋆	0.864	1.000⋆	1.000	0.756⋆	0.988	1.000⋆	1.023	0.411⋆
Avg. Stocks	$h=1$	1.000	5 [0, 1]	0.960	12 [0.1, 1]	0.963	10 [0, 1]	1.000	2 [0, 1]	0.914	11 [0.03, 1]	0.882	9 [0.01, 1]
	$h=5$	1.000	1 [0, 0.25]	0.908	10 [0.02, 1]	0.914	10 [0, 1]	1.000	4 [0, 0.9]	0.860	12 [0.6, 1]	0.928	8 [0, 1]
	$h=22$	1.000	2 [0, 0.3]	0.913	9 [0, 1]	0.912	10 [0, 1]	1.000	8 [0.05, 1]	0.958	12 [0.2, 1]	0.983	9 [0.05, 1]
		Sub-panel A.2: 5 minutes Sampling Frequency						Sub-panel B.2: 5 minutes Sampling Frequency					
SPY	$h=1$	1.000	0.138⋆	0.977	0.305⋆	0.960	1.000⋆	1.000	0.072	0.881	0.972⋆	0.879	1.000⋆
	$h=5$	1.000	0.001	0.942	0.002	0.848	1.000⋆	1.000	0.827⋆	0.957	1.000⋆	0.975	0.827⋆
	$h=22$	1.000	0.000	0.849	0.269⋆	0.826	1.000⋆	1.000	1.000⋆	1.020	0.609⋆	1.030	0.528⋆
Avg. Stocks	$h=1$	1.000	6 [0, 1]	0.965	12 [0.25, 1]	0.966	11 [0.04, 1]	1.000	7 [0, 1]	0.965	11 [0.01, 1]	0.950	12 [0.13, 1]
	$h=5$	1.000	2 [0,0.4]	0.915	11 [0.03, 1]	0.922	10 [0, 1]	1.000	8 [0, 1]	0.918	12 [0.3, 1]	0.980	10 [0.04, 1]
	$h=22$	1.000	1 [0, 0.3]	0.911	9 [0, 1]	0.917	8 [0, 1]	1.000	12 [0.2, 1]	0.987	12 [0.2, 1]	1.015	11 [0.09, 1]
		Sub-panel A.3: 10 minutes Sampling Frequency						Sub-panel B.3: 10 minutes Sampling Frequency					
SPY	$h=1$	1.000	0.034	0.955	0.968⋆	0.960	1.000⋆	1.000	0.848⋆	1.021	0.848⋆	0.966	1.000⋆
	$h=5$	1.000	0.002	0.898	0.121⋆	0.848	1.000⋆	1.000	0.043	0.912	1.000⋆	0.915	0.967⋆
	$h=22$	1.000	0.000	0.830	0.616⋆	0.826	1.000⋆	1.000	0.059	0.929	1.000⋆	1.097	0.059
Avg. Stocks	$h=1$	1.000	6 [0, 1]	0.965	12 [0.22, 1]	0.966	11 [0.02, 1]	1.000	10 [0, 1]	1.050	9 [0.2, 1]	1.001	12 [0.12, 1]
	$h=5$	1.000	2 [0, 0.7]	0.922	11 [0.02, 1]	0.933	9 [0, 1]	1.000	6 [0, 1]	0.937	12 [0.7, 1]	0.991	10[0, 0.7]
	$h=22$	1.000	0 [0, 0.08]	0.906	9 [0.01, 1]	0.926	8 [0, 1]	1.000	10 [0.03, 1]	0.967	12 [0.6, 1]	0.997	9 [0.06, 8]

Note: The table report the out-of-sample standardized QLIKE losses and p-values of the Model Confidence Set (MCS). Each panel is split in three sub-panels that report results for 3 different sampling frequencies. Starred values in the SPY indicate the retained models by the MCS, while for the average stocks the MCS column reports the number of stocks for which each model has been retained by the MCS (in square brackets we report the lowest/highest p-value). The significance level of the MCS is set to $z_\alpha = 10\%$, and use 10,000 bootstrap resamples (with a block length of 20 days).

Table 3.6: Alternative Noise-robust Realized Measures

	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
	Panel A: 1-min			Panel B: 2-min		
SPY						
HAR-PRV	1.261	1.168	1.228	1.379	1.038	**0.934**
HAR-TSRV	1.263	1.234	1.200	1.627	1.225	0.990
HAR-RV	1.062	1.077	1.065	1.062	1.077	1.065
HAR-Q	1.081	1.030	1.021	1.081	1.030	1.021
DBC-HAR	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	1.000
Individual Stocks						
HAR-PRV	1.309	1.203	1.074	1.424	1.298	1.050
HAR-TSRV	1.647	1.480	1.264	2.052	1.564	1.275
HAR-RV	1.051	1.046	1.041	1.051	1.046	1.041
HAR-Q	1.020	**0.988**	**0.991**	1.020	**0.988**	**0.991**
DBC-HAR	**1.000**	1.000	1.000	**1.000**	1.000	1.000

Note: This table reports the out-of-sample forecasts of HAR models using alternative noise-robust realized measures for the SPY and the average of all the individual stocks. The losses of the models are standardized by the loss of the DBC-HAR model, and bold numbers highlight the best model specification. The HAR-PRV and HAR-TSRV models are estimated using 1-min and 2-min on Panel A and B, respectively. On the other hand, the HAR-RV, HAR-Q, and DBC-HAR models rely on 5-min returns.

Table 3.7: Measurement Error Transformations

		$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
		2-min			5-min			10-min		
SPY	HAR	1.044	0.984	1.045	1.062	1.077	1.065	1.036	1.062	1.036
	HAR-Q	1.136	0.980	1.010	1.081	1.030	1.021	1.171	1.034	0.982
	DBC-HAR	**1.000**	1.000	1.000	**1.000**	**1.000**	**1.000**	1.000	**1.000**	**1.000**
	DBC-HAR-log	1.062	**0.966**	1.010	1.035	1.028	1.012	1.050	1.070	1.020
	DBC-HAR-sqrt	1.245	0.998	**0.989**	1.010	1.028	1.048	**0.963**	1.000	1.002
Stock Avg.	HAR	1.077	1.085	1.040	1.051	1.046	1.041	1.033	1.056	1.033
	HAR-Q	1.031	0.998	0.994	1.020	0.988	0.991	1.030	1.014	0.976
	DBC-HAR	**1.000**	1.000	**1.000**	**1.000**	1.000	**1.000**	**1.000**	**1.000**	1.000
	DBC-HAR-log	1.037	1.032	1.002	1.026	1.017	1.008	1.018	1.033	1.001
	DBC-HAR-sqrt	1.002	0.977	1.000	1.000	0.985	1.000	1.010	1.004	**0.987**

Note: This table reports the out-of-sample forecasts losses for the DBC-HAR models using alternative metrics for the measurement error. The losses are standardized by the losses of the DBC-HAR model, and we also include the HAR-Q model to facilitate the comparison. The bottom panel reports the average standardized losses for all the individual stocks under analysis. Bold numbers highlight the best DBC-HAR performance.

Table 3.8: In-Sample Fit SPY – Alternative DBC-HAR Models

	HAR-J	DBC-HAR-J	HAR-CJ	DBC-HAR-CJ	CHAR	DBC-CHAR
β_0	0.124	0.004	0.127	0.014	0.127	0.010
s.e.	(0.073)	(0.072)	(0.074)	(0.073)	(0.072)	(0.074)
β_d	0.231	0.613			0.224	0.599
s.e.	(0.110)	(0.089)			(0.111)	(0.087)
β_w	0.450	0.326			0.469	0.350
s.e.	(0.157)	(0.132)			(0.162)	(0.138)
β_m	0.228	0.118			0.227	0.111
s.e.	(0.107)	(0.123)			(0.108)	(0.118)
β_{C_d}			0.219	0.599		
s.e.			(0.112)	(0.089)		
β_{C_w}			0.484	0.365		
s.e.			(0.162)	(0.139)		
β_{C_m}			0.224	0.118		
s.e.			(0.097)	(0.111)		
β_{J_d}	−0.238	−0.517	0.278	0.159		
s.e.	(0.269)	(0.255)	(0.246)	(0.283)		
β_{J_w}			−1.125	−1.295		
s.e.			(0.641)	(0.684)		
β_{J_m}			0.412	−0.101		
s.e.			(1.779)	(1.450)		
α		−0.013		−0.013		−0.013
s.e.		(0.002)		(0.002)		(0.002)
R^2_{adj}	0.508	0.546*	0.511	0.548*	0.510	0.547*

Note: The table reports the 1-day ahead in-sample parameter estimates with robust standard error in parenthesis for the alternative DBC-HAR models and their benchmark counterpart. The last row reports the adjusted R^2 of the models and starred values indicate that in-sample losses of the alternative DBC-HAR models are significantly smaller relative to their benchmark counterpart based on the Rivers and Vuong (2002) at the 5% confidence level.

Table 3.9: DBC-HAR under Alternative HAR Models

	$h = 1$	MCS	$h = 5$	MCS	$h = 22$	MCS
HAR	1.000	0.541*	1.000	0.282*	1.000	0.034
HAR-Q	1.018	0.541*	0.956	0.282*	0.959	0.555*
DBC-HAR	**0.942**	0.548*	**0.928**	0.282*	**0.939**	0.555*
HAR-J	0.998	0.541*	0.989	0.282*	0.985	0.035
DBC-HAR-J	**0.939****	1.000*	**0.923****	1.000*	**0.920****	1.000*
HAR-CJ	1.087	0.003	1.265	0.000	0.990	0.035
DBC-HAR-CJ	1.006**	0.262*	1.194**	0.004	0.950**	0.555*
CHAR	1.078	0.004	1.161	0.004	1.199	0.000
DBC-CHAR	0.955**	0.548*	1.075**	0.282*	1.124**	0.005

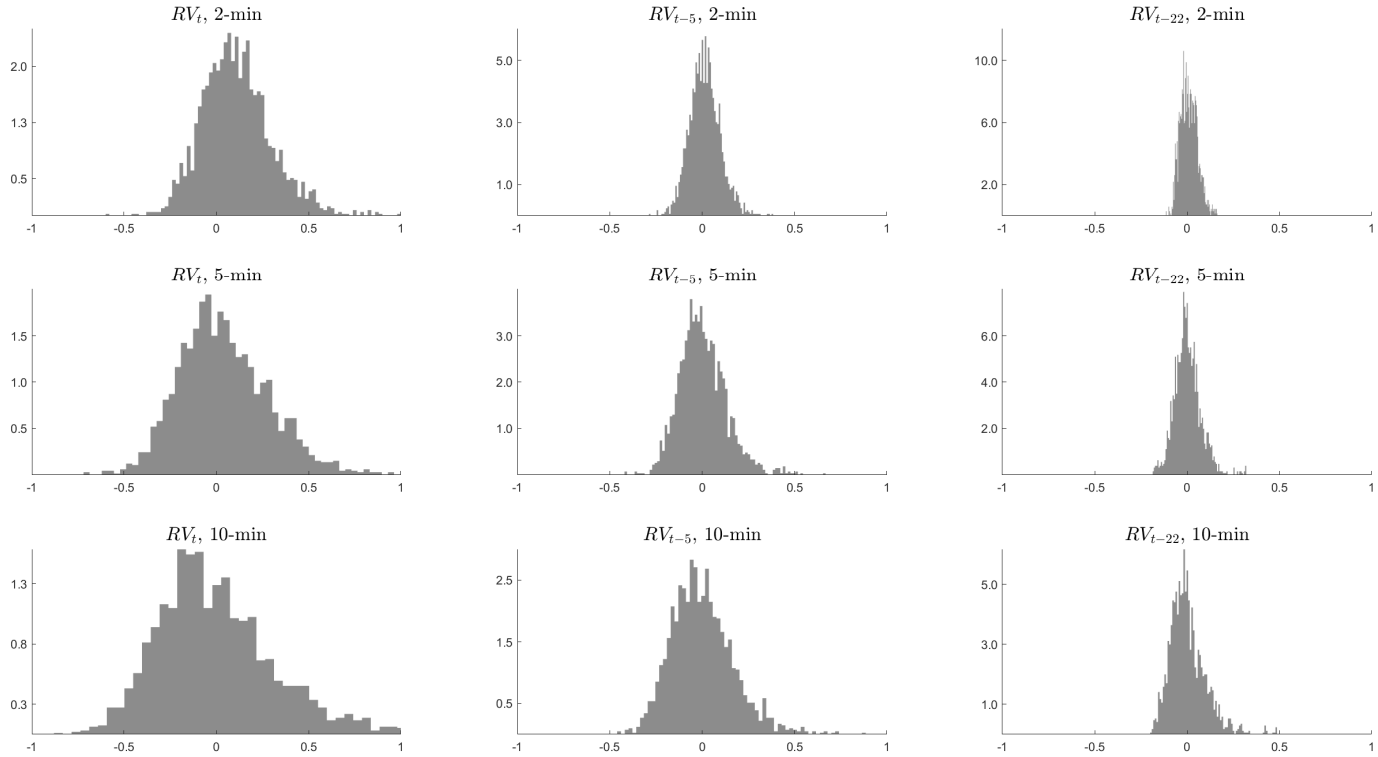
Note: The table reports the QLIKE loss function standardized by the loss of the HAR model for the SPY across forecasting horizon. The realized measures are estimated using 5 minutes returns. Bold numbers highlight the best two performances. The MCS column shows the p-value for each model, and starred values indicate the retained models by the MCS. The significance level of the MCS is set to $z_{\alpha} = 10\%$, and use 10,000 bootstrap resamples (with a block length of 20 days). Finally, double-starred values indicate that the alternative DBC-HAR models significantly outperform their counterparts based on the Giacomini and White (2006) test at the 5% confidence level.

Table 3.10: In- and Out-of-sample Performance of Alternative GARCH Models

	GARCH(1,1)	DBC-GARCH(1,1)	GJR-GARCH(1,1,1)	DBC-GJR-GARCH(1,1,1)
ω	0.018	0.019	0.018	0.016
	(0.000)	(0.002)	(0.000)	(0.000)
β	0.498	0.502	0.587	0.593
	(0.037)	(0.026)	(0.072)	(0.034)
π	0.502	0.498	0.343	0.347
	(0.027)	(0.028)	(0.069)	(0.027)
γ			0.1192	0.120
			(0.001)	(0.002)
α		-0.008		-0.008
		(0.000)		(0.000)
$\mathscr{L}$	-2127.929	-2125.264	-2098.032	-2094.858
BIC	1.551	1.549	1.530	1.528
$QLIKE$	1.000	**0.992**	1.000	**0.990**
Individual Stocks				
$\overline{\mathscr{L}}$	-4786.927	-4778.403	-4740.815	-4738.392
$\overline{BIC}$	3.488	3.482	3.455	3.453
$\overline{QLIKE}$	1.000	**0.980**	1.000	**0.984**

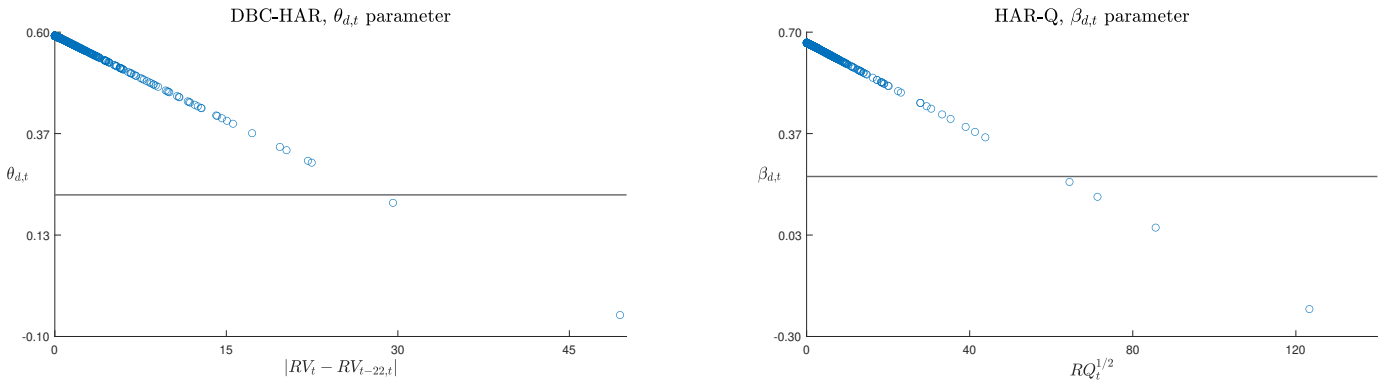
Note: The table reports the 1-day ahead in-sample estimates with robust standard errors in parenthesis, the log-likelihood, and the Bayesian Information Criterion (BIC) for both the GARCH and alternative GARCH specifications. The out-of-sample QLIKE loss is standardized by the benchmark model of each specification, and it is reported at the bottom of the top panel. The bottom panel reports the results for the individual stocks as an average across all of the stocks. $\overline{\mathscr{L}}$, $\overline{BIC}$, and $\overline{QLIKE}$ indicate the average values of the log-likelihood, Bayesian Information Criterion, and the standardized QLIKE loss.

Figure 3.1: Estimation Error of Daily, Weekly, and Monthly RV



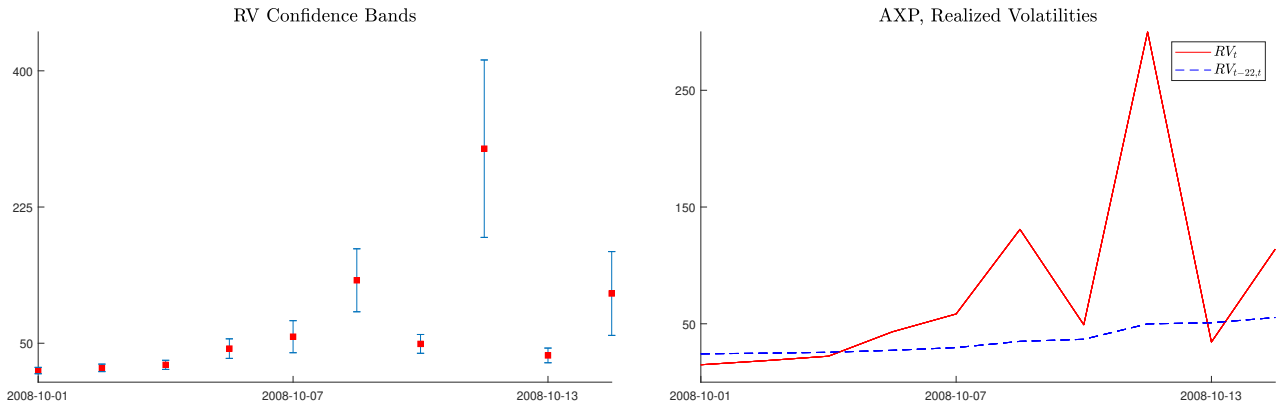
Note: This figure plots the simulated distribution $RV/IV - 1$. The left panel plots the results for daily RV, while middle and right panels show the results for weekly and monthly RV. The top, middle, and bottom panels are based on 2-min, 5-min, and 10-min returns, respectively.

Figure 3.2: DBC-HAR and HAR-Q Time-varying Parameters



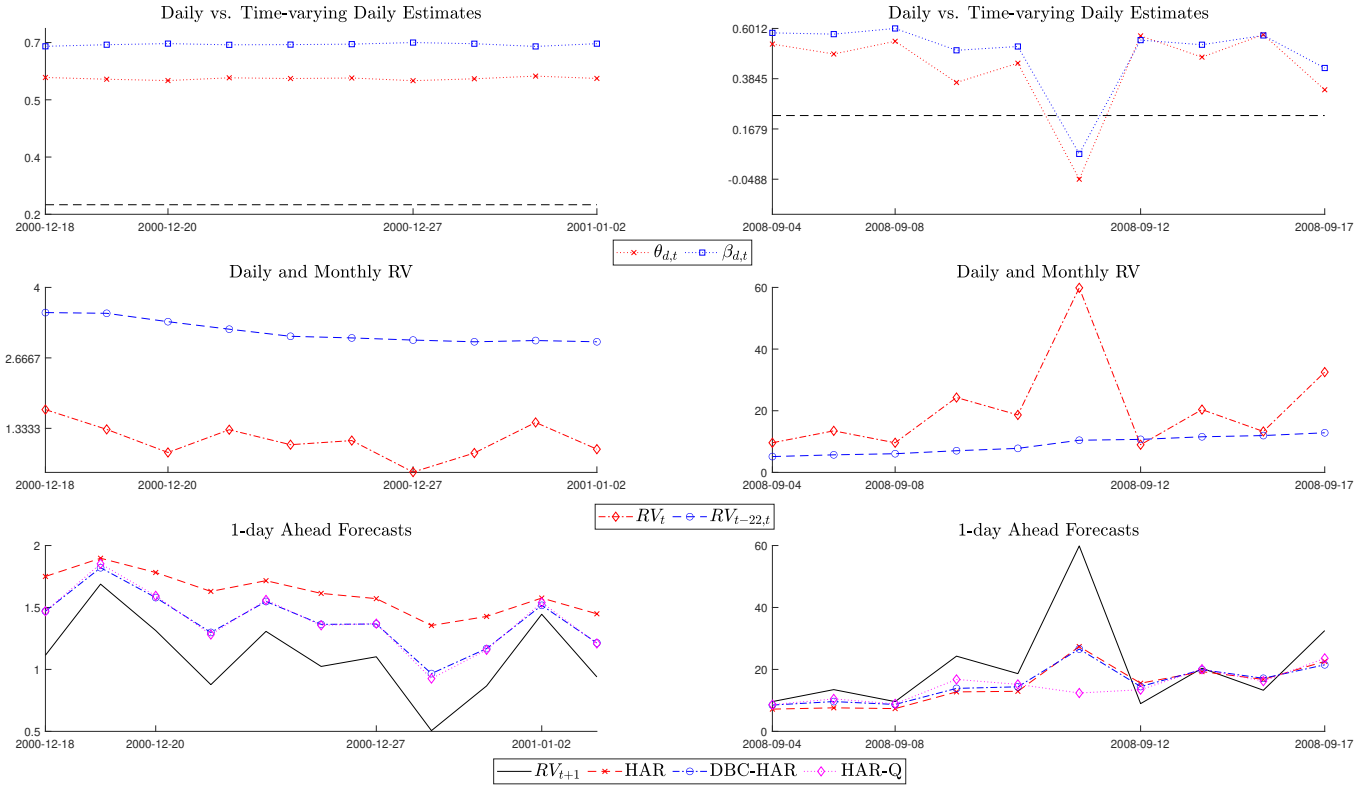
Note: This figure plots the time-varying parameter of the DBC-HAR and HAR-Q models against their respective metric of the measurement error. The black horizontal line plots the constant daily parameter of the HAR model.

Figure 3.3: 95% Confidence Bands for AXP along with Daily and Monthly RV



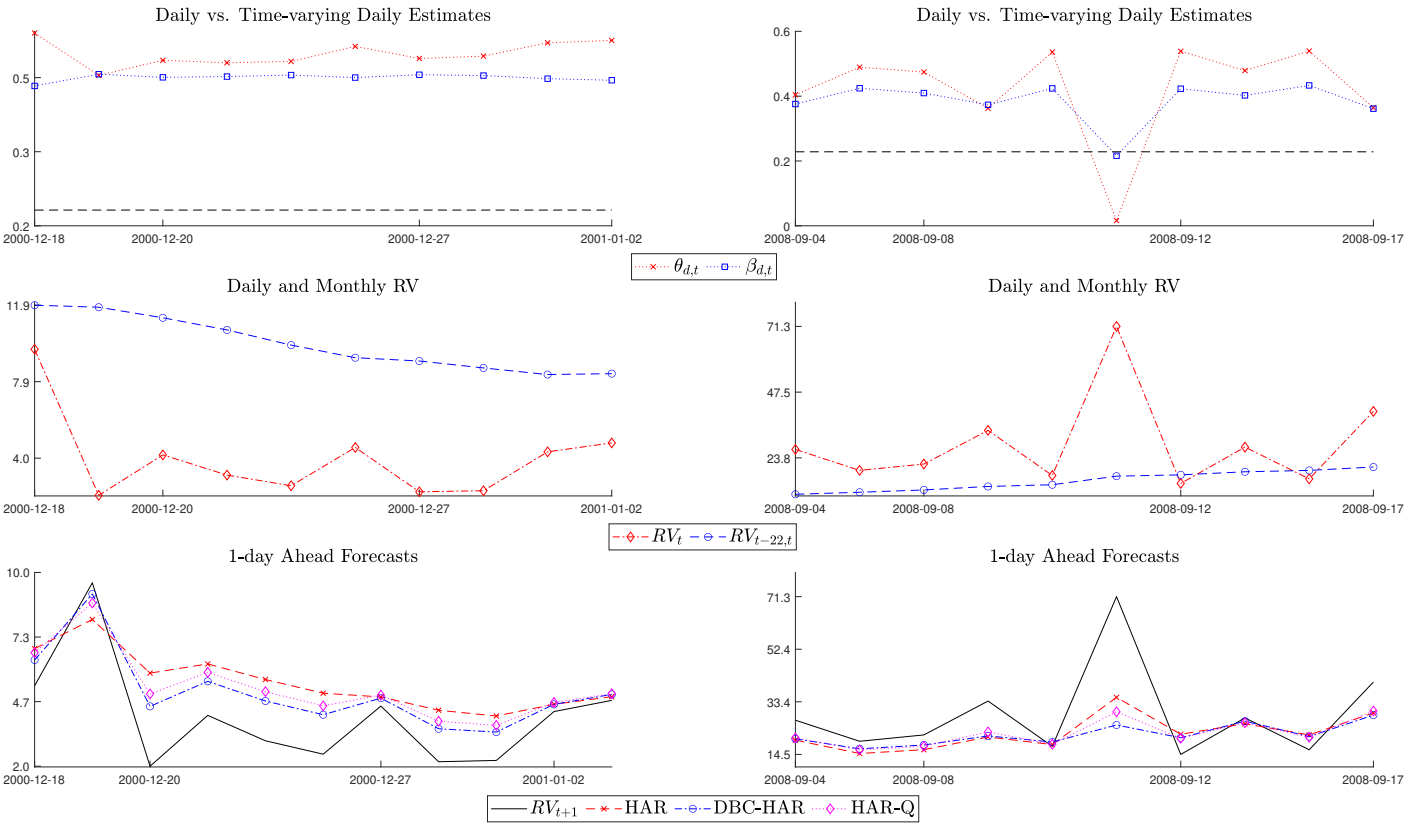
Note: This figure plots the RV of AXP using the 95% confidence bands on the left panel, while the right panel plots the daily and monthly RV. The period correspond to ten consecutive days around the day with the largest RV.

Figure 3.4: Time-Varying vs Constant Daily Estimates – SPY



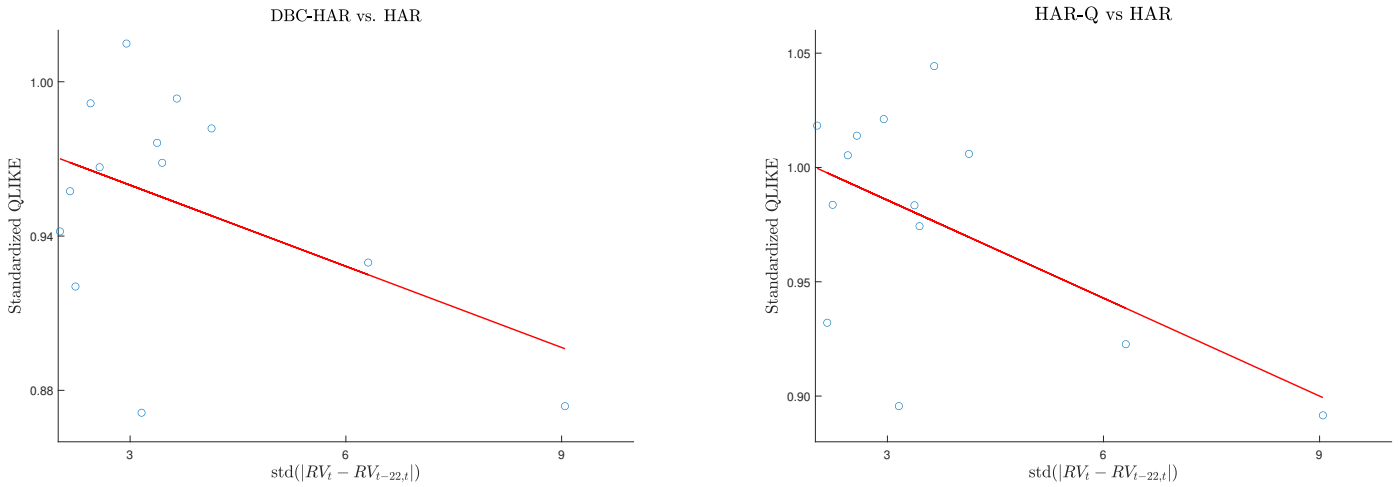
Note: This figure illustrates in the top panel the time-varying dynamics of the the DBC-HAR ($\theta_{d,t}$) and HAR-Q ($\beta_{d,t}$) daily estimates versus the constant daily estimate of the HAR model (black dashed line). The middle panel plots the daily and monthly RV, whereas the bottom panel depicts the 1-day ahead forecasts of the HAR, DBC-HAR and the HAR-Q models compared to the ex-post RV. The left (right) panels plot ten succesive trading days for the pre-crisis (crisis) period.

Figure 3.5: Time-Varying vs Constant Daily Estimates – IBM



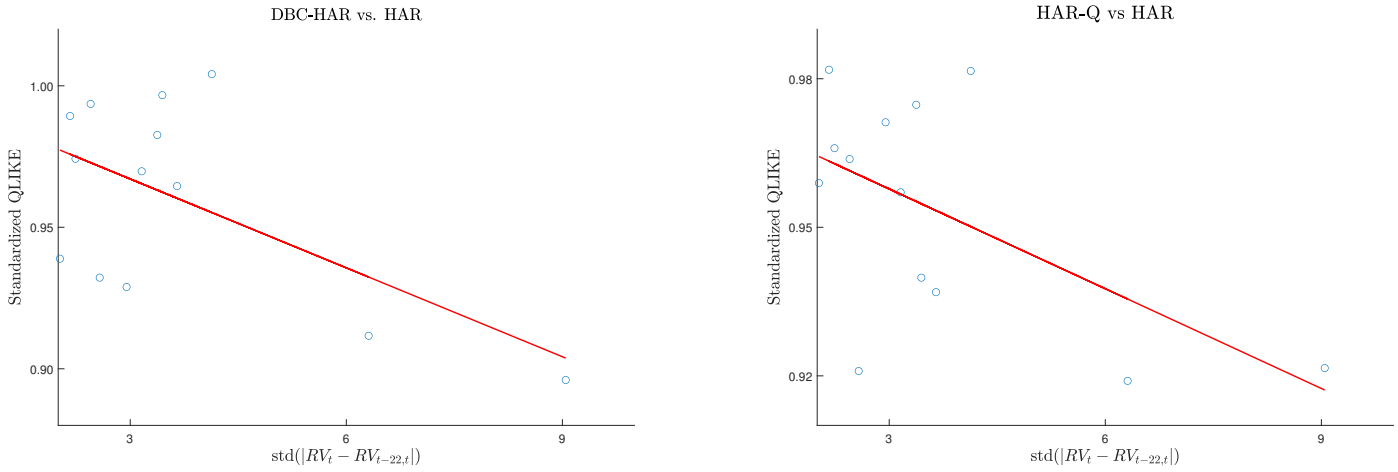
Note: This figure illustrates in the top panel the time-varying dynamics of the the DBC-HAR ($\theta_{d,t}$) and HAR-Q ($\beta_{d,t}$) daily estimates versus the constant daily estimate of the HAR model (black dashed line). The middle panel plots the daily and monthly RV, whereas the bottom panel depicts the 1-day ahead forecasts of the HAR, DBC-HAR and the HAR-Q models compared to the ex-post RV. The left (right) panels plot ten succesive trading days for the pre-crisis (crisis) period.

Figure 3.6: Individual Standardized Losses, $h = 1$



Note: This figure plots the one-day ahead standardized losses of the DBC-HAR and HAR-Q models for each individual stock and the SPY against the standard deviation of the absolute difference between the daily and monthly RV.

Figure 3.7: Individual Standardized Losses, $h = 22$



Note: This figure plots the one-month ahead standardized losses of the DBC-HAR and HAR-Q models for each individual stock and the SPY against the standard deviation of of the absolute difference between the daily and monthly RV.

Chapter 4

Exploiting the Market Factor Information to Improve (Co)Variance Forecasts and Financial Decisions

4.1 Introduction

Accurate forecasts of asset covariances play a key role in many financial economic applications, where the most notable implementation is found in portfolio allocation (Markowitz, 1952). Reducing the errors in the forecasted covariance matrix is a key factor as practitioners rely on several important properties, such as the positive definiteness and well-conditioning to guarantee reasonable portfolio weights. However, the level of idiosyncratic volatility (IdV) has substantially increased in the last 20 years (Campbell et al., 2001; Xu and Malkiel, 2003) and, as shown by Ang et al. (2009), assets with a high level of IdV suffer from low predictability. This issue can further increase the forecasting errors.

In this paper, we consider a factor-type structure for modelling and forecasting both univariate and multivariate volatility. We use the SPDR S&P 500 ETF (SPY) to represent the market factor information. The popular use of observed factors is attributable to their excellent performance and simple economic explanation (e.g. Fama and French, 1993,

1996; King, 1966). The SPY, which is traded at very high-frequency, serves as a useful high frequency market factor (e.g. Fan et al., 2016; Hasbrouck, 2003). The explanation for using the market factor is that asset prices are influenced by market and industry factors, which may not be reflected in the information set of individual asset prices. Thus, the market factor helps assets react faster to new information, especially those with low trading volume.

The use of factor models to model returns and allocate portfolios dates back to Sharpe (1964), who proposed the so-called CAPM model to avoid estimating all the parameters in a covariance matrix (e.g. Lintner, 1965; Treynor, 1962, for similar approaches). Whereas this structure is generally ignored when the object of interest relates to forecasting (co)variances, Fan et al. (2016) show significant improvements in the portfolio allocation based on covariances estimated using a similar factor structure. The findings of Fan et al. (2016), together with the empirical evidence that assets usually show relatively low levels of correlation during bull and tranquil market periods (Longin and Solnik, 2001), suggest that the systematic information is not fully absorbed by the information set of individual asset prices, and therefore specifications incorporating said factor should render more accurate estimates.

The limited multivariate literature, that incorporates factors or exogenous variables for modelling and forecasting purposes, includes Bauer and Vorkink (2011) who forecast realized covariances with estimates of latent factors obtained from a pool of realized continuous and jump measures.[1] Unlike the multivariate literature, the univariate literature has advocated the use of exogenous variables to a much larger extent,[2] where implied volatility (ImpV) is very close to our analysis. Although adding ImpV as an explanatory variable generally improves the fitting of the model, this is largely irrelevant to the out-of-sample prediction (e.g. Busch et al., 2011; Giot and Laurent, 2007). This is because ImpV is shown to be efficient but biased, and therefore leading to poor out-of-sample

[1]The use of principal components analysis to extract latent factors in high-dimensional covariances has also been studied (e.g. Bai and Shi, 2011; Calzolari et al., 2020; Fan et al., 2008, and the list of references therein). Although this technique has proved to be very successful, its obscure our understanding of the relevant variables and difficult the economic interpretation.

[2]Previous literature has also employed macroeconomic factors (e.g. Christiansen et al., 2012; Engle et al., 2007; Paye, 2012) however, the results are not very promising in the out-of-sample exercise.

performance (e.g. Blair et al., 2000; Canina and Figlewski, 1993; Jorion, 1995).

We rely on high-frequency data to construct the latent volatility and covariance measures. Even though high-frequency intraday data allows for the construction of more accurate realized (co)variance measures than lower frequency data (e.g., daily), they are still estimates and as such subject to measurement error (e.g. Andersen et al., 2005). To account for this issue, we incorporate the market factor information into the univariate heterogeneous autoregressive (HAR) model of Corsi (2009) and the HARQ model of Bollerslev et al. (2016). We dubbed these specifications, the HAR-M and HARQ-M models, where "M" stands for the market factor. In a similar fashion, the market factor is added to the multivariate extensions of the HAR (Chiriac and Voev, 2011) and HARQ models (Bollerslev et al., 2018). Whereas these proposed multivariate specifications consider scalar parameters, by assuming that all the unique elements of the covariance matrix have the same dynamics, we allow for different dynamics and therefore model each unique element of the covariance matrix separately.[3]

Our empirical analysis considers 30 S&P 500 constituents for the period 2000–2016. We select a very balanced portfolio of assets with heterogeneous levels of liquidity, proportion of IdV and risk exposure. We consider both the univariate and multivariate setting. For the multivariate setting, we construct 100 random portfolios based on $N = 5$ and $N = 10$ assets, and evaluate their performance using statistical loss functions and global minimum variance portfolios with and without transaction costs.

The main results can be summarized as follows: We show that idiosyncratic volatility comprises on average 71% of the total realized variance, which suggests that most of the price variation is related to idiosyncratic factors. Given that assets with a high level of IdV suffer from low predictability, we incorporate the market factor information in the univariate and multivariate specifications. Results indicate a significant improvement in the in- and out-of-sample performance of both univariate and multivariate models. Although all the "M" specifications improve upon their standard benchmarks, irrespective

[3]Allowing for different dynamics yields to forecasts which may not always be positive definite, so in this paper we correct this issue by using an eigenvalue decomposition as in Hautsch et al. (2015). Please note that less than 2% of the forecasted covariance matrices experience this problem when $N = 10$.

of the forecasting horizon, the HARQ-M model is consistently the best specification. This result is expected as this model accounts for both the market factor information and the heteroskedastic measurement error. Using both simulated and empirical data, we show that the statistical improvements afforded by the models incorporating the market factor directly translates in economic gains. These gains stem from two different sources: the forecasting accuracy and the reduced turnover. Thus, a risk-averse investor is willing to pay up to 157 annual basis points to use one of our strategies and so obtain more accurate financial decisions.

The rest of paper is structured as follows: Section 4.2 sets up the theoretical background and describes the forecasting models and evaluation criteria. Section 4.3 presents the simulation study along with the simulation results. Section 4.4 describes the data along with reporting the proportion of idiosyncratic volatility and realized betas, and reports the empirical results of both the univariate and multivariate models. The global minimum variance results are also reported in this section. Section 4.5 concludes.

4.2 Theoretical Background

We consider a factor-type log-price process, defined on some filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P}\right)$, evolving continuously over time:

$$Y_t = \beta X_t + Z_t, \tag{4.1}$$

where Y_t is an N-dimensional vector of log-prices, X_t is the observable market factor process, Z_t is the idiosyncratic component, and β is an $N \times 1$ vector of constant factor loadings.[4] Following Fan et al. (2016), we assume that the dynamics of the observable

[4]The constant β assumption is very common in the literature (e.g. Aït-Sahalia and Xiu, 2017; Bai and Shi, 2011; Fan et al., 2016), and Reiß et al. (2015) find evidence supportive of this assumption using high-frequency data.

factor and idiosyncratic component follow a continuous Itô semimartingale, that is:

$$X_t = \int_0^t a_s ds + \int_0^t \eta_s dW_s, \tag{4.2}$$

$$Z_t = \int_0^t b_s ds + \int_0^t \gamma_s dB_s, \tag{4.3}$$

where W_s and B_s are respectively a one- and an N-dimensional vector of independent Brownian motions. a_s and b_s are locally bounded predictable drift processes, η_s and γ_s are the spot volatilities, which are adapted and càdlàg. Both b_s and γ_s are N-dimensional vectors. Without loss of generality, we set t equal to 1, which represents one trading day. Therefore, the covariance of Y_t has a factor structure of the form:

$$\int_0^t c_s ds = \beta \left(\int_0^t \eta_s' \eta_s ds \right) \beta' + \int_0^t \gamma_s' \gamma_s ds. \tag{4.4}$$

The first term in the right-hand side of equation (4.4) is the systematic covariance, which allows for common dynamics among assets. The second term is the idiosyncratic covariance which has a diagonal structure, meaning that idiosyncratic information is firm-specific.

The use of observed factors have an established history in empirical applications. This is because of their excellent performance and incredibly simple estimation (e.g. Fama and French, 1993, 1996; Fan et al., 2016; King, 1966, among others). Although several factors can be constructed and employed, we rely on the market factor information to capture the systematic components that are not fully absorbed by the individual asset prices, as evidenced by their high levels of idiosyncratic volatility (e.g. Ang et al., 2009). For instance, illiquid or low volume assets usually have a delay in their reaction to new market information, however, this issue is overcome by the inclusion of a market factor that is more sensitive to economic and financial news. Therefore, we use the SPDR S&P 500 ETF (SPY) as the observable high frequency market factor.

Given that the variance and covariance are latent processes, in this paper we rely on high-frequency data to consistently estimate the realized variance and covariances

(Andersen et al., 2001b, 2003; Barndorff-Nielsen and Shephard, 2004a) as follows:

$$\widehat{s}_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n Y\right)^2, \quad \widehat{S}_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n Y\right)' \left(\Delta_j^n Y\right), \tag{4.5}$$

where $\Delta_j^n Y = \left(\Delta_j^n Y^{(i)}\right)_{i=1,\ldots,N}$, and $\Delta_j^n Y^{(i)} = Y^{(i)}_{j\Delta_n} - Y^{(i)}_{(j-1)\Delta_n}$, $j = 1, \ldots, n$, $\Delta_n = 1/n$ is the sampling intervals, and n is the number of high-frequency increments per day. Of course, when $N = 1$ the realized covariance estimator reduces to the realized variance. In the sequel, we will refer to the realized variance of the stocks using lower case, $\widehat{s}_t$, and to the realized covariance matrix using upper case, $\widehat{S}_t$. The realized variance of the index is estimated using the same approach, and we refer to it as $\widehat{h}_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n X\right)^2$, where $\Delta_j^n X = X_{j\Delta_n} - X_{(j-1)\Delta_n}$.

4.2.1 Forecasting Models and Evaluation Criteria

The most popular method of modelling the quantities outlined in equation (4.5) is the heterogeneous autoregressive (HAR) model of Corsi (2009), which has emerged as the standard univariate benchmark for reduced form realized volatility modelling:

$$\widehat{s}_t = \beta_0 + \beta_d \widehat{s}_{t-1} + \beta_w \widehat{s}_{t-5|t-1} + \beta_m \widehat{s}_{t-22|t-1} + \epsilon_t, \tag{4.6}$$

where $\widehat{s}_{t-h|t-1} = \frac{1}{h}\sum_{i=1}^{h} \widehat{s}_{t-i}$. The most popular extension of the HAR model, is the HAR-Q model proposed by Bollerslev et al. (2016), which accounts for the measurement error observed in realized volatility measures:

$$\widehat{s}_t = \beta_0 + \left(\beta_d + \beta_{d_\pi} \pi_{t-1}^{1/2}\right) \widehat{s}_{t-1} + \beta_w \widehat{s}_{t-5|t-1} + \beta_m \widehat{s}_{t-22|t-1} + \epsilon_t, \tag{4.7}$$

where $\pi_t = \frac{n}{3}\sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n Y\right)^4$ is a consistent estimator of the integrated quarticity as shown in Barndorff-Nielsen and Shephard (2002a).

Given the current evidence that idiosyncratic volatility has increased substantially (e.g. Campbell et al., 2001; Xu and Malkiel, 2003), and that generally stocks with high idiosyncratic volatility suffer from low predictability (e.g. Ang et al., 2009), we propose

simple extensions of the HAR and HARQ models that introduce the daily information of the market factor, giving rise to the HAR-M and HARQ-M models. The inclusion of the market factor contributes with additional sectorial and systematic information, which may not be fully absorbed by the individual asset prices.[5] The HAR-M and HARQ-M models are outlined as:[6]

$$\widehat{s}_t = \beta_0 + \beta_d \widehat{s}_{t-1} + \beta_w \widehat{s}_{t-5|t-1} + \beta_m \widehat{s}_{t-22|t-1} + \beta_h \widehat{h}_{t-1} + \epsilon_t, \tag{4.8}$$

$$\widehat{s}_t = \beta_0 + \left(\beta_d + \beta_{d_\pi} \pi_{t-1}^{1/2}\right) \widehat{s}_{t-1} + \beta_w \widehat{s}_{t-5|t-1} + \beta_m \widehat{s}_{t-22|t-1} + \beta_h \widehat{h}_{t-1} + \epsilon_t. \tag{4.9}$$

Chiriac and Voev (2011) and Bollerslev et al. (2018) have proposed multivariate extensions of the HAR and HARQ models. Both adaptations, consider a simple scalar specification for the vectorized form of the covariance matrix. In this paper, we relax the assumption that all the elements in the covariance matrix share the same dynamics, and instead consider the case in which each variance-covariance term has its own dynamics.[7] Let $\widehat{\mathcal{S}}_t \equiv \text{vech}\left(\widehat{S}_t\right)$, then the vech-HAR and vech-HARQ models are:[8]

$$\widehat{\mathcal{S}}_t = \theta_0 + \theta_d \widehat{\mathcal{S}}_{t-1} + \theta_w \widehat{\mathcal{S}}_{t-5|t-1} + \theta_m \widehat{\mathcal{S}}_{t-22|t-1} + \epsilon_t, \tag{4.10}$$

$$\widehat{\mathcal{S}}_t = \theta_0 + \left(\theta_d + \theta_{d_\Gamma} \Gamma_{t-1}\right) \widehat{\mathcal{S}}_{t-1} + \theta_w \widehat{\mathcal{S}}_{t-5|t-1} + \theta_m \widehat{\mathcal{S}}_{t-22|t-1} + \epsilon_t, \tag{4.11}$$

where $\Gamma_t \equiv \sqrt{\text{diag}\left(\Pi_t\right)}$, and Π_t is a feasible estimator of the asymptotic variance of $\widehat{S}_t$ proposed by Barndorff-Nielsen and Shephard (2004a). Specifically, we define $\Delta_j^n U \equiv$

[5] As mentioned in the introduction, previous literature has advocated the use of implied volatility as a predictor for future stock volatility. However, the evidence is mixed and suggests that even though the implied volatility is an efficient estimator of the future volatility, it is biased and therefore contains little predictive information content.

[6] We also tried correcting for the measurement error in the market factor. The results show that the magnitude of the π parameter for the market factor is negligible and its presence does not improve neither the in- nor the out-of-sample performance.

[7] Bauer and Vorkink (2011) use a similar approach, to forecast realized covariance matrices, that rely on latent factors which are functions of lagged volatility, lagged returns, and other forecasting variables.

[8] Following Bollerslev et al. (2018), for horizons $h = 5$ and $h = 22$, the correction is respectively applied to the weekly and monthly estimates.

vech $\left(\Delta_j^n Y' \Delta_j^n Y\right)$:

$$\Pi_t = n \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n U\right)' \left(\Delta_j^n U\right) - \frac{n}{2} \sum_{j=2}^{\lfloor 1/\Delta_n \rfloor} \left(\left(\Delta_{j-1}^n U\right)' \left(\Delta_j^n U\right) + \left(\Delta_j^n U\right)' \left(\Delta_{j-1}^n U\right) \right). \quad (4.12)$$

The multivariate HAR-M and HARQ-M models are therefore outlined as:

$$\widehat{\mathcal{S}}_t = \theta_0 + \theta_d \widehat{\mathcal{S}}_{t-1} + \theta_w \widehat{\mathcal{S}}_{t-5|t-1} + \theta_m \widehat{\mathcal{S}}_{t-22|t-1} + \theta_h \widehat{h}_{t-1} + \epsilon_t, \quad (4.13)$$

$$\widehat{\mathcal{S}}_t = \theta_0 + \left(\theta_d + \theta_{d_\Gamma} \Gamma_{t-1}\right) \widehat{\mathcal{S}}_{t-1} + \theta_w \widehat{\mathcal{S}}_{t-5|t-1} + \theta_m \widehat{\mathcal{S}}_{t-22|t-1} + \theta_h \widehat{h}_{t-1} + \epsilon_t. \quad (4.14)$$

The forecasts are evaluated using the Frobenius norm, which results in the mean-square error (MSE) for univariate forecasts, and the QLIKE loss function:

$$Frob(H_t, \widehat{S}_t) = \mathrm{Tr}\left[\left(H_t - \widehat{S}_t\right)' \left(H_t - \widehat{S}_t\right) \right], \quad (4.15)$$

$$QLIKE(H_t, \widehat{S}_t) = \mathrm{Tr}\left(H_t^{-1} \widehat{S}_t\right) - \log\left(\left|H_t^{-1} \widehat{S}_t\right|\right) - N, \quad (4.16)$$

where $\widehat{S}_t$ is the ex-post (co)variance proxy, and H_t denote the fitted (co)variance forecasts. As shown by Laurent et al. (2013) and Patton (2011b), the ranking produced by both loss functions based on covariance proxies is consistent with those based on the true latent (co)variance process.

As shown by Callot et al. (2017); Hautsch et al. (2015) the multivariate forecasts may render non-positive definite matrices. To overcome this issue, we follow Hautsch et al. (2015) and apply eigenvalue cleaning to every matrix that has eigenvalues smaller than or equal to 0. The spectral decomposition of the matrix $H_t = V_t' \Lambda_t V_t$, where V_t is the matrix of eigenvectors, and Λ_t is a diagonal matrix with the N eigenvalues $\lambda_{i,t}$ on its diagonal. Let $\lambda_{mp,t} = \min\{\lambda_{i,t} | \lambda_{i,t} > 0\}$, replace all the $\lambda_{i,t} < \lambda_{mp,t}$, and define $\tilde{\Lambda}_t$ as the diagonal matrix with the cleaned eigenvalues on its diagonal. Therefore, the regularized forecast matrix $\tilde{H}_t = V_t' \tilde{\Lambda}_t V_t$ is by construction positive definite.

4.3 Simulation Study

Using a factor-type structure based on the one-factor stochastic volatility model as in Huang and Tauchen (2005), we simulate a d-dimensional process for $dY_{i,t}$, $i = 1, \ldots, 5$ as:

$$\begin{aligned} dY_{i,t} &= \beta_i dX_t + dZ_{i,t}, \\ dX_t &= bdt + \exp\{\kappa \upsilon_t\} dW_t, \\ d\upsilon_t^2 &= \alpha_\upsilon \upsilon_t^2 dt + dW_t^\upsilon, \\ dZ_{i,t} &= \exp\{\gamma \nu_{i,t}\} dB_{i,t}, \\ d\nu_{i,t}^2 &= \alpha_\nu \nu_{i,t}^2 dt + dB_{i,t}^\nu, \end{aligned} \tag{4.17}$$

Where W and W^υ are one-dimensional standard Brownian motions with $\mathbb{E}[dW_t dW_t^\upsilon] = \rho dt$, B and B^ν are d-dimensional Brownian motions, for $i = 1, 2, \ldots, 5$, with $\mathbb{E}[dB_{i,t} dB_{i,t}^\nu] = \varsigma dt$. We choose $(b, \kappa, \alpha_\upsilon, \gamma, \alpha_\nu, \rho, \varsigma) = (0.03, 0.095, -0.15, 0.125, -0.05, -0.4, -0.62)$ the value of these parameters follow closely those of Huang and Tauchen (2005).[9] The loading factor or so-called systematic risk follows a uniform distribution as follows: $\beta_i \sim \mathcal{U}[0.45, 1.5]$ for $i = 1, 2, \ldots, 5$. We use the stationary distribution of ν_t^2 and υ_t^2 to restart the process each day at $\nu_{i,0}^2 \sim \mathcal{N}\left(0, (-2\alpha_\nu)^{-1}\right)$ and $\upsilon_{i,0}^2 \sim \mathcal{N}\left(0, (-2\alpha_\upsilon)^{-1}\right)$, respectively. Employing an Euler scheme, we simulate $T = 2,000$ days, and normalize one second to be $\Delta_n = 1/23400$, so that the interval $[0, 1]$ contains 6.5 hrs, i.e. $n = 23,400$ observations.

Since the choice of the sampling frequency plays a fundamental role in any study using intraday returns, we add microstructure noise so that we can evaluate the performance across various sampling frequencies. We first simulate the log-prices of all the assets at the second frequency, then contaminate the data to simulate the impact of microstructure noise. As is customary in the literature, we add normal random noise with mean zero and variance 0.05^2. We then aggregate the data to lower frequencies such that 5-min, 10-min

[9] We have calibrated these parameters such that the contribution of idiosyncratic volatility to total realized variance oscillates between 35%–80%. Although these values are slightly below and above the range we find in our dataset, it allows us to account for any other variation not captured in our sample. Figure 4.1 displays a realization of the total realized variance and the idiosyncratic volatility across the 2,000 simulated days. In this example, idiosyncratic volatility accounts for approximately 68% of the total realized variance.

and 15-min, which are equivalent to 78, 39 and 26 observations per day. The forecasts are based on a rolling window of 500 days and 1,000 replications.

We report the portfolio turnover defined in equation (4.20), as well as the portfolio standard deviation based on the population covariances, $\sqrt{\widehat{w}_t \Sigma_t \widehat{w}_t'}$. We also assess the distance of the estimated weights to the fundamental weights, based on the population covariances, $d\left(\widehat{w}_t^{(i)}, w_t^{(i)}\right) = \sqrt{\sum_{i=1}^{d}\left(\widehat{w}_t^{(i)} - w_t^{(i)}\right)^2}$.

Table 4.1 displays in two panels the portfolio characteristics (Panel A), and the out-of-sample performance (Panel B) of the HAR-M and HARQ-M models and their standard counterparts across sampling frequencies. By first comparing the results of the models across sampling frequencies, we observe that the portfolio characteristics based on 5-min returns are closer to the optimal estimates. This finding is corroborated by the loss functions, which show that the forecasts based on realized covariances estimated using 5-min returns produce the least amount of deviations.

Turning our attention to the performance of the models, we clearly observe a consistent decrease in the losses of the HAR-M and HARQ-M models, which of course stem from the use of the market factor information. The statistically more accurate covariance forecasts from adding the market factor information should result in portfolio allocations closer to the optimal weights implied by the population covariance matrix. Moreover, models accounting for the attenuation bias should render more stable portfolios and therefore display a reduced turnover (e.g. Bollerslev et al., 2018). As shown in Panel A of Table 4.1, the portfolio turnover, standard deviation and the weights are always closer to the optimal estimates when the market factor information is added into the model. This result holds true irrespective of the sampling frequency under analysis. As expected, the HARQ-(M) models render the smallest turnover and standard deviation, since accounting for the measurement error observed in the realized covariance leads to more stable portfolios. However, the HAR-M model usually provides as good performance as the standard HARQ model, although the HARQ-M model consistently outperforms all the other specifications. The superior performance of the HARQ-M model provides evidence that accounting for the measurement error is important, but further gains can

be attained when the information content of the market factor is exploited.

4.4 Empirical Results

4.4.1 Data

We use 5-minute returns obtained from Tick Data Inc., spanning a time period between January 2000 to December 2016.[10] The five-minute frequency has largely become the standard in the literature due to the trade-off between bias and variance (e.g. Hansen and Lunde, 2006), which is readily shown by the volatility signature plot in Figure 4.2. We consider 30 S&P 500 assets that traded continuously over the time period, along with returns of the SPY as the market factor. We evaluate all individual stocks in the univariate setting, while for the multivariate case we evaluate the forecasting performance of 100 random portfolios based on $N = 5$ and $N = 10$ assets.

Table 4.2 reports the descriptive statistics of our dataset. Each panel shows the mean and standard deviation of the RV, the realized beta (β), the idiosyncratic volatility and trading volume.[11] We consider a balanced panel of assets from different sectors, various levels of trading activity, and different systematic risk exposures. The idiosyncratic volatility represents about 71% of the total volatility across assets, as shown in Figure 4.3, whereas the proportion of idiosyncratic volatility to total variance oscillates between 57.4% to 87.7%, with FTR (Frontier Communications Corp) and GE (General Electric) displaying respectively the least and highest proportion.

[10] The multivariate literature, in an attempt to reduce the impact of the Epps (1979) effect, sample prices less frequent than 5-minutes. However, signature plots of the correlation (see Figure 4.2) show that the 5-minute sampling frequency is as good as 10- and 15-minute returns generally employed.

[11] The RV and idiosyncratic volatility (IdV) are annualized, and the idiosyncratic volatility is estimated as $IdV = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n Y - \beta \Delta_j^n X\right) \left(\Delta_j^n Y - \beta \Delta_j^n X\right)'$. The volume is reported in millions.

4.4.2 Univariate Results

In-sample Results

Table 4.3 reports the parameter estimates as an average across the 30 assets, together with robust standard errors in parentheses, for all the forecasts horizons. As a measure of fit, we report the average adjusted R-squares in the last row of Table 4.3. In line with previous results documented in the literature (Bollerslev et al., 2016; Corsi, 2009), the autoregressive parameter estimates of both the HAR and HARQ models are strongly significant, across all forecasting horizons, with the HARQ model affording a greater level of persistence. For the HAR-M and HARQ-M models both the autoregressive and the market factor estimates are significant irrespective of the forecast horizon. The inclusion of the market factor decreases the load assigned to the daily coefficient by approximately 25% and 12% for the HAR-M and HARQ-M models, respectively. On the other hand, the weekly ($h = 5$) and monthly ($h = 22$) estimates are not significantly impacted by the market factor information.[12] This result indicates that the market factor information helps to react faster to new information that is not always contained in the asset price information set, and therefore it subsumes a portion of the information contained at the daily level of past realized variances.

The R-squares reaffirm the relevance of including a market factor, as the HAR-M and HARQ-M models consistently improve the in-sample fitting of their standard counterparts, specially at longer horizons.

Out-of-sample Results

The out-of-sample performance of the univariate models is reported in Table 4.4. The forecasts are estimated using rolling windows of 1,000 observations, yielding a total of 3,277 out-of-sample forecasts. We report in three panels the losses of the HAR(Q)-M models standardized by the losses of the HAR(Q) models across all forecasting horizons. In brackets we report the proportion of assets that are retained by the Model Confidence

[12] We also tried adding the weekly and monthly levels of the market factor. Although the results are improved irrespective of the horizon at which the market factor is added, the results are consistently superior when the market factor is added at the daily level.

Set (MCS) of Hansen et al. (2011), at the 10% significance level, which is estimated using a block bootstrap with 5,000 replications and a block window of 20 days. Starred values highlight the proportion of assets for which the HAR(Q)-M models significantly outperform their standard counterparts, based on the Giacomini and White (2006) test, at the 5% significance level. A single star shows proportions between $[0, 0.25)$, two stars indicate proportions between $[0.25, 0.5)$, three stars are for proportions between $[0.5, 0.75)$, and four stars indicate proportions between $[0.75, 1.00]$.

As shown in Table 4.4, irrespective of the loss function under analysis, the HAR-M and HARQ-M models consistently outperform their standard counterparts. Whereas both specifications benefit from the market factor information across all forecasting horizons, we observe bigger out-of-sample gains at the month horizon (Panel C), which is inline with the in-sample results. As evidenced by the Giacomini and White test, specifications incorporating the market factor in most cases significantly outperform their counterparts between 0.25–0.50 (0.50–0.75) of the assets based on the MSE (QLIKE).

As expected from the evidence shown by Bollerslev et al. (2016), the models accounting for the heteroskedastic measurement error, which is inherent to any volatility measure, significantly outperform the standard specifications. However, we observe further improvements when we incorporate the market factor. For instance, the MCS always retains the HARQ-M model across all forecasting horizons. On the other hand, the HARQ model is retained for about 77% and 93% of the assets at $h = 1$ and $h = 22$, respectively.

4.4.3 Multivariate Results

The multivariate forecasts are based on 100 random portfolios of dimension $N = 5, 10$, where the assets are randomly chosen from our panel of 30 S&P 500 constituents. The forecasts are generated using the same settings employed for the univariate case (see Section 4.4.2). We allow for different dynamics in the unique elements of the covariance matrix, and therefore they are modelled separately. Whereas for the case of $N = 5$ all the forecasts are positive definite, for the case of $N = 10$ we find that less than 2% of the forecasted covariance matrices are non positive definite. Thus, we employ the eigenvalue

decomposition as in Hautsch et al. (2015), and replace the eigenvalues which are less than or equal to zero with the smaller non-zero and positive eigenvalue. See the end of Section 4.2.1 for more details.

Since the main object of interest is the forecasts of the realized covariances, we focus on the out-of-sample performance of the multivariate extensions of the HAR-M and HARQ-M models. These results are reported in Table 4.5. Similar to the univariate results, we report in three panels the out-of-sample standardized losses at the one-day ($h = 1$), one-week ($h = 5$) and one-month ($h = 22$) horizon. In brackets, we show the proportion of portfolios that are retained by the MCS, and the starred values show the proportion of portfolios for which the HAR-M (HARQ-M) model has significantly outperformed its benchmark model based on the Giacomini and White test at the 5% significance level. A single star shows proportions between $[0, 0.25)$, two stars indicate proportions between $[0.25, 0.5)$, three stars are for proportions between $[0.5, 0.75)$, and four stars indicate proportions between $[0.75, 100]$.

In line with our results presented for the univariate case, the multivariate HAR-M and HARQ-M consistently outperform their counterparts irrespective of the forecasting horizon and whether we consider portfolios of 5 or 10 assets. For instance, for $h = 5$ and $N = 10$, the proportion of portfolios for which the HAR-M model significantly outperforms the HAR model oscillate between 0.75–1.00, while on average the proportion of portfolios for which the HAR(Q)-M model significantly outperforms its counterpart ranges between 0.25–0.50.

To assess the performance across all the models, we rely in the MCS of Hansen et al. (2011). As shown by Bollerslev et al. (2016, 2018), specifications taking into account the heteroskedastic measurement error significantly improve upon the results of standard specifications. Our results show that the HARQ-(M) models are often retained by the MCS, although the HARQ-M model is the only specification that is never excluded. This indicates that whereas accounting for the heteroskedastic measurement error improves the forecasting accuracy of the covariance matrices, the use of the market factor yields to further improvements and therefore its information content cannot be ignored.

Economic Value

We assess the economic value of the different models by constructing Global Minimum Variance (GMV) portfolios.[13] We employ daily, weekly and monthly rebalancing frequencies, and in each period the investor solves the following minimization problem:

$$\begin{aligned} \widehat{w}_t = \arg\min_{w_t} w_t' H_t w_t, \\ \text{s.t.} \quad w_t'\iota = 1, \end{aligned} \tag{4.18}$$

where w_t is an $N \times 1$ vector of GMV portfolios weights, ι is an $N \times 1$ vector of ones, and H_t is the $N \times N$ matrix of forecasted covariances from a particular model. The optimal portfolio weights, $\widehat{w}_t$, are given by:

$$\widehat{w}_t = \frac{H_t^{-1}\iota}{\iota' H_t^{-1}\iota}. \tag{4.19}$$

Given that models using the market factor information generally render more accurate forecasts, we also evaluate important portfolio characteristics that underscore the benefits of our models. For instance, it is well-known that inaccurate forecasts lead to both extreme positions and higher trading costs (e.g. DeMiguel et al., 2014; Han, 2006). Thus, we report the total portfolio turnover, portfolio concentration and portfolio short position. The portfolio turnover (TO) is measured by:[14]

$$TO_t = \sum_{i=1}^{N} \left| \widehat{w}_{t+1}^{(i)} - \widehat{w}_t^{(i)} \frac{1 + r_t^{(i)}}{1 + \widehat{w}_t' r_t} \right|, \tag{4.20}$$

This yields a portfolio excess return net of transaction cost, cTO_t, of:

$$r_{p,t} = \widehat{w}_t' r_t - cTO_t. \tag{4.21}$$

[13] As shown by Jagannathan and Ma (2003) and DeMiguel et al. (2009, 2014) mean-variance optimized portfolios do not perform as well as GMV portfolios in terms of out-of-sample evaluations, as the estimation error in the expected returns tends to distort the positions.

[14] $r_t^{(i)} = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \Delta_j^n Y^{(i)}$ is the daily return of the i-th asset.

The portfolio concentration is estimated as:

$$PC_t = \left(\sum_{i=1}^{N} \left(\widehat{w}_t^{(i)} \right)^2 \right)^{1/2}, \tag{4.22}$$

and the total portfolio short positions:

$$SP_t = \sum_{i=1}^{N} \widehat{w}_t^{(i)} \mathbb{1}_{\{\widehat{w}_t^{(i)} < 0\}}. \tag{4.23}$$

To evaluate the economic significance of the different forecasting models, we consider the utility-based framework of Fleming et al. (2001, 2003). We assume that the investor has quadratic utility with risk aversion γ, therefore the realized utility generated by the portfolio based on the covariance forecasts from model k is expressed as:

$$\mathrm{U}\left(r_{p,t}^{(k)}, \gamma \right) = \left(1 + r_{p,t}^{(k)} \right) - \frac{\gamma}{2(1+\gamma)} \left(1 + r_{p,t}^{(k)} \right)^2. \tag{4.24}$$

The economic value of the different models can therefore be determined by solving Δ_γ in:

$$\sum_{t=1}^{T} \mathrm{U}\left(r_{p,t}^{(k)}, \gamma \right) = \sum_{t=1}^{T} \mathrm{U}\left(r_{p,t}^{(l)} - \Delta_\gamma, \gamma \right), \tag{4.25}$$

where Δ_γ can be interpreted as the return an investor with risk aversion, γ, would be willing to pay to switch from using the model k to using model l.

The daily, weekly and monthly rebalancing portfolio results are respectively shown in Tables 4.6, 4.7 and 4.8. We report in two panels the results for the average 100 portfolios based on $N = 5$ and $N = 10$ assets. As portfolio characteristics, we report the turnover, portfolio short position and portfolio concentration. Since short positions are usually costly and portfolio concentration provides evidence about extreme positions, these two characteristics are good indicators to highlight the improved stability and accuracy of our financial decisions. The bottom of each table shows the economic gains of switching from each standard HAR and HARQ model to the HAR-M and HARQ-M model. We

consider two levels of risk-aversion $\gamma = 1, 10$ and transaction costs $c = 0\%, 0.5\%$, where $c = 0\%$ is equivalent to no transaction costs. To assess the economic significance of our strategies, i.e. whether Δ_γ are significantly different from zero, we use the Reality Check of White (2000). We employ the stationary bootstrap of Politis and Romano (1994), with 5,000 bootstrap samples and an average block length of 20 days. We highlight the significance performance using a single star when the proportion of portfolios with significant economic gains lies between $[0, 0.25)$, two stars indicate proportions between $[0.25, 0.5)$, three stars are for proportions between $[0.5, 0.75)$, and four stars indicate proportions between $[0.75, 1.00]$.

The daily rebalancing portfolio strategies based on the HAR-M and HARQ-M models (Table 4.6) show a consistent decrease in their turnover compared to the standard strategies. For $N = 5$ ($N = 10$), the turnover is reduced up to 11% (5%) for the HAR-M (HARQ-M) model. This result is in line with the simulation study and corroborate the statistical gains reported in Section 4.4.3. The portfolio short positions of the models incorporating the market factor information is bounded by the short positions of the standard models. In other words, the HAR-M and HARQ-M models display a level of SP which is either equal or smaller than that of their benchmark. On the other hand, the portfolio concentration is always smaller for the HAR-M and HARQ-M models. This shows evidence that the market factor leads to more accurate forecasts of the covariance matrix, which therefore reduces the extreme positions that are usually associated with poor out-of-sample predictions.

Given that the gains come from two different sources: forecasting accuracy and reduced turnover, we first ignore transaction costs, so that we can evaluate the gains stemmed from the former. The total gains for $N = 5$ range between 9 and 49 annual basis points, whereas for $N = 10$ the gains range between 11 and 21 annual basis points. In both cases we find that the proportion of portfolios with significant gains usually oscillates between 0–0.25. Since the HAR-M and HARQ-M models result in lower turnover, these gains rise to 122 to 156 for $N = 5$ and to 55 to 142 for $N = 10$ when $c = 0.5\%$.[15]

[15]Increasing the transaction cost levels from 0.5% to, say, 1% or 2% will always increase the performance of the HAR-M and HARQ-M models.

The increase in the economic gains is accompanied by an increase in the significance of these gains, where the proportion of portfolios with significant gains lies usually between 0.5–0.75.

The weekly and monthly rebalancing portfolio results, Tables 4.7 and 4.8, show qualitatively similar findings than those based on daily rebalancing portfolios. The turnover of the HAR-M and HARQ-M models is reduced up to 11% and 14% for respectively weekly and monthly rebalancing strategies. The portfolio short positions and portfolio concentration are also improved in favor of the models using the market factor, with the HARQ-M model displaying the best overall results. The total economic gains, in the absence of transaction costs, for the weekly rebalancing oscillate between 4 to 30 for $N = 5$ and 10 to 15 for $N = 10$, whereas for the monthly rebalancing the total gains range between 5 to 34 for $N = 5$ and 7 to 15 for $N = 10$. When transaction costs are added, we find smaller gains relative to the daily strategies, nonetheless all the Δ_γ's remain positive and statistically significant for all levels of risk-aversion.

The reduction in the performance fee observed at the weekly and monthly horizon is explained by that: i) weekly and monthly strategies are cheaper as we incur in less transaction costs; ii) the level of measurement error is smaller for weekly and monthly realized covariances, and therefore the portfolio weights estimated based on these realized covariances are more stable than those based on daily measures.

4.5 Conclusion

This paper proposes to model and forecast both univariate and multivariate realized variances using the information of the market factor. The use of a market factor is motivated by: i) the fact that asset prices are also influenced by market and industry factors, which cannot be assumed to be contained in the individual asset price information set. Thus, adding the market factor information makes the models react faster to new market information; ii) the low predictability found in assets with high levels of idiosyncratic volatility (see Ang et al., 2009).

When both the univariate and multivariate versions of the HAR and HARQ models are adapted to incorporate the market factor information, the result is more accurate and significant in- and out-of-sample performance of the models, across all forecasting horizons. In particular, we find that the HAR-M and HARQ-M models render significantly bigger out-of-sample gains, and are generally retained by the MCS. These findings hold true using both simulation and empirical data.

In a practical implementation, we assess the economic significance using a global minimum variance portfolio based on the forecasted covariance matrices. We find that a risk-averse investor is willing to sacrifice up to 157 annual basis points for switching to one of the strategies using the market factor information. These gains are obtained from two different sources: the forecasting accuracy and the reduced turnover. Thus, the model accounting for the attenuation bias and the market factor information, HARQ-M model, significantly improves the accuracy of asset pricing models and thus financial decisions.

Appendix 4.A Tables and Figures

Table 4.1: Simulation Based Portfolio Characteristics and Loss Functions

n	∞	78				39				26			
	Optimal	HAR	HAR-M	HARQ	HARQ-M	HAR	HAR-M	HARQ	HARQ-M	HAR	HAR-M	HARQ	HARQ-M
Panel A: Portfolio Characteristics													
Turnover	0.095	0.098	0.096	0.092	0.094	0.115	0.110	0.109	0.104	0.127	0.118	0.118	0.109
StDev	0.936	0.951	0.946	0.948	0.941	0.958	0.956	0.956	0.951	0.960	0.954	0.958	0.953
$d\left(\widehat{w}_t^{(i)}, w_t^{(i)}\right)$	0.000	0.200	0.190	0.190	0.187	0.296	0.285	0.293	0.281	0.376	0.373	0.374	0.373
Panel B: Loss functions													
QLIKE		0.253	0.249	0.254	0.247	0.485	0.481	0.468	0.448	0.725	0.723	0.722	0.703
Frob		1.753	1.746	1.747	1.742	2.204	2.199	2.203	2.197	2.576	2.570	2.476	2.457

Note: The table reports the results from a simulation study pertaining to the out-of-sample global minimum variance portfolios and statistical performance of the forecasted covariance matrix. The covariance matrix forecasts are based on the HAR-M and HARQ-M and their standard counterparts, where the realized measures are estimated with n intraday observations. These observations correspond to 5-, 10- and 15-min returns, while $n = \infty$ shows the results based on the population covariance matrix. Turnover is defined in equation (4.20). StDev refers to the portfolio standard deviation based on the population covariance. $d\left(\widehat{w}_t^{(i)}, w_t^{(i)}\right)$ reports the distance between the portfolio weights and the weights based on the population covariance matrix.

Table 4.2: Descriptive Statistics

	RV		β		Idiosyncratic Vol			Volume	
	mean	std	mean	std	mean	std	proportion	mean	std
ABT	20.727	10.802	0.695	0.360	18.177	9.895	0.784	5.535	3.095
AKS	57.857	26.816	1.613	1.207	52.196	23.608	0.807	5.125	5.229
AMZN	37.549	26.034	1.369	0.568	32.151	23.589	0.762	6.462	5.007
BA	23.548	12.420	0.902	0.344	19.668	10.275	0.695	4.347	2.406
BBY	32.857	19.287	1.048	0.535	28.979	17.757	0.796	6.103	4.373
BFB	19.447	10.283	0.584	0.343	17.532	9.390	0.817	0.875	0.708
C	30.291	27.959	1.229	0.576	24.198	22.814	0.651	16.897	22.306
CVX	20.310	10.952	0.856	0.398	16.175	7.322	0.592	7.188	3.639
DD	23.145	12.656	0.971	0.355	18.430	10.097	0.635	4.519	2.609
DVN	30.082	15.910	0.995	0.741	26.162	12.788	0.732	3.653	2.608
EBAY	33.338	23.012	1.261	0.516	28.132	20.136	0.729	15.869	10.046
EXC	22.382	12.775	0.615	0.373	20.098	10.513	0.775	3.917	2.435
FCX	39.161	22.163	1.225	0.897	34.395	18.870	0.760	15.458	15.092
FTR	31.431	19.483	0.602	0.532	29.382	18.339	0.877	6.017	7.871
GE	22.624	15.788	0.985	0.335	17.393	11.601	0.574	38.656	36.927
HD	24.235	14.118	0.962	0.345	19.784	11.007	0.652	9.746	6.416
HON	24.368	15.038	1.006	0.356	19.401	12.939	0.663	3.884	2.691
KO	17.307	9.686	0.586	0.257	14.959	8.189	0.739	13.275	6.591
MCD	20.027	12.008	0.639	0.307	17.558	10.688	0.775	5.808	3.239
MMM	18.790	10.676	0.821	0.277	14.756	8.636	0.626	3.265	1.813
MRK	21.508	12.251	0.713	0.335	18.689	10.419	0.747	10.259	6.848
MSFT	22.772	12.518	1.015	0.345	17.712	8.954	0.583	54.765	28.667
ORCL	29.646	19.658	1.207	0.494	23.838	15.374	0.636	31.956	19.015
PEG	21.673	12.262	0.613	0.374	19.332	9.875	0.760	2.398	1.369
PFE	21.617	10.977	0.761	0.323	18.563	9.230	0.731	28.672	18.751
SO	18.459	9.941	0.489	0.314	16.734	8.705	0.809	3.203	1.869
TRV	22.120	16.086	0.728	0.373	18.989	13.692	0.733	2.390	1.637
UPS	17.610	10.273	0.698	0.349	14.390	8.176	0.659	3.045	2.079
WFC	24.919	21.525	1.029	0.523	19.663	16.345	0.603	21.680	27.560
WMT	19.500	11.629	0.698	0.298	16.463	9.540	0.702	10.569	6.339

Note: The table reports for each individual asset the averages and time series standard deviations of their volatilities (annualized) of each individual asset, their estimated realized betas with respect to the SPY market index, their idiosyncratic volatility (annualized) together with their proportion to total volatility, and their trading volume. The idiosyncratic volatility is estimated as $IdV = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n Y - \beta \Delta_j^n X\right) \left(\Delta_j^n Y - \beta \Delta_j^n X\right)'$.

Table 4.3: Univariate Models In-sample Estimates

	HAR	HAR-M	HARQ	HARQ-M	HAR	HAR-M	HARQ	HARQ-M	HAR	HAR-M	HARQ	HARQ-M
	$h=1$				$h=5$				$h=22$			
β_0	0.517	0.480	0.129	0.170	0.692	0.658	0.380	0.417	1.085	1.058	0.845	0.869
s.e.	(0.173)	(0.180)	(0.206)	(0.210)	(0.211)	(0.203)	(0.214)	(0.215)	(0.256)	(0.223)	(0.212)	(0.202)
β_d	0.236	0.178	0.566	0.502	0.158	0.110	0.413	0.357	0.091	0.065	0.282	0.249
s.e.	(0.079)	(0.088)	(0.113)	(0.116)	(0.047)	(0.057)	(0.098)	(0.092)	(0.022)	(0.037)	(0.080)	(0.074)
β_w	0.345	0.328	0.275	0.265	0.303	0.289	0.250	0.242	0.238	0.230	0.199	0.194
s.e.	(0.122)	(0.116)	(0.108)	(0.106)	(0.087)	(0.084)	(0.083)	(0.082)	(0.077)	(0.074)	(0.073)	(0.071)
β_m	0.288	0.251	0.168	0.154	0.361	0.331	0.270	0.258	0.384	0.363	0.316	0.308
s.e.	(0.098)	(0.090)	(0.096)	(0.096)	(0.106)	(0.099)	(0.096)	(0.096)	(0.084)	(0.082)	(0.086)	(0.087)
β_h		0.342		0.193		0.299		0.176		0.183		0.106
s.e.		(0.187)		(0.109)		(0.125)		(0.101)		(0.106)		(0.057)
β_{d_π}			−0.002	−0.002			−0.002	−0.002			−0.001	−0.001
s.e.			(0.000)	(0.000)			(0.000)	(0.000)			(0.000)	(0.000)
R^2_{adj}	0.461	0.470	0.487	0.492	0.575	0.589	0.602	0.609	0.552	0.565	0.574	0.581

Note: The table reports in three panels the parameter estimates for 1-day, 1-week, and 1-month ahead forecasts as an average across all the individual assets, together with robust standard errors in parentheses. The robust standard errors are estimated using the Newey-West HAC correction allowing fro serial correlation up to order 5 ($h=1$), 10 ($h=10$) and 44 ($h=22$). The last row reports the adjusted R-squares as an average across all the individual assets.

Table 4.4: Univariate Models Out-of-Sample Results

	HAR	HAR-M	HARQ	HARQ-M
Panel A: $h = 1$				
MSE	1.000	0.927*	1.000	0.989*
MCS	[0.533]	[0.633]	[0.767]	[1.000]
QLIKE	1.000	0.977***	1.000	0.981**
MCS	[0.400]	[0.467]	[0.733]	[1.000]
Panel B: $h = 5$				
MSE	1.000	0.943**	1.000	0.946**
MCS	[0.667]	[0.767]	[0.900]	[1.000]
QLIKE	1.000	0.954***	1.000	0.978***
MCS	[0.500]	[0.567]	[0.833]	[1.000]
Panel C: $h = 22$				
MSE	1.000	0.897**	1.000	0.920**
MCS	[0.667]	[0.800]	[0.930]	[1.000]
QLIKE	1.000	0.952**	1.000	0.968**
MCS	[0.633]	[0.733]	[0.733]	[1.000]

Note: The table reports in three panels the out-of-sample forecast loss for the different models at the 1-day (Panel A), 1-week (Panel B) and 1-month (Panel C) ahead. The loss of the HAR-M (HARQ-M) is standardized by the loss of the HAR (HARQ) model. In brackets we report the proportion of stocks that are retained by the Model Confidence Set (MCS) of Hansen et al. (2011) at the 10% significance level. Starred values indicate the proportion of individual stocks for which the HAR-M (HARQ-M) model has significantly outperformed its benchmark model based on the Giacomini and White (2006) at the 5% level. A single star shows proportions between $[0, 0.25)$, two stars indicate proportion between $[0.25, 0.5)$, three stars is for proportions between $[0.5, 0.75)$, and four stars indicate proportion between $[0.75, 1.00]$.

Table 4.5: Multivariate Models Out-of-Sample Results

	HAR	HAR-M	HARQ	HARQ-M	HAR	HAR-M	HARQ	HARQ-M
	Portfolio $N = 5$				Portfolio $N = 10$			
Panel A: $h = 1$								
Frob	1.000	0.981***	1.000	0.991**	1.000	0.985**	1.000	0.993*
MCS	[0.530]	[0.710]	[0.840]	[1.000]	[0.580]	[0.630]	[0.780]	[1.000]
QLIKE	1.000	0.994**	1.000	0.921***	1.000	0.996*	1.000	0.849***
MCS	[0.200]	[0.450]	[0.730]	[1.000]	[0.270]	[0.320]	[0.630]	[1.000]
Panel B: $h = 5$								
Frob	1.000	0.965***	1.000	0.986***	1.000	0.970****	1.000	0.987***
MCS	[0.440]	[0.650]	[0.860]	[1.000]	[0.170]	[0.410]	[0.700]	[1.000]
QLIKE	1.000	0.985***	1.000	0.888***	1.000	0.991*	1.000	0.804***
MCS	[0.140]	[0.450]	[0.790]	[1.000]	[0.120]	[0.240]	[0.710]	[1.000]
Panel C: $h = 22$								
Frob	1.000	0.973***	1.000	0.993*	1.000	0.974***	1.000	0.993**
MCS	[0.510]	[0.810]	[0.900]	[1.000]	[0.320]	[0.800]	[0.890]	[1.000]
QLIKE	1.000	0.960*	1.000	0.922*	1.000	0.940*	1.000	0.870**
MCS	[0.580]	[0.760]	[0.950]	[1.000]	[0.240]	[0.330]	[0.910]	[1.000]

Note: The table reports in three panels the out-of-sample forecast loss for the different models at the 1-day (Panel A), 1-week (Panel B) and 1-month (Panel C) horizons. The left (right) column shows the average results of 100 random portfolios based on 5 (10) assets. The loss of the HAR-M (HARQ-M) is standardized by the loss of the HAR (HARQ) model. In brackets we report the proportion of portfolios that are retained by the Model Confidence Set (MCS) of Hansen et al. (2011) at the 10% significance level. Starred values indicate the proportion of portfolios for which the HAR-M (HARQ-M) model has significantly outperformed its benchmark model based on the Giacomini and White (2006) at the 5% level. A single star shows proportions between $[0, 0.25)$, two stars indicate proportions between $[0.25, 0.5)$, three stars is for proportions between $[0.5, 0.75)$, and four stars indicate proportions between $[0.75, 1.00]$.

Table 4.6: Minimum Variance Portfolios – Daily Rebalancing

	HAR	HAR-M	HARQ	HARQ-M	HAR	HAR-M	HARQ	HARQ-M
	Portfolio $N = 5$				Portfolio $N = 10$			
TO	0.227	0.202	0.212	0.209	0.402	0.397	0.375	0.358
SP	−0.020	−0.018	−0.018	−0.018	−0.108	−0.088	−0.091	−0.090
PC	0.584	0.575	0.617	0.592	0.497	0.495	0.470	0.430
$c = 0\%$								
Δ_1		49.200**		44.092**		18.809*		21.329*
Δ_{10}		10.762*		8.914*		17.129*		10.745*
$c = 0.5\%$								
Δ_1		141.169****		156.871****		119.211***		142.231***
Δ_{10}		121.796***		147.326***		55.311**		68.739***

Note: The table shows the average results for the global minimum variance portfolio (GMV) based on 100 random portfolios of 5 and 10 assets. We report turnover (TO), portfolio short positions (SP) and portfolio concentration (PC). The table also reports the economic gains from switching from the conventional HAR (HARQ) model to the HAR-M (HARQ-M) model in annual basis points, Δ_γ, for cases with(out) transaction cost $c = 0.5\%$ ($c = 0\%$). Starred values indicate that Δ_γ is significantly different from zero at the 5%, based on the Reality Check of White (2000) using a stationary bootstrap with 999 replications and a block window of 20 days. A single star shows proportions between $[0, 0.25)$, two stars indicate proportions between $[0.25, 0.5)$, three stars are for proportions between $[0.5, 0.75)$, and four stars indicate proportions between $[0.75, 1.00]$.

Table 4.7: Minimum Variance Portfolios – Weekly Rebalancing

	HAR	HAR-M	HARQ	HARQ-M	HAR	HAR-M	HARQ	HARQ-M
	Portfolio $N = 5$				Portfolio $N = 10$			
TO	0.146	0.130	0.123	0.118	0.255	0.246	0.252	0.237
SP	−0.016	−0.015	−0.015	−0.012	−0.070	−0.068	−0.067	−0.065
PC	0.577	0.571	0.605	0.596	0.482	0.480	0.471	0.455
$c = 0\%$								
Δ_1		29.861**		24.530*		11.578*		15.371*
Δ_{10}		5.454*		4.118*		10.261*		10.524*
$c = 0.5\%$								
Δ_1		71.151***		89.026***		80.445***		94.762***
Δ_{10}		42.673**		51.689***		32.792**		53.737***

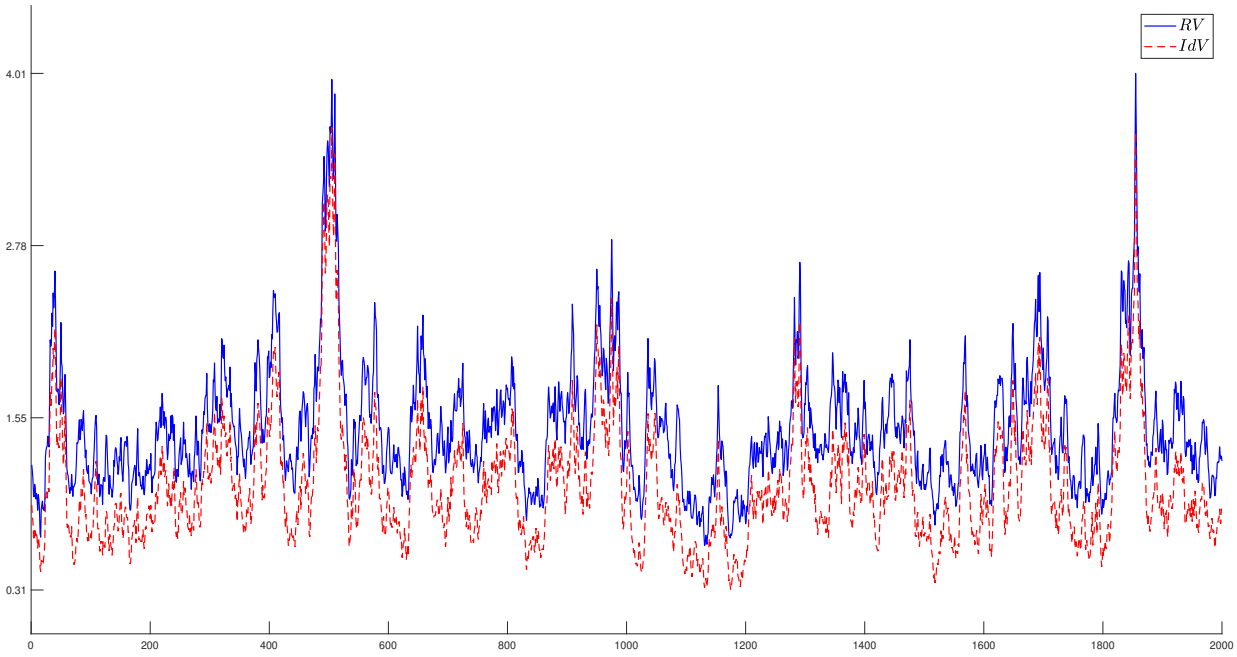
Note: The table shows the average results for the global minimum variance portfolio (GMV) based on 100 random portfolios of 5 and 10 assets. We report turnover (TO), portfolio short positions (SP) and portfolio concentration (PC). The table also reports the economic gains from switching from the conventional HAR (HARQ) model to the HAR-M (HARQ-M) model in annual basis points, Δ_γ, for cases with(out) transaction cost $c = 0.5\%$ ($c = 0\%$). Starred values indicate that Δ_γ is significantly different from zero at the 5%, based on the Reality Check of White (2000) using a stationary bootstrap with 999 replications and a block window of 20 days. A single star shows proportions between $[0, 0.25)$, two stars indicate proportions between $[0.25, 0.5)$, three stars are for proportions between $[0.5, 0.75)$, and four stars indicate proportions between $[0.75, 1.00]$.

Table 4.8: Minimum Variance Portfolios – Monthly Rebalancing

	HAR	HAR-M	HARQ	HARQ-M	HAR	HAR-M	HARQ	HARQ-M
	Portfolio $N = 5$				Portfolio $N = 10$			
TO	0.104	0.089	0.096	0.088	0.177	0.168	0.171	0.161
SP	−0.016	−0.015	−0.010	−0.010	−0.067	−0.065	−0.064	−0.066
PC	0.575	0.570	0.579	0.572	0.479	0.478	0.432	0.422
$c = 0\%$								
Δ_1		34.352**		55.604**		15.121*		14.264*
Δ_{10}		5.632*		11.125*		7.663*		10.844*
$c = 0.5\%$								
Δ_1		57.340**		71.931***		45.292**		53.378**
Δ_{10}		29.573**		38.120**		25.377*		42.775**

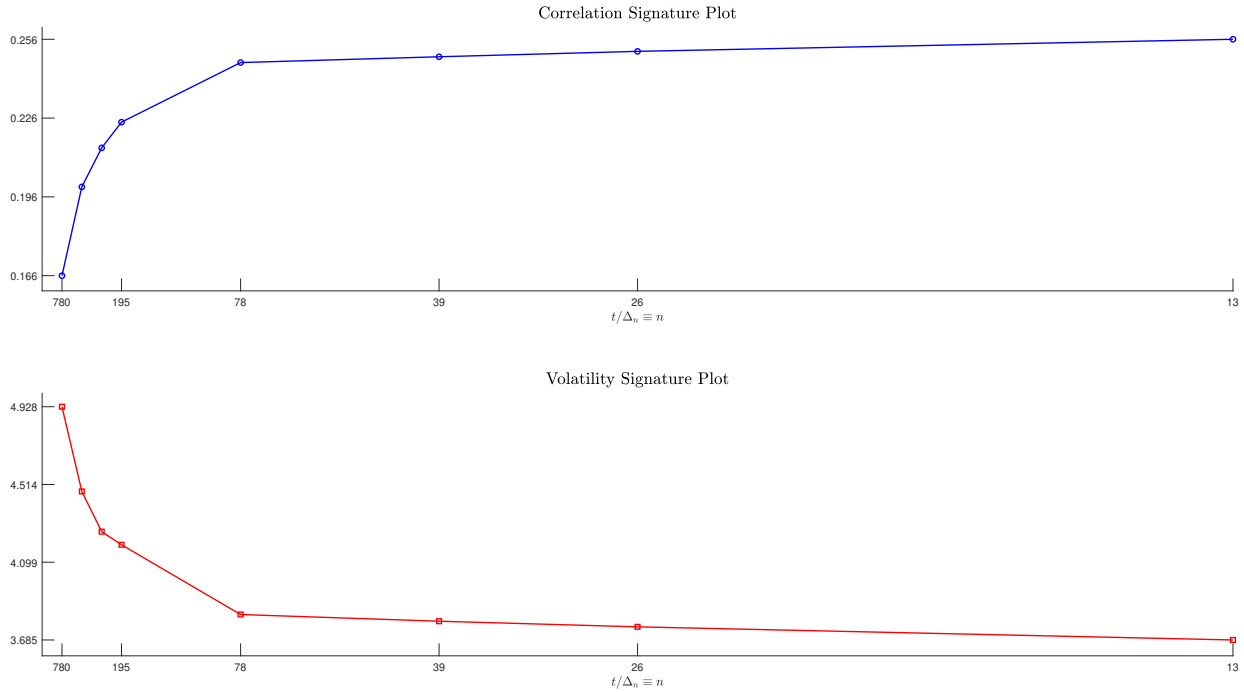
Note: The table shows the average results for the global minimum variance portfolio (GMV) based on 100 random portfolios of 5 and 10 assets. We report turnover (TO), portfolio short positions (SP) and portfolio concentration (PC). The table also reports the economic gains from switching from the conventional HAR (HARQ) model to the HAR-M (HARQ-M) model in annual basis points, Δ_γ, for cases with(out) transaction cost $c = 0.5\%$ ($c = 0\%$). Starred values indicate that Δ_γ is significantly different from zero at the 5%, based on the Reality Check of White (2000) using a stationary bootstrap with 999 replications and a block window of 20 days. A single star shows proportions between $[0, 0.25)$, two stars indicate proportions between $[0.25, 0.5)$, three stars are for proportions between $[0.5, 0.75)$, and four stars indicate proportions between $[0.75, 1.00]$.

Figure 4.1: Simulated Realized and Idiosyncratic Volatility



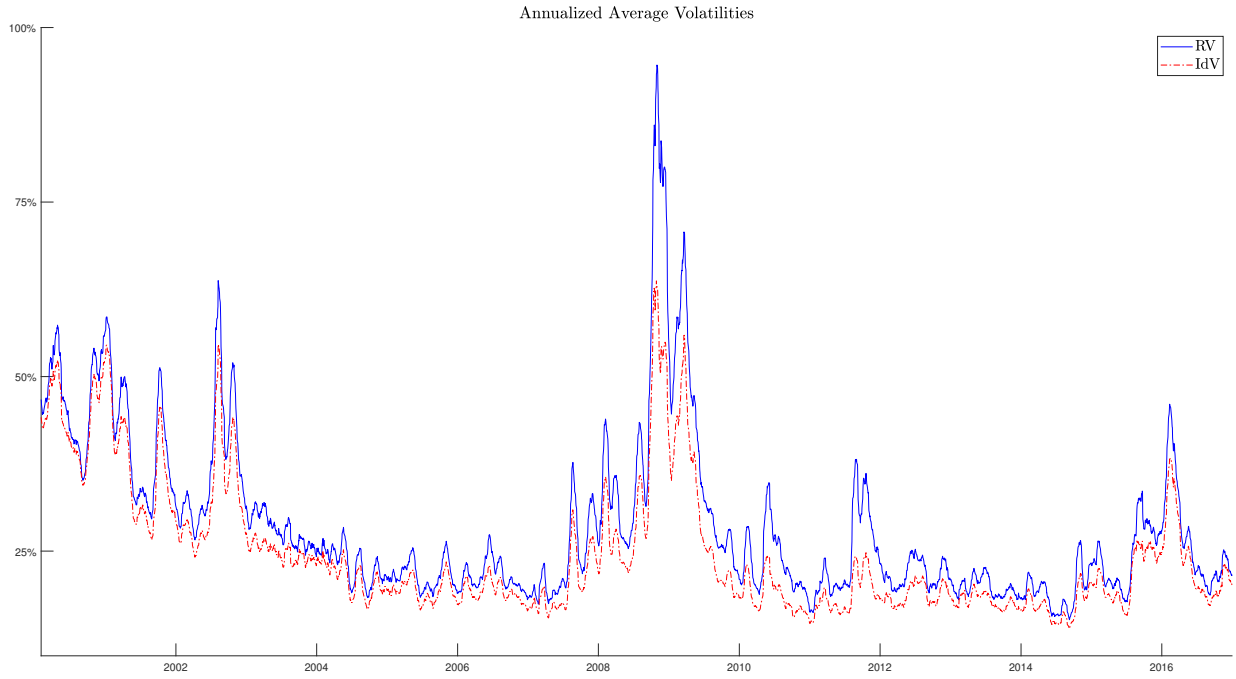
Note: The figure depicts the simulated realized and idiosyncratic volatility across the 2,000 simulated days based on the Monte Carlo set up outlined in Section 3.15.

Figure 4.2: Signature Plots



Note: The figure shows in two panels the correlation and the volatiliy signature plots. The correlation is the average correlation across every unique element below the main diagonal of the covariance matrix, i.e. the average of $N(N-1)/2$ unique elements, while the volatility is the average across the 30 assets in our dataset. The x-axis shows the number of intraday observations for each sampling frequency, i.e. 780 (30-sec), 390 (1-min), 260 (1.5-min), 195 (2-min), 78 (5-min), 39 (10-min), 26 (15-min) and 13 (30-min).

Figure 4.3: Average Annualized Realized and Idiosyncratic Volatility



Note: The figure depicts the average annualized realized and idiosyncratic volatility across all the stocks.

Part II

Identifying the Underlying Components of High-Frequency Data and the Predictive Information Content of Jumps

Chapter 5

Evaluating the Underlying Components of High Frequency Financial Data: Finite Sample Performance and Microstructure Noise Effects

5.1 Introduction

Following Bachelier (1900), a general presumption has been that the continuous evolution of asset prices is driven by Brownian motion processes. However, the failure of Brownian increments to explain heavy-tails observed in the distributions of returns, undermines that presumption. With an alternative approach, Merton (1976) proposes a finite-jump diffusion process,[1] which successfully mimics empirical continuous and jump components. As small jumps eliminate the need of a Brownian component, the attraction of pure jump processes lies with their versatility and flexibility.[2]

[1]The finite-jump diffusion of Merton (1976) considers a compound Poisson process to model the jump part.

[2]Some infinite jump models are the variance gamma model (Madan and Seneta, 1990), the hyperbolic model of (Eberlein and Keller, 1995), the CGMY Carr et al. (2002), the COGARCH model (Klüppelberg et al., 2004), the non Gaussian Ornstein-Uhlenbeck-based models Barndorff-Nielsen and Shephard (2001), the CARMA model (Brockwell, 2001), the normal inverse Gaussian Barndorff-Nielsen (1997); Rydberg (1997), among others.

Advances in computer power and the widening access to high frequency data fuel novel statistical procedures that can test for the properties of the underlying process. There are tests to identify jumps (e.g. Aït-Sahalia et al., 2009; Barndorff-Nielsen and Shephard, 2006; Jiang and Oomen, 2008; Lee and Mykland, 2008; Podolskij and Ziggel, 2010). There are tests to evaluate the presence of a Brownian component (Aït-Sahalia and Jacod, 2010; Cont and Mancini, 2011; Jing et al., 2012a; Kong et al., 2015). There are tests to examine whether jumps have finite or infinite activity (Aït-Sahalia and Jacod, 2011; Cont and Mancini, 2011; Kong, 2019).[3] Most of these procedures work best at ultra high frequencies and assume that microstructure noise is absent. In the main, that literature has investigated the finite sample properties of jump tests (Dumitru and Urga, 2012; Huang and Tauchen, 2005; Maneesoonthorn et al., 2020), yet providing no guidance on the optimal frequency or the effect of microstructure noise on the alternative procedures (Brownian component and infinite vs. finite jump activity tests).

Set against this background, the current paper contributes to the existing literature by examining the finite sample properties of test statistics under different types of market microstructure noise. The paper accommodates temporal heteroskedasticity and dependence, and takes account of bouncebacks typically observed in trade data. We follow Kong et al. (2015) in evaluating the Brownian motion hypothesis. This procedure improves upon the results of Aït-Sahalia and Jacod (2010) and Jing et al. (2012a), for the properties of finite sample in the absence of microstructure noise. To test for finite and infinite activity, we follow Aït-Sahalia and Jacod (2011). To test for jumps, we follow Podolskij and Ziggel (2010) and employ their standard and robustified versions.[4]

The asymptotic distribution of most of these procedures have been derived under the assumption of noiseless prices. However, noisy prices generally skew the distribution of tests, so raising the likelihood of either type I or type II error. Therefore, microstructure noise becomes a relevant consideration as $\Delta_n \to 0$. Thus, it is vital to offset the impact

[3]The literature has also employed the use of the Blumenthal-Getoor index to estimate jump activity however, this estimator serves as an indicator for the presence of finite/infinite activity rather than a formal tool to disentangle these two components, (e.g. Aït-Sahalia and Jacod, 2009; Jing et al., 2012b; Todorov, 2015; Todorov and Tauchen, 2010).

[4]The test of Podolskij and Ziggel (2010) along with the test of Aït-Sahalia et al. (2009) can detect jumps of finite and infinite activity.

of that microstructure noise, as inference about the appropriate model depends on the sampling frequency and testing technique (see, for instance, Todorov and Tauchen, 2010).

Our paper also examines the underlying components of 100 individual S&P 500 constituents and the S&P 500 ETF (SPY) over 17 years. No clear conclusions have emerged from papers that investigate appropriate criteria for models of stock prices. To illustrate, Andersen et al. (2002) compare several diffusion and finite-jump diffusion models, reaching the conclusion that a finite-jump diffusion model is capable of catching the characteristics of the S&P 500 returns. From a contemporaneous investigation of the same index, the conclusion reached by Carr et al. (2002) is that a pure jump process is the most appropriate model.[5] As different techniques and/or sampling frequencies might be the source of such discrepancies,[6] the use of test statistics offers a unified framework, and our paper engages in a comprehensive examination of relevant performances to highlight their advantages and limitations.

From our Monte Carlo experiment, a summary of results is as follows. In the absence of microstructure noise, all the tests had satisfactory finite sample properties, which deviate slowly from the theoretical size and power as the time-interval widened. Conversely, the presence of microstructure noise biases the distributions of all the tests, except for the robustified PZ test, which is derived under the assumption of noisy prices. In the presence of Gaussian noise, the distributions of the tests show bias only at very high frequencies. In the presence of t-distributed and Gaussian-T mixture noise, the performance of all the tests is severely adversely affected. Sampling sparsely decreases the bias of the tests. Under Gaussian noise, the tests display satisfactory results when returns are sampled every 30 seconds. When microstructure noise is t-distributed or Gaussian-T mixture, sampling every 60 seconds give satisfactory results. However, in the presence of non-Gaussian noise, the standard PZ test shows severe upward bias even when return are

[5]While the use of finite-jump diffusion models dates back to Merton (1976) and Jorion (1988), and are backed with their link to macroeconomic announcements (see, Dungey and Hvozdyk, 2012), Carr et al. (2002); Carr and Wu (2007); Cont and Tankov (2004); Daal and Madan (2005); Geman (2002); Huang and Wu (2004), among others support the use of pure jump processes compared to the classical mixture models.

[6]It is well-known that parametric approaches run the risk of incorrect specification for functionals in their chosen models. This is not the case with the non-parametric approaches employed in this paper.

sampled every 90 seconds.

In the empirical analysis, as guided by our Monte Carlo experiment, tests were at a 1% significance level using 1-min returns. The results show strong evidence of the presence of both a Brownian component and a jump part. On average across sectors, jumps occurred on one-third of days. For jumps in the SPY, jumps occurred on one-fifth of days. The smaller proportion recorded for the latter is the result of aggregation. In reducing idiosyncratic risk, aggregation causes stock specific jumps to be 'lost'. Although finite and infinite jumps characterize the jump component, finite jumps contribute more to the total jump part.[7] Finally, we document significant time variation in the rejection rates, with variations ranging between 12–35%. This suggests that the data generating process should allow for time variation, where generally an increase in the rejection of no jumps is accompanied with a decrease in the rejection of infinite jumps.

The remainder of the paper is structured as follows. Section 5.2 presents the theoretical background and outlines the test statistics. Section 5.3 describes the Monte Carlo setup and reports the simulated results. The empirical results including a time-variation exercise are reported in Section 5.4. Section 5.5 concludes.

5.2 Theoretical Background

Let the log-price X_t follow a semimartingale defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$

$$X_t = x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + X_t^d, \tag{5.1}$$

where x_0 is the initial value, b_s is the drift term being continuous and locally bounded, σ_s is a strictly positive and càdlàg stochastic volatility process, W_s is a standard Brownian motion, and X_t^d is a pure-jump component. The Blumenthal-Getoor (BG) index of X_t^d

[7] We also report the jump activity index estimated, $\widehat{\beta}$, as in Jing et al. (2012b). $\widehat{\beta}$ oscillates around 1.0, which confirms the presence of finite and infinite jumps across the individual stocks and SPY.

measures the degree of activity of small jumps and is defined as

$$\beta := \inf\left\{r; \sum_{0\leq s\leq t} |\Delta X_s|^r < \infty\right\}, \tag{5.2}$$

where $\Delta X_s = X_s - X_{s-} \neq 0$ if jumps are present. β serves as an indicator of the activity of jumps contained in X^d. The larger the β, the more active the jumps. A finite activity jump process such as a compound Poisson process has $\beta = 0$, whereas a β-stable process has an index equal to $\beta \in (0, 2)$. Finite variation corresponds to $0 < \beta < 1$ and infinite variation to $1 < \beta < 2$.

To construct the tests we define power, truncated power variations (see, Jacod, 2008; Mancini, 2001, 2009), and the estimator of the continuous part that is robust to infinite jump variation (see, Jacod and Todorov, 2014, JT, hereafter).

Let denote the power variation estimator as $B(p, \infty, \Delta_n)_t$, which is outlined as:

$$B(p, \infty, \Delta_n)_t = n^{p/2-1} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^p \xrightarrow{\mathbb{P}} \begin{cases} \mu_p \int_0^t |\sigma_s|^p ds, & \text{(No Jumps)}, \\ \infty, & \text{(With Jumps)}, \end{cases} \tag{5.3}$$

when $p > 2$, and $\mu_p \equiv E[|U|^p] = \frac{2^{p/2}}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right)$, where $U \sim \mathcal{N}(0, 1)$. $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$, with $\Delta_n = 1/n$ for $0 \leq i \leq n$. It is well-known that,

$$B(2, \infty, \Delta_n)_t = \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^2 \xrightarrow{\mathbb{P}} \underbrace{\int_0^t \sigma_s^2 ds}_{\text{Integrated Variation } (IV_t)} + \underbrace{\sum_{0\leq s\leq t} (\Delta X_s)^2}_{\text{Jump Variation}}. \tag{5.4}$$

We denote the truncated power variation as $B(p, \nu_n, \Delta_n)_t$:

$$B(p, \nu_n, \Delta_n)_t = n^{p/2-1} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^p \mathbb{1}_{\{|\Delta_i^n X| \leq \nu_n\}} \xrightarrow{\mathbb{P}} \int_0^t \mu_p |\sigma_s|^p ds \tag{5.5}$$

where $\nu_n = \alpha \Delta_n^{\varpi}$ is the truncation threshold and $\alpha > 0$ is expressed in units of the standard deviation of the continuous part of the process for a constant $\varpi \in (0, 1/2)$. When the jump of X_t is a Levy process with the Blumenthal-Getoor index $\beta \in [0, 2)$ (as

outlined in equation (5.2)), then the required condition is given by $\varpi \geq \frac{p-2}{2(p-\beta)}$, for $p > 2$.

The JT bias-corrected estimator, $C(u_n)_j^n$, takes the following form:

$$C_0(u_n)_j^n = 2k_n\Delta_n \sum_{j=0}^{\lfloor 1/k_n\Delta_n \rfloor - 1} \left(c_0(u_n)_j^n - \frac{1}{u_n^2 k_n}(\sinh(u_n^2 c_0(u)_j^n))^2 \right) \xrightarrow{\mathbb{P}} \int_0^t \sigma_s^2 ds \tag{5.6}$$

where

$$c_0(u_n)_j^n = -\frac{1}{u_n^2}\log\left(L(u_n)_j^n \vee \frac{1}{\sqrt{k_n}}\right)$$
$$L(u_n)_j^n = \frac{1}{k_n}\sum_{l=0}^{k_n-1}\cos\left(u_n(\Delta_{2jk_n+1+2l}^n X - \Delta_{2jk_n+2+2l}^n X)/\sqrt{\Delta_n}\right)$$

where the following conditions must satisfy $k_n\Delta_n^{1/2} \to 0$, $u_n \to 0$, $\sup_n \frac{k_n\Delta_n^{1/2}}{u_n^4} < \infty$. Possible choice for k_n and u_n are $k_n \asymp 1/\sqrt{\Delta_n}(\log(1/\Delta_n))^x$ and $u_n \asymp 1/(\log(1/\Delta_n))^{x'}$ for $0 < x' \leq x/4$, where $x \in (0, 1]$.

5.2.1 Test Statistics

Pure Jump Test

The pure-jump test of Kong et al. (2015, KLJ hereafter) is based on the realized characteristic function and checks whether the underlying process of a high frequency data set can be modelled as a pure-jump process. In finite sample terms, this test is superior to the Brownian test of Aït-Sahalia and Jacod (2010) and the modified version of the latter proposed by Jing et al. (2012a). The pure jump test is of the following hypotheses,

$$H_0 : \int_0^t \sigma_s^2 ds > 0 \quad \text{v.s.} \quad H_1 : \int_0^t \sigma_s^2 ds = 0.$$

The test takes the following form,

$$\mathcal{T}_t = \frac{C_0(u_n)_j^n - C_1(u_n)_j^n - \gamma_n\Delta^{1/2}}{2I_n^{1/2}\Delta_n^{1/2}} \xrightarrow{\mathcal{L}_s} \mathcal{N}(0, 1), \tag{5.7}$$

where

$$I_n \equiv \frac{1}{2}(I_{n,0} + I_{n,1})$$
$$I_{n,k} = 2k_n\Delta_n \sum_{j=0}^{\lfloor t/k_n\Delta_n \rfloor - 1} \left(c_k(u_n)_j^n - \frac{\sinh(u_n^2 c_k(u_n)_j^n)}{u_n^2(k_n-1)} \right)^2, \quad k = 0, 1$$

where γ_n is some chosen constant satisfying $\gamma_n \to 0$, and can be estimated as $\gamma_n = c^*/\log(u_n^2/\Delta_n)$, where $c^* = 0.2$, when the number of observations is moderate. $C_0(u_n)_j^n$ is estimated as in equation (5.6), whereas $C_1(u_n)_j^n$ can be defined as the $C_0(u_n)_j^n$, where $\Delta^n_{2jk_n+2l+1}X - \Delta^n_{2jk_n+2l}X$ is replaced by $\Delta^n_{2jk_n+2l}X - \Delta^n_{2jk_n+2l-1}X$, for $l = 1, \ldots, k_n - 1$. Finally, H_0 can be rejected if $\mathcal{T}_t < -z_\theta$ where $\mathbb{P}(\mathcal{N}(0,1) > z_\theta) = \theta$ for $\theta \in (0,1)$.

As the authors do not provide analytical results for cases that include microstructure noise, our prior is that, in the presence of microstructure noise, the test would be unable to recognize whether the small and frequent movements are Brownian or pure jump increments.

Infinite Activity Jump Test

The infinite activity jump test proposed by (Aït-Sahalia and Jacod, 2011, ASJ, hereafter) evaluates the following hypothesis:

$$H_0 : \Omega_T^{i\beta} \quad \text{v.s.} \quad H_1 : \Omega_T^f \cap \Omega_T^c,$$

where $\Omega_T^{i\beta}$ and Ω_T^f respectively refer to infinite and finite jump activity, and Ω_T^c is the diffusive part. The ASJ test is outlined as,[8]

$$S_t^{IA} = \frac{B(p', \varphi\nu_n, \Delta_n)_t B(p, \nu_n, \Delta_n)_t}{B(p', \nu_n, \Delta_n)_t B(p, \varphi\nu_n, \Delta_n)_t} \xrightarrow{\mathbb{P}} \varphi^{p'-p}. \tag{5.8}$$

[8] The convergence in probability holds only under the stated null hypothesis.

The CLT of this test takes the following form:[9]

$$(S_t^{IA} - \varphi^{p'-p})/\sqrt{\hat{\sigma}_t^2} \xrightarrow{\mathcal{L}_s} \mathcal{N}(0,1), \tag{5.9}$$

where

$$\begin{aligned}\hat{\sigma}_t^2 = \varphi^{2p'-2p}\Bigg(&\frac{B(2p,\nu_n,\Delta_n)_t}{(B(p,\nu_n,\Delta_n)_t)^2} + (1-2\varphi^{-p})\frac{B(2p,\varphi\nu_n,\Delta_n)_t}{(B(p,\varphi\nu_n,\Delta_n)_t)^2} + \frac{B(2p',\nu_n,\Delta_n)_t}{(B(p',\nu_n,\Delta_n)_t)^2} \\ &+ (1-2\varphi^{-p'})\frac{B(2p',\varphi\nu_n,\Delta_n)_t}{(B(p',\varphi\nu_n,\Delta_n)_t)^2} - 2\frac{B(p+p',\nu_n,\Delta_n)_t}{B(p,\nu_n,\Delta_n)_t B(p',\nu_n,\Delta_n)_t} \\ &- 2(1-\varphi^{-p}-\varphi^{-p'})\frac{B(p+p',\varphi\nu_n,\Delta_n)_t}{B(p,\varphi\nu_n,\Delta_n)_t B(p',\varphi\nu_n,\Delta_n)_t}\Bigg).\end{aligned}$$

We set $p = 3$, $p' = 4$, $\varphi = 2$, $\varpi = 0.2$, and $\alpha = 8$. As shown by ASJ, this test converges to $\varphi^{p'-p}$ (1) when the underlying process has infinitely (finitely) many jumps. When microstructure noise dominates, the test also converges to $\varphi^{p'-p}$. The implication is that in the presence of microstructure noise, the test cannot distinguish whether jumps have finite or infinite activity. z_θ denotes the *upper* θ-quantile of $\mathcal{N}(0,1)$, that is, $\mathbb{P}(\mathcal{N}(0,1) > z_\theta) = \theta$, for $\theta \in (0,1)$, the test rejects H_0 when $S_t^{IA} < \varphi^{p'-p} - z_\theta\sqrt{\hat{\sigma}_t^2}$.

Jump Test

As the difference between the two capture the contribution of jumps, Podolskij and Ziggel (2010, PZ, hereafter) uses the difference between power and truncated power variations to construct their test statistics. The test has the following hypotheses,

$$H_0 : \Omega_T^c \quad \text{v.s.} \quad H_1 : \Omega_T^j,$$

where Ω_T^c and Ω_T^j are respectively the set of a continuous and a discontinuous price path. We outline the test as,

$$S_t^J = \frac{T(\Delta_i^n X, p)_t}{\rho_t} \xrightarrow{\mathcal{L}_s} \mathcal{N}(0,1), \tag{5.10}$$

[9] The convergence in law holds only under the stated null hypothesis.

where,

$$T(\Delta_i^n X, p)_t = n^{(p-1)/2} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_i^n X|^p \left(\mathbb{1} - \eta_i \mathbb{1}_{\{|\Delta_i^n X| \leq \nu_n\}} \right), \quad p \geq 2, \tag{5.11}$$

$$\rho_t^2 = \text{Var}^*(\eta_i) B(2p, \nu_n, \Delta_n)_t. \tag{5.12}$$

η_i is a positive i.i.d. random variables with $\mathbb{E}[\eta_i] = 1$ and $\mathbb{E}[|\eta_i|^2] < \infty$. PZ suggest that η_i can be sampled from the distribution,

$$P^\eta = \frac{1}{2} \left(\delta_{1-\tau} + \delta_{1+\tau} \right),$$

where δ is the Dirac measure, and $\tau = 0.1$ or 0.05. ϖ is chosen such that it satisfies $\varpi \geq \frac{p-2}{2(p-\beta)}$, for $p > 2$.

PZ are amongst the few that account for microstructure noise.[10] Robust jump tests have been generally ignored by subsequent research (e.g Dumitru and Urga, 2012), although more recently Maneesoonthorn et al. (2020) show evidence that both the Aït-Sahalia et al. (2012) and Lee and Mykland (2012) tests lose power very rapidly. Given the evidence that jump tests are very sensitive to microstructure noise effects, in this paper we also study the finite sample properties of the robustified PZ test.

Let $Y_t = X_t + u_t$ be a contaminated price, and u_t an additive i.i.d. process. We assume that u_t has $\mathbb{E}[u_t] = 0$ and $\mathbb{E}[u_t^2] = \omega_t^2$, and $X_t \perp u_t$ ($\perp$ means stochastic independence). We pre-filter the data using the pre-averaging method of Jacod et al. (2009), so that the additive component is eliminated. The pre-averaging returns are defined as,

$$\Delta_i^n \bar{Y} = \sum_{j=0}^{K_n - 1} g\left(\frac{j}{K_n} \right) \Delta_{i+j}^n Y,$$

where $K_n/\sqrt{n} = \Theta + o\left(n^{-1/4}\right)$, for some $\Theta > 0$, and a non-zero real-valued function

[10] Other robustified tests are Aït-Sahalia et al. (2012), Jiang and Oomen (2008), and Lee and Mykland (2012).

$g(x) = (x \wedge 1 - x)$. The test is outlined as,

$$S_t^{Jnoise} = \frac{T^{noise}\left(\Delta_i^n \bar{Y}\right)_t}{\sqrt{\Gamma_t}} \xrightarrow{\mathcal{L}_s} \mathcal{N}(0,1), \tag{5.13}$$

$$T^{noise}\left(\Delta_i^n \bar{Y}, p\right)_t = n^{(p-2)/4} \sum_{i=0}^{\lfloor 1/\Delta_n \rfloor - K_n + 1} |\Delta_i^n \bar{Y}|^p \left(\mathbb{1} - \eta_i \mathbb{1}_{\{|\Delta_i^n \bar{Y}| \leq \nu_n\}}\right), \quad p \geq 2, \tag{5.14}$$

$$\Gamma_t = \text{Var}^*\left(\eta_i\right) n^{(p-2)/2} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor - K_n + 1} |\Delta_i^n \bar{Y}|^{2p} \mathbb{1}_{\{|\Delta_i^n \bar{Y}| \leq \nu_n\}}, \tag{5.15}$$

where $\alpha > 0$, $\varpi \in (0, 1/4)$ and η_i is estimated as described above.[11]. Both versions of the test reject H_0 when $S_t^J(S_t^{Jnoise}) > z_\theta$ where $\mathbb{P}(\mathcal{N}(0,1) > z_\theta) = \theta$ for $\theta \in (0,1)$.

5.3 Monte Carlo Study

All the tests but the robustified PZ test are derived under the assumption of noiseless prices. Therefore, market microstructure noise and finite sample are likely to affect the performance of these procedures. This section evaluates the performance of these tests under different types of microstructure noise and across several sampling frequencies. The aim is to find a trade-off between performance and bias.[12]

We follow Aït-Sahalia and Jacod (2010, 2011); Jing et al. (2012a) and use a Heston stochastic volatility model that allows for both finite and infinite jumps, while for the pure-jump process $\nu_t \equiv 0$, i.e. $dX_t = \theta_L dL_t$. The model is described as,

$$\begin{aligned} dX_t &= \nu_t^{1/2} dW_t^{(1)} + \theta_L dL_t \\ d\nu_t &= k(\eta - \nu_t)dt + \gamma \nu_t^{1/2} dW_t^{(2)}, \end{aligned} \tag{5.16}$$

with $\mathbb{E}[dW_t^{(1)} dW_t^{(2)}] = \rho dt$, $\eta = 0.25^2$, $\gamma = 0.5$, $k = 5$, $\rho = -0.5$, and the pure jump process is either a finite activity compound Poisson process or a β-stable infinite activity

[11] For more details about the pre-averaging methods employed here, the reader can consult (Podolskij and Ziggel, 2010, Section 4)

[12] Another reason to avoid very high sampling frequencies is related to price staleness. Less active stocks have higher proportions of zero returns, which might distort the finite sample properties of these type of tests.

process. The compound Poisson has intensity $\lambda = 1$ and jumps that are uniformly distributed on $\nu_t^{1/2}\sqrt{m}([-2,-1]\cup[1,2])$, where $m = 0.7$. When using compound Poisson Jumps, we set $\theta_L = 1.0$. We set the following values for $\beta = \{1.00,\ 1.25,\ 1.50\}$, and $\theta_L = 0.5$.

We add the measurement error, so that we do not observe the true price X_t, but rather we observe the contaminated price as follows,

$$Y_t = X_t + u_t, \tag{5.17}$$

where Y_t and X_t are respectively the contaminated and true log-price processes, and u_t is the measurement error with $\mathbb{E}[u_t] = 0$, and $\mathbb{E}[u_t^2] = \omega_t^2$. We follow Aït-Sahalia et al. (2012) and consider four settings,

$$u_t = \begin{cases} 0, & \text{(No noise)} \\ 2\nu_t^{1/2}\Delta_n^{1/2}u_t^A, & \text{(Gaussian noise)} \\ 2\nu_t^{1/2}\Delta_n^{1/2}u_t^B \Big/ \sqrt{\frac{df}{df-2}}, & \text{(T-distributed noise)} \\ 2\nu_t^{1/2}\Delta_n^{1/2}\left(u_t^A + u_t^B \Big/ \sqrt{\frac{df}{df-2}}\right), & \text{(Gaussian-T mixture noise),} \end{cases} \tag{5.18}$$

where u_t^A and u_t^B are mutually independent i.i.d. drawn from an $\mathcal{N}(0,1)$ distributed and a t-distribution with degree of freedom $df = 2.5$, respectively. The instantaneous standard deviations of the Gaussian noise and the t-distributed noise are twice that of the diffusive increment, i.e., $(\nu_t\Delta_n)^{1/2}$, and allow for temporal heteroskedasticity and dependence in u_t. The t-distributed noise captures the large bouncebacks commonly observed in transaction data as shown in Figure 5.1. We generate data for 50 days and use 3,000 replications, which encompass 150,000 simulated days.

Figure 5.2 plots the distribution of the tests using a diffusion process (equation (5.16)) contaminated with different types of microstructure noise. Of course, in the absence of noise effects (Figure 5.2a) all procedures are well-behaved, so giving good finite sample properties. When noise is added, there is a decrease in performance in all the tests, but

the robustified PZ test. Gaussian noise (Figure 5.2b), which is the most popular type of noise in the literature, produces the least distortions in all the tests statistics. By contrast, t-distributed (Figure 5.2c) and Gaussian-T mixture (Figure 5.2d) noise severely affect the standard procedures.

By comparison, the standard PZ test is upward biased, which results in a high rate of type I error. However, the bias induced by Gaussian noise is present only at high sampling frequencies, while the bias from t-distributed and Gaussian-T mixture noise distorts the distribution of the test even when returns are sampled every 90 seconds. On the other hand, the distribution of the robustified PZ test is largely the same whether or not the underlying process is contaminated with microstructure noise. The KLJ and ASJ tests are also more affected when the noise is non-Gaussian. Nevertheless, the distributions of both tests under the different types of microstructure noise suggest that sampling sparsely, and not at very low frequencies, can solve this issue without losing much power.

Figure 5.3 shows the distribution of the tests when the model is an infinite-jump diffusion process with $\beta = 1.0$. The distribution of both versions of the PZ test explodes to infinity in the presence of infinite jumps, which confirms the ability of these tests to identify jumps of infinite variation. The noise does not impact much the finite sample performance of the PZ tests. This is mainly because the noise-robust version is not affected by any type of noise as shown in Figure 5.2, and the standard PZ test is oversized in the presence of noise, which does not affect the power of the test but increases the spurious detection of jumps. The distribution of the KLJ test shows a similar behavior to the jump diffusion case, that is the test is downward biased in the presence of microstructure noise. This means that the test is unable to distinguish between the null and alternative hypothesis. As expected, the ASJ test is centered around 2 in the absence of noise, while in the presence of t-distributed and Gaussian-T mixture noise the distribution is somewhat shifted to the left of 2, which indicates an increase in the type II error.

The subsequent analysis of this section focuses on the finite sample properties of these procedures across different sampling frequencies using a significance level of $\theta = 0.01$. We sample the simulated data every 5, 30, 60, and 90 seconds, which correspond to

$\lfloor 1/\Delta_n \rfloor = \{4680,\ 780,\ 390,\ 260\}$ observations per day.

Pure Jump Test

Tables 5.1, 5.2 and 5.3 report the empirical sizes of the KLJ test given that all the underlying processes contain a diffusive component. When the underlying model is a diffusion process (Table 5.1), the KLJ test is slightly upward biased in the absence of microstructure effects. In presence of microstructure noise, as shown in Figure 5.2, the test shows a downward bias, which worsens when the noise is t-distributed or Gaussian-T mixture. However, the KLJ test gets closer to the theoretical size around 30 to 60 seconds, i.e $\lfloor 1/\Delta_n \rfloor = 780$ and 390, respectively. When the underlying process is a finite-jump diffusion process (Table 5.2), results are very similar to those found in Table 5.1, i.e. the KLJ test is slightly oversized in the absence of microstructure noise, while when any type of noise is contaminating the true price process, the test is undersized and approaches to the theoretical size around 30 to 60 second sampling frequency.

Table 5.3 reports the results using an infinite-jump diffusion process with jump activity index equal to $\beta = \{1.00,\ 1.25,\ 1.50\}$. In the absence of microstructure noise, the KLJ test is very close to its theoretical size. However, when using a $\beta = 1.50$, the test is downward biased. Our explanation to this finding is that as $\beta \to 2$, the increments are closer to those of a Brownian motion making it difficult for the test to recognize the true Brownian increments. When microstructure noise contaminates the true underlying process, we observe a decrease in the size of the test, which is recovered by sampling more sparsely, i.e. every 780 to 390 observations.

Table 5.4 reports the power of the KLJ test using a pure jump process. The microstructure noise is simulated as outlined in equation (5.18), where ν_t is replaced by the JT estimator, so that the variance of the noise is constant within a day, but changes from day-to-day. In the absence of microstructure noise, the power of the KLJ test is close to the theoretical power, and decreases both as the time interval increases and $\beta \to 2$. Unlike the noiseless case, the presence of Gaussian noise marginally decreases the power of the test when $\beta = 1.0$. However, for values of $\beta \in [1.25, 1.5]$ there is a sharp reduction in

the power of the test, which wanes away as $\lfloor 1/\Delta_n \rfloor = 780$ or smaller. By contrast, when the underlying process is contaminated by either t-distributed and Gaussian-T mixture noise, the KLJ test displays very low power at higher frequencies, which is magnified when $\beta \in [1.25, 1.5]$. The impact of these types of noise, unsurprisingly, diminishes by sampling more sparsely, with the power of the test being of similar magnitude to the no noise case when $\lfloor 1/\Delta_n \rfloor = 390$.

All in all, the KLJ test displays very good power in the absence of microstructure noise, and the power increases as $\Delta_n \to 0$. When microstructure noise is added, the KLJ test is undersized under the null and increases the type II error under the alternative hypothesis. Although this is observed for all types of microstructure noise, the Gaussian noise produces less severe distortions that are only observed at high sampling frequencies and as $\beta \to 2$. The distortions of t-distributed and Gaussian-T mixture noise are less obvious around $\lfloor 1/\Delta_n \rfloor = 390$, which suggests that sampling returns every 1-min results in a good trade-off between bias and performance.[13]

Infinite Activity Jump Test

Tables 5.1 and 5.2 report the power of the ASJ test, since the data generating process is a diffusion and finite-jump diffusion, respectively. In absence of microstructure noise, the power of the test declines as the time interval increases. For instance, when the data generating process is a diffusion and a finite-jump diffusion, the power of the test is respectively 0.940 and 0.936 using 1-min returns, i.e. $\lfloor 1/\Delta_n \rfloor = 390$. This suggests that the loss of power due to sampling more sparsely when prices are noiseless is very limited.

As shown by Aït-Sahalia and Jacod (2011), when microstructure noise dominates, the ASJ test converges to $\varphi^{p'-p}$, see also Figure 5.2, which means that the test is unable to distinguish between its null and alternative hypotheses. Nevertheless, we observe that irrespective of whether the process is a diffusion or finite-jump diffusion, all types of microstructure noise increase the type II error of the test. Whereas the effect of Gaussian

[13] As a robustness-check we also tried the test of Jing et al. (2012a), which is completely oversized for time intervals lower than 4,680 (5-seconds). Since this comparison is made in Kong et al. (2015) and our results are in line with theirs, we omit them here for brevity.

noise declines rapidly after $\lfloor 1/\Delta_n \rfloor = 4680$, the t-distributed and Gaussian-T mixture noise's distortions are of bigger magnitude, which are considerably reduced by sampling from $\lfloor 1/\Delta_n \rfloor = 390$ onwards.

Table 5.3 reports the size of the ASJ test, i.e. the results when the underlying process is an infinite-jump diffusion process. In the absence of microstructure noise effects, the size of the ASJ test is very close to the theoretical size. As $\beta \to 2$ and the number of intraday observations decreases, the test increases the type I error, as it is harder to distinguish between infinite jumps and Brownian increments. When the underlying process is contaminated with Gaussian noise and $\beta \in [1.0, 1.25]$, there are few differences with the noiseless case, which become more apparent with $\beta = 1.5$. The t-distributed and Gaussian-T mixture noise considerably increase the spurious number of rejections in the ASJ test. This is observed irrespective of the value of β, and as the sampling frequency decreases we observe a reduction in the number of spurious rejections, which bounces back at $\lfloor 1/\Delta_n \rfloor = 260$.

To summarize, we find that the ASJ test performs relatively well in the presence of Gaussian noise, but when the noise is t-distributed or Gaussian-T mixture, the type I error significantly increases. However, sampling sparsely provides a simple remedy at around 1-min returns, as we find that sampling lower than 1-min increases the number of spurious rejections.

Jump Test

The null hypothesis $H_0 : \Omega_T^c$ is satisfied when the underlying model follows a diffusion process, and therefore Table 5.1 reports the size of the PZ tests. When prices have no noise, the size of both versions of the PZ tests lie close to the theoretical size across all sampling frequencies. Whereas the noise-robust version of the test performs extremely well irrespective of the type of noise employed, the standard PZ test is very oversized under microstructure effects. Nonetheless, the standard PZ test seems to be somewhat robust to Gaussian noise, with a type I error close to the theoretical size, once returns are sampled every 30 seconds onwards. This highlights the sensitivity of the standard

version to the presence of microstructure noise, which increases the spurious detection of jumps making it impossible to identify true jump days.

Tables 5.2 and 5.3 report the results when the underlying process follows a finite- and infinite-jump diffusion, respectively. Given the above results, it is no surprise that the standard PZ shows power value of 1 across all sampling frequencies. This is mainly because microstructure noise causes the test to explode to infinity, so that it is unable to distinguish between the null and alternative hypotheses. For instance, the size of the test when $\lfloor 1/\Delta_n \rfloor = 780$ under t-distributed noise, is 0.744. This suggests that the test is likely to identify almost every day as a jump day, even though the true process has no discontinuities. On the other hand, the noise-robust PZ test, S_t^{Jnoise}, is not affected by the microstructure noise effects, showing almost identical finite sample performance relative to the no noise case.

Despite the capacity of the PZ tests to detect jumps of finite and infinite activity, it is surprising to find such a power level when $\beta = 1.50$. As $\beta \to 2$, the infinite jumps are akin to Brownian increments, whereby one would expect the tests to struggle in disentangling the small jumps from the Brownian increments. However, as shown in Figure 5.3, the distribution of both tests is completely shifted to the right, confirming their ability to detect small jumps.

In summary, we find that while the Gaussian noise is the least problematic type of microstructure effect for the standard PZ jump test, the t-distributed and the Gaussian-T mixture noise significantly sully the finite sample properties of the standard test, increasing the spurious detection of jumps. Conversely, the noise-robust PZ test performs well irrespective of both the presence and the type of microstructure noise, and succeed in detecting jumps of finite and infinite activity.[14]

[14] As a robustness-check we also tried the test of Aït-Sahalia et al. (2009) and their noise-robust version (Aït-Sahalia et al., 2012). Both tests show a sharp decrease in power after 5-seconds, with the standard test being badly affected by any type of noise. Results are available upon request.

5.4 Empirical Study

Data

Our empirical application considers 100 individual stocks from the S&P 500 basket and the SPDR S&P 500 ETF (SPY) for the period January 3, 2000 to December 30, 2016.[15] 10 representative stocks taken from each sector, vary in terms of liquidity and market capitalization, so ensuring a heterogeneous representation on each sector. Taking direction from our simulations, we sample data every 1-min, i.e. $\lfloor 1/\Delta_n \rfloor = 390$. The use of the robustified PZ test throughout our empirical analysis is motivated by its excellent performance both in the absence and presence of microstructure noise, and because the standard PZ test is very sensitive to non-Gaussian noise even when the data are sampled every 1-min. The subsequent analysis uses a significance level of $\theta = 0.01$.

Empirical Rejections by Sector and Market Capitalization

Panel A of Table 5.5 reports total number of rejections standardized by the number of days as an average for all the stocks on each sector and the SPY. Panel B shows the contribution of the continuous and discontinuous component to the total variance estimated as $B\left(2, \infty, \Delta_n\right)_t$, and the index of jump activity, $\widehat{\beta}$, estimated as in Jing et al. (2012b). We estimate the variables of Panel B as follows: $C_t = B\left(2, \infty, \Delta_n\right)_t \cdot \mathbb{1}(no\ jumps) + B\left(2, \nu_n, \Delta_n\right)_t \cdot \mathbb{1}(jumps)$. Therefore, $J_t = B\left(2, \infty, \Delta_n\right)_t - C_t$. Finally, $C_T = \frac{\sum_{t\in(0,T]} C_t}{\sum_{t\in(0,T]} C_t + J_t}$ and $J_T = \frac{\sum_{t\in(0,T]} J_t}{\sum_{t\in(0,T]} C_t + J_t}$. The contribution of finite and infinite jumps to the total jump component are obtained as $FJ_t = J_t \cdot \mathbb{1}(finite\ jumps)$ and $IJ_t = J_t \cdot \mathbb{1}(infinite\ jumps)$. Thus, $FJ_T = \frac{\sum_{t\in(0,T]} FJ_t}{\sum_{t\in(0,T]} J_t}$ and $IJ_T = \frac{\sum_{t\in(0,T]} IJ_t}{\sum_{t\in(0,T]} J_t}$.[16]

The empirical results in Table 5.5 show strong evidence for the presence of a Brownian component with an average rejection rate in favor of a pure jump process of about 2%. Although those results are above the theoretical size ($\theta = 0.01$), the values are in line with our Monte Carlo exercise, which shows that the KLJ test is mildly oversized under the

[15] Our data is obtained from TickData, Inc. The data provider uses proprietary data filters. We sample down to 1-min using the previous tick interpolation.

[16] The indicator function for jumps and infinite jumps were obtained using the robustified version of the PZ test, and the ASJ test, respectively.

null, for relatively low frequencies such as 1-min. This finding rules out the theory of a pure jump process, and confirms the relevance of the Brownian component for modelling the diffusive part.

The first column of Panel A reports the rejection rates of the robustified PZ test, which confirms the presence of jumps in both the SPY and individual stocks across sectors. Whereas jumps are observed in 32% of the days as an average across the sectors, the number of days with jumps in the SPY account for only 19% of the sample data. The low rejection rate in the SPY is due to the fact that aggregation of information reduces the idiosyncratic risk, whereby stock specific jumps are diversified away.

Having established evidence for the presence of a Brownian motion and a jump component, we now evaluate whether the jump part is of finite and/or infinite activity. The second column of Table 5.5 reports the rejection rate of the ASJ test for both the SPY and the individual stocks classified by sector. The ASJ test provides evidence for the presence of both types of jumps, where the rejection rate for the SPY and the sector average is 80.1% and 64.2%, indicating stronger support for finite activity jumps. The index of jump activity, $\widehat{\beta}$, oscillates around 1.0 for all the sectors and SPY. This reaffirms that both finite and infinite jump activity characterize the jump component of our 1-min stock data.

Provided that the alternative hypothesis of the ASJ could mean finite jumps or just a Brownian motion, we report on Panel B (Table 5.5) the contribution of these variables to total variance. The Brownian component contributes about 86% of the total variance as an average across all sectors and SPY. From the 14% of the jump part, 75% corresponds to finite activity jumps. The evidence of both types of jumps is supported such that finite jumps are normally related to macroeconomic announcements and stock specific news, which are more likely to produce spillovers. On the other hand, infinite jumps can be linked to continuously adjusting market microstructure dynamics, which can be attributable but not limited to inventory allocation, price-contingent trading, stop-loss and limit-profit orders, among others.

Table 5.6 reports the rejection rates and contribution of the Brownian and jump

component to total variance, where the stocks are classified by sector and market capitalization. In general, big and large cap have the greatest trading liquidity, therefore our analysis aims to examine the sensitivity of the rejection rates to the size of the company and level of liquidity. Small cap stocks display a rejection rate for the null of no jumps and infinite activity jumps that is about 5% bigger and 4% smaller than big cap companies. This directly translates in a bigger contribution of jumps to total variance, which has important implications for asset allocation and risk management. For instance, a risk averse investor might be expected to avoid investments with large unforeseeable movements. Furthermore, Jiang and Yao (2013) show that small and illiquid stocks have higher jump returns to the extent that cross-sectional differences in jumps fully account for the size and illiquidity effects.

These findings provide new empirical evidence about the frequency of jumps and their contribution to total variance. Previous studies have generally reported that the proportion of jumps is found to be around 10%, while the contribution to total variance does not exceed 7% (e.g. Huang and Tauchen, 2005). By contrast, our results indicate that jumps are not rare events; therefore, the inclusion of both infinite and finite jumps affords a better explanation for the excess kurtosis observed in intraday stock data than only considering compound Poisson jumps. These results shed also light about the appropriate specification for modelling stock price data, which has important implications for empirical asset pricing and equity research, as the diffusive and jump part imply different hedging strategies (e.g Bollerslev and Todorov, 2011a,b).

Time-varying Rejections

Across the length of our dataset (2000-2016), financial markets experienced several crises periods (dot-com, sub-prime, European sovereign debt) and the Brexit referendum. These along with the development of efficient trading systems, including electronic platforms and algorithmic trades, have contributed to more frequent and faster reactions to changes in the market, which could easily generate time variation in statistical properties. To further investigate this issue, we follow Erdemlioglu et al. (2015) and fit a probit

model to the daily series of 1% level rejection indicators $(0, 1)$ for the PZ and ASJ test. We omit the KLJ test, as the low rejection rate refrains any time variation, with fitted values usually fluctuating less than 5%. We use a sixth order polynomial of time in a probit model, where regressors are a constant, and time trends up to a sixth power. We have orthogonalized and standardized them to have unit variance.

Figure 5.4 plots the fitted values of one representative stock from each sector and the SPY using 1% rejection rates for the robustified PZ and ASJ tests. We find significant time variation in the PZ and ASJ tests, with on average 3 significant polynomial coefficients per regression. There is generally a negative correlation in the evolution of the time-varying rejections. In other words, an increment in the number of jump days, results usually in an increase of infinite jumps vis-à-vis finite jumps.

The most notorious feature of these series is the increase in the rejection of the null of no jumps around 2002, 2008 and 2016. These periods experienced the dot-com bubble with the NASDAQ falling by 78% in October 2002, the 2008 crisis with the Dow Jones losing 777.68 points in September 2008, and on January 2016 due to crude oil low prices the Dow Jones fell 565 points.

5.5 Conclusions

This paper examines the sensitivity of novel test statistics to different types of microstructure noise. These procedures test for the presence of a Brownian motion, jumps, and whether jumps have finite or infinite activity. Knowing the existence of these components facilitate the correct modelling of stock price data, which is crucial for option pricing, risk management, and portfolio allocation. We perform a thorough Monte Carlo experiment that allows for Gaussian, t-distributed, and Gaussian-T mixture noise, as well as various sampling frequencies. As most of these tests are not robust to microstructure effects, our simulations exploit their finite sample properties, providing guidance about the bias induced by each type of noise and the right choice of sampling frequency.

The Monte Carlo experiment shows that the presence of microstructure noise skews

the distribution of the test statistics, with the exception of the robustified PZ test, as their asymptotic distributions are derived under the assumption of noiseless prices. This results in an increase in the type I or type II error. Specifically, we find that Gaussian noise only affects the distribution of the tests at very high frequencies, while t-distribution and Gaussian-T mixture noise completely distort the performance of the tests. These findings do not apply for the robustified PZ test, which performs extremely well under any type of microstructure noise. Sampling returns every 30- and 60-seconds when the noise is respectively Gaussian and t-distributed or Gaussian-T mixture reduces considerably the microstructure noise effects.

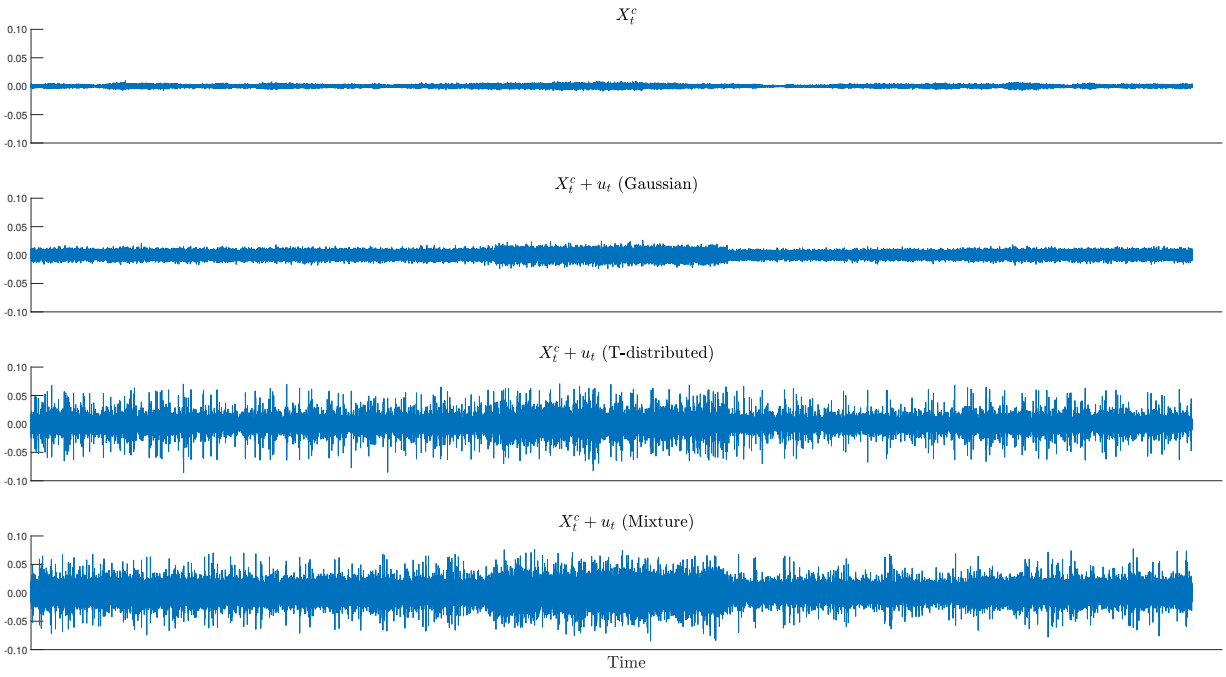
We apply these tests to 100 individual stocks and SPY using 17 years of data sampled every 1-min. The empirical results indicate strong evidence for the presence of a Brownian motion, ruling out the theory of a pure jump process. We also find evidence for the presence of jumps, which are not rare events. For instance, jumps account for one-third of the days as an average across sectors compared to one-fifth of the days for SPY. The smaller proportion of jump days in the SPY is because idiosyncratic jumps are diversified away from the Index. The Jump part is characterized by finite and infinite jumps, with finite activity jumps comprising about 75% of the total jump component. However, the contribution of jumps to total variance ranges between 10–22% for the sectors and is close to 5% for SPY. This suggests that the data are more consistent with both finite and infinite activity jumps, where finite/infinite jumps are generally associated to news releases/continuously adjusting market microstructure dynamics.

We also find significant time variation in rejection rates. These variations range between 12–35%, where generally an increase in the rejection of no jumps is associated with a decrease in the rejection of infinite jumps.

Finally, we recommend that the most appropriate specification for modelling stock price data should allow for a Brownian motion and jumps of finite and infinite activity, with the jump component having time-varying both activity and intensity.

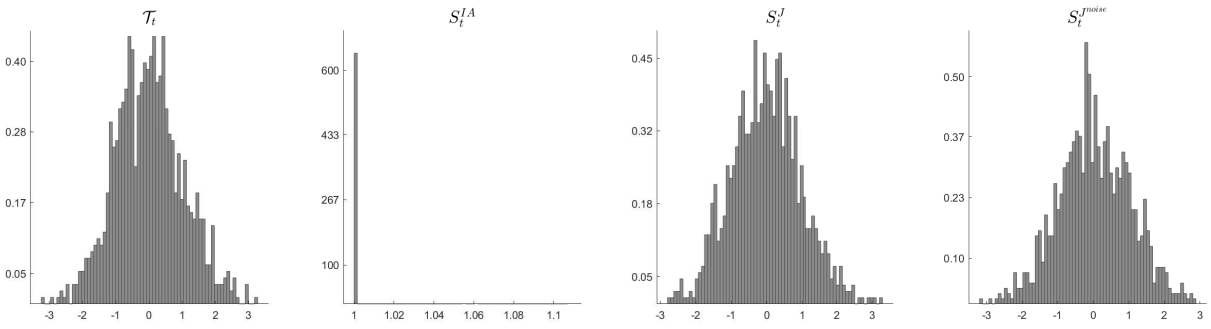
Appendix 5.A Tables and Figures

Figure 5.1: A Realization of the True and Contaminated Continuous Part of the log-price Increments

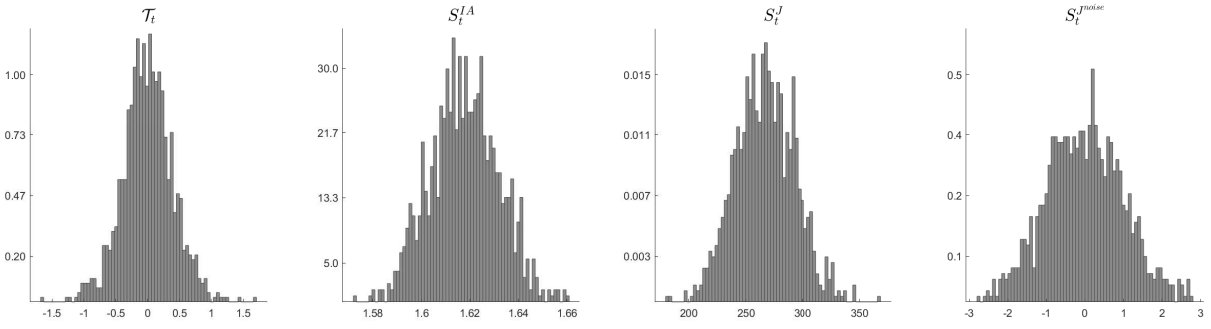


Note: The figures depicts a realization of the noisy continuous part of the log price $(X_t^c + u_t)$. From top to bottom, the continuous part is contaminated with no noise, Gaussian noise, t-distributed noise, and Gaussian-T mixture noise.

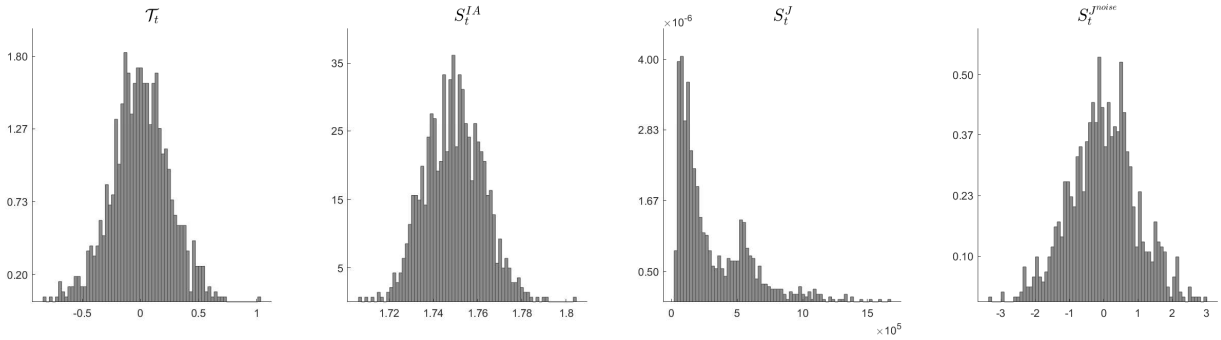
Figure 5.2: Distribution of the Test Statistics using a Diffusion Process with Different Types of Noise



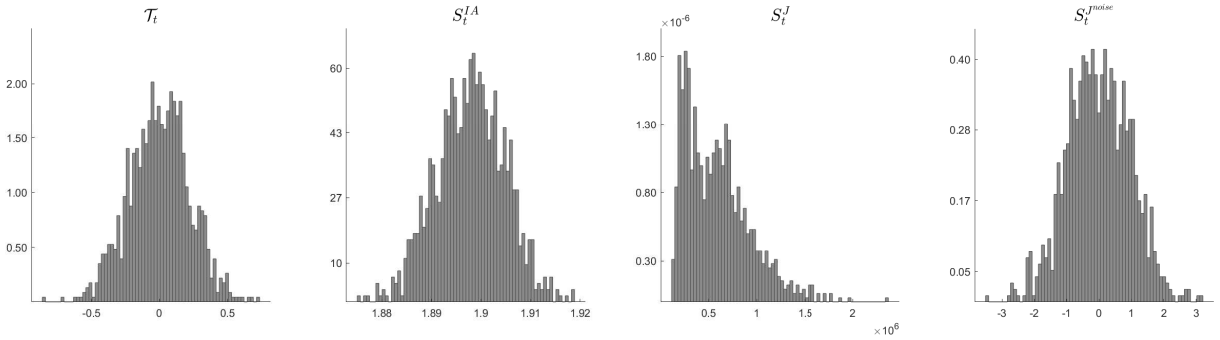
(a) X_t^c $(u_t = 0)$



(b) $X_t^c + u_t$ (Gaussian)



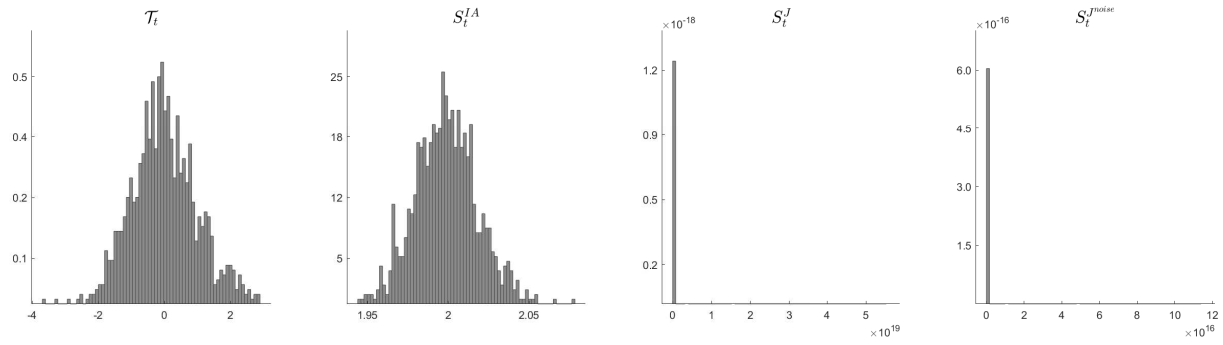
(c) $X_t^c + u_t$ (t-distributed, $df = 2.5$)



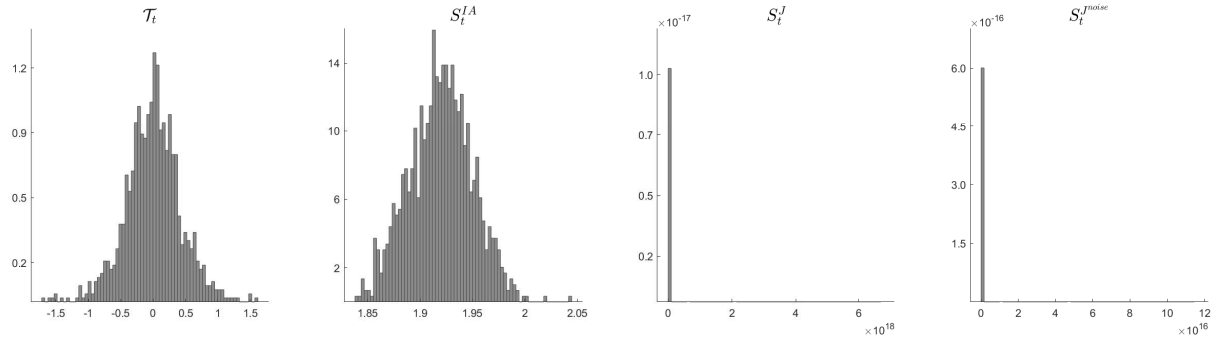
(d) $X_t^c + u_t$ (Gaussian-T Mixture)

Note: The figure plots the simulated distribution of the different tests under a diffusion process (equation (5.16)) with (a) no noise, (b) Gaussian noise, (c) t-distributed noise with 2.5 degrees of freedom, and (d) Gaussian-T mixture noise. X_t^c denotes the diffusive component.

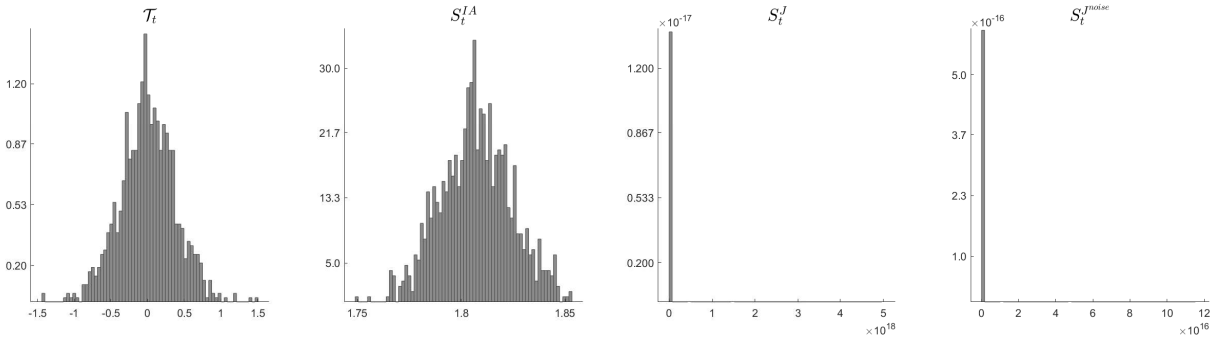
Figure 5.3: Distribution of the Test Statistics using an Infinite Jump Diffusion Process with Different Types of Noise



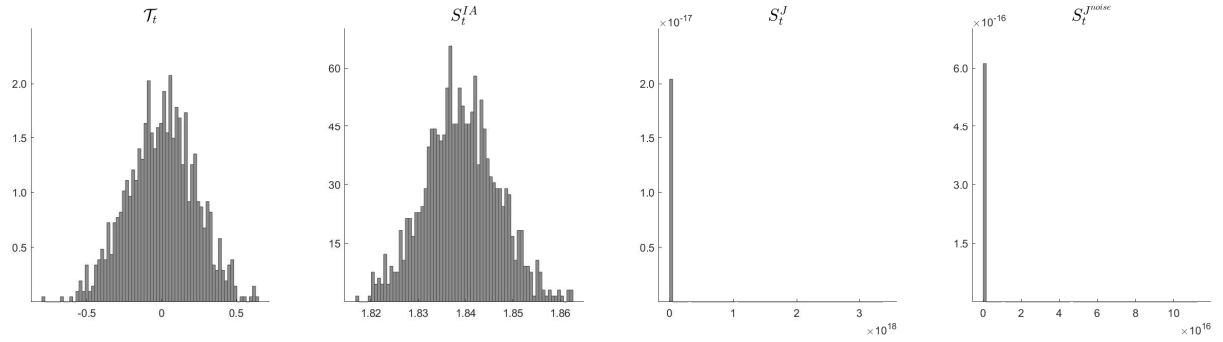
(a) $X_t^c + X_t^d$ $(u_t = 0)$



(b) $X_t^c + X_t^d + u_t$ (Gaussian)



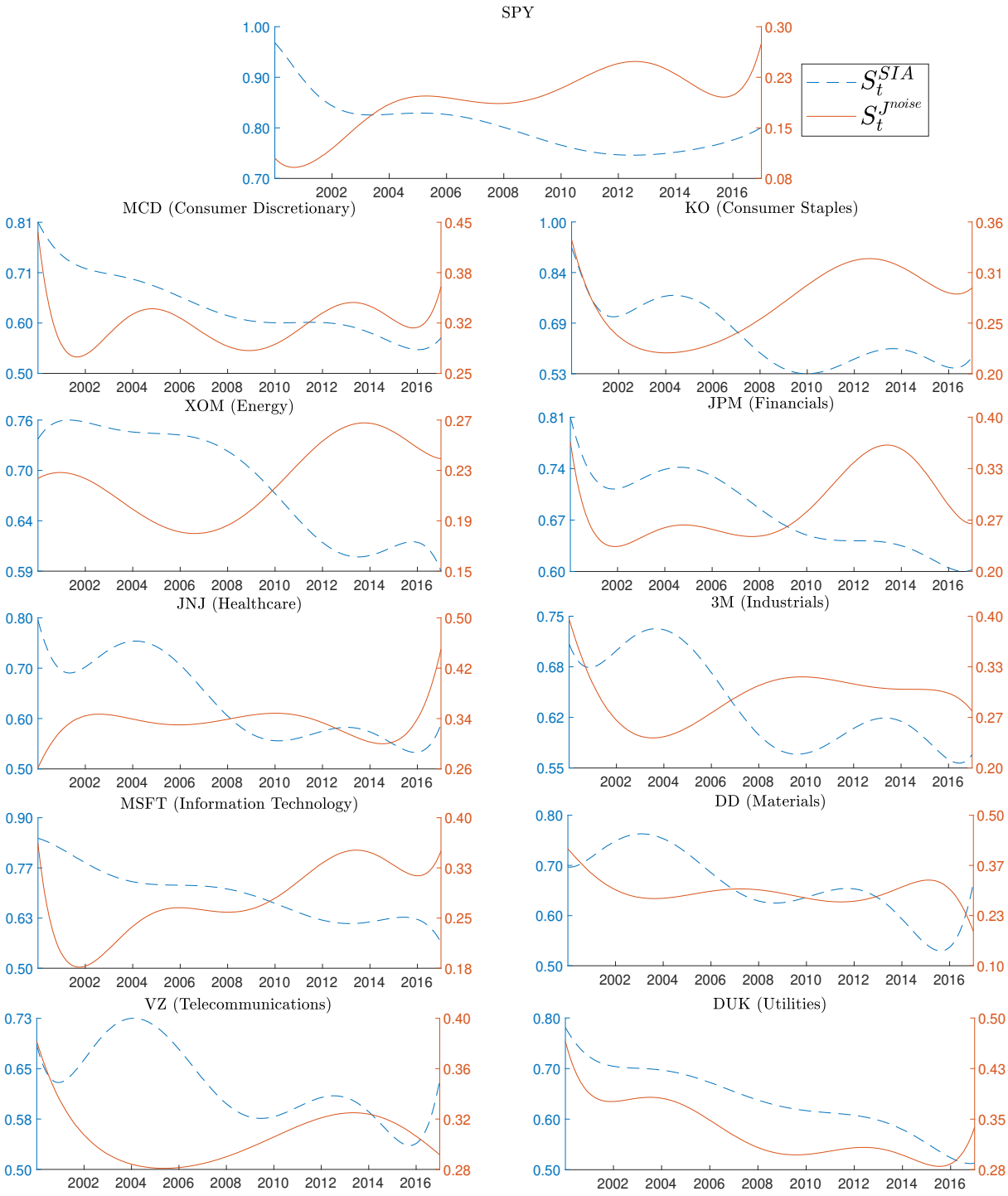
(c) $X_t^c + X_t^d + u_t$ (t-distributed, $df = 2.5$)



(d) $X_t^c + X_t^d + u_t$ (Gaussian-T Mixture)

Note: The figure plots the simulated distribution of the different tests under a infinite jump diffusion process (equation (5.16) and $\beta = 1.0$) with (a) no noise, (b) Gaussian noise, (c) t-distributed noise with 2.5 degrees of freedom, and (d) Gaussian-T mixture noise. X_t^c and X_t^d denote the diffusive and pure jump components.

Figure 5.4: Time Variation in Rejection Rates



Note: The figure depicts the time variation in rejection rates predicted by a probit model with a 6th order polynomial in time. The left (right) y-axis of each subplot denotes the probability of rejection over time for the ASJ (robustified PZ test).

Table 5.1: Monte Carlo Rejection Rates under a Diffusion Process

$\lfloor 1/\Delta_n \rfloor$	4,680	780	390	260	4,680	780	390	260
	No Noise				Gaussian Noise			
$\mathcal{T}_t$	0.028	0.032	0.035	0.038	0.003	0.017	0.025	0.030
S_t^J	0.012	0.014	0.016	0.015	0.204	0.023	0.019	0.019
$S_t^{J^{noise}}$	0.010	0.009	0.007	0.008	0.010	0.008	0.007	0.009
S_t^{SIA}	0.999	0.971	0.940	0.922	0.850	0.977	0.947	0.926
	t-distributed Noise				Gaussian-T Mixture Noise			
$\mathcal{T}_t$	0.005	0.014	0.019	0.024	0.000	0.004	0.009	0.012
S_t^J	1.000	0.744	0.307	0.163	1.000	0.878	0.388	0.204
$S_t^{J^{noise}}$	0.010	0.009	0.007	0.009	0.009	0.008	0.007	0.009
S_t^{SIA}	0.285	0.687	0.912	0.925	0.662	0.817	0.917	0.927

Note: The table reports rejection rates across sampling frequencies for the four test statistics outlined in Section 5.2.1. Under a diffusion process the KLJ ($\mathcal{T}_t$) and PZ ($S_t^{J^{noise}}$, S_t^J) tests report their empirical size, while the ASJ test (S_t^{SIA}) reports its empirical power. $\lfloor 1/\Delta_n \rfloor$ represents the number of intraday observations per day and the significance level is $\theta = 0.01$.

Table 5.2: Monte Carlo Rejection Rates under a Finite Jump-Diffusion Process

$\lfloor 1/\Delta_n \rfloor$	4,680	780	390	260	4,680	780	390	260
	No Noise				Gaussian Noise			
$\mathcal{T}_t$	0.028	0.030	0.032	0.035	0.003	0.016	0.023	0.027
S_t^J	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.999
$S_t^{J^{noise}}$	0.995	0.971	0.956	0.928	0.995	0.971	0.943	0.910
S_t^{SIA}	0.999	0.964	0.936	0.916	0.795	0.954	0.937	0.920
	T-Distributed Noise				Gaussian-T Mixture Noise			
$\mathcal{T}_t$	0.004	0.014	0.019	0.021	0.000	0.004	0.008	0.011
S_t^J	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.999
$S_t^{J^{noise}}$	0.995	0.971	0.955	0.928	0.993	0.963	0.945	0.908
S_t^{SIA}	0.286	0.684	0.908	0.910	0.560	0.762	0.899	0.908

Note: The table reports rejection rates across sampling frequencies for the four test statistics outlined in Section 5.2.1. Under a finite-jump diffusion process the KLJ test ($\mathcal{T}_t$) reports its empirical size, while the PZ ($S_t^{J^{noise}}$, S_t^J) and the ASJ (S_t^{SIA}) tests report their empirical power. $\lfloor 1/\Delta_n \rfloor$ represents the number of intraday observations per day and the significance level is $\theta = 0.01$.

Table 5.3: Monte Carlo Rejection Rates under an Infinite Jump-Diffusion Process

	$\lfloor 1/\Delta_n \rfloor$	4,680	780	390	260	4,680	780	390	260
		No Noise				Gaussian Noise			
$\beta = 1.00$	$\mathcal{T}_t$	0.017	0.013	0.013	0.012	0.003	0.010	0.011	0.012
	S_t^J	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$S_t^{J^{noise}}$	1.000	0.995	0.986	0.977	1.000	0.995	0.987	0.977
	S_t^{SIA}	0.010	0.015	0.025	0.032	0.018	0.016	0.028	0.039
$\beta = 1.25$	$\mathcal{T}_t$	0.012	0.011	0.011	0.011	0.003	0.010	0.010	0.010
	S_t^J	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$S_t^{J^{noise}}$	0.999	0.985	0.968	0.950	0.999	0.985	0.968	0.950
	S_t^{SIA}	0.010	0.014	0.023	0.038	0.020	0.023	0.026	0.049
$\beta = 1.50$	$\mathcal{T}_t$	0.009	0.004	0.002	0.001	0.001	0.002	0.002	0.001
	S_t^J	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$S_t^{J^{noise}}$	0.990	0.959	0.935	0.910	0.990	0.959	0.935	0.910
	S_t^{SIA}	0.025	0.036	0.049	0.068	0.142	0.047	0.040	0.042
		t-distributed Noise				Gaussian-T Mixture Noise			
$\beta = 1.00$	$\mathcal{T}_t$	0.004	0.010	0.011	0.010	0.000	0.004	0.007	0.008
	S_t^J	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$S_t^{J^{noise}}$	1.000	0.994	0.987	0.977	1.000	0.995	0.987	0.977
	S_t^{SIA}	0.179	0.054	0.030	0.048	0.101	0.067	0.045	0.083
$\beta = 1.25$	$\mathcal{T}_t$	0.002	0.008	0.009	0.009	0.000	0.003	0.005	0.007
	S_t^J	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$S_t^{J^{noise}}$	0.999	0.984	0.969	0.949	0.999	0.984	0.969	0.950
	S_t^{SIA}	0.195	0.040	0.037	0.049	0.115	0.050	0.036	0.057
$\beta = 1.50$	$\mathcal{T}_t$	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000
	S_t^J	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$S_t^{J^{noise}}$	0.992	0.960	0.909	0.909	0.992	0.960	0.908	0.910
	S_t^{SIA}	0.638	0.083	0.041	0.060	0.535	0.125	0.059	0.067

Note: The table reports rejection rates across sampling frequencies for the four test statistics outlined in Section 5.2.1. Under an infinite-jump diffusion process the KLJ ($\mathcal{T}_t$) and ASJ (S_t^{SIA}) tests report their empirical size, while the PZ tests ($S_t^{J^{noise}}$, S_t^J) report their empirical power. $\lfloor 1/\Delta_n \rfloor$ represents the number of intraday observations per day, β is the jump activity index, and the significance level is $\theta = 0.01$.

Table 5.4: Monte Carlo Rejection Rates under a Pure Jump Process

$\lfloor 1/\Delta_n \rfloor$	4,680	780	390	260	4,680	780	390	260
	No Noise				Gaussian Noise			
$\beta = 1.00$	0.999	0.980	0.942	0.939	0.984	0.976	0.938	0.937
$\beta = 1.25$	0.997	0.946	0.909	0.877	0.854	0.927	0.883	0.861
$\beta = 1.50$	0.994	0.901	0.880	0.809	0.307	0.836	0.814	0.764
	T-Distributed Noise				Mixture Noise			
$\beta = 1.00$	0.578	0.942	0.927	0.920	0.211	0.883	0.914	0.892
$\beta = 1.25$	0.283	0.825	0.816	0.851	0.037	0.646	0.743	0.814
$\beta = 1.50$	0.109	0.598	0.666	0.748	0.007	0.265	0.505	0.678

Note: The table reports rejection rates across sampling frequencies for the KLJ test. Under a pure jump process the KLJ test reports its empirical power. $\lfloor 1/\Delta_n \rfloor$ represents the number of intraday observations, β is the jump activity index per day, and the significance level is $\theta = 0.01$.

Table 5.5: Empirical Test Rejections and Contribution Total Variance by Sector

	Panel A: Test Rejections			Panel B: Components				
	S_n^{Jnoise}	S_n	$\mathcal{T}_n$	C_T	J_T	FJ_T	IJ_T	$\widehat{\beta}$
SPY	0.194	0.801	0.021	0.952	0.048	0.702	0.298	1.056
Consumer Discretionary	0.333	0.623	0.027	0.837	0.163	0.775	0.225	0.905
Consumer Staples	0.322	0.616	0.035	0.820	0.180	0.722	0.278	1.151
Energy	0.292	0.655	0.023	0.882	0.118	0.745	0.255	0.794
Financials	0.299	0.666	0.027	0.878	0.122	0.821	0.179	0.831
Healthcare	0.342	0.625	0.031	0.831	0.169	0.743	0.257	1.032
Industrials	0.322	0.643	0.029	0.864	0.136	0.702	0.298	0.956
Information Technology	0.297	0.684	0.025	0.903	0.097	0.784	0.216	0.800
Materials	0.326	0.655	0.024	0.855	0.145	0.790	0.210	0.797
Telecommunications	0.344	0.639	0.021	0.800	0.200	0.767	0.233	0.957
Utilities	0.344	0.612	0.027	0.778	0.222	0.747	0.253	1.267

Note: The table reports in two panels the rejection rates and the contribution of the continuous and discontinuous part to total variance estimated as $B\left(2,\infty,\Delta_n\right)_t$. Panel A presents the number of rejections for each test, which is standardized by the total number of days in the sample data. The rejection rate is the average across the 10 stocks of each sector. Panel B depicts the contribution of the continuous and discontinuous part to total variance, as well as the contribution of finite and infinite activity jumps to the total jump component, J_T. $\widehat{\beta}$ is an estimate of the Glumenthal-Gettor index as in Jing et al. (2012b). $C_t = B\left(2,\infty,\Delta_n\right)_t \cdot \mathbb{1}(no\ jumps) + B\left(2,\nu_n,\Delta_n\right)_t \cdot \mathbb{1}(jumps)$. $J_t = B\left(2,\infty,\Delta_n\right)_t - C_t$. Hence, $C_T = \frac{\sum_{t\in(0,T]} C_t}{\sum_{t\in(0,T]} C_t + J_t}$ and $J_T = \frac{\sum_{t\in(0,T]} J_t}{\sum_{t\in(0,T]} C_t + J_t}$. The contribution of finite and infinite jumps to the total jump component are obtained as $FJ_t = J_t \cdot \mathbb{1}(finite\ jumps)$ and $IJ_t = J_t \cdot \mathbb{1}(infinite\ jumps)$. Thus, $FJ_T = \frac{\sum_{t\in(0,T]} FJ_t}{\sum_{t\in(0,T]} J_t}$ and $IJ_T = \frac{\sum_{t\in(0,T]} IJ_t}{\sum_{t\in(0,T]} J_t}$

Table 5.6: Empirical Rejection Rates and Contribution to Total Variance Classified by Market Capitalization and Sector

	Panel A: Rejections			Panel B: Components			
	S_n^J	S_n	$\mathcal{T}_n$	C_T	J_T	FJ_T	IJ_T
	Big Market Cap Companies						
Consumer Discretionary	0.308	0.649	0.031	0.875	0.125	0.774	0.226
Consumer Staples	0.284	0.650	0.037	0.849	0.151	0.732	0.268
Energy	0.277	0.667	0.026	0.896	0.104	0.722	0.278
Financials	0.287	0.697	0.024	0.898	0.102	0.832	0.168
Healthcare	0.323	0.631	0.035	0.854	0.146	0.688	0.312
Industrials	0.305	0.660	0.030	0.879	0.121	0.683	0.317
Information Technology	0.284	0.692	0.027	0.918	0.082	0.745	0.255
Materials	0.310	0.668	0.024	0.865	0.135	0.778	0.222
Telecommunications	0.294	0.675	0.022	0.845	0.155	0.754	0.246
Utilities	0.329	0.628	0.028	0.834	0.166	0.721	0.279
	Small Market Cap Companies						
Consumer Discretionary	0.358	0.597	0.024	0.800	0.200	0.777	0.223
Consumer Staples	0.360	0.582	0.032	0.791	0.209	0.711	0.289
Energy	0.310	0.641	0.019	0.864	0.136	0.773	0.227
Financials	0.312	0.635	0.030	0.857	0.143	0.811	0.189
Healthcare	0.361	0.619	0.027	0.808	0.192	0.799	0.201
Industrials	0.340	0.626	0.027	0.849	0.151	0.720	0.280
Information Technology	0.310	0.677	0.023	0.888	0.112	0.822	0.178
Materials	0.343	0.642	0.024	0.845	0.155	0.803	0.197
Telecommunications	0.395	0.603	0.019	0.756	0.244	0.780	0.220
Utilities	0.360	0.596	0.026	0.721	0.279	0.774	0.226

Note: See Notes to Table 5.5. The top (bottom) panel reports the results for the biggest (smallest) 5 companies of each sector selected by market capitalization.

Chapter 6

The Contribution of Jump Signs and Activity to Forecasting Stock Price Volatility

6.1 Introduction

Modelling and forecasting asset return volatility is central to asset pricing, portfolio optimization and risk management. The introduction and use of high-frequency data provide a framework for directly measuring and capturing the main stylized facts of volatility. Realized volatility (RV), a non-parametric measure calculated as the sum of intraday squared returns, provides a consistent estimator of the quadratic variation when the price process contain discontinuities or jumps.[1]

In relation to volatility forecasting, the seminal work of Andersen et al. (2007a) suggests that the jump component is both highly important and distinctly less persistent than the continuous component. Thus, treating rough jumps separately results in significant improvements in out-of-sample volatility forecasts, not least because many significant jumps are associated with specific macroeconomic announcements. However, recent empirical evidence that classifies jumps into finite and infinite activity jumps (e.g. Aït-

[1]Early adoption of RV in modelling and forecasting featured in the work of Andersen and Bollerslev (1998a), Andersen et al. (2001b, 2003, 2005), inter alios.

Sahalia and Jacod, 2012), presents a new question as to whether such different types of jumps are equally important in the prediction of future volatility.[2]

A large literature examines the role of jumps in volatility forecasting. However, much of that literature focuses on signed jumps, and does not separate finite jumps from infinite jumps. It also tends to use 300-second returns, rather than higher frequencies such as 5- or 60-second returns, in order to mitigate the impact of the market microstructure noise. Whether for jumps or signed jumps, the literature provides mixed evidence regarding their value added in forecasting. One side of the literature reports gains in forecasting from incorporating jumps. Andersen et al. (2007a) find that separating jumps from the continuous volatility component improves out-of-sample forecasts. Corsi et al. (2010) show that the use of a threshold bipower estimator to calculate the jump component affords substantial out-of-sample gains. Patton and Sheppard (2015) argue that volatility is strongly related to the volatility of past negative returns, and show that models incorporating signed jumps lead to significantly better out-of-sample forecast performance. Duong and Swanson (2015) identify large and small jumps using higher order power variations, and find that small jumps are more important for forecasting volatility than large jumps.

Another side of the literature finds that jumps do not significantly improve volatility forecasts. For instance, Forsberg and Ghysels (2007), Giot and Laurent (2007), Martens et al. (2009), Busch et al. (2011), Sévi (2014), Prokopczuk et al. (2016) review the use of jumps and signed jumps to forecast future volatility. Their results suggest that the inclusion of jumps and signed jumps improves the in-sample fit of models, but generate no significant out-of-sample forecasting gains.

The current paper contributes to the literature in a number of ways. First, we show how jumps may be decomposed into signed, finite and infinite activity jumps. We identify the finite and infinite jump components using the intersection of the ABD jump test and the SFA finite activity jump tests (Aït-Sahalia and Jacod, 2011; Andersen et al., 2007b).

[2]Other research considering the role of finite jumps can be found in Huang and Tauchen (2005), Lee and Mykland (2008), Aït-Sahalia (2004), Aït-Sahalia and Jacod (2009), Dumitru and Urga (2012). For infinite jumps see Todorov and Tauchen (2010), Aït-Sahalia and Jacod (2009, 2014), and the extensive references therein.

Duong and Swanson (2015) use higher order power variations to decompose jumps into large and small jumps, and examine their role in predicting the volatility of returns. By contrast, we use a more robust tests based decomposition of days with significant jumps into ones with finite or infinite activity jumps. As noted by Aït-Sahalia and Jacod (2014), in finite samples the estimated jumps based on higher-order power variations are often poor measure of actual jumps. Second, we develop versions of the ABD test and realized semivariance measures that are robust to microstructure noise, and perform well at high-frequency. The noise robust semivariance measures are modifications of the two-scale realized variance measure of Zhang et al. (2005). Third, we present new empirical evidence showing the contribution of the various types of signed, finite and infinite activity, jumps to improving volatility forecasts at different forecast horizons. We examine the choice of sampling frequency and sampling scheme, as well as the use of noise-robust realized measures. Volatility forecasts using transaction-time based measures are dominated by those using regular clock-time based measures. Fourth, as most jumps are idiosyncratic, no single forecasting model dominates, so better forecasts are obtained with simple model averages using 300-second jump measures.

Our application uses high-frequency data from 2000 to 2016. Using extended HAR models, we forecast the volatility of SPY, the SPDR S&P 500 ETF, as well as 20 constituents of the S&P 100 index which vary by sector and volume. We show that jumps contribute significantly to the volatility of SPY and the 20 stocks we examine. As expected, we find the SPY volatility forecasts to be more accurate, since aggregation helps to identify more informative jumps which improves the out-of-sample mean square prediction error (MSPE) performance.

To preview our findings, when jumps classified by sign and activity are used as additional predictors in HAR models, we find significant improvements with both in- and out-of-sample performance. We focus on the MSPE results from pseudo, out-of-sample forecasts using rolling window regressions. In terms of our classification of jumps by activity, infinite jumps are relatively more important at shorter horizons, whereas finite jumps dominate at longer horizons. Adding signed finite and infinite jumps to the fore-

casting model often generates significantly better forecasts than the standard HAR-RV model. However, no single extended model dominates.

The use of noise-robust estimators substantially improves the out-of-sample performance of our extended HAR models, especially at higher frequencies. The gains are greater for individual stocks than for the SPY index. This is unsurprising since SPY is the most liquid asset with a low level of microstructure noise. One might have expected standard volatility measures to deliver more accurate forecasts at the 300-second frequency, since microstructure noise should be small. However, this only holds true for SPY. For individual stocks, the forecasting gains are quite similar using noise-robust and standard volatility measures. In line with Ghysels and Sinko (2011), noise-robust measures only improve forecasting performance when the level of market microstructure noise is significant.

The greatest gains in real-time forecasting performance are generally found using returns sampled at 300-second intervals, rather than at 5- or 60-second intervals, irrespective of whether noise-robust or standard volatility measures are used.[3] Since the forecasting performance of no single model dominate across sampling frequency and forecasting horizon, we investigate model averaging using the model confidence set approach of Hansen et al. (2011) to reduce the set of retained models in the averages. Simple model averaging, including averages with time-varying weights, generally results in significant out-of-sample forecasting performance (e.g. Aiolfi et al., 2011; Aiolfi and Timmermann, 2006; Elliott and Timmermann, 2016; Timmermann, 2006). These gains arise using both SPY and individual stocks across different horizons. The gains are greatest using the returns sampled every 300-seconds. We assess the predictive accuracy of model averaging using the pair-wise test of Diebold and Mariano (1995). The results show that model averaging produces significantly smaller MSPEs, even at longer horizons of 66 days / 3 months.

These results are in line with Giacomini and Rossi (2010), where the relative fore-

[3]This result is inline with Liu et al. (2015) who find that 300-second/5-min RV is very difficult to beat. Across a range of different asset classes, they find that 5-minute returns volatilities obtained from the two-scale realized volatility (TSRV) subsampling approach of Zhang et al. (2005) is the preferred method of estimating daily volatility.

casting performance of individual models often changes over time. Here, we identify the incidence of cojumps in our data using the co-exceedance rule of Gilder et al. (2014). The cojumps results indicate that the jumps in our data are mainly idiosyncratic, reflecting stock specific differences in the arrival of news and the reaction to that news.[4] The fact that the timing, size and sign of most jumps are stock specific is the main reason why no single forecast model dominates.

As a robustness check, we consider alternative, transaction-time sampled volatility measures. To the best of our knowledge, only Patton and Sheppard (2015) have considered an alternative sampling scheme for forecasting and their focus is on signed jumps. They do not examine the role of finite and infinite jumps, nor do they compare their results with those using the popular clock-time sampling scheme. In the case of SPY, we find that the share of jumps in transaction-time based RV measures is far smaller than for clock-based measures, and any jumps are predominantly finite activity jumps. In terms of forecasting performance, we conclude that forecasts using volatility and jump measures based on transaction sampling are inferior to the forecasts from clock-based sampling.

The remainder of the paper is as follows. The theoretical background is set out in Section 6.2. The estimation of signed, finite and infinite activity jumps is described in Section 6.3. Noise-robust volatility measures are also discussed. Section 6.4 sets out the forecasting framework, including the extended HAR forecasting model and forecast evaluation criteria. The data used in this study are described in Section 6.5, where the incidence of various types of jumps is tabulated. The forecasting gains from adding different types of signed, finite and infinite activity jumps to HAR models are documented in Section 6.6. Model averaging results are presented in Section 6.7. Volatility forecasting results using transaction-time sampled volatility measures are presented in Section 6.8. Finally, Section 6.9 summarizes the paper and presents our conclusion.

[4]Similar qualitative conclusion are obtained using the multijump test of Caporin et al. (2017). The number of detected cojumps is also similar to the numbers reported in Caporin et al. (2017) and Mukherjee et al. (2020).

6.2 Theoretical Background

Let X_t denote the log-price of an equity or an equity index. We assume X is an Itô-semimartingale process defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, with the following representation:

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s + J_t, \quad t \in [0, T] \tag{6.1}$$

where a is a locally bounded and predictable drift term, σ is the adapted, càdlàg spot volatility, W_t is a standard Brownian motion, and J_t is a pure jump process with finite and infinite activity components, $J_t = J_t^F + J_t^I$. The finite activity J_t^F and infinite activity J_t^I jump processes are:

$$J_t^F := \int_0^t \int_{|x|>\varepsilon} x\mu(dx, ds), \tag{6.2}$$

$$J_t^I := \int_0^t \int_{|x|\leq\varepsilon} x(\mu(dx, ds) - \nu(dx)ds), \tag{6.3}$$

where μ is the jump measure of X with compensator ν, and $\varepsilon > 0$ is an arbitrary number. For more details on Itô-semimartingale processes, see Aït-Sahalia and Jacod (2014) and the references therein. As Aït-Sahalia and Jacod (2012) note, the continuous part of the X process captures the normal risk of an asset that can be hedged using standard methods. The large, finite jumps capture default risk or big news-related events, while small jumps capture price moves which impact high-frequency prices but wash out at the daily level, e.g. the price impact of large transactions.

Since volatility is a latent variable, realized measures are widely employed to give consistent estimates of the quadratic variation (QV) of the process using high-frequency data. The QV of the price process is defined as:

$$QV_t = \underbrace{\int_0^t \sigma_s^2 ds}_{\text{Integrated Variation (IV)}} + \underbrace{\sum_{0\leq s\leq t} (\Delta X_s)^2}_{\text{Jump Contribution}} \tag{6.4}$$

where $\Delta X_s := X_s - X_{s-}$ when X jumps at time s. The widely used, realized volatility (RV) measure converges in probability to the QV as the sampling interval $\Delta_n \to 0$:

$$RV_t = \sum_{i=1}^{n} (\Delta_i^n X)^2 \xrightarrow{p} QV_t, \tag{6.5}$$

where the day is split into $n = \lfloor 1/\Delta_n \rfloor$ equally spaced intervals of length Δ_n with n, $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ is the log-return in interval i, and $\lfloor x \rfloor$ denotes the integer part of x.

To separate the integrated variation component of QV from the jump component, we use the threshold bipower variation (TBPV) measure proposed by Corsi et al. (2010), a modified version of the so-called bipower variation measure of Barndorff-Nielsen and Shephard (2004b). The TBPV, which is robust to jumps in both the stochastic limit and the asymptotic distribution, converges in probability to the integrated variance as the sampling interval $\Delta_n \to 0$:

$$TBPV_t = \mu_1^{-2} \frac{n}{n-1} \sum_{i=2}^{n} |\Delta_i^n X| \mathbb{1}_{\{(\Delta_i^n X)^2 \le \vartheta_i\}} |\Delta_{i-1}^n X| \mathbb{1}_{\{(\Delta_{i-1}^n X)^2 \le \vartheta_{i-1}\}} \xrightarrow{p} \int_0^t \sigma_s^2 ds, \tag{6.6}$$

where $\mu_1 = \sqrt{2/\pi} \approx 0.7979$, $n/(n-1)$ is a small sample correction, and ϑ is the threshold estimator defined as in Corsi et al. (2010, appendix B).

Barndorff-Nielsen et al. (2010) introduced positive and negative realized semivariance (RS) estimators to capture upside and downside risk:

$$RS_t^+ = \sum_{i=1}^{n} (\Delta_i^n X)^2 \mathbb{1}_{\{\Delta_i^n X > 0\}} \xrightarrow{p} \frac{1}{2} \int_0^t \sigma_s^2 ds + \sum_{0 \le s \le t} (\Delta X_s)^2 \mathbb{1}_{\{\Delta X_s > 0\}} \tag{6.7}$$

$$RS_t^- = \sum_{i=1}^{n} (\Delta_i^n X)^2 \mathbb{1}_{\{\Delta_i^n X < 0\}} \xrightarrow{p} \frac{1}{2} \int_0^t \sigma_s^2 ds + \sum_{0 \le s \le t} (\Delta X_s)^2 \mathbb{1}_{\{\Delta X_s < 0\}}. \tag{6.8}$$

6.3 Identifying and Decomposing Jumps by Sign and Activity

To identify days with significant jumps, we employ the intra-day jump test proposed by Andersen et al. (2007b, ABD). If the largest intra-daily value of the test exceeds the critical value, we classify the day as a jump day. The $\mathcal{J}_t$ indicator for a day with significant jumps is 1 if $\max_i \left(|\Delta_i^n X| / \sqrt{\Delta_n TBPV} \right) > \Phi^{-1}_{1-\beta/2}$ and 0 otherwise, where $\Phi^{-1}_{(\cdot)}$ is the inverse of the standard normal distribution function, α is the significance level and $\beta = 1 - (1-\alpha)^{\Delta_n}$ is the Šidàk multiple testing correction. Hence, the estimated continuous and jump components of QV are:

$$\widehat{C}_t = RV_t \cdot (1 - \mathcal{J}_t) + TBPV_t \cdot \mathcal{J}_t, \tag{6.9}$$

$$\widehat{J}_t = (RV_t - TBPV_t, 0)^+ \cdot \mathcal{J}_t. \tag{6.10}$$

To identify days with significant finite and infinite activity jumps, we employ the SFA test proposed by Aït-Sahalia and Jacod (2011). The test statistic uses the ratio of two truncated realized power variation measures to eliminate the large jumps. The truncated realized power variation $B(p, \upsilon_n, \Delta_n)_t = \sum_{i=1}^n |\Delta_i^n X|^p \mathbb{1}_{\{|\Delta_i^n X \leq \upsilon_n\}}$, with $\upsilon_n = \varrho \Delta_n^{\varpi}$, $\varrho > 0$, $\varpi \in (0, 1/2)$, is the sum of truncated absolute returns, $|\Delta_i^n X| \leq \upsilon_n$, raised to the power p over different sampling frequencies Δ_n. The SFA test statistics has different limits depending on whether the jumps in X_t are finite or infinite activity jumps: $SFA_t = \frac{B(p,\upsilon_n,k\Delta_n)_t}{B(p,\upsilon_n,\Delta_n)_t} \xrightarrow{p} k^{p/2-1}$ in the finite activity case and 1 in the infinite activity case. Under the finite activity null, the statistic $\left(SFA_t - k^{p/2-1}\right) \Big/ \sqrt{\widehat{V}_t} \xrightarrow{L} \mathcal{N}(0,1)$, where $\widehat{V}_t = N(p,k) \frac{B(2p,\upsilon_n,\Delta_n)_t}{B(p,\upsilon_n,\Delta_n)_t^2}$. For further details on $N(p,k)$, and other settings, see Aït-Sahalia and Jacod (2011). We set $k = 2$ and $p = 4$, and use the indicator function $F_t = \mathbb{1}\left\{ SFA_t < k^{p/2-1} - \Phi^{-1}_{1-\alpha} \sqrt{\widehat{V}_t} \right\}$ to identify days with finite activity jumps.

Our classification of jumps by sign and activity is described below.

Use	Measure	Formula
QV Contributions	Finite Activity Jumps	$\widehat{FJ}_t = \widehat{J}_t \cdot F_t$
	Infinite Activity Jumps	$\widehat{IJ}_t = \widehat{J}_t \cdot (1 - F_t)$
	Positive Jumps	$\widehat{PJ}_t = \left(RS_t^+ - \frac{1}{2}TBPV_t, 0\right)^+ \cdot \mathcal{J}_t$
	Negative Jumps	$\widehat{NJ}_t = \left(RS_t^- - \frac{1}{2}TBPV_t, 0\right)^+ \cdot \mathcal{J}_t$
Forecasting Models	Signed Jumps	$\widehat{SJ}_t = \widehat{PJ}_t - \widehat{NJ}_t$
	Positive Signed Jumps	$\widehat{J}_t^+ = \widehat{SJ}_t \cdot \mathcal{P}_t$
	Negative Signed Jumps	$\widehat{J}_t^- = \widehat{SJ}_t \cdot (1 - \mathcal{P}_t)$
	Positive Signed Finite Jumps	$\widehat{FJ}_t^+ = \widehat{J}_t^+ \cdot F_t$
	Negative Signed Finite Jumps	$\widehat{FJ}_t^- = \widehat{J}_t^- \cdot F_t$
	Positive Signed Infinite Jumps	$\widehat{IJ}_t^+ = \widehat{J}_t^+ \cdot (1 - F_t)$
	Negative Signed Infinite Jumps	$\widehat{IJ}_t^- = \widehat{J}_t^- \cdot (1 - F_t)$

We classify jumps by activity using the jump $\mathcal{J}_t$ and finite activity F_t indicators. The contribution of positive and negative jumps to overall QV are based on $(RS_t^+ - \frac{1}{2}TBPV_t, 0)^+ \cdot \mathcal{J}_t$ and $(RS_t^- - \frac{1}{2}TBPV_t, 0)^+ \cdot \mathcal{J}_t$ respectively. When forecasting volatility using our extended HAR models, we use daily (net) signed jumps, $\widehat{SJ}_t$, the difference between the positive and negative measures (e.g. Patton and Sheppard, 2015). The corresponding positive and negative signed jumps are $\widehat{J}_t^+ = \widehat{SJ} \cdot \mathcal{P}_t$ and $\widehat{J}_t^- = \widehat{SJ} \cdot (1 - \mathcal{P}_t)$ respectively, where $\mathcal{P}_t = \mathbb{1}\left\{\widehat{SJ}_t > 0\right\}$. Their finite/infinite counterparts are identified using the finite activity F_t indicator.

6.3.1 Market Microstructure Noise

Market microstructure noise can distort realized volatility measures, and hence the identification of jumps. We know that the contribution of jumps varies by sampling frequency (Table 6.3), and that the level of market microstructure noise increases as the sampling interval $\Delta_n \to 0$. As a result, standard high-frequency realized volatility measures tend to be biased, distorting jump test statistics (e.g. Hansen and Lunde, 2006;

Huang and Tauchen, 2005).[5] This suggests that noise-robust volatility measures should be used at high frequencies (e.g. 5 and 60 seconds), and possibly lower frequencies. Although Aït-Sahalia and Xiu (2019) suggest that improvements in stock market liquidity mean that the common practice of treating the 5-minute returns of S&P 100 constituents as noise-free is a reasonably safe choice for data sampled after 2009, it is problematic before then. They also suggest that the 5-minute returns of a large portion of the S&P 500 index constituents cannot be treated as noise-free.

We assume that the observed log price process, Y_t, is contaminated by additive, microstructure noise:[6]

$$Y_t = X_t + u_t, \tag{6.11}$$

where X_t is the process described in equation (6.1), u_t is an i.i.d. noise process with $\mathbb{E}[u_t] = 0$ and $\mathbb{E}[u_t^2] = \omega^2$, and $u_t \perp\!\!\!\perp X_t$. Jacod et al. (2009) and Christensen et al. (2014) propose pre-averaging estimators for the RV and a consistent estimator of the IV. The pre-averaging returns are estimated as a weighted average of returns within a local neighborhood of L log-prices:

$$\Delta_i^n X^* = \sum_{j=1}^{L-1} g\left(\frac{j}{L}\right) \Delta_{i+j}^n Y, \tag{6.12}$$

where $g = \min(x, 1-x)$, $L = \theta\sqrt{n}$ with $\theta = 1/3$ for 5 and 60 seconds return or $\theta = 1$ for 300 seconds returns. With these choices, the noise-robust estimator for the realized

[5] The bias is due to $\mathbb{E}[|\Delta_i^n X|] \leq \mathbb{E}[|\Delta_i^n X + \eta_i|]$, where $\eta_i = u_i - u_{i-1}$, and its presence produces poor measures of the true volatility, as well as induces an attenuation bias in the autoregressive estimates (e.g. Bollerslev et al., 2016).

[6] The mechanics of trading generate a diverse array of market microstructure effects including bid-ask spread and corresponding bounce, the gradual response of prices to a block trade, and the strategic component of order flow inventory control effects (Aït-Sahalia and Jacod, 2014). Additive noise is the simplest and most common market microstructure model.

variance and the bipower variation are:[7]

$$RV_t^* = \frac{n}{n-L+2}\frac{1}{L\psi_2^L}\sum_{i=0}^{n-L+1}|\Delta_i^n X^*|^2 - \frac{\psi_1^L\hat{\omega}_{AC}^2}{\theta^2\psi_2^L} \tag{6.13}$$

$$BPV_t^* = \frac{n}{n-2L+2}\frac{1}{L\psi_2^L\mu_1^2}\sum_{i=0}^{n-2L+1}|\Delta_i^n X^*||\Delta_{i+L}^n X^*| - \frac{\psi_1^L\hat{\omega}_{AC}^2}{\theta^2\psi_2^L}, \tag{6.14}$$

where the leading $n/(n-L+2)$ and $n/(n-2L+2)$ terms are small sample corrections, and the trailing term $\frac{\psi_1^L\hat{\omega}_{AC}^2}{\theta^2\psi_2^L}$ is a bias-correction to remove residual noise not eliminated by the pre-averaging, and $\psi_1^L = L\sum_{j=1}^{L}\left[g\left(\frac{j}{L}\right) - g\left(\frac{j-1}{L}\right)\right]^2$ and $\psi_2^L = \frac{1}{L}\sum_{j=1}^{L-1} g^2\left(\frac{j}{L}\right)$ are constants associated with $g(\cdot)$ (e.g. Christensen et al., 2014; Jacod et al., 2009, Appendix A). The unknown noise variance ω^2 can be approximated using either the Bandi and Russell (2006) estimator $\hat{\omega}_{RV}^2 = \frac{1}{2n}RV_t$, or Oomen (2006a) estimator $\hat{\omega}_{AC}^2 = -\frac{1}{n-1}\sum_{i=2}^{n}\Delta_{i-1}^n Y\Delta_i^n Y$, the negative of the first order autocovariance of (log)-returns. We use the latter procedure.

The ABD test in Andersen et al. (2007b) can be modified to yield a test that is robust to the presence of market microstructure noise. To do this we use the asymptotic distribution of pre-averaged returns (see, for instance Christensen et al., 2014; Jacod et al., 2009; Podolskij and Vetter, 2009, and the references therein):

$$n^{1/4}\Delta_i^n X^*\big|\mathcal{F}_{i/n} \sim \mathcal{N}\left(0, \frac{\theta\sigma^2}{12} + \frac{\omega^2}{\theta}\right). \tag{6.15}$$

Thus, we can define a threshold for identifying jumps as follows:

$$\tau = \frac{q_\beta}{n^\varpi}\sqrt{\psi_2^L\theta\hat{\sigma}^2 + \frac{\hat{\omega}^2}{\theta}}, \tag{6.16}$$

where $q_\beta = \Phi_{1-\beta/2}^{-1}$ is the inverse of the standard normal distribution, α is the significance level, and $\beta = 1-(1-\alpha)^{\Delta_n}$ is the Šidàk multiple testing correction. We use the BPV_t^* to estimate $\hat{\sigma}^2$ and $\hat{\omega}_{AC}^2$ to estimate $\hat{\omega}^2$. We set $\varpi = 1/4$ and $\theta = 1/3$. Therefore, we reject the null of no jumps whenever $\max_i(|\Delta_i^n X^*|) > \tau$.

[7] We also tried the threshold bipower variation measure proposed by Christensen et al. (2018), but the differences were negligible.

Noise-robust versions of the realized semivariances, which capture upside and downside risk, are constructed by appropriately modifying the two-scale realized variance measure of Zhang et al. (2005):

$$TSRS_t^+ = \frac{1}{K}\sum_{k=1}^{K} RS_{t,k}^+ - \frac{n}{\bar{n}} RS_t^+ \xrightarrow{\mathbb{P}} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0\leq s\leq t} (\Delta X_s)^2 \mathbb{1}_{\{\Delta X_s>0\}}, \tag{6.17}$$

$$TSRS_t^- = \frac{1}{K}\sum_{k=1}^{K} RS_{t,k}^- - \frac{n}{\bar{n}} RS_t^- \xrightarrow{\mathbb{P}} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0\leq s\leq t} (\Delta X_s)^2 \mathbb{1}_{\{\Delta X_s<0\}}, \tag{6.18}$$

where $RS_{t,k}^+$ and $RS_{t,k}^-$ are subsample, slower time scale, realized semivariance measures; RS_t^+ and RS_t^+ are the full sample, faster time scale, realized semivariance measures; $\bar{n} = \frac{n-K+1}{K}$ is the average number of observations in the subsamples; $K = \lfloor cn^{2/3} \rfloor$ and c is the optimal bandwidth as in Zhang et al. (2005). The two-time scale estimators average the realized semivariances over K subsamples, and apply a bias correction from the highest possible frequency.[8]

6.3.2 Noise-Robust ABD Test and Two-Time Scale Realized Semivariance – Monte Carlo Results

We examine the performance of our noise-robust ABD test statistic and two-time scale realized semivariance estimators using Monte Carlo simulations, where the log-price X is simulated as:

$$\begin{aligned} dX_t &= \sqrt{\nu_t} dW_t + \theta_L dL_t \\ d\nu_t &= \kappa(\eta_\nu - \nu_t)dt + \gamma_\nu \nu_t^{1/2} dB_t, \end{aligned} \tag{6.19}$$

where W_t and B_t are standard Brownian motions with covariance $\mathbb{E}[dW_t, dB_t] = \rho dt$, and L_t is either a finite activity compound Poisson process or an infinite activity Cauchy process (a β-stable process with $\beta = 1$).

Following Aït-Sahalia and Jacod (2011), we set $\kappa = 5$, $\eta_\nu = 1/16$, $\rho = -0.5$. The compound Poisson process has intensity λ, and jumps that are uniformly distributed on

[8] Aït-Sahalia et al. (2012) develop a noise-robust, pre-averaging, version of the Aït-Sahalia et al. (2009) jump test, while Li and Xiu (2016) develop general GMM procedures that address measurement error in realized volatility measures.

$\nu_t^{1/2}\sqrt{m}([-2,-1]\cup[1,2])$. We set $m = 0.7$ and $\lambda = 1.0$ such that there is on average one jump every day. When jumps are of finite activity we set $\theta_L = 1$, while for infinite jumps we set $\theta_L = 0.5$. Following Barndorff-Nielsen et al. (2008), we add noise to the $X_{t,i}$ process:

$$Y_{t,i} = X_{t,i} + u_{t,i},$$

where Y is the noisy, observed log price, ξ is the noise-to-signal ratio used to simulate market microstructure noise, $u_{t,i} \sim \mathcal{N}(0, \omega_t^2)$ and $\omega_t^2 = \xi^2 \int_0^t \nu_s ds$. With this design, the variance of the noise is constant throughout the day, but changing from day to day.

The price process is simulated via an Euler scheme where we normalize one second to be $\Delta_n = 1/23,400$. Thus, the interval $[0,1]$ contains the usual 6.5hrs of trading activity. To generate the observed prices, we discretize $[0,1]$ into a number $n = 23,400$ of intervals. We then contaminate the prices with market microstructure noise and aggregate the data to the 5-, 60- and 300 seconds, which are equivalent to 4,680, 390 and 78 intraday observations per day. We simulate 5 trading days and use 5,000 replications.

Table 6.1 shows the results of our Monte Carlo exercise exploring the size and power of the two versions of the ABD test under finite and infinite jumps, with a moderate and higher level of noise-to-signal ratio. The tests are evaluated at the 5% level. The noise-robust ABD test is more powerful at higher, 5-second and 60-second, frequencies and when the noise-to-signal ratio is higher. The standard ABD test is undersized (oversized) at higher (lower) frequencies, irrespective of the level of noise-to-signal ratio, whereas the noise-robust test displays very decent size levels which decrease with the sampling frequency. This result is expected as the level of microstructure noise decreases when the data is sampled more sparsely and therefore pre-averaged methods are less efficient.

The second and third panels show the power of the tests under finite and infinite activity jumps. With finite activity jumps and a small noise-to-signal ratio, both tests perform quite well with the noise-robust test outperforming (underperforming) the standard test at higher (lower) frequencies.[9] Finally, when jumps are infinite activity, the

[9] Maneesoonthorn et al. (2020) show, using a similar data generating process, that the Lee and

standard ABD test is badly affected by the noise-to-signal levels.

Table 6.2 compares the finite sample MSEs of the realized semivariance and two-time scale realized semivariance measures. The results show that the realized semivariance is very sensitive to market microstructure noise, resulting in large MSEs even when the noise-to-signal ratio is moderate and the sampling frequency is low. On the other hand, the performance of the two-time scale realized semivariance is very good overall.

6.4 Forecasting Models and Forecast Comparisons

The HAR-RV in Corsi (2009) models current and future RV as a linear function of lagged daily, weekly and monthly values of RV. Andersen et al. (2007a) originally added jumps to the HAR-RV model. Our forecasting models extend the HAR-RV model further by adding signed, finite and infinite activity jumps. The benchmark HAR-RV model is

$$RV_{t,t+h} = \beta_0 + \beta_d RV_t + \beta_w RV_{t-5,t} + \beta_m RV_{t-22,t} + \epsilon_{t+h}, \tag{6.20}$$

where h is the forecast horizon, and $RV_{t,t+h-1} = \frac{1}{h}\sum_{i=1}^{h} RV_{t+1-i}$. We examine nine different, extended HAR models. The first three forecasting models include daily, weekly and monthly jumps in addition to the daily, weekly and monthly continuous component of RV. The next three models replace the jump variables in previous models with their finite activity counterparts. The final three models replace the jump part with their infinite activity jumps. We estimate separate models for unsigned, positive and negative jumps:

Mykland (2012) and Aït-Sahalia et al. (2012) tests –which are also noise-robust versions– have very poor power at lower frequencies.

Jumps, Signed and Unsigned Models:

HAR-CJ: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{J_d}\widehat{J}_t + \beta_{J_w}\widehat{J}_{t-5,t} + \beta_{J_m}\widehat{J}_{t-22,t} + \epsilon_{t,t+h}$$

HAR-CJ$^+$: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{J_d^+}\widehat{J}_t^+ + \beta_{J_w^+}\widehat{J}_{t-5,t}^+ + \beta_{J_m^+}\widehat{J}_{t-22,t}^+ + \epsilon_{t,t+h}$$

HAR-CJ$^-$: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{J_d^-}\widehat{J}_t^- + \beta_{J_w^-}\widehat{J}_{t-5,t}^- + \beta_{J_m^-}\widehat{J}_{t-22,t}^- + \epsilon_{t,t+h}$$

Finite Jumps, Signed and Unsigned Models:

HAR-CFJ: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{FJ_d}\widehat{FJ}_t + \beta_{FJ_w}\widehat{FJ}_{t-5,t} + \beta_{FJ_m}\widehat{FJ}_{t-22,t} + \epsilon_{t,t+h}$$

HAR-CFJ$^+$: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{FJ_d^+}\widehat{FJ}_t^+ + \beta_{FJ_w^+}\widehat{FJ}_{t-5,t}^+ + \beta_{FJ_m^+}\widehat{FJ}_{t-22,t}^+ + \epsilon_{t,t+h}$$

HAR-CFJ$^-$: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{FJ_d^-}\widehat{FJ}_t^- + \beta_{FJ_w^-}\widehat{FJ}_{t-5,t}^- + \beta_{FJ_m^-}\widehat{FJ}_{t-22,t}^- + \epsilon_{t,t+h}$$

Infinite Jumps, signed and Unsigned Models:

HAR-CIJ: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{IJ_d}\widehat{IJ}_t + \beta_{IJ_w}\widehat{IJ}_{t-5,t} + \beta_{IJ_m}\widehat{IJ}_{t-22,t} + \epsilon_{t,t+h}$$

HAR-CIJ$^+$: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{IJ_d^+}\widehat{IJ}_t^+ + \beta_{IJ_w^+}\widehat{IJ}_{t-5,t}^+ + \beta_{IJ_m^+}\widehat{IJ}_{t-22,t}^+ + \epsilon_{t,t+h}$$

HAR-CIJ$^-$: $$RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{IJ_d^-}\widehat{IJ}_t^- + \beta_{IJ_w^-}\widehat{IJ}_{t-5,t}^- + \beta_{IJ_m^-}\widehat{IJ}_{t-22,t}^- + \epsilon_{t,t+h}$$

The realized continuous and jump measures in the models are estimated using the formulae outlined in Section 6.3. We also have an additional nine models where all the right-hand volatility measures are the noise-robust measures discussed in Section 6.3.1. Although additional variants of these models could be developed and evaluated, we do not believe that it is worthwhile doing so since model averages should encompass these variants.

Our primary interest is in the performance of pseudo out-of-sample forecasts. We consider horizons $h = 1,\ 5,\ ,22$, and 66, corresponding to one day, one week, one month, and one quarter ahead. We also use rolling window regressions of size 1000, or approximately four years, to estimate the models. The out-of-sample performance is evaluated using the mean squared prediction error (MSPE) loss function and, to a lesser extent, the out-of-sample R^2_{oos}. The MSPE, which has been shown to be robust to noise in the proxy for volatility in Patton (2011b) is:

$$MSPE = S_h^{-1} \sum_{s=1}^{S_h} \left(RV_s^h - \widehat{RV}_s^h \right)^2, \tag{6.21}$$

where RV_s^h and $\widehat{RV}_s^h$ are respectively the actual and pseudo out-of-sample forecasts of $RV_{t,t+h}$, and S_h is the total number of out-of-sample forecasts from the series of rolling window models. Additionally, we carry out pairwise tests of the null of equal predictive ability using Diebold and Mariano (1995, DM,hereafter) tests with a MSPE loss criterion and HAC standard errors.

The Model Confidence Set (MCS) procedure of Hansen et al. (2011) is used to identify the subset of models with significantly lower MSPEs than the other models. We use the MCS procedure with a quadratic loss function. We denote by $\mathcal{M}$ the set of all the HAR models. We define $d_{h,i,j} = L(RV_{t,t+h}, \widehat{RV}^{(i)}_{t,t+h}) - L(RV_{t,t+h}, \widehat{RV}^{(j)}_{t,t+h})$ as the difference in the loss of model i and model j. We use a quadratic loss function as L. Finally, we

construct the average loss difference, $\bar{d}_{h,i,j}$, and define the test statistics as follows

$$t^h_{i,j} = \frac{\bar{d}_{h,i,j}}{\sqrt{\widehat{\text{Var}}(\bar{d}_{h,i,j})}}, \quad \forall i,j \in \mathcal{M} \tag{6.22}$$

The MCS test statistics are given by $T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t^h_{i,j}|$ and have the null hypothesis, H_0 that all models have the same expected loss. The alternative hypothesis is that there is some model i with a MSPE that is greater than the MSPE's of all the other models $j \in \mathcal{M} \backslash i$. When the null is rejected the worst performing model is eliminated, and this process is iterated until no further model can be eliminated. The surviving models denoted by $\mathcal{M}_{MCS}$ are retained with a confidence level $\alpha = 0.05$. We implement the MCS via a block bootstrap using a block length of 10 days and 5000 bootstrap replications.[10]

6.5 Data

For our forecasting exercise, we use the SPDR S&P 500 ETF (SPY) and 20 individual stocks in the S&P 500 index. The data are for the years 2000 to 2016, a total of 4277 trading days. The 20 individual stocks were chosen based on their jump activity index, and the relative contributions of finite and infinite jumps. The data are sourced from the TickData database.[11] We follow Hansen and Lunde (2006) and use previous tick interpolations to aggregate the ticks to the required frequency.

Mean daily RV for SPY and the 20 stocks ranges from 1.037 to 8.284, while the average number of shares traded per day ranges from 0.875 to 98.972 million. Since we are interested in the role of realized measures using different sampling frequencies in forecasting realized volatility, we sample returns every 5, 60, and 300 seconds. The choice of 300 second is standard in high-frequency finance studies, and is motivated by

[10] Qualitatively similar results were obtained using different block sizes (20 and 50 days), and additional bootstrap replications (10,000 and 20,000).

[11] TickData provides pre-cleaned and filtered price series. The algorithmic data filters identify bad prints, decimal errors, transposition errors and other data irregularities. The filters take advantage of the fact that, since we are not producing data in real time, we have the capacity to look at the tick following a suspected bad tick before we decide whether or not the tick is valid. The filters are proprietary and are based upon recent tick volatility, moving standard deviation windows, and time day. For a more detailed explanation, see the high-frequency data filtering white paper on the TickData resources page TickData.

the trade-off between bias and variance (see Aït-Sahalia et al., 2005; Bandi and Russell, 2006; Zhang et al., 2005, inter alios for a more detailed discussion).

The contribution of the different types of jumps to total QV are shown in Table 6.3. The contribution of jumps decreases as the sampling interval increases from 5 to 300 seconds. For SPY, the share of jumps decreases from 43.2% (5 seconds) to 14.3% (300 seconds).[12] For the 20 stocks, the average jump share decreases from 67.6% to 29.8%. In both cases, the decline is mainly due to the drop in the share of infinite jumps. The share of infinite jumps in SPY drops from 32.6% using 5-second returns to 0.1% using 300-second returns, and for the 20 stocks, the average share of infinite jumps drops from 34.2% to 0.2%. Hence, when returns are sampled every 300 seconds, the vast majority of jumps in SPY and the 20 stocks are finite activity jumps. At this frequency, the small variations that characterize jumps are close to Brownian increments. We find little evidence of asymmetry in the shares of signed jumps. The Blumenthal-Getoor index or jump activity index ($\hat{\beta}_{IJA}$),[13] which measures the activity of small increments, are consistent with the estimated shares of finite and infinite jump components. In the case of SPY, the index is 1.45 using 5-second returns and 0.78 using 300-second returns, which implies that infinite jumps are more important at higher frequencies.

Figure 6.1 plots the continuous and jump components of RV for SPY and the three stocks – AMZN, HD and KO – with the largest, smallest and average RV. The days with jumps are shown in red, and other days in blue. It is clear that there is considerable heterogeneity in the level and timing of volatility. Although the highest spikes in volatility occur around the dot-com and sub-prime crises (shaded areas), many other spikes in volatility are idiosyncratic. The 5- and 300-second autocorrelation functions of the SPY realized measures based on noise-robust and standard measures are displayed in Figure 6.2. The SPY RV_t and $\widehat{C}_t$ measures appear to be long memory processes since their autocorrelations do not decline exponentially. The ACF of the 5-second RV_t and $\widehat{C}_t$ measures (left-panel) lie below their 300-second counterparts (right-panel) – a hint that

[12] The contribution of jumps to total QV is in line with those reported by Aït-Sahalia and Jacod (2012), who show that the level of continuous component of the 30 stocks within the Dow Jones Industrial Average oscillate between 65% to 85%, and between 85% to 95% for the overall index.

[13] The jump activity index is estimated as in Jing et al. (2012b).

volatility forecasts using 300-second realized measures may perform better than ones using 5-second realized measures.

6.6 Empirical Findings

6.6.1 SPY Forecasting Results

Since we use the HAR-RV model as a benchmark for assessing the forecasting performance of our extended HAR models, Table 6.4 sets out the in-sample coefficients, as well as the in- and out-of-sample R^2s and MSPEs, of the HAR-RV model for four forecast horizons – $h = 1$ (day), $h = 5$ (week), $h = 22$ (month), $h = 66$ (quarter), using returns sampled every 300 seconds. The significance of the coefficients is evaluated using Newey-West HAC-robust standard errors, allowing for serial correlation of up to 5 ($h = 1$), 10 ($h = 5$), 44 ($h = 22$), and 132 ($h = 66$), since the random error term in the models is serially correlated at least up to order $h - 1$. In following Andersen et al. (1999) and Patton and Sheppard (2015), we estimate R^2_{oos} as 1 minus the ratio of the out-of-sample models-based MSPE to the out-of-sample MSPE from a forecast including only a constant. The MSPE results are based on a pseudo out-of-sample rolling regression forecast using a 1000 day window.

All the coefficients in Table 6.4 are significant even at the three month horizon, confirming the high persistence of volatility. The magnitude of the daily and weekly coefficients decrease as we lengthen the forecast horizon. Although, the magnitude of the monthly coefficient changes little with the horizons, its relative importance increases at longer horizons.[14]

Summary forecasting results for extended HAR-CJ (jumps), HAR-CFJ (finite jumps), and the HAR-CIJ (infinite jumps) models are presented in Table 6.5, also using 300 second returns. In- and out-of-sample R^2s and the MSPEs are presented for unsigned jumps, positive signed jumps and negative signed jumps. Full results are available on request. A few points about the coefficients estimates are worth noting. The restrictions that the

[14]These results are well-documented in the literature, see Andersen et al. (2007a), Corsi (2009), and Corsi et al. (2010) among others.

coefficients on finite and infinite jumps are the same, and that the coefficients on positive and negative jumps are the same, are decisively rejected. In line with Andersen et al. (2007a) and Patton and Sheppard (2015), overall jumps tend to reduce future volatility, negative jumps tend to increase it and positive jumps to decrease it. Finite (infinite) jumps tend to decrease (increase) future volatility.

Unsurprisingly, the in-sample R-squared statistics (R^2_{is}) in Table 6.5 suggest that incorporating jumps as predictors results in a better fit for our models, outperforming the benchmark HAR-RV across the four forecasting horizons under examination. The out-of-sample R-squared statistics (R^2_{oos}) show that extended HAR models outperform the benchmark model at one day and one week horizons, and about half the time at longer horizons. The models with positive jumps have higher R^2_{oos}'s at all horizons. Turning to the MSPE results, the forecasting performance of the extended HAR models is significantly better at one day and one week horizons, and better (significantly better) about half (one quarter) of the time at the one-month and three-month horizons. Note that no single extended HAR model outperforms all the others, a finding also reported in Patton and Sheppard (2009), which suggests that model averages combining the information contained in the different volatility forecasting models may generate further forecast gains. See Section 6.7 below.

6.6.2 SPY Forecasting Results Using Standard and Noise-Robust Realized Measures

We know that microstructure noise is important at higher frequencies, and the resulting attenuation bias may generate less accurate volatility forecasts than forecasts using noise-robust measures, such as the ones discussed in Section 6.3.1 above. We examined this issue in detail. Table 6.6 compares the forecasting performance of SPY extended HAR volatility models using standard versus noise-robust realized measures identifying models with significantly lower MSPEs than the benchmark HAR-RV model. The entries in the top panel are based on forecasts using standard realized jump measures as explanatory variables; the bottom panel entries are based on noise-robust measures. The entries

are relative MSPEs –The ratio of the MSPE of the proposed model to the MSPE of the corresponding benchmark model– so ratios below one indicate more accurate rolling regression forecasts.[15] Models with significantly lower MSPE than the benchmark model, based on pair-wise Diebold and Mariano (1995, DM) tests, are starred. The DM tests show that many of the extended HAR models in Table 6.6 forecast as well as, or better, than the HAR-RV models, although there is considerable variation across sampling frequencies and time horizon.

At the 5 and 60 second frequencies, the forecasts from models using noise-robust realized jump measures are somewhat more accurate than forecasts based on regular realized jump measures. Many models using 5 and 60 second standard volatility measures are excluded from the MCS at longer horizons, confirming the importance of taking account of microstructure noise at higher frequencies. Nevertheless, the MSPE numbers for the benchmark HAR-RV model in the final row of Table 6.6 suggest that models using 300-second volatility measures tend to give better forecast than models using 5- or 60-second returns, irrespective of whether standard or noise-robust volatility measures are used.

6.6.3 Extended HAR Model Forecasting Results for the Twenty S&P Stocks

Some results for the 20 S&P 500 stocks are presented in Table 6.7. The relative MSPE entries (averaged across the 20 stocks) are shown in the body of the table, while the average MSPEs for the benchmark HAR-RV models using standard realized measures are shown in the final row of the table. The entries for models which are not retained in the MCS at least 15 times (out of 20) are suffixed with a dagger (†). The relative MSPE entries are more clustered around one than in Table 6.6.[16] In addition, with the majority of the models retained in the MCS at least 15 times, this indicates that

[15]The MSPE results are based on pseudo out-of-sample, rolling regression forecast using 1,000 day window. Most models are retained in the model confidence set (MCS); the small number of entries for models that are not retained in the MCS are identified with a dagger (†). The MCS results are generated using a 10-day block bootstrap and 5,000 replications.

[16]The entries are also less dispersed, in part because we are reporting averages.

the improvement in the forecasting performance of extended models with jumps is less clearcut for the 20 stocks, than it is for the SPY. At the 5 and 60-second frequencies, the results show that noise-robust volatility measures work best. This is because noise-robust measures provide more efficient estimators of the latent volatility process, thereby reducing the attenuation bias on the autoregressive coefficients (see, e.g. Andersen et al., 2005; Bollerslev et al., 2016). However, consistent with the results for SPY, forecasts using 300 second volatility measures are generally better than forecasts using 5 or 60 second-based volatility measures.[17] In addition, the relative MSPEs of the standard volatility measures are often lower than those of the noise-robust measures.

No single extended HAR model with jumps dominates all the other models – the main reason being the small number of systematic jumps across the 20 stocks.[18] We find that, on average, cojumps only contribute to 9% of the total jump component, which means that most jumps are idiosyncratic. To illustrate, the left panel of Figure 6.3 shows the returns on May 06, 2010, the day of the so-called Flash Crash, one of the few days when the stocks jumped together. The movement in returns on that date is very different from returns on a typical day such as December 23, 2003 (right-panel) in which only idiosyncratic jumps are present. Since the idiosyncratic jumps are stock specific reactions to news, what it is perceived as negative news for one stock might be positive news for another stock, so generating jumps of different size and directions. Aït-Sahalia and Xiu (2016) suggest that co-jumps stem from surprising news announcements that occur primarily before the opening of the U.S. market. Amengual and Xiu (2018) note that downward intraday volatility jumps in the S&P 500 index are often associated with a resolution of policy uncertainty, mostly through statements from the FOMC meetings and speeches by the chair of the Federal Reserve. Aït-Sahalia et al. (2020) find that idiosyncratic jumps are related to idiosyncratic events such as earning disappointments.

[17] The improvements of the 300-second based realized measures vis-à-vis 5- and 60-second returns are due to noise-robust measures are sometimes derived under some (strong) assumptions about the microstructure noise, and whenever (some of) these assumptions are not met in practice, the estimators turn out to be inconsistent. Therefore, the 300-second returns offers enough statistical power that seems to avoid distortions that could arise from microstructure noise.

[18] We identify jumps using the co-exceedance procedure of Gilder et al. (2014), which relies on the intersection of the univariate jump tests.

Given the rich information content of the different jump classifications and since no single extended HAR model dominates, the next section focuses on whether model averages forecasts consistently outperform the forecasts from the benchmark HAR-RV and the best extended HAR models across sampling frequencies and forecasting horizons.

6.7 The Gains from Model Averaging

Hitherto, we have shown that a variety of extended HAR volatility models, that account for the nature and sign of jumps, generate significant improvements in forecasting performance. However, no single specification consistently outperforms the other models across horizons and frequencies, which suggests that model averaging might generate further forecasting gains. Four simple approaches to assigning model averaging weights are considered.[19] The aim of model averaging is to exploit relevant information embedded in the different forecasts, and produce an ensemble model that outperforms the benchmark HAR-RV model and, more importantly, the best single, extended HAR-RV jump model. Our approaches follow the literature closely (see, e.g. Aiolfi et al., 2011; Aiolfi and Timmermann, 2006; Bates and Granger, 1969; Elliott and Timmermann, 2016, and the references therein).

We present model averaging results for the four sets of weights tabulated below – weights minimizing the estimated variance of the prediction errors, inverse MSPE weights, inverse MSPE rank weights and equal weights. In the first three cases, the weights are recalculated every time a new set of rolling forecasts are generated, and we prune the set of models under consideration by only averaging models that are retained in the model confidence set.

[19] We experimented with more complicated model averaging procedures, but the results were similar to those presented here. To conserve space, we do not report these experiments, but the details are available on request.

Weight	Formula	Models
Min. Prediction Error Variance	$w_t^h = \underset{w}{\text{argmin}}\ \ w'\widehat{\Sigma}_t^h w \text{ s.t. } \iota' w = 1$	MCS
Inverse MSPE	$w_{t,i}^h = \frac{(MSPE_{t,i}^h)^{-1}}{\sum_{i \in \mathcal{M}_J}(MSPE_{t,i}^h)^{-1}}$	MCS
Inverse Rank	$w_{t,i}^h = \frac{(Rank_{t,i}^h)^{-1}}{\sum_{i \in \mathcal{M}_J}(Rank_{t,i}^h)^{-1}}$	MCS
Equal Weights	$w_{i,t}^h = \frac{1}{N}$	All

Note: $\widehat{\Sigma}_t^h$ is the estimated, rolling window variance-covariance matrix of the set of MCS retained horizon h volatility forecasting models at time t. ι is a vector of ones representing each retained model. $MSPE_{t,i}^h$ and $Rank_{t,i}^h$ are the rolling window MSPEs and MCS Ranks for the MCS retained horizon h forecasting model at time t. Finally, N represents all the jump specifications used in this study.

We present model averaging results for SPY and four individual stocks chosen by the level of their jump activity. All the stocks have estimated Blumenthal-Getoor index in the range 0 to 1, so their returns include finite and infinite activity jumps, with finite jumps dominating. BA and KO with jump activity of 0.58 and 0.91 are the extreme cases.

The relative MSPEs for the best extended HAR-RV model and the four model averaging approaches are shown in Table 6.8. The MSPEs for each index or stock and forecast horizon are measured relative to the MSPE of the corresponding HAR-RV model. The bold entries are model averages with lower MSPEs than the MSPEs of both the HAR-RV and best extended HAR models. The starred entries denote model averages with significantly lower MSPEs than the MSPEs of the HAR-RV models. Double starred entries identify models whose MSPEs are significantly lower than the MSPEs of both the benchmark HAR-RV and the best extended HAR model. The four model averages generate forecasts that typically outperform the benchmark model for the four forecast horizons examined: $h = 1$ (on-day), $h = 5$ (one week), $h = 22$ (one month), $h = 66$ (one quarter). For example, in the case of SPY with 300-second returns, the one-week relative MSPE of the best extended HAR model is 0.753 as compared with a range of 0.693 to 0.715 for the four model averages. The largest MSPE reductions are generally found at the one-week

horizon, followed by the one-month horizon.

We also compare the model averaging results for SPY using 60 and 300 second returns. The 300-second model average forecasts dominate the forecasts using 60-second returns, generating significantly lower MSPEs. The 300-second forecasts also dominate the unreported model average forecasts using 5-second returns. These results also hold for the four stocks reported here, and for the other 16 stocks. The 300-second model averaged MSPEs are generally lower than the MSPEs of both the benchmark HAR-RV and best extended HAR models. In about a quarter of the cases, the MSPEs from the 300-second model average are significantly lower than the MSPEs of the best extended HAR model.

In conclusion, model averaging the forecasts from extended HAR-RV models generally result in lower MSPEs. Forecasting 300-second returns dominate forecasts using higher frequency returns. The MCS procedure for pruning dominated models and the use of time varying weights for the model averages are helpful. Simple weighting schemes, e.g. the use of inverse MSPES of inverse MSPE ranks, work as well as schemes that are more complicated (e.g. Patton and Sheppard, 2009).

6.8 A Robustness Check using Transaction-Time Sampled Volatility Measures

In this section, we examine the volatility forecasting performance of alternative jump measures based on a transaction-based sampling scheme. Relatively few studies have considered alternative sampling schemes. For instance, Griffin and Oomen (2008) and Oomen (2006b) study the properties of alternative RV measures using clock/calendar, transaction and business time sampling, but they do not consider jumps. To the best of our knowledge, only Patton and Sheppard (2015) examine the forecasting performance of jump measures using transaction time sampling, but they do not compare the clock and transaction time-based volatility components and the forecasting performance thereof. We contribute to this literature in two ways. Firstly, we decompose clock and transaction-

based RV measures into their continuous and jump components, including their signed and finite/infinite activity jump components. Secondly, we compare the volatility forecasting performance of the clock and transaction time-based measures, using our extended HAR model averaging frameworks.

For brevity, we only report results for SPY. The transaction-based volatility measures are calculated using a 78 intraday returns sampling scheme as in Patton and Sheppard (2015). This is the transaction-based equivalent of the 300-second/5-minute sampling scheme, which is widely used in the literature. Intraday returns are calculated by fixing the opening and closing prices, and recording the prices at business time $\lfloor ik \rfloor$, where $i = 1, \ldots, 79$, $k = \frac{N-1}{79}$, N is the number of unique date stamps per day, and $\lfloor . \rfloor$ denotes rounding down to the nearest integer.[20]

Table 6.9 shows that the transaction-based RV measure is primarily driven by its continuous part: the contribution of jumps to total QV is about 4.6% versus 14.3% for the clock-based measures. Almost all the jumps are finite jumps, the same as for clock time, and there is little difference in the contribution of positive and negative jumps. Although most jumps are finite activity jumps, the smaller contribution of transaction time based jumps to total QV implies a somewhat smaller jump activity index $\hat{\beta}_{IJA}$ (0.708 versus 0.778).

The relative MSPEs in Table 6.10 suggest that the forecasting performance of extended HAR models using transaction-based measures is comparable to that of the benchmark HAR-RV model, in sharp contrast to forecasting performance of extended HAR models using clock-based measures. Similar to the clock-time results, the MSPEs of most of the extended models are lower than the MSPE of the benchmark model at the one-day horizon, although only three forecasts have significantly lower MSPEs. By contrast, as the horizon increases, we only obtain a handful of statistically significant reductions in MSPEs. Consequently, the model confidence set now includes all the models; since the forecasting performance of the models is broadly similar, we cannot identify a set of superior models.

[20]Note that clock- and transaction-based RV descriptive statistics for SPY are very similar.

A comparison of clock- and transaction-time based SPY model averaging results is presented in Table 6.11. Results are presented for daily, weekly, monthly, and quarterly horizons. With transaction-based sampling, simple model averaging procedures (using MSPE, rank or equal weights) generate statistically significant improvements in the MSPEs. However, the MSPE improvements are far smaller than those obtained with clock-based sampling, so the transaction-time based MSPEs are always higher than their close-based counterparts. Based on these SPY results, as well as results for the 20 stocks that are not reported, we conclude that forecasts using volatility measures from transaction-based sampling of returns are inferior to forecasts from clock-based sampling.

6.9 Conclusion

We examine the gains in forecasting the volatility of equity prices by decomposing jumps by activity (finite/infinite) and by sign using high-frequency data for SPY and 20 individual stocks. Our key findings are as follows. Quadratic variation contains a significant jump component, even at the 300-second frequency. The contribution of infinite jumps is greater than that of finite jumps at higher frequencies. However, at the 300-second frequency, jumps are mainly of finite activity.

Extended HAR style models, incorporating a variety of jump activity and sign measures, generate statistically significant in- and out-of-sample improvements for both SPY and the 20 individual stock we examined. The use of noise-robust realized measures improve the forecasts of future volatility at higher frequencies. However, since market microstructure noise declines as the sampling interval increases, the forecasting advantage of the noise-robust jump volatility measures also diminishes.

The rolling window, out-of-sample forecast results suggest that the lowest MSPE forecasts are obtained using returns sampled every 300 seconds, rather than 5 or 60 seconds. This result holds for all of the horizons we examined – a day, a week, a month and a quarter – irrespective of whether noise-robust volatility measures are, or are not, used. In terms of MSPEs, there is little to choose between standard or noise-robust

measures at this frequency.

We also examine the volatility forecasting performance of alternative jump measures based on a transaction time-based sampling scheme. The transaction-based RV measures are mainly driven by their continuous component, and finite jumps dominate infinite jumps. Using transaction-based volatility measures, the overall forecasting performance of extended HAR models is similar to that of the benchmark HAR-RV model. Our conclusion is that forecasts using realized volatility and jump measures based on transaction sampling are inferior to forecasts using clock-based sampling measures. As our findings relate to the role of jumps using transaction time versus calendar time based sampling, this underscores the importance of the appropriate choice of the sampling scheme.

In the absence of a single dominant forecasting model, we investigate whether various model averaging procedures generate significant forecasting gains. In many cases, we prune the set of models using the MCS of Hansen et al. (2011) to eliminate dominated models. We find that simple model averaging procedures generally result in significant gains in forecasting performance vis-à-vis the single best extended HAR model, which in turn outperforms the benchmark HAR-RV model. For example, model averaged results using equal weights, or the normalized inverse MSPE weights in Bates and Granger (1969) perform as well as model averaged results where the weights minimize the variance of the prediction error.

Appendix 6.A Tables and Figures

Table 6.1: Noise-Robust ABD Test – Size and Power Simulations

	$\xi = 0.01$			$\xi = 0.1$		
	5-Sec.	60-Sec.	300-Sec.	5-Sec.	60-Sec.	300-Sec.
	Size					
ABD Noise-robust	0.059	0.047	0.035	0.051	0.021	0.016
ABD	0.030	0.055	0.128	0.029	0.046	0.084
	Power – Compound Poisson (Finite Jumps)					
ABD Noise-robust	0.999	0.991	0.941	0.963	0.910	0.892
ABD	0.989	0.992	0.988	0.394	0.546	0.622
	Power – Cauchy Process (Infinite Jumps)					
ABD Noise-robust	0.956	0.815	0.746	0.910	0.717	0.546
ABD	0.736	0.770	0.768	0.482	0.572	0.616

Note: The table reports the empirical size and power of the ABD test of Andersen et al. (2007b), and our modified, noise-robust version. ξ is the noise-to-signal ratio used to simulate market microstructure noise. The theoretical size of the tests is 5% ($\alpha = 0.05$). The models and Monte Carlo settings are laid out in Section 6.3.2 of the paper.

Table 6.2: Standard vs. Noise-Robust Realized Semivariances – Finite Sample MSE Performance

	$\xi = 0.01$			$\xi = 0.1$		
	5-Sec.	60-Sec.	300-Sec.	5-Sec.	60-Sec.	300-Sec.
RS^+	9.568	0.067	0.003	967.498	6.737	0.274
RS^-	9.589	0.069	0.004	968.441	6.801	0.287
$TSRS^+$	0.001	0.001	0.002	0.112	0.014	0.008
$TSRS^-$	0.001	0.001	0.002	0.113	0.016	0.009

Note: The table entries are the MSEs of the realized and two-scale realized semivariances in the simulation described in Section 6.3.2 of the paper. The DGP is a Heston model augmented with a finite activity, compound Poisson jumps. ξ represents the noise-to-signal ratio used to simulate the market microstrcture noise. Second-by-second prices were simulated 5,000 times for 5 days with 6.5 trading hours per day.

Table 6.3: Estimated Contribution of Signed, Finite and Infinite Activity Jumps to QV

	SPY			Avg. Stocks			AMZN	BA	BFB	CAT	CHL	COST	CVX
	5s	60s	300s	5s	60s	300s	300s	300s	300s	300s	300s	300s	300s
Continuous	56.798	88.474	85.725	32.399	65.612	70.198	73.426	72.586	55.143	74.899	62.182	69.525	80.277
Jumps	43.202	11.526	14.275	67.601	34.388	29.802	26.574	27.414	44.857	25.101	37.818	30.475	19.723
Pos. Jumps	21.847	6.450	8.257	33.946	16.535	14.992	15.208	14.362	22.474	12.574	17.978	15.963	9.849
Neg. Jumps	21.355	5.075	6.018	33.653	17.853	14.810	11.366	13.052	22.383	12.527	19.841	14.512	9.874
Finite Jumps	10.602	10.419	14.156	33.394	32.417	29.597	26.410	27.228	44.649	24.852	37.314	30.357	19.576
Infinite Jumps	32.600	1.106	0.118	34.207	1.971	0.205	0.165	0.187	0.208	0.249	0.504	0.118	0.147
Pos. Finite Jumps	5.584	5.941	8.219	17.028	15.539	14.883	15.127	14.248	22.380	12.465	17.681	15.892	9.766
Neg. Finite Jumps	5.017	4.478	5.937	16.366	16.878	14.714	11.283	12.979	22.269	12.387	19.633	14.465	9.810
Pos. Infinite Jumps	16.263	0.509	0.038	16.918	0.996	0.108	0.081	0.114	0.093	0.110	0.296	0.070	0.083
Neg. Infinite Jumps	16.338	0.597	0.080	17.287	0.975	0.096	0.084	0.073	0.115	0.140	0.208	0.047	0.064
$\hat{\beta}_{IJA}$	1.454	1.056	0.778	1.455	1.040	0.723	0.461	0.576	0.802	0.621	0.763	0.697	0.748

	DOW	EXC	GILD	GS	HD	JNJ	JPM	KO	OKE	PG	SO	UPS	WMT
	300s	300s	300s	300s	300s	300s	300s	300s	300s	300s	300s	300s	300s
Continuous	68.881	69.488	63.203	75.979	73.935	70.611	76.122	74.208	59.168	71.147	70.791	68.292	74.102
Jumps	31.119	30.512	36.797	24.021	26.065	29.389	23.878	25.792	40.832	28.853	29.209	31.708	25.898
Pos. Jumps	15.029	15.506	18.911	12.311	13.875	12.919	12.926	12.498	19.059	15.416	14.486	15.477	13.013
Neg. Jumps	16.090	15.006	17.886	11.710	12.190	16.470	10.952	13.294	21.773	13.438	14.723	16.231	12.885
Finite Jumps	30.849	30.400	36.458	23.941	25.940	29.279	23.822	25.519	40.602	28.777	28.642	31.527	25.802
Infinite Jumps	0.270	0.112	0.339	0.080	0.125	0.111	0.056	0.273	0.230	0.076	0.568	0.181	0.096
Pos. Finite Jumps	14.830	15.434	18.670	12.297	13.843	12.832	12.899	12.341	18.982	15.365	14.274	15.373	12.968
Neg. Finite Jumps	16.019	14.966	17.788	11.644	12.097	16.447	10.923	13.178	21.620	13.413	14.368	16.154	12.834
Pos. Infinite Jumps	0.198	0.072	0.241	0.014	0.032	0.088	0.028	0.157	0.077	0.051	0.213	0.104	0.045
Neg. Infinite Jumps	0.071	0.040	0.098	0.066	0.093	0.023	0.029	0.116	0.153	0.025	0.355	0.077	0.051
$\hat{\beta}_{IJA}$	0.579	0.725	0.522	0.610	0.665	0.971	0.606	0.913	0.645	0.955	0.878	0.895	0.824

Note: The table reports the estimated percentage contribution of the different jump measures to QV. Results using 5-, 60-, and 300-second returns are shown for SPY and the average of the 20 stocks. The results for the individual stocks were estimated using 300-second returns. $\hat{\beta}_{IJA}$ is the estimated Blumenthal-Getoor index of jump activity (see, Jing et al., 2012b, for more details and settings).

Table 6.4: HAR-RV Benchmark – SPY, 300 Second Returns

	$h = 1$	$h = 5$	$h = 22$	$h = 66$
β_0	0.095^{*}	0.148^{**}	0.288^{***}	0.527^{***}
β_d	0.246^{**}	0.184^{***}	0.103^{***}	0.061^{***}
β_w	0.422^{***}	0.347^{***}	0.322^{***}	0.200^{***}
β_m	0.238^{**}	0.323^{***}	0.290^{***}	0.215^{***}
$R^2_{(in)}$	0.512	0.629	0.562	0.337
$R^2_{(oos)}$	0.443	0.673	0.707	0.470
MSPE	3.102	1.322	0.944	1.262

Note: The table reports the OLS coefficient estimates and in- and out-of-sample R-squared for HAR-RV forecasting regressions for SPY RV at the daily ($h = 1$), weekly ($h = 5$), monthly ($h = 22$) and quarterly ($h = 66$) horizons. The RV measures are calculated using 300 second returns. The significance of the coefficients is based on Newey-West HAC standard errors, allowing for serial correlation up to order 5, 10, 44 or 132 for horizons $h = 1, 5, 22$ and 66 trading days. The superscripts *,**, and *** denote statistical significance at the 10%, 5% or 1% levels. The out-of-sample R-squared, R^2_{oos}, is calculated as one minus the ratio of the MSPE from the HAR-RV model to the MSPE from a model that only has an intercept.

Table 6.5: SPY Extended HAR Regressions Using Total, Positive and Negative Signed Jumps

	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$
	HAR-CJ				HAR-CJ$^+$				HAR-CJ$^-$			
$R^2_{(in)}$	**0.555**	**0.666**	**0.572**	**0.338**	**0.541**	**0.668**	**0.578**	**0.341**	**0.523**	**0.664**	**0.612**	**0.362**
$R^2_{(oos)}$	**0.493**	**0.747**	**0.728**	0.465	**0.450**	**0.754**	**0.739**	**0.489**	**0.511**	**0.724**	0.690	0.445
MSPE	**2.821***	**1.017***	**0.872***	1.274	**3.059**	**0.995***	**0.840***	**1.218***	**2.720***	**1.110***	0.994	1.318
	HAR-CFJ				HAR-CFJ$^+$				HAR-CFJ$^-$			
$R^2_{(in)}$	**0.555**	**0.666**	**0.572**	**0.338**	**0.541**	**0.668**	**0.577**	**0.341**	**0.523**	**0.665**	**0.614**	**0.363**
$R^2_{(oos)}$	**0.493**	**0.747**	**0.728**	0.464	**0.449**	**0.753**	**0.734**	**0.478**	**0.511**	**0.724**	0.684	0.446
MSPE	**2.822***	**1.018***	**0.874***	1.276	**3.066**	**0.998***	**0.857***	**1.243**	**2.721***	**1.112***	0.994	1.317
	HAR-CIJ				HAR-CIJ$^+$				HAR-CIJ$^-$			
$R^2_{(in)}$	**0.512**	**0.630**	**0.563**	**0.340**	**0.512**	**0.630**	**0.576**	**0.381**	**0.512**	**0.629**	**0.563**	**0.339**
$R^2_{(oos)}$	**0.511**	**0.709**	0.644	0.452	**0.509**	**0.711**	0.652	**0.475**	**0.512**	**0.712**	0.651	0.454
MSPE	**2.722***	**1.173***	1.151	1.316	**2.731***	**1.168***	1.125	1.264	**2.714***	**1.162***	1.121	1.299

Note: See Notes to Table 6.4. Bold in-sample and out-of-sample R-squared entries indicate that the fit of the proposed models is better than that of the benchmark HAR-RV model in Table 6.4. Bold MSPE entries are lower than the MSPEs of the benchmark models. Significantly lower MSPE entries at the 5% level are starred. The complete table of coefficient estimates is available on request.

Table 6.6: SPY Relative MSPEs by Frequency – Standard vs. Noise-Robust Measures

	$h = 1$ (day)			$h = 5$ (week)			$h = 22$ (month)			$h = 66$ (quarter)		
	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.
						Panel A: Standard Jump Measures						
HAR-RV	1.000	1.000	1.000	1.000	1.000	1.000†	1.000†	1.000	1.000	1.000†	1.000	1.000
HAR-CJ	1.253	0.755*	0.909*	1.029	0.990	0.770*	0.980*	1.172	0.924*	0.968	1.167†	1.010
HAR-CFJ	0.871*	0.752*	0.910*	1.181	0.992	0.770*	1.051	1.178	0.926*	1.010†	1.171†	1.011
HAR-CIJ	1.124	1.060	0.878*	1.022	1.034	0.888*	0.969*	1.001	1.220	0.940*	0.993	1.043
HAR-CJ$^+$	0.903*	0.993	0.986	1.165	0.969	0.753*	1.147†	0.894*	0.891*	1.074†	0.977	0.965*
HAR-CJ$^-$	0.848*	0.969	0.877*	1.124	1.017	0.840*	0.841*	0.936*	1.053	0.917*	1.020	1.045
HAR-CFJ$^+$	0.925*	0.993	0.988	1.175	0.971	0.755*	1.198†	0.877*	0.908*	1.096†	0.959	0.985
HAR-CFJ$^-$	0.915*	0.969	0.877*	1.215	1.035	0.841*	0.982	0.959*	1.054	1.035†	1.020	1.044
HAR-CIJ$^+$	0.910*	1.055	0.881*	1.151	1.020	0.884*	1.086†	0.964*	1.192	1.136†	0.940*	1.002
HAR-CIJ$^-$	0.729*	1.059	0.875*	0.996	1.030	0.879*	1.054†	0.921*	1.189	0.939*	0.977*	1.029
						Panel B: Noise-Robust Jump Measures						
HAR-RV	0.843*	0.907*	1.009	0.882*	0.976	0.962	0.821*	1.031	1.154	0.893*	1.013	1.014
HAR-CJ	0.768*	0.966	1.015	0.865*	1.010	0.962	0.977	1.044	1.145	0.988	0.996	0.906*
HAR-CFJ	0.775*	0.960*	1.015	0.867*	1.060	0.958*	0.987	1.031	1.143	0.921*	0.925*	1.032
HAR-CIJ	0.791*	0.980	1.018	0.890*	1.025	0.965	0.803*	1.073	1.179	0.875*	1.016	0.998
HAR-CJ$^+$	0.851*	0.684*	1.015	0.884*	0.930*	0.960	0.838*	0.907*	1.145	0.926*	1.037	0.991
HAR-CJ$^-$	0.870*	0.852*	1.013	0.828*	0.889*	0.953*	0.772*	0.912	1.135	0.899*	0.968	0.997
HAR-CFJ$^+$	0.866*	0.677*	1.015	0.895*	0.889*	0.960	0.861*	0.938*	1.145	0.919*	1.037	0.990
HAR-CFJ$^-$	1.111	0.852*	1.013	0.882*	0.894*	0.953*	0.786*	0.902*	1.135	0.931*	0.953	0.753*
HAR-CIJ$^+$	0.794*	0.972	1.026	0.875*	1.005	0.977	0.841*	1.166	1.164	0.930*	1.038	0.994
HAR-CIJ$^-$	1.009	0.958	1.016	0.793*	1.015	0.961	0.794*	0.947*	1.137	0.852*	0.941*	1.000
Memo:												
HAR-RV MSPE	3.364	4.550	3.102	1.553	1.350	1.322	1.443	1.025	0.944	1.778	1.344	1.262

Note: The relative MSPE ratios are the ratios of the MSPEs of the extended HAR models using standard volatility measures (top panel) or noise-robust measures (bottom panel) relative to the benchmark HAR-RV models employing standard measures. The starred MSPE entries indicate statistically significant reductions in the MSPEs at the 5% level. Entries with a dagger, †, denote models not in the MCS. The MSPE and MCS results are respectively based on rolling regression using 1,000 observations and a 10-day block bootstrap with 5,000 replications.

Table 6.7: Twenty Stock averages of Relative MSPEs – Standard vs. Noise-Robust Measures

	$h = 1$ (daily)			$h = 5$ (week)			$h = 22$ (month)			$h = 66$ (quarter)		
	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.
	Panel A – Standard Jump Measures											
HAR-RV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000†	1.000	1.000
HAR-CJ	0.999	0.991	0.972	0.950	0.916	0.933	0.929	0.942	0.970	0.928	0.958	0.995
HAR-CFJ	1.057	0.984	0.973	1.048†	0.916	0.934	1.064†	0.943	0.974	1.043	0.952	0.997
HAR-CIJ	1.010	0.973	0.940	0.986	0.955	0.942	1.035	1.010	1.063	1.007†	1.014†	1.037
HAR-CJ$^+$	1.044	1.000	0.968	1.098†	0.939	0.945	1.203†	0.994	1.038	1.127†	1.004†	1.033
HAR-CJ$^-$	1.063	1.018	0.932	1.038	0.943	0.934	1.144†	0.970	1.026	1.078†	0.997†	1.018
HAR-CFJ$^+$	1.055†	0.999	0.969	1.153†	0.940	0.945	1.267†	0.994	1.038	1.173†	1.004	1.031
HAR-CFJ$^-$	1.103†	0.984	0.932	1.115†	0.938	0.937	1.228†	0.970	1.030	1.144†	0.997†	1.016
HAR-CIJ$^+$	1.044	0.979	0.939	1.090†	0.966	0.946	1.189†	1.010	1.080	1.129†	1.004†	1.042
HAR-CIJ$^-$	1.011	0.982	0.947	1.071†	0.960	0.945	1.213†	1.005	1.091	1.137†	1.006†	1.062
	Panel B – Noise-Robust Jump Measures											
HAR-RV	0.966	0.916	0.969	0.975	1.017	0.998	0.975	1.081	1.138	0.956†	1.050	1.032†
HAR-CJ	0.958	0.935	0.975	0.934	0.975	0.990	0.958	1.077	1.135	0.949	1.040	0.962
HAR-CFJ	0.980	0.939	0.976	0.962	1.003	0.996	0.966	1.082	1.075	0.882	0.963	0.994
HAR-CIJ	0.969	0.926	0.970	0.956	1.022	0.985	0.943	1.064	1.122	0.905	1.042	1.021
HAR-CJ$^+$	0.955	0.986	0.978	0.962	1.008	0.991	0.981	1.082	1.092	0.956	1.042	1.017†
HAR-CJ$^-$	0.973	0.943	0.961	0.950	0.984	0.994	0.938	1.043	1.126	0.936†	1.030	1.019
HAR-CFJ$^+$	0.947	0.987	0.980	0.952	1.010	0.993	0.967	1.086	1.091	0.924†	1.044	1.014
HAR-CFJ$^-$	0.963	0.938	0.962	0.962	0.984	0.994	0.948	1.047	1.107	0.945†	1.031	1.024
HAR-CIJ$^+$	0.972	0.926	0.950	0.957	1.022	0.994	0.966	1.073	1.091	0.952†	1.045	1.008
HAR-CIJ$^-$	0.964	0.935	0.948	0.948	1.025	0.986	0.969	1.061	1.116	0.943†	1.037	1.033
Memo:												
HAR-RV MSPE	373.1364	54.8865	22.7444	85.5808	16.9684	9.9258	27.0842	8.8123	6.3931	17.2674	7.9011	6.2917

Note: The relative MSPE entries are the 20 stock average ratios of the MSPEs of the extended HAR models using standard volatility measures (top panel) or noise-robust measures (bottom-panel) to the MSPEs of HAR-RV models employing standard measures. The entries with a dagger, †, denote models which were retained in the MCS for fewer than 15 stocks. The MSPE and MCS results are respectively based on rolling regression using 1,000 observations and a 10-day block bootstrap with 5,000 replications.

Table 6.8: Model Averaging Results – Relative MSPEs at Different Horizons for SPY, BA, BFB, COST and KO

	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$
	SPY – 300 seconds				SPY – 60 seconds			
Best Extended HAR	0.875*	0.753*	0.891*	0.965*	0.752*	0.969	0.877	0.940*
Avg. – Min Var Weights	0.987	**0.693****	0.895*	0.966*	0.812*	0.977	0.940*	0.971*
Avg. – MSPE Weights	0.879*	**0.706****	**0.862****	**0.919****	0.875*	**0.914****	**0.850***	0.965*
Avg. – Rank Weights	0.910*	**0.715***	**0.845****	**0.873****	0.880*	**0.923***	**0.846***	0.986
Avg. – Equal Weights	**0.873***	**0.712***	**0.876***	**0.928***	0.877*	**0.914****	**0.852***	0.964*
Memo: HAR-RV MSPE	3.102	1.322	0.944	1.262	4.550	1.350	1.025	1.344
	BA – 300 seconds				BFB – 300 seconds			
Best Extended HAR	0.981	0.937	0.993	0.864*	0.924*	0.836*	0.822*	0.876*
Avg. – Min Var Weights	0.992	**0.905****	1.083	1.001	0.969*	0.845*	**0.751****	**0.812****
Avg. – MSPE Weights	**0.972***	**0.906***	**0.915****	0.959*	0.926*	**0.823***	**0.814***	**0.856****
Avg. – Rank Weights	**0.976***	**0.923***	**0.928****	0.980	0.936*	**0.820***	**0.810****	**0.847****
Avg. – Equal Weights	**0.972***	**0.906***	**0.919****	0.961*	0.926*	**0.823***	**0.816***	0.878*
	COST – 300 seconds				KO – 300 seconds			
Best Extended HAR	0.958*	0.879*	0.925*	0.957*	0.814*	0.709*	0.882*	0.939*
Avg. – Min Var Weights	1.016	0.985	**0.881****	**0.950***	0.923*	**0.695****	**0.837****	**0.916***
Avg. – MSPE Weights	0.962*	**0.871***	**0.920***	0.958*	0.817*	0.713*	0.888*	0.975*
Avg. – Rank Weights	0.969*	**0.856***	**0.907****	**0.945****	**0.811***	**0.686***	**0.829****	0.950*
Avg. – Equal Weights	0.962*	**0.873***	**0.922***	0.960*	0.817*	0.723*	0.914*	0.983*

Note: The table reports the relative MSPE, the ratio of MSPE of the model indicated in the first column to the MSPE of the benchmark HAR-RV, in both cases using standard volatility measures as opposed to noise-robust measures. The best models refers to the min. MSPE model from the set of extended HAR models presented in Section 6.4. The bold entries are model averages with lower MSPEs than the MSPEs of both the HAR-RV and the best extended HAR models. The starred entries denote model averages with significantly lower MSPEs than the benchmark HAR-RV models, whereas doubled starred (superscript **) entries identify models whose MSPEs are significantly lower than the MSPEs of both the benchmark HAR-RV and the best extended HAR model.

Table 6.9: Estimated Contribution of Jumps to QV – Comparison of Clock and Transaction Time Sampling Results

	Clock Time Sampling	Transaction Time Sampling
Continuous	85.725	95.413
Jumps	14.275	4.587
Pos. Jumps	8.257	2.279
Neg. Jumps	6.018	2.308
Finite Jumps	14.156	4.503
Infinite Jumps	0.118	0.084
Pos. Finite Jumps	8.219	2.232
Neg. Finite Jumps	5.937	2.271
Pos. Infinite Jumps	0.038	0.047
Neg. Infinite Jumps	0.080	0.038
$\hat{\beta}_{IJA}$	0.778	0.708

Note: The table reports the contribution of the different realized jumps to QV using 300 second clock and transaction-based (78 ticks per interval) sampling.

Table 6.10: SPY Volatility Forecasting Performance – Transaction-Based Sampling Results

	$h = 1$ (day)	$h = 5$ (week)	$h = 22$ (month)	$h = 66$ (quarter)
HAR-RV	1.000	1.000	1.000	1.000
HAR-CJ	0.973*	1.114	1.030	1.023
HAR-CFJ	0.973*	1.114	1.030	1.022
HAR-CIJ	0.981	0.999	1.061	1.017
HAR-CJ$^+$	1.037	1.119	0.956*	0.971*
HAR-CJ$^-$	0.990	1.003	1.036	1.012
HAR-CFJ$^+$	1.037	1.119	0.956*	0.971*
HAR-CFJ$^-$	0.990	1.003	1.036	1.012
HAR-CIJ$^+$	0.981*	0.996	1.052	1.011
HAR-CIJ$^-$	0.980*	0.997	1.064	1.016
Memo: HAR-RV MSPE	3.724	1.500	1.071	1.349

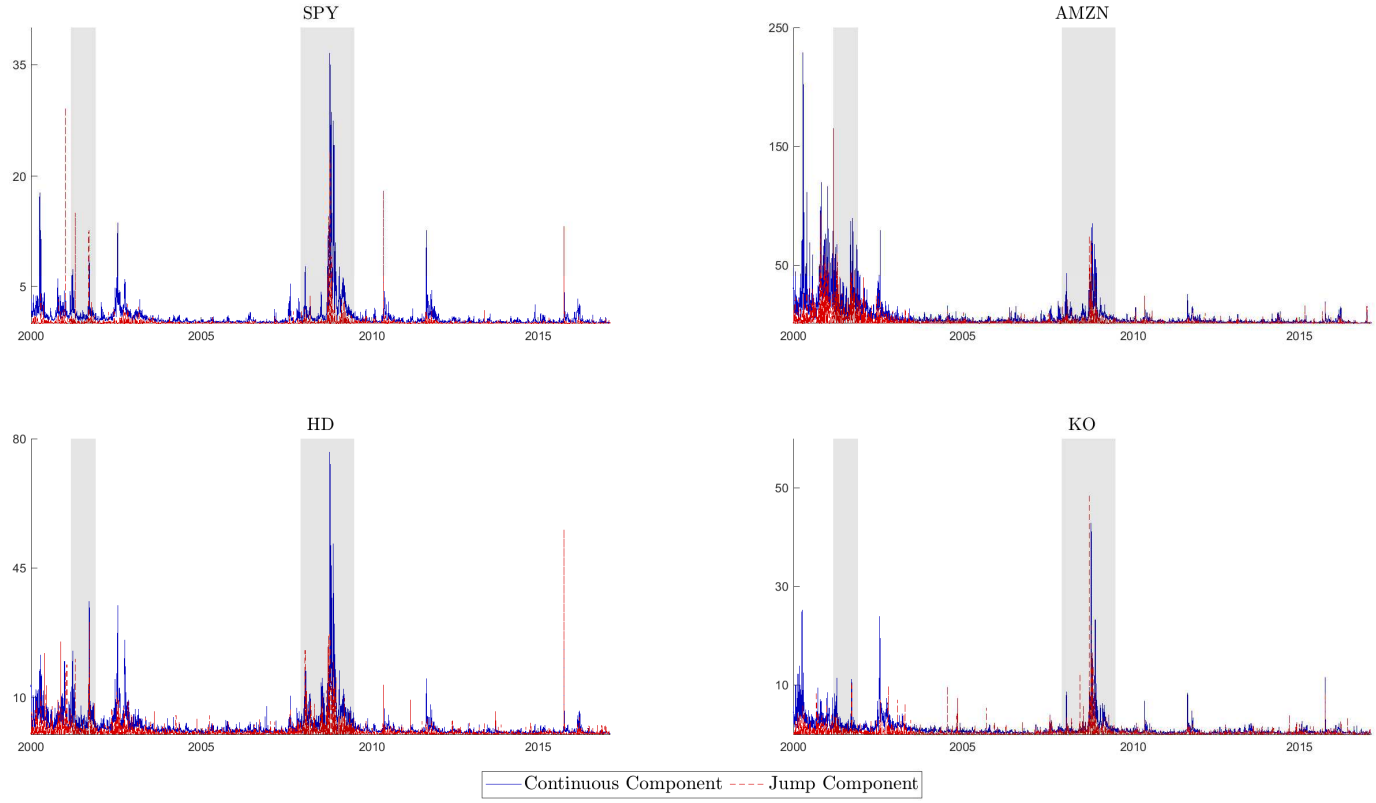
Note: The Table reports the relative MSPE of the extended HAR SPY volatility forecasting models at different horizons. The relative MSPEs are the ratio of the MSPEs of the extended HAR models relative to the benchmark HAR-RV model. The starred entries indicate statistically significant reductions in MSPE identified by the Diebold and Mariano (1995) test using a 5% significance level.

Table 6.11: SPY Model averaging Relative MSPEs – Comparison of Clock and Transaction-Based Sampling Results

	300 second, Clock-Based Sampling				Transaction-Based Sampling			
	$h = 1$	$h = 5$	$h = 22$	$h = 66$	$h = 1$	$h = 5$	$h = 22$	$h = 66$
HAR-RV benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Best Extended HAR	0.875*	0.753*	0.891*	0.965*	0.973*	0.996	0.956*	0.971*
Avg. – Min Var Weights	0.987	**0.693****	0.895*	0.966*	1.009	**0.995**	**0.921****	1.001
Avg. – MSPE Weights	0.879*	**0.706****	**0.862****	**0.919****	**0.926****	**0.950****	**0.889****	**0.961***
Avg. – Rank Weights	0.910	**0.715***	**0.845****	**0.873****	**0.969***	**0.957****	**0.855****	**0.943****
Avg. – Equal Weights	**0.873***	**0.712***	**0.876***	**0.928****	**0.937****	**0.954****	**0.914****	**0.963***
Memo:								
HAR-RV MSPE	3.102	1.322	0.944	1.262	3.724	1.500	1.071	1.349

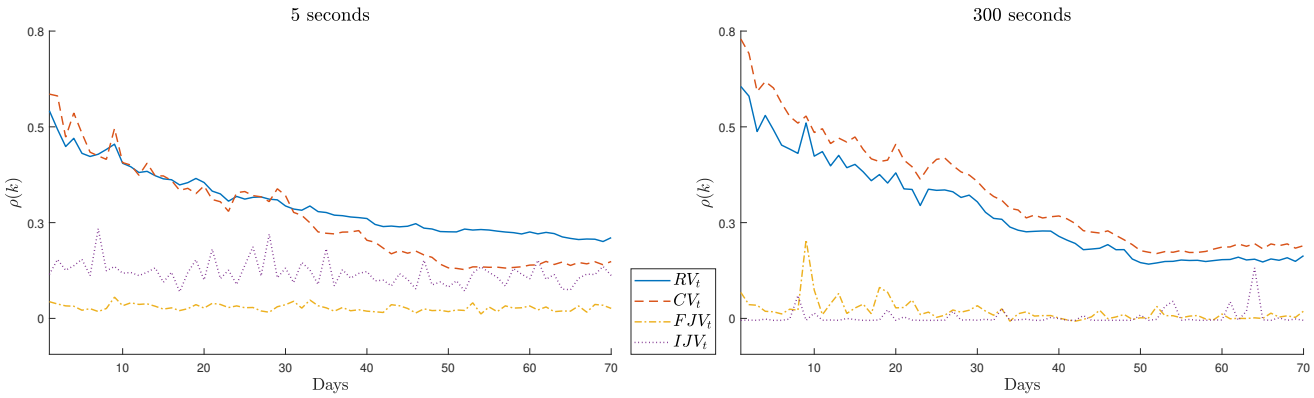
Note: The table compares the forecasting performance of the extended HAR SPY volatility forecasting models at different horizons h using clock and transaction based realized measures. The clock-based results use 300 second returns. The relative MSPEs are the ratio of the MSPEs of the models indicated in the first column to the MSPE of the benchmark HAR-RV model. The bold entries are models averages with lower MSPEs than the MSPEs of both the HAR-RV and the best extended model. The starred entries denote model averages with significantly lower MSPEs than the benchmark HAR-RV models, whereas doubles starred (**) entries identify models whose MSPEs are significantly lower then the MSPEs of both the benchmark HAR-RV and the best extended HAR model.

Figure 6.1: Time Series of Realized Volatility – Jump and Continuous Components



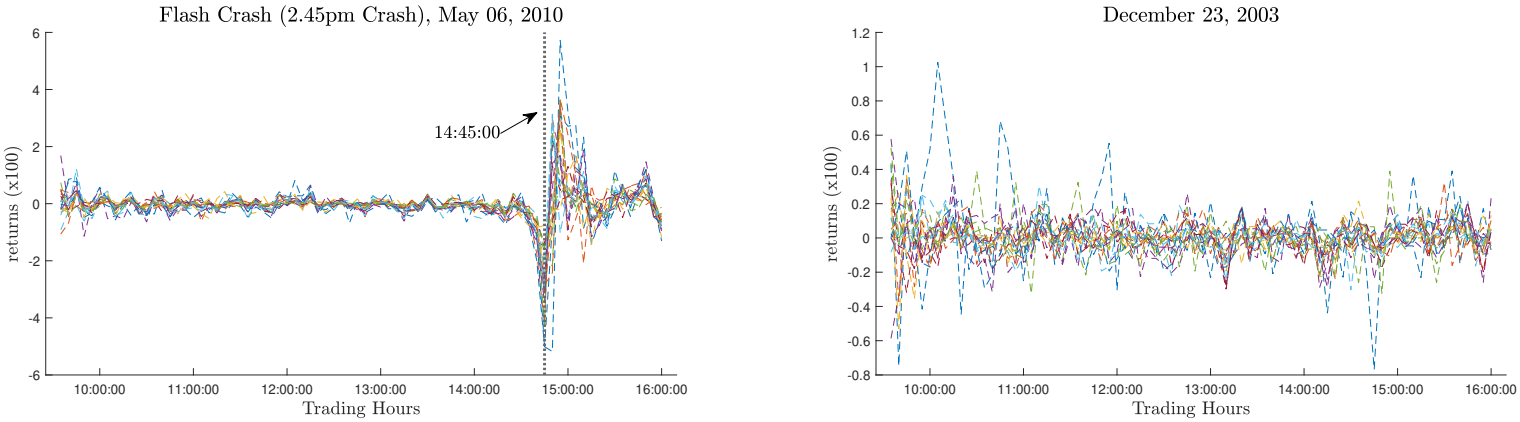
Note: This figure depicts the elements of the realized volatility for SPY and three individual stocks estimated at the 300 second frequency. The three individual stocks have the largest, smalles and average RV. NBER dated U.S. recession are shaded grey.

Figure 6.2: Autocorrelation Function of SPY Realized Measures



Note: The figure graphs the autocorrelation of the realized variance and its elements. The autocorrelations at the 5 and 300 second frequencies were estimated using noise-robust and raw estimators, respectively.

Figure 6.3: Systematic versus Idiosyncratic Jumps



Note: The figure depicts in two plots the intraday returns of the 20 individual stocks across two different trading days. The left plot displays the behavior of the stocks during the Flash Crash of May 06, 2010, where all the stocks jump together, whereas the right panel show a normal day on December 23, 2003, where all jumps are idiosyncratic.

Chapter 7

Modelling and Forecasting Realized Covariances using (Directional) Common Jumps

7.1 Introduction

Market crashes and sudden reactions to major financial news generally trigger the occurrence of common jumps in several stocks, thereby raising statistical correlations among asset prices in consequence of enhanced market-wide information. As this phenomenon raises short-term predictability, it increases (decreases) the persistence of covariances when the common jumps are associated with bad (good) news. The high levels of correlation among common jumps, and the changes in the persistence of covariances, shed light on their rich information content for modelling and forecasting realized covariance matrices.

This paper proposes a robust non-parametric framework for measuring separately the common jumps and continuous components of the quadratic covariation matrix. Our approach builds directly on the theoretical results of Barndorff-Nielsen and Shephard (2004a) and Mancini and Gobbi (2012) that involve the use of so-called realized and truncated realized covariation. The divergence between these two estimators leads to a

matrix of common jumps, as identified with the multi-jump test of Caporin et al. (2017). By employing the approach of the latter, we are able to detect days where all elements in the common jump matrix are distinctly different from zero. We then demonstrate significant forecasting and economic gains at the daily, weekly and monthly horizons, as attained by models that utilize the information of common jumps. Furthermore, we construct measures of directional common jumps, and investigate whether the sign of 'news' contains further explanatory power. Metrics are estimated as the difference between the positive and negative realized semicovariances (e.g. Bollerslev et al., 2020).

Common jumps have many implications for portfolio allocation, risk management, and forecasting. As noted by Das and Uppal (2004) and Longin and Solnik (2001) the increased correlation, that is associated with a general market crash, reduces the diversification potential of portfolio and risk managers. Common jumps are also likely to affect the aggregate attitude to risk, with obvious effects upon risk premia. For instance, Bollerslev and Todorov (2011b); Bollerslev et al. (2015) show that the risk compensation for large jumps is also large and time-varying. Separating the impact of the continuous and common jumps components is also crucial for forecasting covariances. The different explanatory factors for these distinct sources of risk have to be considered with different coefficients in order to account for the Brownian correlations and common jumps (e.g. Andersen et al., 2007a; Corsi et al., 2010, for a similar rationale in a univariate framework).

Contrasting with the volatility forecasting literature, where the role of jumps has been extensively studied (e.g. Andersen et al., 2007a; Busch et al., 2011; Corsi et al., 2010; Duong and Swanson, 2015; Patton and Sheppard, 2015, inter alios), the literature on covariance forecasting largely ignores common jumps. To the best of our knowledge, only Asai et al. (2020) have considered the relevance of common jumps in forecasting bi-variate volatility. Yet, for asset allocations, it is vital to understand the role of common jumps for a large set of assets, as the effect of common jumps in a pair of assets is negligible in large portfolios. A related study (Caporin et al., 2017) that focuses on univariate volatility, finds common jumps to have a greater impact upon volatility than univariate jumps. The procedure of that study allows the detection of simultaneous jumps for a large number

of assets. It thereby demonstrates the multi-jump test to be more powerful than the co-exceedance rule (e.g. Gilder et al., 2014),[1] which fails to identify common jumps that are realized with small lags, and which suffer from both, the slow incorporation of news of low volume stocks and the volatility spikes that are generally associated with common jumps.[2]

Our empirical application considers 20 individual Dow Jones stocks for the period 2000–2016. A preview of our results is as follows: we find common jumps to be strongly associated with major financial and economic news. In particular, we find that FOMC announcements generally trigger simultaneous common jumps, with jump sizes between 0.8–2.0%. Alternatively, flash crashes are associated with jump sizes between 1.5–5% (see Figure 7.1). These results are in line with the findings of Aït-Sahalia and Xiu (2016), Dungey and Hvozdyk (2012) and Lahaye et al. (2011), that macroeconomic announcements are sufficient to explain the occurrence of common jumps, as they significantly change the probability of observing common jumps.

In extending the vech-HAR model of Chiriac and Voev (2011) to account for common jumps, we propose the vech-HARJ and vech-HARCJ models, which are multivariate extensions of the HARJ and HARCJ of Andersen et al. (2007a).[3] Whereas the HARJ model augments the vech-HAR by incorporating a daily common jump variable, the HARCJ model uses the daily, weekly, and monthly levels of the continuous and common jump variables to model future covariance matrices. The incorporation of common jumps results in large in- and out-of-sample improvements vis-à-vis the HAR model; but the HARCJ model delivers larger forecasting gains across all horizons. In general, models based on directional common jumps, deliver forecasts that improve upon the in- and out-

[1]The co-exceedance rule identifies common jumps by intersecting univariate intraday jump tests. For univariate intraday jump tests see Andersen et al. (2007b); Lee and Mykland (2008), while for daily jump tests see Andersen et al. (2012); Barndorff-Nielsen and Shephard (2004b, 2006); Corsi et al. (2010), among others.

[2]The test of Jacod and Todorov (2009), Mancini and Gobbi (2012) and Bibinger and Winkelmann (2015) also identify common jumps, but they are limited to a pair of assets. The test of Bollerslev et al. (2008) identifies common jumps using an aggregate market index.

[3]Multivariate GARCH models are popular alternatives available in the literature for modelling and forecasting covariances (e.g. Bollerslev, 1990; Engle, 2002a; Engle and Kroner, 1995; Noureldin et al., 2012). However, the curse of dimensionality is of relevant consideration, as the number of parameters to be estimated grows very rapidly when the number of assets is large.

of-sample performance of the HAR model. However, their forecasts are inferior to that of models using common jumps.

To assess the relative economic value of the different models, we construct global minimum variance portfolios, which we evaluate using a utility-based approach, as in Fleming et al. (2001, 2003). The use of common and directional common jumps delivers statistical improvements, with economic gains arising from the enhanced accuracy associated with stable covariance matrices. Reduced turnover lowers trading costs: an investor with a risk-aversion of $\gamma = 6$, would be willing to sacrifice up to 100 annual basis points to switch to the models that utilize the common or directional common jumps.

Finally, using simulations of realistic price processes that accommodate for the presence of idiosyncratic and common jumps, we show that our framework successfully disentangles the continuous and common jump parts of the quadratic covariation, and their use in forecasting significantly outperform the forecasts of the standard multivariate HAR model.

The remainder of the paper is organized as follows: Section 7.2 presents the theoretical framework, where the multi-jump test and methodology employed for decomposing the covariance matrix into its continuous and common jumps parts are outlined. Multivariate models and the forecasting evaluation criteria are set out in Section 7.3. The Monte Carlo exercise is described in Section 7.4, and the simulated results are also presented. The data, occurrence of simultaneous jumps and their link with major financial and economic news are reported in Section 7.5. Here, we also report the in- and out-of-sample performance of the multivariate forecast models. Section 7.6 reports the incidence of directional common jumps in forecasting realized covariances. Section 7.7 presents the economic evaluation of the different models based on a utility-based approach. Section 7.8 concludes.

7.2 Theoretical Framework

Let $X = \left(X^{(i)}\right)_{i=1,\dots,N}$ be the log-prices of an N-dimensional vector of assets. We assume that stock prices evolve continuously on a filtered probability space $(\Omega, \mathcal{F}_t, \mathcal{F}_{t\geq 0}, \mathbb{P})$,

and that dynamics for X are as follows:

$$X_t = x_0 + \int_0^t \mu_s ds + \int_0^t \Sigma_s dW_s + dJ_t \tag{7.1}$$

where μ_s is an N-dimensional drift term which is bounded and predictable, $\Sigma_s \equiv \sigma_s'\sigma_s$ is the instantaneous covariance, and W_s is an N-dimensional vector of independent Brownian motions. The unit time interval is normalized to a day. The jump component is of finite activity of the form $J_t^{(i)} = \sum_{s=1}^{N_t^{(i)}} \gamma_{\tau_s^{(i)}}$, for $i = 1, \ldots, N$, where $N_t^{(i)}$ is a non-explosive counting process and $\gamma_{\tau_s}^{(i)}$ are jump sizes at times $\tau_s^{(i)}$. Finally, we assume that jump sizes are such that $\forall s = 1, \ldots,$ we have $\mathbb{P}\left[\gamma_{\tau_s^{(i)}}^{(i)} = 0\right] = 0$, $i = 1, \ldots, N$.

We estimate the realized covariance (e.g. Barndorff-Nielsen and Shephard, 2004a) of the process as follows:

$$RC_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n X\right)' \left(\Delta_j^n X\right) \xrightarrow{\mathbb{P}} \int_0^t \Sigma_s ds + \sum_{0 \leq s \leq t} \left(\Delta X_s\right)' \left(\Delta X_s\right), \tag{7.2}$$

where $\Delta_j^n X = \left(\Delta_j^n X^{(i)}\right)_{i=1,\ldots,N}$ is an N-dimensional vector containing the jth intraday return, and $\Delta_j^n X^{(i)} = X_{j\Delta_n}^{(i)} - X_{(j-1)\Delta_n}^{(i)}$, where $j = 1, \ldots, n$, $\Delta_n = 1/n$ is the sampling interval, and n is the number of high frequency increments per day. ΔX_s denote the N-dimensional vector of jumps occurring at time s, if a jump occurred, and set to zero if no jump occurred at time s.[4]

In presence of only idiosyncratic jumps the matrix of common jumps has a spherical form with (some) non-zero diagonal elements representing the univariate jumps of each stock. By contrast, when stocks co-jump, the matrix of common jumps is formed by non-zero elements. We estimate the integrated covariation (IC) using the threshold realized covariance estimator of Mancini and Gobbi (2012). This estimator is the multivariate extension of the so-called threshold realized variance (e.g. Jacod, 2008; Mancini, 2001,

[4]To illustrate, if in a portfolio, comprised by three assets, all the assets jump jointly, we have that $\Delta X_s = \Delta X_s^{(1)} \Delta X_s^{(2)} \Delta X_s^{(3)} = \gamma_{\tau_s}^{1(23)} \gamma_{\tau_s}^{2(13)} \gamma_{\tau_s}^{3(12)} \Delta_n N_t^{123}$.

2009):

$$TRC_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(\Delta_j^n X \cdot \mathbb{1}_{\{|\Delta_j X| \leq \upsilon_n\}}\right)' \left(\Delta_j^n X \cdot \mathbb{1}_{\{|\Delta_j X| \leq \upsilon_n\}}\right) \xrightarrow{\mathbb{P}} \int_0^t \Sigma_s ds, \tag{7.3}$$

where $\upsilon_n = \left(\upsilon_n^{(i)}\right)_{i=1,\ldots,N} = \alpha^{(i)} \Delta_n^{\varpi}$, for $\alpha^{(i)} > 0$ and $\varpi \in (0, 1/2)$.[5] The multivariate jump matrix can then be obtained as the difference between the realized covariance and the threshold realized covariance as follows:

$$MJ_t = RC_t - TRC_t \xrightarrow{\mathbb{P}} \sum_{0 \leq s \leq t} (\Delta X_s)' (\Delta X_s). \tag{7.4}$$

Since we are interested in cases where all the elements of MJ_t are different from zero, i.e. presence of common jumps, we employ the test of Caporin et al. (2017) to identify only the common jumps that are significantly different from zero.

7.2.1 Multi-jump Test

The multi-jump test of (Caporin et al., 2017, CKR, hereafter) is defined in the following sets:

$$\Omega_t^{MJ,N} = \left\{ \omega \in \Omega \middle| \prod_{i=1}^{N} \Delta X_s^{(i)} \neq 0 \right\}$$
$$\overline{\Omega}_t^N = \Omega \backslash \Omega_t^{MJ,N}.$$

The set $\Omega_t^{MJ,N}$ contains all the trajectories with multi-jumps among all N assets, whereas the complementary set $\overline{\Omega}_t^N$ contains trajectories without multi-jumps in N stocks. However, it can contain jumps and multi-jumps up to $N-1$ stocks. Therefore, the null and

[5] As it is customary in the literature (e.g. Aït-Sahalia and Jacod, 2014; Todorov and Bollerslev, 2010; Todorov and Tauchen, 2014, among others), we choose $\alpha^{(i)} = 3\sqrt{BV_t^{(i)}}$ and $\varpi = 0.49$. $BV_t^{(i)} = \mu_1^{-2} \frac{n}{n-1} \sum_{j=2}^{\lfloor 1/\Delta_n \rfloor} |\Delta_j^n X^{(i)}||\Delta_{j-1}^n X^{(i)}|$ is the so-called bipower variation of Barndorff-Nielsen and Shephard (2004b), and $\mu_1 \equiv \mathbb{E}[|\mathcal{N}(0,1)|] = \sqrt{2/\pi}$.

alternative hypotheses are:

$$H_0 : (X_t(\omega))_{t\in[0,t]} \in \overline{\Omega}_t^N \quad \text{v.s.} \quad H_1 : (X_t(\omega))_{t\in[0,t]} \in \Omega_t^{MJ,N}.$$

The CKR test is based on two jump-robust integrated variance estimators, which generalize the truncated realized variance estimator of Mancini (2001, 2009), and are named smoothed realized variance (SRV). The first SRV takes the following form:

$$SRV_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_j^n X^{(i)}|^2 \cdot K\left(\frac{\Delta_j^n X^{(i)}}{H_{j\Delta_n,n}^{(i)}}\right), \tag{7.5}$$

where $X^{(i)}$ and $H^{(i)}$ are the respective i-th component of the vectors X and H. $K(\cdot)$ is kernel estimator,[6] and $H_{t,n}$ is the bandwidth defined as:

$$H_{t,n}^{(i)} = h_n \cdot \hat{\sigma}_t^{(i)} \sqrt{\frac{T}{n}}, \tag{7.6}$$

where $\hat{\sigma}_t^{(i)}$ is a point estimator of the local standard deviation of the i-th stock, $i = 1, \ldots, N$. h_n is the bandwidth parameters, where its role is to gauge the largeness of high-frequency returns with respect to the local volatility.[7]

The second SRV is outlined as:

$$\widetilde{SRV}_t^N = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_j^n X^{(i)}|^2 \cdot \left(K\left(\frac{\Delta_j^n X^{(i)}}{H_{j\Delta_n,n}^{(i)}}\right) + \prod_{k=1}^{N}\left(1 - K\left(\frac{\Delta_j^n X^{(k)}}{H_{j\Delta_n,n}^{(k)}}\right)\right)\right). \tag{7.7}$$

Returns in the above estimator are smoothed twice. Whereas the first term, $K\left(\frac{\Delta_j^n X^{(i)}}{H_{j\Delta_n,n}^{(i)}}\right)$, has the same effect as in the first SRV, the second term, $\left(1 - K\left(\frac{\Delta_j^n X^{(k)}}{H_{j\Delta_n,n}^{(k)}}\right)\right)$, leaves the corresponding return similar to the raw returns when all multivariate returns are big. Although both smoothing procedures are meant to eliminate jumps, the smoothing in

[6] As pointed out by the authors, when the kernel function is $K(x) = \mathbb{1}_{|x|\leq\epsilon}$ this estimator is equivalent to the truncated realized variance of Mancini (2001, 2009).

[7] We follow the authors and use the same h_n across all the stocks, as the normalization is respect to each stock volatility. As pointed out by Caporin et al. (2017), the advantage of replacing the indicator function with a kernel is that it provides an estimator that depends smoothly on the bandwidth, which stabilizes the procedure in small samples.

the second SRV allows multi-jumps to survive.

The test statistic is based on the difference between the two SRV estimators. In the absence of multi-jumps, this difference tends to zero, while under the alternative of multi-jumps this difference becomes large and positive. However, the authors need to randomize one of them to obtain a non-degenerate limit distribution under the null. To do so, they apply the wild bootstrap technique suggested in Podolskij and Ziggel (2010), and replace the first SRV by the smoothed randomized realized variance (SRRV):

$$SRRV_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_j^n X^{(i)}|^2 \cdot K\left(\frac{\Delta_j^n X^{(i)}}{H^{(i)}_{j\Delta_n,n}}\right) \cdot \eta_j^i, \tag{7.8}$$

where $\left(\eta_j^i\right)_{1\leq i\leq N, 1\leq j\leq n}$ is an $N \times n$ matrix of independent and identically distributed (i.i.d.) draws with $\mathbb{E}[\eta_j^i] = 1$ and $\text{Var}[\eta_j^i] = V_n < \infty$. In our application we follow the authors by allowing η_j^i to take values in $\{1+\tau, 1-\tau\}$ with equal probability, so that $V_n = \tau^2$. We set $\tau = 0.05$ so that, in practice, SRRV is virtually indistinguishable from SRV.

The test statistics to detect multi-jumps is described as:

$$S_t^N = \frac{1}{V_n} \sum_{i=1}^{N} \frac{(SRRV_{t,i} - \widetilde{SRV}^N_{t,i})^2}{SQX_{t,i}}, \tag{7.9}$$

where:

$$SQX_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} |\Delta_j^n X^{(i)}|^4 \cdot K^2\left(\frac{\Delta_j^n X^{(i)}}{H^{(i)}_{j\Delta_n,n}}\right), \quad i = 1, \ldots, N. \tag{7.10}$$

The authors show that if $\left(\eta_j^i\right)_{1\leq i\leq N, 1\leq j\leq n}$ is pairwise independent, as $n \to \infty$, it holds that:

$$\begin{cases} S_t^N \xrightarrow{d} \chi^2_N, & on\ \bar{\Omega}_T^N, \\ S_t^N \xrightarrow{d} +\infty & on\ \Omega_T^{MJ,N} \end{cases}, \tag{7.11}$$

where χ^2_N denotes the χ-square distribution with N degrees of freedom.

7.2.2 Disentangling the Continuous and Discontinuous Component

The common jump estimator in equation (7.4) provides consistent estimates of the multivariate jump component as $\Delta_n \to 0$. In practice, the difference between the realized and threshold realized covariance can be non-zero owing to finite sample problems. As we are interested only in cases where the full multivariate jump matrix is non-zero, we use the multi-jump test (defined above), to disentangle significant common jumps as follows:

$$\begin{aligned} IC_t &= (1 - Z_t) \cdot RC_t + Z_t \cdot TRC_t, \\ CJ_t &= RC_t - IC_t, \end{aligned} \tag{7.12}$$

where $Z_t = \mathbb{1}\left(S_t^N > z_\theta\right)$ and $\mathbb{P}(\chi_N^2 > z_\theta) = \theta$, for $\theta \in (0, 1)$. In this paper we use $\theta = 0.01$, and $h_n = h_0 + \frac{c}{N}$, where $h_0 = 1.4$ and $c = 9.57$.[8] The matrix IC_t is equal to the realized covariance when there are no common jumps on day t, while in the presence of common jumps IC_t is equal to the threshold realized covariance (e.g. Andersen et al., 2007a, for a similar approach in a univariate framework). We ignore the impact of idiosyncratic jumps,[9] as they have a negligible effect in large portfolios. This is because a specific non-major financial news might be perceived as either good or bad news across some stocks, resulting in respective positive and negative jumps, whose impact is offset in large portfolios.

7.3 Forecasting Models and Evaluation Criteria

Following Chiriac and Voev (2011),[10] we use the vector heterogeneous autoregressive (vech-HAR) model to estimate and forecast the realized covariance matrix. This model extends the so-called HAR model of Corsi (2009), so that realized covariance is expressed

[8] Those values are recommended by the authors as conservative choices.

[9] Aït-Sahalia et al. (2020) find that idiosyncratic jumps are related to stock specific events, such as earning disappointments.

[10] Similar to the HAR model (Corsi, 2009), the vech-HAR model approximates long-memory in a parsimonious way. The model involves a fixed number of parameters regardless of the number of assets, which makes it very easy-to-estimate. This model features in the work of Bauer and Vorkink (2011); Bollerslev et al. (2018); Hautsch et al. (2015), inter alios.

as a linear combination of past daily, weekly and monthly covariances:

$$S_{t+h} = \theta_0 + \theta_d S_t + \theta_w S_{t-5|t} + \theta_m S_{t-22|t} + \epsilon_{t+h}, \tag{7.13}$$

where $S_t \equiv \text{vech}(RC_t)$ denote the $N^* = N(N+1)/2$ dimensional vectorized version of the realized covariance matrix of interest RC_t. $S_{t-h|t} = \frac{1}{h}\sum_{i=1}^{h} S_{t-i}$ denote the vectorized version of the h-day realized covariance matrix. The intercept θ_0 is an $N^* \times 1$ dimensional vector, while θ_d, θ_w and θ_m parameters are all assumed to be scalar. This simply and extremely parsimonious specification ensures that the covariance matrix forecasts are positive definite.[11]

The standard vech-HAR formulation in equation (7.13) does not distinguish between Brownian correlation and common jumps. In order to capture those different sources of risk, they must be separately modelled, so that the distinct explanatory factors have their own coefficients. To achieve this, we propose two simple modifications in the spirit of Andersen et al. (2007a). The first specification extends the vech-HAR model by incorporating a daily measure of the common jumps, this specification is the vech-HARJ model:

$$S_{t+h} = \theta_0 + \theta_d S_t + \theta_w S_{t-5|t} + \theta_m S_{t-22|t} + \theta_{J_d} J_t + \epsilon_{t+h}. \tag{7.14}$$

The second specification uses the continuous and multi-jump parts of the realized covariances as shown in Section 7.2.2. This structure fully incorporates both Brownian correlation and common jumps, because the information content of idiosyncratic jumps remaining in the continuous part is negligible when the number of assets is large. The

[11] A generalization of the vech-HAR model considers that each variance-covariance term has its own dynamics. Therefore, each element in the covariance is estimated separately, which increases the number of parameters from 4 to $4 \times N^*$. As a result, these models yield to forecasts which are not positive definite, and most of the time lead to worse forecasting performance, especially when the number of assets is large.

vech-HARCJ model is outlined as follows:

$$S_{t+h} = \theta_0 + \theta_{C_d} C_t + \theta_{C_w} C_{t-5|t} + \theta_{C_m} C_{t-22|t} + \theta_{J_d} J_t + \theta_{J_w} J_{t-5|t} + \theta_{J_m} J_{t-22|t} + \epsilon_{t+h}, \tag{7.15}$$

where $J_t \equiv \text{vech}(CJ_t)$ and $C_t \equiv \text{vech}(IC_t)$.

We evaluate the forecasting capability of models based on the Frobenius distance, L_t^F, which extends the mean squared error loss function to the multivariate space, and the Euclidean loss function, L_t^E, computed by equally-weighting all the unique elements of the forecast error matrix:

$$L_t^F = \sqrt{\text{Tr}\left[\left(\widehat{S}_t - RC_t\right)\left(\widehat{S}_t - RC_t\right)'\right]}, \tag{7.16}$$

$$L_t^E = \sqrt{\text{vech}\left(\widehat{S}_t - RC_t\right)' \text{vech}\left(\widehat{S}_t - RC_t\right)}, \tag{7.17}$$

where $\widehat{S}_t$ denote the fitted covariance matrices, and RC_t is the ex-post realized covariances. As discussed in Laurent et al. (2013) and Patton (2011b) the ranking produced by both loss functions based on covariance proxies is consistent with those based on the true latent covariance matrix.

We employ the conditional predictive ability (CPA) test of Giacomini and White (2006), to identify models whose losses are significantly smaller than those of the vech-HAR model. Although this approach was developed to asses forecasts in the univariate setting, it directly translates into a multivariate setting when the loss function generates a scalar measure.

7.4 Monte Carlo Evidence

Using the setup of Barndorff-Nielsen et al. (2011), we simulate a multivariate factor stochastic volatility model for $X_{i,t}$, $i = 1, \ldots, 10$ as:

$$\begin{aligned} dX_t^{(i)} &= \mu^{(i)}dt + \sigma_t^{(i)}\left(\rho^{(i)}dB_t^{(i)} + \sqrt{1-(\rho^{(i)})^2}dW_t\right) + dq_t^{(i)}J_t^{(i)} + dk_t L_t^{(i)} \\ \sigma_t^{(i)} &= \exp\left\{\beta_0 + \beta_1 \nu_t^{(i)}\right\}, \\ d\nu_t^{(i)} &= \alpha\nu_t^{(i)}dt + dB_t^{(i)}. \end{aligned} \tag{7.18}$$

The elements of $B_t^{(i)}$ are independent standard Brownian motions and are also independent of W_t. Following Barndorff-Nielsen et al. (2011), we set the parameters to $(\mu, \beta_0, \beta_1, \alpha, \rho) = (0, -5/16, 1/8, -1/40, -0.83)$. The true spot correlation of $X_t^{(i)}$ and $X^{(k)}$ is constant and equals $\sqrt{\left(1-(\rho^{(i)})^2\right)\left(1-(\rho^{(k)})^2\right)}$ for $i \neq k$. The leverage between $X_t^{(i)}$ and $\nu_t^{(i)}$ is $\rho^{(i)}$. The fact that ρ is set equal for all i leads to an equicorrelation structure with common correlation of 0.31.[12] The stationary distribution of ν_t is used to restart the process each day at $\nu_0^{(i)} \sim \mathcal{N}\left(0, \left(-2\alpha^{(i)}\right)^{-1}\right)$.

The idiosyncratic jumps are modelled as independent compound Poisson processes, $dq_t^{(i)}$, with intensity $\lambda_J = 0.2$, and jump sizes $\mathcal{N}(0, 0.628)$. The common jumps are determined by a unique compound Poisson process, dk_t, with jump intensity $\lambda_L = 0.1$ and jump sizes $\mathcal{MN}(0, \Pi)$, where $\Pi = \text{diag}(\varrho)\Gamma\,\text{diag}(\varrho)$. Γ is an equicorrelation matrix with common correlation of 0.75, where ϱ is the diagonal matrix containing the standard deviations, which are equal to 0.756. The idiosyncratic and common jump intensities are chosen such that all discontinuities account for about 30% of the sample. As we simulate $T = 2,000$ days, there are approximately 600 jumps in total, from which about 200 are common jumps. The jump sizes are set to account for about 30% of the total quadratic variation of each asset, with common jumps contributing up to 10% (out of the 30%).

The price process is simulated via an Euler scheme, and we normalize one second to be $\Delta_n = 1/23400$, so that the interval $[0, 1]$ contains 6.5 hrs. In generating the observed

[12] This level of correlation among assets is similar to that found across the asset used in our empirical analysis.

price, we discretize $[0,1]$ into a number $n = 23,400$ of intervals. The prices are then aggregated to the 5-min, which is equivalent to 78 observations per day.[13] The forecasts are based on a rolling window of 500 days and 1,000 replications.[14]

Table 7.1 reports the simulated daily, weekly and monthly in- and out-of-sample forecasts of the HAR, HARJ, and HARCJ models. The results of Panels A and B are based on common jumps; and those in Panels C and D are based on directional common jumps (see Section 7.6 for details about the estimation of directional common jumps). As shown in Panels A and B, the incorporation of common jumps not only improves the in-sample fits of the realized covariances, but also the out-of-sample forecast accuracy. The use of directional common jumps also improves on the in- and out-of-sample forecasts of the HAR model irrespective of the forecast horizon.

Across all forecast horizons, the HARCJ model consistently outperforms both the HAR and HARJ models, with forecasting gains increasing slightly as the horizon lengthens. This result suggests that separating out these two sources of risk increases the persistence of the models relative to that of the HAR model. Hence, the inference that models explicitly accounting for the presence of common jumps provide more accurate predictions of covariance matrices at longer horizons.

7.5 Empirical Results

7.5.1 Data

We consider 20 individual Dow Jones stocks from the period 2000–2016. The data are sourced from the TickData database. We use the previous tick interpolation to aggregate the data down to the required sampling frequency. We sample returns every 5-minutes, which results in 78 observations per day. This sampling frequency is customary in the high-frequency literature, as it provides a trade-off between achieving enough statistical power and avoiding distortions that may arise from microstructure noise (e.g. Hansen and

[13] Monte Carlo results for 10- and 15-min are qualitatively similar to those obtained at the 5-min sampling, and therefore not reported here.

[14] The rolling window size in our simulation represents 25% of the full sample size ($T = 2,000$), which is equivalent to the rolling window size used in our empirical study.

Lunde, 2006; Patton, 2011a). To validate the latter statement, we perform the Hausman test for the presence of market microstructure noise of Aït-Sahalia and Xiu (2019). We use the third test ($H_{3,n}$) proposed by the authors as this test is robust to jumps, and as a robustness check we employ the first-order autocorrelation in log-returns test (AC_n).[15] The last two columns of Table 7.2 report the proportion of rejections of both $H_{3,n}$ and AC_n across all individual stocks. The average proportion of rejections is respectively 0.025 and 0.017 for $H_{3,n}$ and AC_n, suggesting that the level of microstructure noise in our dataset is negligible at the 5-minutes sampling frequency, and therefore we can treat our dataset as noise-free.

Table 7.2 reports the descriptive statistics of the 20 individual stocks together with the number of identified common jumps and their contribution to the total variance. The second panel shows the descriptive statistics of the common jumps – based on days where the null of no simultaneous jumps is rejected. American Express (AXP) and JP Morgan Chase (JPM) display the highest average (co)jump size, with peaks detected during the global financial crisis. The contribution of common jumps to the total variance, displayed in the last column of Table 7.2, shows values ranging between 1.4–5%, with an average of 2.2%. Previous findings in the literature document a contribution of jumps to the total variance of about 6% (Huang and Tauchen, 2005), which means that common jumps make up a significant portion of the total jump part.

Table 7.3 displays in the lower (upper) triangular section, the correlations of the common jumps (continuous component) across all the individual stocks. Unlike idiosyncratic jumps, the simultaneous arrival of common jumps triggered by the wide-market economic information results in highly correlated jump measures, with an average correlation of 0.67. Surprisingly, the level of correlation found in common jumps is slightly higher than that of the continuous component whose average correlation is 0.65. This finding is in line with Das and Uppal (2004) and Longin and Solnik (2001) who note that the increase in correlation, after a collective crash, is mainly explained by the occurrence of common jumps.

[15] These two tests are recommended by the authors, however our results are qualitatively similar under the alternative Hausman tests proposed by Aït-Sahalia and Xiu (2019).

7.5.2 Common Jumps and Major Financial and Economic News

Over 149 days, our empirical results show stocks to have jumped simultaneously.[16] Our results indicate that over 85% of these common jumps are strongly associated with major financial and economic news. However, days with common jumps account for less than 5% of the total major financial and economic news available for the period under analysis. For instance, FOMC meetings are scheduled every 6 weeks, which means, between 2000–2016, at least 136 meetings took place.[17]

To illustrate some of these findings, Figure 7.1 lists 6 days when the CKR test detected simultaneous jumps: i) May 06, 2010 – in a flash crash (aka the crash of 14:45 hrs) US stocks lost one trillion dollar in 36 minutes, with the Dow Jones losing 998.5 points or 9%. However, the loses were rapidly recovered; ii) April 23, 2013 – a flash crash was associated with a false report of White House explosions. This news triggered a 143-point fall in the Dow Jones. However, the fake tweet was immediately corrected, allowing the stock market to recover the big losses within minutes; iii) February 03, 2016 – US stock market indices gains reflected an 8% rise in oil prices as Russia and OPEC cut production. This effort translated in 183 points gained by S&P 500 index; iv) November 06, 2002 – The Fed cut both the target federal reserve funds rate, and overnight bank lending rate, by half a percentage point to 1.25 percent, a new 40-year low; v) August 09, 2011 – large stock market gains followed the Fed announcement that interest rates would remain low until 2013; vi) January 27, 2016 – as disappointing quarterly reports renewed concerns over economic growth the Dow Jones Industrial average closed lower by 222 points.

In line with Lahaye et al. (2011) and Dungey and Hvozdyk (2012), we find that macroeconomic announcements are generally sufficient to produce simultaneous jumps, with returns jumping between 0.8–2.0%. The resulting sign of the simultaneous jumps triggered by these type of news is often negative. This result is consistent with that of Amengual and Xiu (2018), as they find that downward intraday volatility jumps in

[16] Using different bandwidth settings, Caporin et al. (2017) identify between 20–101 simultaneous jumps.

[17] Other important news are, among others, the Consumer Price index (CPI), Producer Price Index (PPI), Manufacturing Composite Index (MCI) and employee on non-farm payrolls, all released monthly, while DGP-related news are released quarterly. For a full list of the scheduled macroeconomic announcements, the reader can consult Lahaye et al. (2011).

the S&P 500 index are often associated with a resolution of policy uncertainty, mostly through statements from the FOMC meetings and speeches by the chair of the Federal Reserve. Our results also indicate that simultaneous jumps are realized within 30 minutes around the schedule time.[18] On the other hand, flash crashes or market sell-offs usually spark bigger simultaneous jumps, with sizes between 1.5–5%.[19]

7.5.3 In-Sample Estimates

Table 7.4 reports the in-sample parameter estimates with robust standard errors in parentheses. The last three rows report the goodness-of-fit measures for the different models across three forecast horizons: one-day ($h = 1$), one-week ($h = 5$) and one-month ($h = 22$). As expected, the daily, weekly and monthly parameter estimates of the realized covariance and its continuous component are strongly significant across all forecasting horizons. The magnitudes of the parameters are similar across all the models; however, the HAR-J and HAR-CJ give greater weight to the daily and weekly estimates, and thus those forecasting models react faster to new information.

In respect of jump variables, for the HAR-J model the jump estimate is generally negative and significant across all forecasting horizons. This result is in line with the findings obtained in the univariate framework (e.g. Andersen et al., 2007a; Corsi et al., 2010). Contrasting with the results for the univariate literature, where most of the jump estimates of the HARCJ model are negative and insignificant (e.g. Andersen et al., 2007a), common jumps estimates in the multivariate HARCJ model are generally positive and strongly significant across all forecasting horizons. This supports the view that the information that is implicit from common jumps (increasing both correlation among assets and the persistence of stock (co)variances) can improve the accuracy of multivariate forecasts. The greater persistence from the HARJ and HARCJ directly translates into 'better fits' for in-sample models.

[18] The manufacturing report released on July 1, 2011 is an example of another macroeconomic announcement that triggers simultaneous jumps at the released time (10:00 hrs), with returns jumping more than 1%.

[19] Other relevant stock market sell-off identified in our sample are on August 18 and 24 and 25, 2015 due to fear of a lack of liquidity in the market, and on Jun 24, 2016 due to the Brexit Referendum.

7.5.4 Out-of-Sample Forecasts

Out-of-sample results across three different forecast horizons are shown in Table 7.5. Bold numbers indicate losses of the HARJ and HARCJ models outperforming those of the HAR models. Starred p-values of the CPA test of Giacomini and White (2006) indicate that losses of the HARJ and HARCJ models are smaller than those from HAR model at the 5% significance level.

As with the in-sample results, and as confirmed by the CPA test, the HARJ and HARCJ models consistently outperform those of the HAR model irrespective of the loss function and forecast horizon considered.[20] For instance, the null of equal predictive ability is rejected across all forecasting horizons for the HARCJ and at the one-month horizon for the HARJ model.

The bigger forecasting gains from the HARCJ model are in line with the univariate forecasting literature (e.g. Andersen et al., 2007a; Duong and Swanson, 2015). The associated rationale is as follows: As the HARCJ model relies upon full decomposition of the covariance matrix (continuous and common jumps), it captures the two distinct sources of risk and their different dependencies across various horizons. Moreover, some financial and economic news might have a longer impact in the stocks, and therefore the inclusion of the weekly and monthly common jumps variables increase the predictability of covariance matrices, as the coefficients of these measures capture this residual information.

7.6 Directional Common Jumps

This section investigates the incidence of directional common jumps in forecasting realized covariance matrices. Bollerslev et al. (2020) show that the realized covariance matrix may be decomposed into four distinct elements: two based on concordant signs

[20] This finding is also corroborated by our Monte Carlo experiment in Section 6.3.2.

and two based on discordant signs as follows:[21]

$$\begin{aligned} \mathtt{P}_t &= \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} p\left(\Delta_j^n X\right)' p\left(\Delta_j^n X\right), \quad \mathtt{N}_t = \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} n\left(\Delta_j^n X\right)' n\left(\Delta_j^n X\right), \\ \mathtt{M}_t &= \sum_{j=1}^{\lfloor 1/\Delta_n \rfloor} \left(p\left(\Delta_j^n X\right)' n\left(\Delta_j^n X\right) + n\left(\Delta_j^n X\right)' p\left(\Delta_j^n X\right)\right), \end{aligned} \tag{7.19}$$

where $p(x) \equiv \max\{x, 0\}$ and $n(x) \equiv \min\{x, 0\}$ denote the component-wise positive and negative elements of the real vector x. Therefore, $\mathtt{P}_t$ and $\mathtt{N}_t$ correspond to the positive and negative realized semicovariance matrices, while $\mathtt{M}_t$ is the sum of the two discordant elements. In the presence of jumps, these three elements contain both diffusive and jump covariation components. As shown in Bollerslev et al. (2020, Section 2), the limiting behavior of $\mathtt{P}_t$ and $\mathtt{N}_t$, derived for a pair of assets, only differs in the jump component:

$$\mathtt{P}_t \xrightarrow{\mathbb{P}} \int_0^t \phi_{j,s}\phi_{k,s}\psi(\rho_{jk,s})ds + \sum_{0\leq s\leq t} p\left(\Delta X_s\right)' p\left(\Delta X_s\right), \tag{7.20}$$

$$\mathtt{N}_t \xrightarrow{\mathbb{P}} \int_0^t \phi_{j,s}\phi_{k,s}\psi(\rho_{jk,s})ds + \sum_{0\leq s\leq t} n\left(\Delta X_s\right)' n\left(\Delta X_s\right), \tag{7.21}$$

$$\mathtt{M}_t \xrightarrow{\mathbb{P}} -2\int_0^t \phi_{j,s}\phi_{k,s}\psi(-\rho_{jk,s})ds + \sum_{0\leq s\leq t} \left(p\left(\Delta X_s\right)' n\left(\Delta X_s\right) + n\left(\Delta X_s\right)' p\left(\Delta X_s\right)\right), \tag{7.22}$$

where $\phi_{i,t} = \Sigma_{ii,t}^{1/2}$ is the spot volatility of asset i, $\rho_{ik,t} = \Sigma_{ik,t}/(\phi_{j,t}\phi_{k,t})$ denotes the spot correlation coefficient between assets i and k, and $\psi(\rho) = (2\pi)^{-1}\left(\rho\arccos(-\rho) + \sqrt{1-\rho^2}\right)$.[22] Given that the diffusive component of the positive and negative semicovariances is the same, the intuition that these measures carry distinct economic information about the good and bad news over each day resides only in their jump component. Therefore, the directional common jumps are estimated as follows:[23]

$$DCJ_t = \mathtt{P}_t - \mathtt{N}_t \xrightarrow{\mathbb{P}} p\left(\Delta X_s\right)' p\left(\Delta X_s\right) - n\left(\Delta X_s\right)' n\left(\Delta X_s\right). \tag{7.23}$$

[21] The realized semicovariances proposed by Bollerslev et al. (2020) can be seen as a multivariate extension of the realized semivariance pioneered by Barndorff-Nielsen et al. (2010).

[22] $\psi(\rho)$ corresponds to $\mathbb{E}[Z_1, Z_2 \mathbb{1}_{\{Z_1<0, Z_2<0\}}]$ bivariate standard normally distributed with correlation ρ.

[23] Significant directional common jumps are identified by intersecting this measure with the indicator function of the CKR test as, $DCJ_t \cdot Z_t$.

To illustrate the impact of common jumps in the positive and negative semicovariances, Figure 7.2 depicts the intraday returns on the day of the common jumps, and the positive and negative semivariances for a five-day trading period around the date of the common jumps. Two subplots (top panel) show the intraday returns (left) on June 29, 2006 together with the positive and negative semicovariances (right) between 26 and 30 June 2006. On June 29, the Fed raised the short-term rate by a quarter-percentage point. There followed a positive jump in the S&P 500 of about 0.7% intraday and a daily change of 2.2%. This positive simultaneous jumps is fully absorbed by the positive semicovariance, and therefore its magnitude is much bigger than that of the negative semicovariance.

The bottom panel plots the intraday returns (left) on August 5, 2014, and the concordant elements of the realized covariance (right) during the week beginning on August 4, 2014. On August 5, 2014, Russian troops were reported lining on the borders of Ukraine. This news triggered a negative simultaneous jumps, which is completely absorbed by the negative semicovariance as shown in Figure 7.2. This evidence suggests that the information content and dynamic dependencies of the directional common jumps might improve the forecasting accuracy of realized covariance matrices.

7.6.1 In-Sample Estimates

Table 7.6 reports in-sample parameter estimates (robust standard errors in parentheses) for daily ($h = 1$), weekly ($h = 5$) and monthly ($h = 22$) forecasts, together with the respective measures of fit. As with the results in Table 7.4, the parameter estimates of the realized covariance and its continuous component are strongly significant across all forecasting horizons. However, directional common jumps show a different pattern to that of common jumps. The estimates are generally significantly negative across all horizons. As negative/positive common jumps are fully absorbed by the negative/positive semicovariances, their difference reflects the direction of the common jumps, and therefore the negative estimates indicate that common jumps increase the persistence and future level of the (co)variance. This finding corroborates those of Patton and Sheppard (2015), that

negative semivariances are more important than their positive counterparts for modelling and forecasting in the univariate framework.

The measures of fit (Table 7.6, last 3 rows) indicate that the use of directional common jumps increases the in-sample fitting of realized covariance matrices. Larger improvements are at longer horizons, with the HARCJ model achieving the best in-sample performance.

7.6.2 Out-of-Sample Forecasts

Out-of-sample forecast results using directional common jumps are reported in Table 7.7. The losses of the HARJ and HARCJ models (indicated by bold numbers) are smaller than those of the HAR models. For the CPA test p-values (Giacomini and White, 2006), starred values indicate (at the 5% significance level) significantly smaller losses than those of the HAR model. This is wholly consistent with our analysis. The biggest out-of-sample gains are attined by the HARCJ models, which outperform both the HAR and HARJ models irrespective of the forecast horizons and loss function under analysis. However, the CPA test indicates that only the losses of the HARCJ model, at the one-day ($h = 1$) and one month ($h = 22$) horizons, are significantly smaller than those of the HAR models. Whereas at the one-week ($h = 5$) horizon the losses of the HAR model are significantly smaller (10% level) to those of the HAR model, the HARJ model fails to outperform the HAR model across all horizons. This suggests that, although directional common jumps improve the predictability of the realized covariance matrices, their contribution is more limited than that of common jumps.

7.7 Minimum Variance Portfolios

This section assesses the economic value of the different models by constructing Global Minimum Variance (GMV) portfolios. The GMV approach relies solely upon returns covariances, which makes it a 'clean' framework for evaluating the merits of the different covariance forecasts. This is because the estimation errors in sample means are large

and the corresponding portfolios perform poorly compared to the GMV portfolio (e.g. DeMiguel et al., 2009; Jagannathan and Ma, 2003). We employ daily, weekly and monthly rebalancing frequencies, and in each period the investor solves the following minimization problem:

$$\begin{aligned} w_t^* = \underset{w_t}{\text{argmin}} \quad & w_t' \widehat{S}_t w_t, \\ \text{s.t} \quad & w_t' \iota = 1, \end{aligned} \tag{7.24}$$

where w_t is an $N \times 1$ vector of GMV portfolio weights, ι is an $N \times 1$ vector of ones, and $\widehat{S}_t$ is the $N \times N$ matrix of forecasted covariances from a particular model. The optimal portfolio weights, w_t^*, are given by:

$$w_t^* = \frac{\widehat{S}_t^{-1}\iota}{\iota' \widehat{S}_t^{-1}\iota}. \tag{7.25}$$

It is well-known that inaccurate estimates of the covariance matrix lead to worse portfolio performance, with higher turnover and trading costs (e.g. DeMiguel et al., 2014; Han, 2006). Thereby, we incorporate these features in our analysis and define the total portfolio turnover from day t to day $t+1$ as:

$$TO_t = \sum_{i=1}^{N} \left| w_{t+1}^{*(i)} - w_t^{*(i)} \frac{1 + r_t^{(i)}}{1 + w_t^{*'} r_t} \right|. \tag{7.26}$$

The portfolio excess return net of transaction cost is therefore:

$$r_{p_t} = w_t^{*'} r_t - cTO_t, \tag{7.27}$$

where c is the transaction cost.

We follow Fleming et al. (2001, 2003), and evaluate the economic significance of the different strategies using a utility-based framework, in which the investor has quadratic utility with risk aversion γ. The realized daily utility generated by the portfolio based on the covariance forecasts from model k is:

$$U\left(r_{p_t}^{(k)}, \gamma\right) = \left(1 + r_{p_t}^{(k)}\right) - \frac{\gamma}{2(\gamma+1)} \left(1 + r_{p_t}^{(k)}\right)^2. \tag{7.28}$$

The economic value of the different models can be determined by solving:

$$\sum_{t=1}^{T} U\left(r_{p_t}^{(k)}, \gamma\right) = \sum_{t=1}^{T} U\left(r_{p_t}^{(q)} - \Delta_{\gamma}, \gamma\right), \tag{7.29}$$

where Δ_{γ} can be interpreted as the return an investor with risk-aversion γ is willing to sacrifice to switch from using model k to using model q.

The GMV strategies, using a risk-aversion $\gamma = 6$ and a transaction cost $c = 0.5\%$, for the different models based on daily, weekly and monthly rebalancing are reported in Table 7.8. We use a dagger (†) to differentiate the HARJ and HARCJ models that utilize the directional common jumps, and the starred values indicate that Δ_6 is significantly different from zero.[24]

The results in Table 7.8 show that separating common jumps from the continuous component not only increases the accuracy of the forecasted covariances, but leads to substantial gains in portfolio performance. This is because the increased accuracy of the forecasted covariances leads to more stable portfolio strategies, thereby reducing the trading costs. Although the HARJ and HARCJ strategies, using both common and directional common jumps, improve on the performance of the HAR model, the portfolio gains of the HARCJ strategy are significantly different from zero across all horizons. In line with our previous results, common jumps provide superior performance than directional common jumps.

7.8 Conclusion

We propose a robust non-parametric framework that builds on the recent theoretical developments of Barndorff-Nielsen and Shephard (2004a), Mancini and Gobbi (2012), and Caporin et al. (2017), and provide an easy-to-implement approach for measuring separately the multivariate continuous and common jumps components of quadratic co-variation matrices. We further investigate the incidence of directional common jumps,

[24] To evaluate the performance of our strategies, we follow Bandi et al. (2008) and Engle (2002a) and create a null hypothesis that examines whether the performance is equal to zero, i.e. $H_0 : \Delta_6 = 0$ and $H_1 : \Delta_6 > 0$. Therefore, we apply a one-sided t-test with a robust variance-covariance estimator.

which are estimated as the difference between the positive and negative semicovariances. The sign of the directional common jumps depends upon the direction of the simultaneous jumps, as the latter are fully absorbed by the positive or negative semicovariances (e.g. Bollerslev et al., 2020).

Applying the theory to 20 individual stocks for a period of 17 years, we find that common jumps are strongly associated with major financial and economic news. As common jumps are highly correlated, their occurrence generally increases both the dependence across stocks and the persistence of stock (co)variances.

The inclusion of the continuous and common jumps in a vectorized heterogeneous autoregressive (vech-HAR) model results in significant in- and out-of-sample forecasting gains, which are attained at the daily, weekly and monthly horizons. When the continuous and the common jumps variables are entered separately in the vech-HAR model, common jumps estimates are generally positive, and lead to an increase in future covariances. On the other hand, estimates of directional common jumps are usually negative, which means that negative (positive) common jumps increase (decrease) the persistence and future level of (co)variances.

Finally, by using a utility-based approach to assess the economic value of the forecasted covariance matrices, we find that the improved accuracy of the HARJ and HARCJ models yields cheaper portfolio allocations, as the more stable portfolio strategies lead to lower trading costs.

Appendix 7.A Tables and Figures

Table 7.1: Simulated In- and Out-of-sample Forecast Results

	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ
	$h = 1$			$h = 5$			$h = 22$		
	Panel A: In-Sample Forecasts (Common Jumps)								
R^2_{adj}	0.336	0.350	0.365	0.622	0.646	0.674	0.695	0.718	0.746
L^F	2.813	2.770	2.677	1.968	1.921	1.810	1.649	1.599	1.498
L^E	2.376	2.302	2.274	1.666	1.621	1.543	1.442	1.397	1.330
	Panel B: Out-of-Sample Forecasts (Common Jumps)								
L^F	2.827	2.767	2.758	1.977	1.908	1.892	1.657	1.595	1.554
L^E	2.387	2.340	2.316	1.672	1.616	1.577	1.448	1.401	1.376
	Panel C: In-Sample Forecasts (Directional Common Jumps)								
R^2_{adj}	0.336	0.337	0.365	0.622	0.623	0.674	0.695	0.695	0.746
L^F	2.813	2.793	2.704	1.968	1.965	1.829	1.649	1.618	1.513
L^E	2.376	2.346	2.297	1.666	1.650	1.559	1.442	1.412	1.344
	Panel D: Out-of-Sample Forecasts (Directional Common Jumps)								
L^F	2.827	2.802	2.786	1.977	1.979	1.910	1.657	1.659	1.583
L^E	2.387	2.365	2.339	1.672	1.673	1.591	1.448	1.449	1.385

Note: The table reports the in- and out-of-sample results for the different models based on daily ($h = 1$), weekly ($h = 5$) and monthly ($h = 22$) horizons. L^F and L^E denote the Frobenius and Euclidean loss function, respectively. The results are generated using the Monte Carlo simulation outlined in Section 6.3.2.

Table 7.2: Descriptive Statistics

Name	Ticker	Variance		Common Jumps			Common Jumps	Hausman Test	
		Mean	St. Dev.	Mean	St. Dev	Identified	to Total Variance	$H_{3,n}$	AC_n
American Express	AXP	4.020	9.195	2.619	5.495	149	2.178	0.022	0.015
Boeing	BA	2.812	3.900	1.592	2.961	149	1.879	0.029	0.024
Catterpillar	CAT	3.239	4.889	2.081	3.512	149	2.133	0.022	0.016
Disney	DIS	2.966	5.015	1.973	4.462	149	2.177	0.030	0.023
Dow	DOW	3.976	7.353	2.601	6.061	149	2.080	0.023	0.019
Dupont	DD	2.761	4.076	1.554	2.814	149	1.843	0.024	0.016
Home Depot	HD	3.121	4.938	2.101	5.592	149	2.188	0.022	0.015
IBM	IBM	2.026	3.527	1.454	4.845	149	2.434	0.019	0.014
Intel	INTC	4.075	5.754	1.867	4.564	149	1.511	0.018	0.015
Johnson & Johnson	JNJ	1.385	3.482	2.047	14.222	149	5.010	0.031	0.019
JP Morgan Chase	JPM	4.615	10.848	4.231	11.760	149	2.937	0.024	0.014
Coca-Cola	KO	1.561	2.535	0.781	1.389	149	1.685	0.027	0.017
MacDonald	MCD	2.164	4.333	0.943	1.850	149	1.417	0.036	0.024
Merck and Co.	MRK	2.431	5.227	1.570	4.004	149	2.219	0.028	0.015
3M	MMM	1.853	3.216	1.297	3.647	149	2.307	0.026	0.017
Microsoft	MSFT	2.679	3.854	1.279	2.014	149	1.630	0.021	0.015
Procter & Gamble	PG	1.492	2.907	1.106	3.332	149	2.445	0.027	0.017
United Technologies	UTX	2.300	3.793	1.497	3.009	149	2.191	0.025	0.017
WalMart	WMT	2.045	3.277	0.992	2.001	149	1.609	0.026	0.018
Exxon Mobil	XOM	1.987	3.955	1.424	3.720	149	2.379	0.020	0.014

Note: The table reports the descriptive statistics of the variance and common jumps together with the contribution of the common jumps to total variance. The latter is estimated as the ratio of sum of the common jumps (of stock i) to the sum of the total variance (of stock i). The descriptive statistics of the common jumps only considers the days with significant common jumps, depicted in the third column. The last two columns report respectively the proportion of rejections of the Hausman test for the presence of market microstructure noise and first-order autocorrelation in log-returns test proposed by Aït-Sahalia and Xiu (2019).

Table 7.3: The Correlation of the Common Jumps and Continuous Components

	AXP	BA	CAT	DIS	HD	IBM	INTC	JNJ	JPM	KO	MCD	MRK	MMM	MSFT	PG	UTX	WMT	XOM	DD	DOW
AXP	–	0.699	0.822	0.693	0.792	0.752	0.633	0.662	0.802	0.686	0.555	0.539	0.796	0.681	0.610	0.715	0.653	0.802	0.791	0.554
BA	0.554	–	0.754	0.745	0.795	0.742	0.697	0.743	0.639	0.727	0.540	0.491	0.776	0.737	0.616	0.798	0.689	0.727	0.761	0.492
CAT	0.722	0.854	–	0.686	0.773	0.774	0.632	0.692	0.717	0.758	0.528	0.561	0.820	0.721	0.671	0.777	0.682	0.808	0.846	0.593
DIS	0.603	0.820	0.845	–	0.755	0.715	0.696	0.704	0.629	0.679	0.556	0.467	0.740	0.705	0.586	0.751	0.666	0.702	0.695	0.489
HD	0.678	0.517	0.611	0.541	–	0.809	0.740	0.757	0.716	0.774	0.584	0.537	0.820	0.769	0.662	0.796	0.776	0.785	0.800	0.505
IBM	0.833	0.531	0.630	0.538	0.570	–	0.821	0.728	0.688	0.766	0.566	0.506	0.798	0.829	0.697	0.755	0.806	0.755	0.779	0.504
INTC	0.496	0.877	0.773	0.810	0.431	0.573	–	0.679	0.604	0.705	0.532	0.433	0.692	0.848	0.620	0.677	0.776	0.600	0.679	0.474
JNJ	0.299	0.253	0.280	0.280	0.858	0.193	0.192	–	0.635	0.769	0.583	0.519	0.751	0.724	0.665	0.746	0.728	0.753	0.734	0.485
JPM	0.802	0.635	0.722	0.556	0.658	0.765	0.427	0.276	–	0.629	0.501	0.500	0.674	0.652	0.530	0.622	0.594	0.680	0.692	0.553
KO	0.760	0.667	0.842	0.682	0.739	0.580	0.469	0.443	0.821	–	0.583	0.559	0.775	0.737	0.717	0.772	0.775	0.756	0.770	0.476
MCD	0.700	0.558	0.756	0.676	0.847	0.544	0.486	0.669	0.626	0.819	–	0.384	0.598	0.542	0.519	0.572	0.590	0.602	0.560	0.331
MRK	0.498	0.850	0.831	0.827	0.668	0.430	0.824	0.521	0.506	0.665	0.711	–	0.541	0.482	0.458	0.498	0.480	0.560	0.555	0.396
MMM	0.426	0.866	0.817	0.814	0.411	0.411	0.865	0.233	0.420	0.537	0.526	0.898	–	0.762	0.717	0.837	0.749	0.854	0.833	0.528
MSFT	0.677	0.676	0.667	0.592	0.819	0.623	0.631	0.662	0.606	0.689	0.718	0.674	0.544	–	0.663	0.733	0.762	0.699	0.753	0.496
PG	0.631	0.667	0.705	0.636	0.870	0.570	0.594	0.767	0.612	0.708	0.781	0.822	0.701	0.820	–	0.668	0.706	0.672	0.688	0.425
UTX	0.639	0.833	0.863	0.779	0.624	0.589	0.691	0.276	0.800	0.810	0.686	0.771	0.686	0.605	0.617	–	0.741	0.773	0.787	0.504
WMT	0.761	0.567	0.775	0.607	0.737	0.712	0.542	0.462	0.677	0.782	0.857	0.630	0.556	0.704	0.739	0.671	–	0.697	0.748	0.443
XOM	0.527	0.860	0.861	0.857	0.499	0.476	0.838	0.310	0.496	0.663	0.653	0.917	0.939	0.596	0.739	0.709	0.615	–	0.802	0.481
DD	0.613	0.910	0.883	0.805	0.523	0.524	0.780	0.219	0.707	0.747	0.607	0.804	0.792	0.644	0.622	0.876	0.554	0.818	–	0.570
DOW	0.542	0.790	0.743	0.713	0.722	0.479	0.779	0.607	0.489	0.583	0.645	0.853	0.757	0.810	0.812	0.633	0.552	0.760	0.750	–

Note: The table reports in the lower (upper) triangular matrix the correlation of the common jumps (continuous component) across the 20 individual stocks using the full sample size.

Table 7.4: In-Sample Estimates using Common Jumps

	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ
		$h=1$			$h=5$			$h=22$	
θ_0	0.087	0.084	0.092	0.133	0.130	0.139	0.241	0.240	0.242
s.e.	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
θ_d/θ_{C_d}	0.255	0.294	0.268	0.170	0.207	0.195	0.097	0.102	0.111
s.e.	(0.018)	(0.023)	(0.023)	(0.006)	(0.007)	(0.007)	(0.004)	(0.005)	(0.005)
θ_w/θ_{C_w}	0.377	0.357	0.431	0.326	0.308	0.341	0.267	0.264	0.238
s.e.	(0.023)	(0.024)	(0.027)	(0.016)	(0.016)	(0.020)	(0.018)	(0.018)	(0.020)
θ_m/θ_{C_m}	0.281	0.275	0.204	0.371	0.365	0.310	0.396	0.395	0.402
s.e.	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.016)	(0.018)	(0.018)	(0.020)
θ_{J_d}		−0.285	0.119		−0.267	−0.018		−0.033	0.022
s.e.		(0.030)	(0.016)		(0.015)	(0.016)		(0.021)	(0.007)
θ_{J_w}			−0.487			−0.170			0.490
s.e.			(0.055)			(0.062)			(0.077)
θ_{J_m}			1.443			1.450			0.436
s.e.			(0.148)			(0.182)			(0.113)
R^2_{adj}	0.517	0.520	0.525	0.643	0.647	0.651	0.609	0.609	0.610
L^F	14.998	14.978	14.861	11.509	11.491	11.429	11.584	11.580	11.553
L^E	11.867	11.852	11.761	9.188	9.174	9.127	9.190	9.186	9.165

Note: The table reports the in-sample parameter estimates with robust standard errors in parentheses, along with measures of fit for the different models estimated using forecast horizons equal to one-day ($h = 1$), one-week ($h = 5$) and one-month ($h = 22$). L^F and L^E denote the respective Frobenius and Euclidean loss function. The standard errors are computed using methods that are robust to model misspecification.

Table 7.5: Out-of-sample Forecast Results using Common Jumps

	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ
		$h = 1$			$h = 5$			$h = 22$	
L^F	13.481	**13.472**	**12.782**	10.720	**10.703**	**10.427**	11.705	**11.443**	**11.354**
CPA	–	0.237	0.001*	–	0.746	0.021*	–	0.041*	0.035*
L^E	10.548	**10.543**	**10.011**	8.428	**8.416**	**8.205**	9.044	**8.841**	**8.783**
CPA	–	0.252	0.001*	–	0.773	0.014*	–	0.042*	0.030*

Note: The table reports the out-of-sample forecast loss for the different models, along with the p-value of the CPA test of Giacomini and White (2006) based on a 5% significance level. Bold numbers indicate that the forecast losses are smaller than that of the HAR model, while starred numbers highlight the forecasts whose losses are significant smaller relative to the HAR model.

Table 7.6: In-Sample Estimates using Directional Common Jumps

	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ
		$h=1$			$h=5$			$h=22$	
θ_0	0.087	0.087	0.068	0.133	0.132	0.106	0.241	0.238	0.216
s.e.	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.007)	(0.007)	(0.008)
θ_d/θ_{C_d}	0.255	0.255	0.272	0.170	0.171	0.189	0.097	0.099	0.101
s.e.	(0.018)	(0.019)	(0.023)	(0.008)	(0.008)	(0.010)	(0.007)	(0.007)	(0.008)
θ_w/θ_{C_w}	0.377	0.377	0.423	0.326	0.327	0.370	0.267	0.268	0.285
s.e.	(0.023)	(0.023)	(0.028)	(0.017)	(0.017)	(0.023)	(0.018)	(0.018)	(0.022)
θ_m/θ_{C_m}	0.281	0.281	0.278	0.371	0.371	0.377	0.396	0.395	0.434
s.e.	(0.014)	(0.014)	(0.016)	(0.016)	(0.016)	(0.019)	(0.016)	(0.016)	(0.019)
θ_{J_d}		−0.026	0.144		−0.071	0.242		−0.201	0.058
s.e.		(0.026)	(0.019)		(0.022)	(0.022)		(0.023)	(0.019)
θ_{J_w}			−0.257			−1.121			−0.950
s.e.			(0.071)			(0.103)			(0.107)
θ_{J_m}			−1.477			−1.437			−1.198
s.e.			(0.182)			(0.203)			(0.174)
R^2_{adj}	0.517	0.517	0.523	0.643	0.643	0.655	0.609	0.610	0.612
L^F	14.998	14.998	14.707	11.509	11.506	11.132	11.584	11.556	11.030
L^E	11.867	11.867	11.662	9.188	9.186	8.916	9.190	9.168	8.774

Note: The table reports the in-sample parameter estimates with robust standard errors in parentheses, along with measures of fit for the different models estimated using forecast horizons equal to one-day ($h = 1$), one-week ($h = 5$) and one-month ($h = 22$). L^F and L^E denote the respective Frobenius and Euclidean loss function. The standard errors are computed using methods that are robust to model misspecification. The jump variable used to generate these models is the directional common jumps estimated as in equation (7.23). To ease the notation we prefer not to change the name given to the specifications.

Table 7.7: Out-of-sample Forecast Results using Directional Common Jumps

	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ	HAR	HARJ	HARCJ
		$h = 1$			$h = 5$			$h = 22$	
L^F	13.481	**13.464**	**12.904**	10.720	**10.716**	**10.552**	11.705	**11.649**	**10.355**
CPA	–	0.446	0.011*	–	0.930	0.056	–	0.162	0.037*
L^E	10.548	**10.533**	**10.129**	8.428	**8.424**	**8.328**	9.044	**9.001**	**8.013**
CPA	–	0.422	0.015*	–	0.975	0.097	–	0.156	0.034*

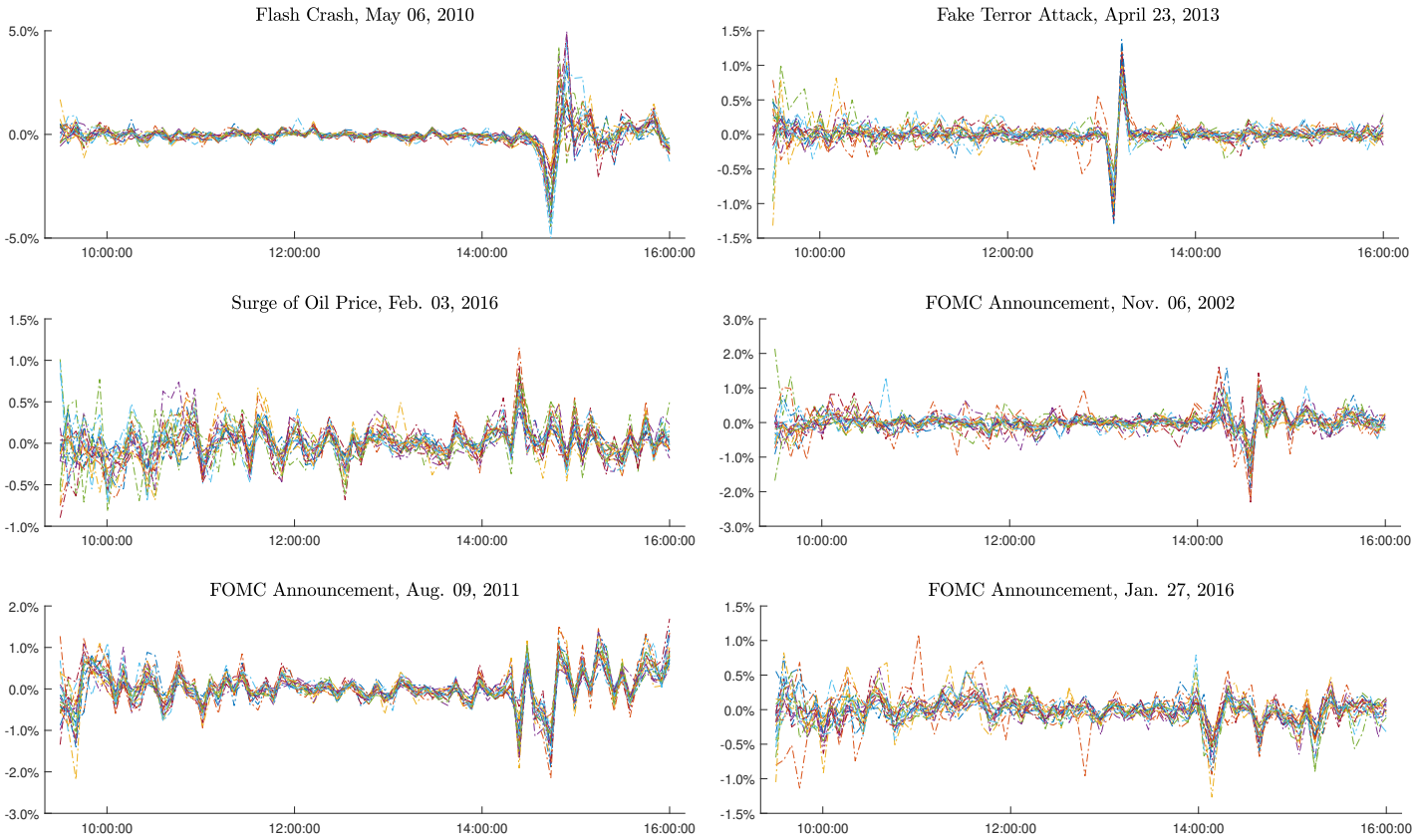
Note: The table reports the out-of-sample forecast loss for the different models, along with the p-value of the CPA test of Giacomini and White (2006) based on a 5% significance level. Bold numbers indicate that the forecast losses are smaller than that of the HAR model, while starred numbers highlight the forecasts whose losses are significant smaller relative to the HAR model.

Table 7.8: Minimum Variance Portfolios

	Daily Rebalancing		Weekly Rebalancing		Monthly Rebalancing	
	Δ_6	TO	Δ_6	TO	Δ_6	TO
HAR		0.733		0.524		0.302
HARJ	39.810*	0.711	8.131	0.518	7.291	0.297
HARCJ	96.709*	0.669	54.092*	0.490	100.256*	0.254
HARJ†	11.978	0.729	4.895	0.521	3.071	0.301
HARCJ†	74.038*	0.700	23.711*	0.514	50.681*	0.267

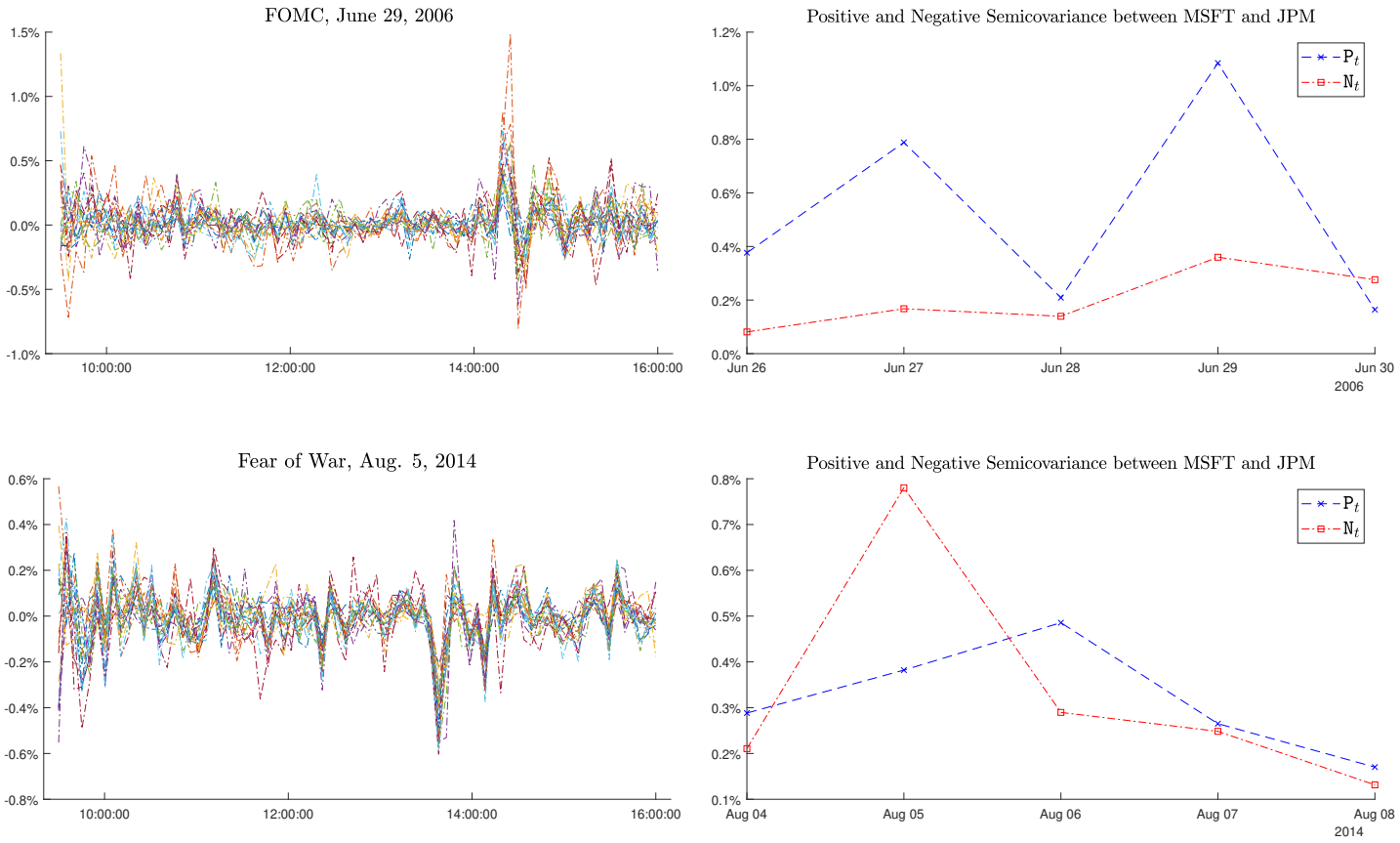
Note: The table reports the results for the global minimum variance (GMV) portfolio using daily, weekly and monthly rebalancing. The results show the average turnover (TO) and the annual basis points (Δ_6) that an investor is willing to sacrifice to switch from the HAR model to one of the models that utilize the common jumps or directional common jumps. The dagger (†) denote the use of directional common jumps, and starred values indicate that Δ_6 is significantly different from zero using a one-sided test. The analysis uses a risk-aversion $\gamma = 6$ and transaction cost $c = 0.5\%$.

Figure 7.1: Intraday Returns on Days with Simultaneous Jumps



Note: The figure depicts the 5 minutes intraday returns of the 20 individual stocks on 6 different days where simultaneous jumps where identified in our sample. The dates correspond to well known Flash Crashes, FOMC meetings and the surge of the WTI oil price which increases by 8% in a single day.

Figure 7.2: Semicovariance and Directional Common Jumps



Note: The figure depicts the 5 minutes intraday returns of the 20 individual stocks and the positive and negative realized semicovariances between MSFT and JPM for two different days. The top (bottom) panel shows results based on June 29, 2006 (August 5, 2014).

References

Aiolfi, M., Capistrán, C., and Timmermann, A. (2011). Forecasts combinations. In Clements, M. P. and Hendry, D. F., editors, *The Oxford Handbook of Economic Forecasting*. OUP.

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1-2):31–53.

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics*, 74(3):487–528.

Aït-Sahalia, Y. and Jacod, J. (2009). Estimating the degree of activity of jumps in high frequency data. *The Annals of Statistics*, 37(5A):2202–2244.

Aït-Sahalia, Y. and Jacod, J. (2010). Is brownian motion necessary to model high-frequency data? *The Annals of Statistics*, 38(5):3093–3128.

Aït-Sahalia, Y. and Jacod, J. (2011). Testing whether jumps have finite or infinite activity. *The Annals of Statistics*, 39(3):1689–1719.

Aït-Sahalia, Y. and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature*, 50(4):1007–50.

Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.

Aït-Sahalia, Y., Jacod, J., et al. (2009). Testing for jumps in a discretely observed process. *The Annals of Statistics*, 37(1):184–222.

Aït-Sahalia, Y., Jacod, J., and Li, J. (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics*, 168(2):207–222.

Aït-Sahalia, Y., Kalnina, I., and Xiu, D. (2020). High-frequency factor models and regressions. *Journal of Econometrics*, 216(1):86–105.

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. *Review of Financial Studies*, 18(2):351–416.

Aït-Sahalia, Y. and Xiu, D. (2016). Increased correlation among asset classes: Are volatility or jumps to blame, or both? *Journal of Econometrics*, 194(2):205–219.

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Aït-Sahalia, Y. and Xiu, D. (2019). A hausman test for the presence of market microstructure noise in high frequency data. *Journal of Econometrics*, 211(1):176–205.

Amengual, D. and Xiu, D. (2018). Resolution of policy uncertainty and sudden declines in volatility. *Journal of Econometrics*, 203(2):297–315.

Andersen, T. G., Benzoni, L., and Lund, J. (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance*, 57(3):1239–1284.

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*, pages 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. (2001a). 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., and 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(4):701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.

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 iid noise: Theory and testable distributional implications. *Journal of Econometrics*, 138(1):125–180.

Andersen, T. G., Bollerslev, T., and Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance*, 6(5):457–477.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73(1):279–296.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160(1):220–234.

Andersen, T. G., Dobrev, D., and Schaumburg, E. (2010). Integrated quarticity estimation: Theory and practical implementation. *Working paper, Northwestern University, Federal Reserve Board of Governors and Federal Reserve Bank of New York.*

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.

Andreou, E. and Ghysels, E. (2002). Rolling-sample volatility estimators: some new theoretical, simulation and empirical results. *Journal of Business & Economic Statistics*, 20(3):363–376.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, 91(1):1–23.

Asai, M., Gupta, R., and McAleer, M. (2020). Forecasting volatility and co-volatility of crude oil and gold futures: Effects of leverage, jumps, spillovers, and geopolitical risks. *International Journal of Forecasting*.

Bachelier, L. (1900). Théorie de la spéculation. In *Annales Scientifiques de l'École Normale Supérieure*, volume 17, pages 21–86.

Bai, J. and Shi, S. (2011). Estimating high dimensional covariance matrices and its applications. *Annals of Economics and Finance*, 12:199–215.

Bandi, F. M. and Russell, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3):655–692.

Bandi, F. M. and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, 75(2):339–369.

Bandi, F. M., Russell, J. R., and Zhu, Y. (2008). Using high-frequency data in dynamic portfolio choice. *Econometric Reviews*, 27(1-3):163–198.

Barndorff-Nielsen, O. E. (1997). Processes of normal inverse gaussian type. *Finance and Stochastics*, 2(1):41–68.

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., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, pages 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169.

Barndorff-Nielsen, O. E., Kinnebrock, S., and Shephard, N. (2010). Measuring downside risk– realised semivariance. In Bollerslev, T., Russell, J., and Watson, M., editors, *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, pages 117–136. Oxford University Press, Oxford.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241.

Barndorff-Nielsen, O. E. and Shephard, N. (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477.

Barndorff-Nielsen, O. E. and Shephard, N. (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

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

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

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.

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93–101.

Bibinger, M. and Winkelmann, L. (2015). Econometrics of co-jumps in high-frequency data with noise. *Journal of Econometrics*, 184(2):361–378.

Blair, B., Poon, S., and Taylor, S. (2000). Modelling s&p 100 volatility: The information content of stock returns, forthcoming. *Journal of Banking & Finance*, 25(9):1665–1679.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics*, pages 498–505.

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., Law, T. H., and Tauchen, G. (2008). Risk, jumps, and diversification. *Journal of Econometrics*, 144(1):234–256.

Bollerslev, T., Li, J., Patton, A. J., and Quaedvlieg, R. (2020). Realized semicovariances. *Econometrica*, 88(4):1515–1551.

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.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2018). Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions. *Journal of Econometrics*, 207(1):71–91.

Bollerslev, T. and Todorov, V. (2011a). Estimation of jump tails. *Econometrica*, 79(6):1727–1783.

Bollerslev, T. and Todorov, V. (2011b). Tails, fears, and risk premia. *The Journal of Finance*, 66(6):2165–2211.

Bollerslev, T., Todorov, V., and Xu, L. (2015). Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1):113–134.

Boudt, K., Cornelissen, J., Croux, C., and Laurent, S. (2012). *Nonparametric Tests for Intraday Jumps: Impact of Periodicity and Microstructure Noise*, chapter Eighteen, pages 447–463. John Wiley & Sons, Ltd.

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.

Brockwell, P. J. (2001). Continuous time ARMA model. In Shanbhag, D. N. and Rao, C. R., editors, *Stochastic Processes: Theory and Methods*, volume 19, pages 249–276. North-Holland, Amsterdam.

Buccheri, G. and Corsi, F. (2019). Hark the shark: Realized volatility modeling with measurement errors and nonlinear dependencies. *Journal of Financial Econometrics*.

Busch, T., Christensen, B. J., and Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160(1):48–57.

Callot, L. A., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158.

Calzolari, G., Halbleib, R., and Zagidullina, A. (2020). A latent factor model for forecasting realized variances. *Journal of Financial Econometrics (in press)*, pages 1–50.

Campbell, J. Y., Lettau, M., Malkiel, B. G., and Xu, Y. (2001). Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. *The Journal of Finance*, 56(1):1–43.

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.

Canina, L. and Figlewski, S. (1993). The informational content of implied volatility. *The Review of Financial Studies*, 6(3):659–681.

Caporin, M., Kolokolov, A., and Renò, R. (2017). Systemic co-jumps. *Journal of Financial Economics*, 126(3):563–591.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *The Journal of Business*, 75(2):305–332.

Carr, P. and Wu, L. (2007). Stochastic skew in currency options. *Journal of Financial Economics*, 86(1):213–247.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

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., Hounyo, U., and Podolskij, M. (2018). Is the diurnal pattern sufficient to explain intraday variation in volatility? a nonparametric assessment. *Journal of Econometrics*, 205(2):336–362.

Christensen, K., Oomen, R. C., and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599.

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

Christiansen, C., Schmeling, M., and Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics*, 27(6):956–977.

Cipollini, F., Gallo, G. M., and Otranto, E. (2020). Realized volatility forecasting: Robustness to measurement errors. *International Journal of Forecasting*, 37(1):44–57.

Conley, T. G., Hansen, L. P., Luttmer, E. G., and Scheinkman, J. A. (1997). Short-term interest rates as subordinated diffusions. *The Review of Financial Studies*, 10(3):525–577.

Cont, R. and Mancini, C. (2011). Nonparametric tests for pathwise properties of semimartingales. *Bernoulli*, 17(2):781–813.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall, Boca Raton, FL.

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

Corsi, F. and 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(3):368–380.

Daal, E. A. and Madan, D. B. (2005). An empirical examination of the variance-gamma model for foreign currency options. *The Journal of Business*, 78(6):2121–2152.

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*.

Das, S. R. and Uppal, R. (2004). Systemic risk and international portfolio choice. *The Journal of Finance*, 59(6):2809–2834.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.

DeMiguel, V., Nogales, F. J., and Uppal, R. (2014). Stock return serial dependence and out-of-sample portfolio performance. *The Review of Financial Studies*, 27(4):1031–1073.

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):253–63.

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.

Dungey, M. and Hvozdyk, L. (2012). Cojumping: Evidence from the us treasury bond and futures markets. *Journal of Banking & Finance*, 36(5):1563–1575.

Duong, D. and Swanson, N. R. (2015). Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics*, 187(2):606–621.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299.

Elliott, G. and Timmermann, A. (2016). *Economic Forecasting*. Princeton University Press.

Engle, R. (2002a). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. (2002b). New frontiers for arch models. *Journal of Applied Econometrics*, 17(5):425–446.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, pages 987–1007.

Engle, R. F., Ghysels, E., and Sohn, B. (2007). On the economic sources of stock market volatility. *Unpublished Working Paper*.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, pages 122–150.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a):291–298.

Erdemlioglu, D., Laurent, S., and Neely, C. J. (2015). Which continuous-time model is most appropriate for exchange rates? *Journal of Banking & Finance*, 61:S256–S268.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.

Fama, E. F. and French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, 51(1):55–84.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197.

Fan, J., Furger, A., and Xiu, D. (2016). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business & Economic Statistics*, 34(4):489–503.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, 67(3):473–509.

Forsberg, L. and Ghysels, E. (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics*, 5(1):31–67.

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

Geman, H. (2002). Pure jump lévy processes for asset price modelling. *Journal of Banking & Finance*, 26(7):1297–1316.

Ghysels, E. and Sinko, A. (2011). Volatility forecasting and microstructure noise. *Journal of Econometrics*, 160(1):257–271.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Gilder, D., Shackleton, M. B., and Taylor, S. J. (2014). Cojumps in stock prices: Empirical evidence. *Journal of Banking & Finance*, 40:443–459.

Giot, P. and Laurent, S. (2007). The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 27(4):337–359.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Griffin, J. E. and Oomen, R. C. (2008). Sampling returns for realized variance calculations: tick time or transaction time? *Econometric Reviews*, 27(1-3):230–253.

Han, Y. (2006). Asset allocation with a high dimensional latent factor stochastic volatility model. *The Review of Financial Studies*, 19(1):237–271.

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

Hansen, P. R. and Lunde, A. (2014). Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory*, 30(1):60–93.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

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

Hasbrouck, J. (2003). Intraday price formation in us equity index markets. *The Journal of Finance*, 58(6):2375–2400.

Hautsch, N., Kyj, L. M., and Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30(2):263–290.

Huang, J.-z. and Wu, L. (2004). Specification analysis of option pricing models based on time-changed lévy processes. *The Journal of Finance*, 59(3):1405–1439.

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

Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and their Applications*, 118(4):517–559.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications*, 119(7):2249–2276.

Jacod, J. and Todorov, V. (2009). Testing for common arrivals of jumps for discretely observed multidimensional processes. *The Annals of Statistics*, 37(4):1792–1838.

Jacod, J. and Todorov, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *The Annals of Statistics*, 42(3):1029–1069.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

Jiang, G. J. and Oomen, R. C. (2008). Testing for jumps when asset prices are observed with noise–a "swap variance" approach. *Journal of Econometrics*, 144(2):352–370.

Jiang, G. J. and Yao, T. (2013). Stock price jumps and cross-sectional return predictability. *Journal of Financial and Quantitative Analysis*, 48(5):1519–1544.

Jing, B.-Y., Kong, X.-B., and Liu, Z. (2012a). Modeling high-frequency financial data by pure jump processes. *The Annals of Statistics*, 40(2):759–784.

Jing, B.-Y., Kong, X.-B., Liu, Z., and Mykland, P. (2012b). On the jump activity index for semimartingales. *Journal of Econometrics*, 166(2):213–223.

Jorion, P. (1988). On jump processes in the foreign exchange and stock markets. *The Review of Financial Studies*, 1(4):427–445.

Jorion, P. (1995). Predicting volatility in the foreign exchange market. *The Journal of Finance*, 50(2):507–528.

King, B. F. (1966). Market and industry factors in stock price behavior. *The Journal of Business*, 39(1):139–190.

Klüppelberg, C., Lindner, A., and Maller, R. (2004). A continuous-time garch process driven by a lévy process: stationarity and second-order behaviour. *Journal of Applied Probability*, 41(3):601–622.

Kong, X.-B. (2019). Lack of fit test for infinite variation jumps at high frequencies. *Statistica Sinica*, 29:81–95.

Kong, X.-B., Liu, Z., and Jing, B.-Y. (2015). Testing for pure-jump processes for high-frequency data. *The Annals of Statistics*, 43(2):847–877.

Lahaye, J., Laurent, S., and Neely, C. J. (2011). Jumps, cojumps and macro announcements. *Journal of Applied Econometrics*, 26(6):893–921.

Laurent, S., Rombouts, J. V., and Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*, 173(1):1–10.

Lee, S. S. and Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies*, 21(6):2535–2563.

Lee, S. S. and Mykland, P. A. (2012). Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics*, 168(2):396–406.

Li, J. and Xiu, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica*, 84(4):1613–1633.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, 20(4):587–615.

Liu, L. Y., Patton, A. J., and Sheppard, K. (2015). Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293–311.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676.

Madan, D. B. and Seneta, E. (1990). The variance gamma (vg) model for share market returns. *Journal of Business*, pages 511–524.

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping brownian motion. *Giornale dell'Istituto Italiano degli Attuari*, 64(19-47):44.

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(2):270–296.

Mancini, C. and Gobbi, F. (2012). Identifying the brownian covariation from the co-jumps given discrete observations. *Econometric Theory*, pages 249–273.

Maneesoonthorn, W., Martin, G. M., and Forbes, C. S. (2020). High-frequency jump tests: Which test should we use? *Journal of Econometrics*.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

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.

Martens, M., Van Dijk, D., and De Pooter, M. (2009). Forecasting s&p 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting*, 25(2):282–303.

Meddahi, N. (2003). Arma representation of integrated and realized variances. *The Econometrics Journal*, 6(2):335–356.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144.

Mukherjee, A., Peng, W., Swanson, N. R., and Yang, X. (2020). Financial econometrics and big data: A survey of volatility estimators and tests for the presence jumps and co-jumps. In Vinod, H. D. and Rao, C., editors, *Financial, Macro and Micro Econometrics Using R*, volume 42, pages 3–59. North Holland.

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.

Nielsen, H. B. and Rahbek, A. (2014). Unit root vector autoregression with volatility induced stationarity. *Journal of Empirical Finance*, 29:144–167.

Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 27(6):907–933.

Oomen, R. C. A. (2006a). Comment on: "realized variance and market microstructure noise" by Peter R. Hansen and Asger Lunde. *Journal of Business & Economic Statistics*, 24(2):195–202.

Oomen, R. C. A. (2006b). Properties of realized variance under alternative sampling schemes. *Journal of Business & Economic Statistics*, 24(2):219–237.

Patton, A. J. (2011a). Data-based ranking of realised volatility estimators. *Journal of Econometrics*, 161(2):284–303.

Patton, A. J. (2011b). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Patton, A. J. and Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25(2):218–238.

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697.

Paye, B. S. (2012). 'déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106(3):527–546.

Podolskij, M. and Vetter, M. (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and their Applications*, 119(9):2803–2831.

Podolskij, M. and Ziggel, D. (2010). New tests for jumps in semimartingale models. *Statistical Inference for Stochastic Processes*, 13(1):15–41.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Prokopczuk, M., Symeonidis, L., and Wese Simen, C. (2016). Do jumps matter for volatility forecasting? evidence from energy markets. *Journal of Futures Markets*, 36(8):758–792.

Reiß, M., Todorov, V., and Tauchen, G. (2015). Nonparametric test for a constant beta between itô semi-martingales based on high-frequency data. *Stochastic Processes and their Applications*, 125(8):2955–2988.

Rivers, D. and Vuong, Q. (2002). Model selection tests for nonlinear dynamic models. *The Econometrics Journal*, 5(1):1–39.

Rydberg, T. H. (1997). The normal inverse gaussian lévy process: simulation and approximation. *Communications in Statistics. Stochastic Models*, 13(4):887–910.

Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, 235(3):643–659.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Timmermann, A. (2006). Forecast combinations. In Elliot, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, pages 135–196. Elsevier.

Todorov, V. (2015). Jump activity estimation for pure-jump semimartingales via self-normalized statistics. *The Annals of Statistics*, 43(4):1831–1864.

Todorov, V. and Bollerslev, T. (2010). Jumps and betas: A new framework for disentangling and estimating systematic risks. *Journal of Econometrics*, 157(2):220–235.

Todorov, V. and Tauchen, G. (2010). Activity signature functions for high-frequency data analysis. *Journal of Econometrics*, 154(2):125–138.

Todorov, V. and Tauchen, G. (2014). Limit theorems for the empirical distribution function of scaled increments of itô semimartingales at high frequencies. *The Annals of Applied Probability*, 24(5):1850–1888.

Treynor, J. L. (1962). Toward a theory of market value of risky assets. *Unpublished Manuscript*.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

Xu, Y. and Malkiel, B. G. (2003). Investigating the behavior of idiosyncratic volatility. *The Journal of Business*, 76(4):613–645.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411.