

Evidence on the Importance of Volatility Density Forecasting for Financial Risk Management*

Arpita Mukherjee¹ and Norman R. Swanson²

¹Fort L.P. and ²Rutgers University

November 2021

Abstract

In this paper, we provide new empirical evidence of the relative usefulness of interval (density) and point forecasts of asset-return volatility, in the context of financial risk management using high frequency data. In our evaluation we use both statistical criteria (i.e., accuracy of directional volatility predictions) and economic criteria (i.e., profitability of trading strategies based on said predictions). We construct interval forecasts using nonparametric kernel estimators, while point forecasts are based on “linear” heterogeneous autoregressive models as well as “nonlinear” deep-learning recurrent neural network models. For our profitability analysis, we utilize a novel trading strategy that builds on the contemporaneous return-volatility relationship and leads to new insights related to linkages between economic and statistical methods of forecast evaluation. Our empirical findings indicate that interval forecasts can improve on point forecasts in terms of trading profitability (as measured using Sharpe and Sortino Ratios), regardless of the “linear” or “nonlinear” nature of the point-forecasting model. Moreover, linear (nonlinear) model-based point forecasts perform worse (marginally better) than interval forecasts when it comes to directional predictive accuracy. These findings are consistent with a number of hypotheses concerning both nonlinear volatility dynamics and the ability of interval forecasts to accurately estimate “large price jump” induced future volatility movements. A follow-up series of Monte Carlo experiments is motivated by our finding that for translation of statistical improvements into economic gains, the choice of volatility estimation technique is crucial. Our experiments reveal that the inability of certain nonparametric high frequency volatility estimators to accurately predict “pseudo true” volatility densities for specific magnitudes of “price jumps” and “microstructure noise” in the price process can explain why these same estimators are less profitable when used in our empirical trading strategies.

Keywords: Integrated Volatility, Point, Interval, and Density Forecast, Directional Predictive Accuracy, Trading Profitability, Deep Learning, Price and Volatility Jumps, Microstructure Noise.

JEL Classification: C14, C15, C22, C52, C53, C58, G11, G17.

*Arpita Mukherjee, Fort L.P., 280 Park Avenue, New York, NY 10017, USA, arpitmukherjee812@gmail.com; Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@economics.rutgers.edu. The authors are grateful to Valentina Corradi, Walter Distaso, Dobrislev Dobrev, Jeffrey Gerlach, John Landon Lane, Yuan Liao, Michiel Pooter, Xiye Yang, and seminar participants at the Federal Reserve Board, the Federal Reserve Bank of Richmond, the Indian Institute of Technology at Kanpur, Ashoka University, the Indira Gandhi Institute of Development Research, and the Citigroup Economic Forecasting Division for comments that have been utilized in the preparation of this paper. The contents herein solely represent the views of the authors, and do not represent the views of any institutions with which the authors are engaged.

1 Introduction

Successful risk management, asset pricing, and asset allocation all depend significantly on asset return volatility dynamics and the accuracy of predictions of integrated volatility (see e.g., Christoffersen and Diebold (2000) and Christoffersen and Diebold (2006)). When modelling uncertainty, it is also important to note that the economic relevance of predictions in general depends on whether densities or point forecasts are used. In particular, Granger and Pesaran (2000) find that the use of predictive distributions is more appropriate than point forecasts. These authors argue that distributions that provide a comprehensive description of uncertainty associated with the future values of a variable allow practitioners to more readily specify linkages between economic and statistical methods of forecast evaluation. Additionally, when it comes to taking advantage of such linkages in investment contexts, the empirical finance literature has found that measures of directional accuracy are more suitable than traditional measures of forecasting performance, such as quadratic loss functions based on point forecast errors. For example, Leitch and Tanner (1991) find that the relationship between directional predictive accuracy and profits from an interest-rate based trading strategy is stronger than the relationship between root mean squared error (RMSE) accuracy and profits. In this paper we ask questions that are related to the findings of both Granger and Pesaran (2000) and Leitch and Tanner (1991), in the context of volatility forecasting. Namely, what is the relevance of directional volatility forecasting, when used in return and volatility based trading strategies. Additionally, in this context, does it make sense to utilize point or interval (density) forecasts; and does the use of statistical or economic evaluation methods make a difference when choosing between forecast types? Finally, what role does the use of different non-parametric integrated volatility estimators play when evaluating the trade-offs associated with using interval versus point forecasts.

We undertake to answer the above questions by first assessing the marginal predictive gains associated with using “interval” (density) forecasts and “point” forecasts, when predicting the directional change in integrated volatility. We then utilize directional volatility predictions from both interval and point forecasting techniques in the construction of novel trading strategies, based on the contemporaneous causal relationship between daily returns and volatility. The strategies are particularly useful, as they lead to new insights about the linkage between predictive accuracy and trading profitability. Additionally, in order to allow for the existence of nonlinear dynamics in the relationship between volatility and lags thereof, we use both linear Heterogeneous Autoregressive (HAR) and nonlinear Recurrent Neural Network (RNN) models when constructing point predictions.¹ Our interval forecasts are constructed using the nonparametric Nadarya-Watson kernel density type estimators introduced by Corradi et al. (2011). Finally, we build on the seminal work of Jacod (2018) (which was originally published as a 1994 working paper) that has led to the development of numerous nonparametric measures of volatility, some of which are robust to microstructure noise and some of which are robust to jumps. Our contribution to this literature consists of a set of detailed Monte Carlo simulations, in which alternative estimators are used in the context of predictive interval estimation accuracy, in order to shed new light on the usefulness of said estimators. In closing, we note that we are aware of no study that has been carried out in order to evaluate the relative usefulness of interval (density) versus point predictions (of volatility) in the context of risk management.

One ingredient in our analysis of the marginal predictive gains associated with alternative point-forecasting models is the linear HAR model, which was first introduced by Corsi (2009) and later extended by Corsi et al. (2010), Duong and Swanson (2015), and Patton and Sheppard (2015). We also utilize, as mentioned above, nonlinear deep-learning Recurrent Neural Networks (RNN), which are relatively novel in the empirical asset

¹Empirical evidence presented in Cao and Tay (1992) suggests that clusters of returns are associated with clustered outliers, indicating that volatility dynamics may be nonlinear.

pricing (see e.g., Feng et al. (2018)) and volatility forecasting literature. Another important paper that analyzes the usefulness of such deep learning models for problems in financial prediction and classification is Heaton et al. (2016). In this paper, we employ a specific variant of the RNN model called the Long Short Term Memory Network (LSTM). The differentiating factor between our model and alternative neural network models used in the literature (as in Gallant and White (1992) and Gu et al. (2019)), is that the LSTM can capture both long run as well as short run dynamics of the variables being modeled, which makes LSTMs more suitable for volatility prediction.

Another important component of our analysis involves addressing whether superior statistical forecast accuracy necessarily translates into economic profitability. This is a relevant question, given that industry practitioners often base their decisions on financial rather than statistical criteria. Related literature in this area points to mixed results. Fleming et al. (2003) show that predictability captured by volatility modeling is economically significant and Bollerslev et al. (2016) show that it is possible to translate predictive improvements into sizable economic gains for a risk-averse investor. On the other hand, Satchell and Timmermann (1995) and Fuertes et al. (2015) do not find evidence of strong correlation between the statistical significance and economic value of forecasts. Moreover, none of the above studies utilize volatility density forecasts for economic evaluation, or consider directional accuracy as the key measure of statistical forecast accuracy. In this paper, we do both of these things, by extending the application of the *Volatility Feedback Effect* theory of Bollerslev and Zhou (2006) to a daily-framework, and utilizing contemporaneous causal relationships between returns and estimated volatility to develop empirical trading strategies. In our strategies, trading signals (buy or sell) are generated by directional volatility forecasts. Moreover, in order to reduce the number of daily trades and limit potential losses caused by large price drops, we also use simple-moving-average (SMA) and double-crossing-moving-average (DCMA) trading signals in our strategies. These strategies differ from the “volatility trading” strategy that treats the VIX index as an asset class (see Konstantinidi et al. (2008)) or “volatility timing” strategy that uses covariance forecasts for portfolio construction (see Fleming et al. (2003)).

Finally, it should be stressed that the framework of this paper offers a natural route to assessing the efficacy of using alternative nonparametric estimators of volatility in forecasts and trading contexts. In our experiments and simulations, we explore the usefulness of various estimators, such as realized volatility (RV, Andersen et al. (2001)) which is robust to neither microstructure noise nor price jumps, noise-robust two scale realized volatility (TSRV, Zhang et al. (2005)), and price-jump robust truncated realized volatility (TRV, Aït-Sahalia and Jacod (2009))². In particular, we consider ten such volatility estimators. Although there are a number of studies comparing the point accuracy of volatility estimators (see, e.g. Liu et al. (2015)), only one paper that we know of includes a discussion of such estimators in the context of predictive interval estimation accuracy (i.e., Corradi et al. (2011)), and these authors only discuss the results of a very limited set of simulations. We build on Corradi et al. (2011), by incorporating a much larger collection (variety) of volatility estimators as well as simulated financial market conditions or data generating processes (DGPs).

Our empirical findings, based on the evaluation of high-frequency asset-price data sampled at 5, 2.5, 1.25 minutes, and on a series of Monte-Carlo simulations can be summarized as follows. First, and most importantly, interval (density) predictions can improve upon point predictions in terms of trading profitability, regardless of the “linear” or “nonlinear” nature of the point-forecasting model. Additionally, linear (nonlinear) model-based point forecasts perform worse (marginally better) than interval forecasts when it comes to directional predictive accuracy. More specifically, we find that interval-forecasts perform better than (linear) HAR point-forecasts 97% of the time, in terms of directional predictive accuracy (with 5% higher directional predictive

²Refer to Barndorff-Nielsen et al. (2008) and Barndorff-Nielsen et al. (2011) for a discussion of kernel based noise-robust estimators. Jump-robust estimators are discussed in Barndorff-Nielsen and Shephard (2004), Corsi et al. (2010), and Andersen et al. (2012).

accuracy or DPA rates, on average) and in terms of generation of higher risk-adjusted returns (29% higher Sharpe Ratios, on average). On the other hand, interval forecasts perform marginally worse than nonlinear RNN point forecasts, in terms of directional predictive accuracy, by outperforming RNN forecasts 25% of the time and generating a 0.8% lower DPA rates, on average. Still, even in this case, RNN-point forecasts result in lower overall mean-variance return ratios compared to interval forecasts (30% lower Sharpe ratios, on average). In summary, the fact that the (linear) HAR model performs the worst in terms of predictive accuracy and trading profitability, in comparison to the other forecasting methodologies highlights the likely existence of a nonlinear relationship between estimated volatility and lags thereof. Additionally, it is worth noting that predictive intervals more accurately estimate future volatility movements (in comparison to RNN forecasts) when there are “large jumps” in the price process. In particular, stocks for which interval forecasts outperform RNN forecasts contain a higher percentage of “large jump variations” in comparison to other stocks.

Second, the above findings suggest that there is a linkage between directional predictive accuracy and economic gains. However, the strength of such linkage depends on the volatility estimator used, which in turn is affected by the magnitude of price jumps and/or microstructure noise in the price process. In particular, Sharpe (Sortino) ratios for HAR-forecasts, relative to those associated with interval forecasts, are larger for noise-robust volatility estimators such as TSRV than for jump-robust estimators such as TRV or even RV (which is robust to neither noise nor jumps). Indeed, our empirical evidence also suggests that both TRV and RV have a stronger causal relationship with daily returns (in comparison to TSRV), in the sense of resulting in superior risk management (i.e., higher trading profits).

Finally, our Monte-Carlo experiments reveal that, in terms of predictive (interval) accuracy, RV performs better than all noise-robust estimators, as well as some jump-robust ones, even when the data generating process contains small-jumps and microstructure noise. This suggests that in the presence of “small” jumps, noise-robust measures tend to overestimate the level of noise, while some jump-robust estimators fail to distinguish between “large” and “small” jumps. This superior predictive performance of RV points to the relevance of parsimony in the construction of volatility estimators. It also explains why RV displays a strong causal linkage with daily returns and consequently generates higher risk-adjusted returns, in our empirical analysis.

The rest of the paper is organized as follows. Section 2 summarizes the theoretical setup, Sections 3 and 4 introduce interval and point forecasting methodologies, respectively, and Section 5 lists the realized volatility measures examined in the paper. Section 6 contains a discussion of the data, and Section 7 discusses empirical trading strategies. Finally, Sections 8 and 9 outline the empirical setup and summarize our empirical findings, while Section 10 contains the results from our Monte Carlo simulations. Section 11 concludes.

2 Theoretical Setup

Let Y_t be the log-price of financial asset at time t . Assume Y_t follows a jump-diffusion model (almost surely, its paths are right continuous with left limits) and has the following representation

$$Y_t = Y_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + J_t, \quad (1)$$

where μ_s is the drift term, σ_s is the diffusion term, W_s is a standard Brownian motion³. J_t is a pure jump process, equivalent to the sum of all discontinuous log price movements, $\sum_{s \leq t} \Delta Y_s$. When the jump component

³The setup used here follows that used in Mukherjee et al. (2020).

is a finite activity jump process (i.e. a compound poisson process (CPP)), then

$$J_t = \sum_{j=1}^{N_t} c_j \quad (2)$$

where N_t is a poisson process with intensity λ , jumps occur at the corresponding times given as $(\tau_j)_{j=1,\dots,N_t}$, and c_j refers to *i.i.d* random variables measuring the size of the jumps at time τ_j ⁴. Quadratic variation (QV) denotes the variance of Y_t and is written as

$$QV_t = [Y, Y]_t = [Y, Y]_t^c + [Y, Y]_t^d. \quad (3)$$

The two components of QV are respectively, (i) variation due to the continuous component $[Y, Y]_t^c = \int_0^t \sigma_s^2 ds$ and (ii) variation due to the discontinuous jump component $[Y, Y]_t^d = \sum_{j=1}^{N_t} c_j^2$. The continuous part of QV, the main variable of interest, is known as (daily) integrated volatility (IV), defined as

$$IV_t = \int_{t-1}^t \sigma_s^2 ds, \quad t = 1, \dots, T. \quad (4)$$

The presence of market frictions in high frequency financial data has been well documented in the literature. Hence the observed log price X is considered to be the sum of the latent log price Y and market microstructure noise, ϵ . If M equi-spaced intra-daily observations for each of T days for X are considered, then

$$X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}, \quad t = 0, \dots, T \quad \& \quad j = 1, \dots, M. \quad (5)$$

Since IV is a latent variable, consistent estimators thereof are utilized in the sequel. These estimators or realized measures (RM), use M high-frequency intra-day observations to estimate IV for a fixed time period (e.g., one trading day) and satisfy

$$RM_{t,M} = IV_t + N_{t,M} \xrightarrow{u.c.p} \int_{t-1}^t \sigma_s^2 ds = IV_t, \quad (6)$$

where $N_{t,M}$ denotes the measurement error and $\xrightarrow{u.c.p}$ denotes convergence in probability, uniformly in time⁵.

3 Interval (Density) Forecasts

This section details the interval (density) forecasting methodology used in this study. Tay and Wallis (2000) refer to the density forecast of a realization of a random variable at some future time as “an estimate of the probability distribution of the possible future values of that variable”. Consequently, these density forecasts provide a comprehensive description of the uncertainty associated with a prediction and can be harnessed to construct predictive intervals. These predictive intervals specify the probability with which the given random variable falls within a stated interval at a future time point. In this paper, our objective is construct estimators of (density based) predictive intervals for IV at time $T+h$ conditional on past and present values of estimated IV. For simplicity, we confine our attention to one step ahead prediction, so that $h = 1$. Our one-step ahead interval estimators are model-free or nonparametric, are introduced in Corradi et al. (2009) and Corradi et al.

⁴For more details on continuous time asset price modeling, see Aït-Sahalia and Jacod (2014), and references cited therein.

⁵For regularity conditions under which standard volatility estimators are consistent, refer to Aït-Sahalia and Jacod (2014)

(2011), and are briefly discussed in the next section.

3.1 Predictive Intervals

The one-step ahead predictive IV interval (or conditional confidence interval), $CI_{T+1}(u_1, u_2 | RM_{T,M})$, can be defined as the probability that IV at time ($T+1$), falls within a given interval, say $[u_1, u_2]$, conditional on the values of estimated IV at all time points up to T . Namely

$$CI_{T+1}(u_1, u_2 | RM_{T,M}) = Pr(u_1 \leq IV_{T+1} \leq u_2 | RM_{T,M}). \quad (7)$$

Since IV is a latent variable, confidence intervals are based on realized measures of IV, which leads to the following expression

$$\begin{aligned} CI_{(T+1),M}(u_1, u_2 | RM_{T,M}) &= Pr(u_1 \leq RM_{T+1,M} \leq u_2 | RM_{T,M}) \\ &= F_{RM_{T+1,M}|RM_{T,M}}(u_2 | RM_{T,M}) \\ &\quad - F_{RM_{T+1,M}|RM_{T,M}}(u_1 | RM_{T,M}). \end{aligned} \quad (8)$$

Given that the moments of the measurement error, $N_{t,M}$, decay to zero at a fast enough rate as $M \rightarrow \infty$, it is feasible to assume that

$$Pr(u_1 \leq RM_{T+1,M} \leq u_2 | RM_{T,M}) \approx Pr(u_1 \leq IV_{T+1} \leq u_2 | RM_{T,M}). \quad (9)$$

We follow Corradi et al. (2011) and use nonparametric Nadaraya-Watson estimators to estimate $CI_{(T+1),M}$, so that

$$\begin{aligned} \widehat{CI}_{(T+1),M}(u_1, u_2 | RM_{T,M}) &= \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_2 | RM_{T,M}) \\ &\quad - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_1 | RM_{T,M}) \\ &= \frac{\frac{1}{T\xi} \sum_{t=0}^{T-1} 1_{(u_1 \leq RM_{t+1} \leq u_2)} K(\frac{RM_{t,M} - RM_{T,M}}{\xi})}{\frac{1}{T\xi} \sum_{t=0}^{T-1} K(\frac{RM_{t,M} - RM_{T,M}}{\xi})}, \end{aligned} \quad (10)$$

where K is a given kernel function and ξ is the bandwidth parameter. $\widehat{CI}_{(T+1),M}(u_1, u_2 | RM_{T,M})$ is consistent and asymptotically normally distributed given mild assumptions of the underlying diffusion process ⁶. Namely, if the following assumptions hold: (i) σ_t^2 is a strictly stationary & α -mixing process (spot volatility process is strong mixing); (ii) kernel K is a symmetric, nonnegative, continuous function (standard assumptions in the nonparametric density estimation literature); and (iii) there exists a sequence b_M , with $b_M \rightarrow \infty$, such that as $M \rightarrow \infty$, $E(|N_{t,M}|^k) = O(b_M^{-K/2})$ for $k \geq 2$, then

$$\sqrt{T\xi}(\widehat{CI}_{(T+1),M}(u_1, u_2 | RM_{T,M}) - CI_{T+1}(u_1, u_2 | RM_{T,M})) \xrightarrow{d} N(0, V(u_1, u_2)), \quad (11)$$

where

$$\widehat{V}(u_1, u_2) = \frac{\int K^2(u) du}{\widehat{f}_{RM_{T,M}}(RM_{T,M})} \widehat{CI}_{(T+1),M}(u_1, u_2 | RM_{T,M})(1 - \widehat{CI}_{(T+1),M}(u_1, u_2 | RM_{T,M})), \quad (12)$$

⁶Refer to Corradi et al. (2011), Assumptions A1-A4 (pp. 1497) for more details.

and $\widehat{f}_{RM_{T,M}}(RM_{T,M}) = \frac{1}{T\xi} \sum_{t=0}^{T-1} K\left(\frac{RM_{t,M} - RM_{T,M}}{\xi}\right)$. It should be noted here that for (11) to hold, M, T, ξ should follow either: (i) $T\xi^3 \rightarrow \infty, T\xi^5 \rightarrow \infty$ & $\max\{b_M^{-1/2}T^{\frac{3}{2k}}, T^{\frac{3+k}{k}}\xi b_M^{-1}\} \rightarrow 0$ or (ii) $T\xi \rightarrow \infty, T\xi^5 \rightarrow 0$ & $\max\{b_M^{-1/2}T^{\frac{3}{2k}}, T^{\frac{3+k}{k}}\xi b_M^{-1}, b_M^{-1}\xi^{-2}\} \rightarrow 0$. Additionally, b_M differs according to the choice of the IV estimation technique. Given the above conditions, reasonable choices for M, T , and ξ are $M = \{78, 156, 312\}$, $T = 100$, a Gaussian kernel, and $\xi = cT^{-1/5}$, where c follows Silverman's 1986 rule ⁷.

3.2 Directional Predictions

Our motivation for using confidence interval estimators is to predict volatility movement (i.e., directional change in estimated IV). In particular, we are interested in gauging whether future volatility rises or falls. Thus, we want to estimate the following probability (predictive interval)

$$Pr(0 \leq RM_{T+1} \leq RM_T | RM_{T,M}). \quad (13)$$

Using the confidence interval estimator given in (10) leads to the following interval estimator

$$p = \widehat{Pr}(0 \leq RM_{T+1} \leq RM_T | RM_{T,M}) = \frac{\frac{1}{T\xi} \sum_{t=0}^{T-1} \mathbb{1}_{(0 \leq RM_{t+1} \leq RM_T)} K\left(\frac{RM_{t,M} - RM_{T,M}}{\xi}\right)}{\frac{1}{T\xi} \sum_{t=0}^{T-1} K\left(\frac{RM_{t,M} - RM_{T,M}}{\xi}\right)}. \quad (14)$$

It follows that if the estimated probability, p , is above a given benchmark, then future volatility will fall within the pre-specified interval. Following standard convention in the literature related to probability and direction-of-change forecasts, we set $p = 0.5$ ⁸, and define

$$PD_{T+1} = \begin{cases} up, & p = \widehat{Pr}(0 \leq RM_{T+1} \leq RM_T | RM_{T,M}) < 0.5 \\ down, & p = \widehat{Pr}(0 \leq RM_{T+1} \leq RM_T | RM_{T,M}) \geq 0.5, \end{cases}$$

where PD_{T+1} denotes the predicted *direction of change* of volatility. If the predictive interval p has a value greater or equal to 0.5, then the predicted direction of change is *down* while a value less than 0.5 for the predictive interval p signifies an “upward” movement of future volatility.

4 Point Forecasts

This section details the point forecasting methodologies used in the sequel. In particular, we focus on two specific types of models (i) Heterogeneous Autoregressive models (HAR) and (ii) Recurrent Neural Network models (RNN). HAR models, which are standard in the volatility forecasting literature, assume linearity ⁹. On the other hand, deep learning neural network models are capable of extracting nonlinear factors and provide a powerful alternative for feature selection and shrinkage. We use a specific type of Recurrent Neural network called the Long-Short-Memory Network (LSTM). The LSTM network has the capacity to store a long running memory of the sequence of input data as well as the short run memory of recent network outputs.

⁷Details about the construction of infeasible conditional confidence intervals based on “pseudo true” IV are given in Section 10.

⁸Refer to Christoffeersen et al. (2007) for more details on probability and direction-of-change forecasts when modeling equity-returns and volatility.

⁹Refer to Duong and Swanson (2015) and Patton and Sheppard (2015) for more details

4.1 Heterogeneous Autoregressive (HAR) Model

After the introduction of the HAR model in Corsi (2009), the basic model was augmented by various authors with the inclusion of a variety of jump variation related variables¹⁰. For our comparison with interval (density) forecasts, we use the baseline version of the model

$$RM_{t+h} = \alpha_0 + \alpha_1 RM_t + \alpha_2 RM_{t-4} + \alpha_3 RM_{t-21} + \epsilon_{t+h}, \quad (15)$$

where RM_{t+h} is h -step ahead estimated IV, h is set equal to 1, and $RM_{t,t-p}$ is the average of estimated IV over the most recent ($p+1$) days. In particular

$$RM_{t,t-p} = \frac{1}{p+1} \sum_{i=0}^p RM_{t-i} \quad (16)$$

Here, $RM_{t,t-4}$ refers to average estimated IV over the last one week and $RM_{t,t-21}$ refers to the average estimated IV over the last one month. Model estimation and volatility prediction is carried out each day using a rolling window estimation scheme, prior to the construction of each new daily IV prediction. The length of the rolling window is the in-sample period, T_1 days. The out-of-sample period consists of T_2 days, with $T_1 + T_2 = T$. Our rolling window estimation scheme is carried out as follows. First, the model is estimated using $1, \dots, T_1$ observations and one-day ahead forecasts are constructed for the $(T_1 + 1)^{th}$ day. Then, in order to forecast the volatility on the $(T_1 + 2)^{th}$ day, we first estimate the data using trading days $2, \dots, T_1 + 1$. This procedure is continued until the end of the dataset is reached.

4.2 Recurrent Neural Network (RNN) Model

Recurrent Neural Networks (RNN) are a specific variety of deep (supervised) learning models¹¹. Unlike traditional deep learning models (like feed forward networks) where signals only travel one way (from input to output), recurrent neural networks allow signals to travel in both directions (predictions derived from the input of a previous layer are fed back into the network) by introducing loops in the network. Additionally, the particular form of the RNN used in this paper, called the Long Short Term Memory Network (LSTM), incorporates multiple memory units in the network architecture and solves the issue of the “vanishing and exploding gradient” which the traditional RNNs suffer from. As the RNN (or LSTM) is an adaptation of the standard deep learning architecture (the feed-forward type), it is important to first understand how deep learning models work in the context of volatility forecasting.

In a standard deep learning network, L nonlinear transformations are used to predict an outcome variable y via predictors, x . Given that the original input is x , the output from the first of the L transformations represents the first layer (and so on) while the output of the $(L + 1)^{th}$ transformation is the outcome variable y . The number of layers L where $l \in \{1, \dots, L\}$ (also referred to as the hidden layers) represents the depth of the network architecture. Each of the L layers has a univariate activation function attached to it, say f_1, \dots, f_L . Activation functions are simply nonlinear transformations of the weighted data. One commonly used activation functions is the sigmoid function, $\left(\frac{1}{1+exp(-x)}, tanh(x)\right)$. Now, $Z^{(l)}$ denotes the $l - th$ layer, which is vector with the same length as the number of neurons in that layer, and with input $x = Z^{(0)}$. Finally, the

¹⁰More recently Bollerslev et al. (2016) further improved volatility forecasting by allowing the HAR type model coefficients to evolve according to the degree of measurement error.

¹¹Refer to Feng et al. (2018) for a detailed description of the application of deep learning models in empirical asset pricing.

“deep prediction rule” includes univariate semi-affine functions and can be written as

$$F^{W,b} = F_1^{W(1),b(1)} \circ \dots \circ F_L^{W(L),b(L)}, \quad (17)$$

where

$$F_l^{W(l),b(l)} := f_l \left(W^{(l)} Z^{(l)} + b^{(l)} \right) = f_l \left(\sum_{i=1}^{N_l} W_i^{(l)} Z_i^{(l)} + b_i^{(l)} \right), \quad (18)$$

and N_l is the number of neurons, $w^{(l)}$ denotes weight matrices, and $b^{(l)}$ denotes the threshold or activation level contributes to the output of the hidden unit. In terms of traditional financial modeling, the $F^{(l)}$ can be viewed as latent factors. One feature of this model is that with the deep learning structure, a convolution of factors is used instead of the traditional additive structure seen in the case of traditional econometric models.

For IV prediction the deep learning model can be adapted in the following way. Let $RM_{t+1} \in \mathbb{R}^{T \times 1}$ be the vector of estimated IV and let $x_t \in \mathbb{R}^{T \times p}$ be the set of predictor variables. Factors are extracted from the dataset using the dual goals of good out-of-sample prediction and in-sample-fit. In particular, define the following hierarchical model

$$RM_{t+1} = \alpha + \beta x_t + \beta_f F_t + \epsilon_{t+1} \quad (19)$$

$$F_t = F^{W,b}(x_t), \quad (20)$$

where $F^{W,b}$ is defined in (17). Traditionally researchers estimate factors, F_t , and then “learn” the coefficients α and β in a two step procedure. Deep learning networks estimate the coefficients α and β jointly. To “train” the model, one generally uses the following loss function

$$Loss = \frac{1}{T} \sum_{t=1}^T (RM_{t+1} - \widehat{RM}_{t+1})^T (RM_{t+1} - \widehat{RM}_{t+1}) + \lambda \phi(\beta, W, b), \quad (21)$$

where $\lambda \phi(\beta, W, b)$ is a regularization penalty used for predictor selection that controls for over-fitting and λ is the regulation parameter. We use the stochastic gradient descent (SGD) algorithm in order to jointly estimate α , β , β_f , and F_t using this loss function¹². This is the basic framework of the RNN models from which LSTMs can be specified and estimated. For a complete discussion of these models, refer to (Heaton et al. (2016)). Finally, note that in order to allow for a fair comparison with the HAR model, the input x_t at time point t for the LSTM model is kept the same as the regressor variables in the HAR model in eq. (15). Additionally, the in-sample dataset for training the LSTM model for parameter estimation is the same as the dataset used for the initial in-sample fit of the HAR model.

In our empirical analysis. we use one-step ahead point forecasts from both the HAR and RNN (LSTM) models to predict the *direction of change* of volatility, just as was done using our interval (density) forecasts. Namely, we define

$$PD_{T+1} = \begin{cases} up, & RM_{T+1}^{predicted} \geq RM_T^{actual} \\ down, & RM_{T+1}^{predicted} < RM_T^{actual} \end{cases}$$

where PD_{T+1} denotes the predicted *direction of change* of volatility. If the predicted volatility value for period (T+1) is greater than actual volatility of period T, it signifies an “upward” movement. On the other hand a predicted volatility value less than the current period actual volatility denotes a “downward” movement.

¹²Refer to Le et al. (2015) for a discussion of the stochastic gradient descent algorithm.

5 Integrated Volatility Estimators

This section details the nonparametric IV estimators that we use in our empirical analysis (see Table 1 for a summary of these estimators)¹³. The “jump robust” estimators that we use are consistent estimators of IV which make use of jump-truncation methodologies to remove variation due to price jumps from the quadratic variation of the price process. When the magnitude of jump variation (i.e. the difference between QV and IV) is negligible, jump-robust estimators have a tendency to over-estimate the level of jumps (as evidenced by the Monte-Carlo experiments discussed in Section 10 of this paper). In such cases, to allow for more accurate volatility prediction, one option is to utilize non jump-robust estimators, the simplest of which is RV. There is also a class of non-jump robust estimators which are robust to microstructure noise. We examine the usefulness of three *noise-robust estimators*, six *jump-robust estimators*, and RV in the sequel. A summary of these estimators follows.

5.1 Realized Volatility

The RV estimator was introduced in Andersen et al. (2001), and was the first empirical measure that used high-frequency intra-day returns to compute daily return variability without explicitly modeling the intra-day data. The authors showed that under suitable conditions RV is an unbiased and efficient estimator of QV. More specifically, in the absence of jumps or when jumps populate the data infrequently, RV converges in probability to IV, as $M \rightarrow \infty$, where

$$RV_{t,M} = \sum_{j=1}^{M-1} (X_{t+(j+1)/M} - X_{t+j/M})^2. \quad (22)$$

5.2 Jump-Robust Estimators

(i) Bipower Variation (BPV: Barndorff-Nielsen and Shephard (2004)) is an estimator that is robust to rare jumps, and is defined as

$$BPV_{t,M} = (\mu_1)^{-2} \sum_{j=2}^{M-1} |\Delta_j X| |\Delta_{j-1} X|, \quad (23)$$

where

$$\Delta_j X = X_{t+(j+1)/M} - X_{t+j/M}, \quad (24)$$

and $\mu_1 = 2^{\frac{1}{2}} \frac{\Gamma(1)}{\Gamma(\frac{1}{2})}$.

(ii) Tripower Variation (TPV: Barndorff-Nielsen and Shephard (2004)) is an extension of BPV which is both more efficient and more vulnerable to microstructure noise, and is defined as

$$TPV_{t,M} = (\mu_{\frac{2}{3}})^{-3} \sum_{j=3}^{M-1} |\Delta_j X|^{2/3} |\Delta_{j-1} X|^{2/3} |\Delta_{j-2} X|^{2/3}, \quad (25)$$

where $\mu_{\frac{2}{3}} = 2^{\frac{1}{3}} \frac{\Gamma(\frac{5}{6})}{\Gamma(\frac{1}{2})}$.

(iii) Truncated Realized Volatility (TRV: Aït-Sahalia and Jacod (2009)) removes price jumps greater than a

¹³Refer to Mukherjee et al. (2020) for a detailed review of nonparametric IV estimators.

given threshold, with the threshold chosen in a data-driven manner, and is defined as

$$TRV_{t,M} = \sum_{j=1}^{M-1} |\Delta_j X|^2 1_{\{|\Delta_j X| \leq \alpha \Delta_M^\varpi\}}, \quad (26)$$

where $\alpha = 5\sqrt{\sum_{j=1}^{M-1} |\Delta_j X|^2 1_{\{|\Delta_j X| \leq \Delta_M^{1/2}\}}}$, $\varpi = 0.47$, and $\Delta_M = 1/M$.

(iv) Threshold Bipower Variation (TBPV: Corsi et al. (2010)) is robust to the choice of threshold function, v , and is defined as

$$TBPV_{t,M} = \mu_1^{-2} \sum_{j=2}^{M-1} |\Delta_{j-1} X| |\Delta_j X| I_{\{|\Delta_{j-1} X|^2 \leq v_{j-1}\}} I_{\{|\Delta_j X|^2 \leq v_j\}}, \quad (27)$$

where $v_j = c_v^2 \hat{V}_j$ and $\hat{V}_j^z = \frac{\Sigma_{i=-L}^L \kappa(\frac{i}{L}) (\Delta_{j+i} X)^2 I_{\{(\Delta_{j+i} X)^2 \leq c_v^2 \hat{V}_{j+i}^{z-1}\}}}{\Sigma_{i=-L}^L \kappa(\frac{i}{L}) I_{\{(\Delta_{j+i} X)^2 \leq c_v^2 \hat{V}_{j+i}^{z-1}\}}}$. Here we take $L = 25$, $c_v = 3$, $\hat{V}^0 = +\infty$. v_j is

the threshold for removal of large returns at each j . \hat{V}_j^z gives estimated local variance in the presence of jumps at each iteration z for any j . Large returns are removed at each iteration according to $\{(\Delta_j X)^2 \leq c_v^2 \hat{V}_j^{z-1}\}$ and the estimated variance at that iteration is multiplied by c_v^2 to get the threshold for the next iteration. When large returns cannot be removed any more, the iterations stop. Typically z is taken to be 2.

(iv) MedRV, which is due to Andersen et al. (2012), uses “nearest neighbor truncation” methodology (doesn’t require an ex-ante threshold), is robust to both jumps and microstructure noise, and is defined as

$$MedRV_{t,M} = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{M}{M-2} \right) \sum_{j=2}^{M-1} med(|\Delta_{j-1} X|, |\Delta_j X|, |\Delta_{j+1} X|)^2. \quad (28)$$

(v) MinRV, which is due to Andersen et al. (2012)), uses “nearest neighbor” methodology and minimum of adjacent absolute returns, is robust to both jumps and noise, and is defined as

$$MinRV_{t,M} = \frac{\pi}{\pi - 2} \left(\frac{M}{M-1} \right) \sum_{j=1}^{M-1} min(|\Delta_j X|, |\Delta_{j+1} X|)^2. \quad (29)$$

5.3 Noise-Robust Estimators

(i) Two Scale Realized Volatility (TSRV: Zhang et al. (2005)) combines estimators obtained over two different time scales, avg and M , and is defined as

$$TSRV_{t,M} = [X, X]^{avg} - \frac{1}{K} [X, X]^M, \quad (30)$$

where $[X, X]^i = \sum_{j=0}^{m-2} (X_{t+((j+1)K+i)/M} - X_{t+(jK+i)/M})^2$, $i = 1, \dots, K$ & $m \approx \frac{M}{K}$. Additionally $[X, X]^{avg} = \frac{1}{K} \sum_{i=1}^K [X, X]^i$ and $[X, X]^M = \sum_{j=1}^{M-1} (\Delta_j X)^2$. $K = cM^{2/3}$ is the number of subsamples and $c > 0$ is a constant.

(ii) Multi Scale Realized Volatility (MSRV: Zhang et al. (2006)) uses N different time scales, and is defined as

$$MSRV_{t,M} = \sum_{n=1}^N a_n [X, X]^{(M, K_n)}, \quad n = 1, \dots, N, \quad (31)$$

where $a_n = 12 \frac{n}{N^2} \frac{n/N-1/2-1/(2N)}{1-1/N^2}$, $\sum_{n=1}^N a_n = 1$, and $\sum_{n=1}^N a_n/n = 0$. Additionally $[X, X]^{(M, K_n)} = \frac{1}{K_n} \sum_{l=1}^{K_n} \sum_{j=0}^{m_{n,l}-2} (X_{t+((j+1)K_n+l)/M} - X_{t+(jK_n+l)/M})^2$. Here $l = 1, \dots, K_n$ and $m_{n,l} = \frac{M}{K_n}$. We take $N = 3$, $K_1 = 1$, $K_2 = 2$, and $K_3 = 3$.

(iii) Realized Kernel (RK: Barndorff-Nielsen et al. (2008)) is a kernel function based estimator which for particular kernel functions can be asymptotically more efficient than MSRV, and is defined as

$$RK_{t,M} = \gamma_0(X) + \sum_{h=1}^H \kappa\left(\frac{h-1}{H}\right)\{\gamma_h(X) + \gamma_{-h}(X)\}, \quad (32)$$

where $\gamma_h(X) = \sum_{j=1}^{M-1} (X_{t+(j+1)/M} - X_{t+j/M})(X_{t+(j+1-h)/M} - X_{t+(j-h)/M})$. Here c is a constant. For our analysis we consider a Turkey-Hanning₂ kernel which gives $\kappa(x) = \sin^2\{\pi/2(1-x)^2\}$ and $H = cM^{1/2}$.

6 Data Description

This section details the data used in our empirical analysis. The dataset consists of high frequency intraday stock price data retrieved from NYSE's Trade and Quote database (TAQ) for a period of eight years, 2010-2017. In particular, we consider six stocks representing different sectors of the market; Boeing Company (BA), General Electric (GE), International Business Machines Corporation (IBM), JP Morgan Chase & Co. (JPM), Microsoft Corporation (MSFT), Proctor and Gamble Co. (PG). Elimination of trading days with missing data leads to a final dataset that contains 2003 days. Transaction prices for stocks are recorded for the regular trading hours 9.30 a.m to 4.00 p.m (overnight returns are excluded from our sample). We use three different sampling intervals - namely 5 minutes, 2.5 minutes and 1.25 minutes. These represent 78, 156 and 312 observations per day, respectively. Data cleaning and sub-sampling all follow standard procedures as described in Ait-Sahalia and Jacod (2012).

In our prediction experiments we divide the dataset into an in-sample training period of 1250 days (2010-2014) and an out-of-sample testing period of 753 days (2015-2017). The in-sample dataset serves two purposes: (i) training the forecasting models for out-of-sample prediction and (ii) design of empirical trading strategies. We also carry out empirical “jump tests” on in-sample, out-of-sample and the entire dataset to estimate the level of “price jumps” in the price process of the representative stocks. This is crucial in our subsequent analysis, when we discuss the results of our predictive accuracy and trading profitability comparisons. In the following sub-section, we describe the level of “jump-activity” in our data.

6.1 Price Jump Activity

Table 2 (Panels A-C) summarize jump-activity test findings. Our testing approach follows Duong and Swanson (2011), and uses the adjusted jump ratio test statistic developed by Huang and Tauchen (2005) to test the null hypothesis of “no price jumps” (i.e., $H_0 : J_t = 0$). In particular, H_0 is rejected if the test statistic $Z_{t,M}(\alpha)$ is more than the critical value, Φ_α , where

$$Z_{t,M} = \frac{\sqrt{M}}{\sqrt{\vartheta \max(t^{-1}, \widehat{IQ}_t / (\widehat{IV}_t)^2)}} \left(1 - \frac{\widehat{IV}_t}{\sum_{j=1}^M (\Delta_j X)^2} \right) \xrightarrow{D} N(0, 1), \quad (33)$$

with $\widehat{IQ}_t = M \mu_{\frac{4}{3}}^{-3} \sum_{j=3}^M |\Delta_j X|^{\frac{4}{3}} |\Delta_{j-1} X|^{\frac{4}{3}} |\Delta_{j-2} X|^{\frac{4}{3}}$. We test for the presence (absence) of jumps and construct the following measures:

- (i) Percentage of days identified as have price jumps, $PJD = \frac{\sum_{i=0}^T I(Z_{t,M}(\alpha) > \Phi_\alpha)}{T} * 100$
- (ii) Relative contribution of Jump Variation to the Total Variation,
 $JV = \frac{\sum_{i=0}^T \frac{VJ_t}{RV_{t,M}}}{T} * 100$ (where $VJ = \max\{0, RV_{t,M} - \widehat{IV}_t\} * I_{jump,t}$)
- (iii) Relative contribution of “Large” Jump Variation to Total Variation,
 $LJV = \frac{\sum_{i=0}^T \frac{VLJ_t(\gamma)}{RV_{t,M}}}{T} * 100$ (where “large jumps” are defined as large price movements above a given truncation level γ and $VLJ = \min\{VJ_t, (\sum_{j=1}^M (\Delta_j X)^2 I_{|\Delta_j X| \geq \gamma} * I_{jump,t})\}$)¹⁴.

Panels A to C in Table (2) denote jump-test based results (i.e., PJD, JV and LJV values for the six representative stocks across the (i) in-sample period; (ii) out-of-sample period (iii) entire data set) for a given sampling frequency. Our key findings can be summarized as below. First, notice that although the percentage of days identified as having jumps (PJD) does not differ appreciably across different stocks, there are sometimes considerable difference in terms of the “large jump” variation. For instance although PJD is 35.60% for MSFT and 33.44% for GE (for a sampling frequency of 5 minutes using the in-sample period), the large jump variation for MSFT is 1.58% and for GE is (0.89%). So MSFT’s large jump variation is almost 1.7 times that of GE, while the percentage of jump days is only 1.06 times that of GE. Second, the level of price jump activity does not differ much across different sampling periods within a given sampling frequency. This is crucial as it indicates that if a trading strategy is based on the in-sample period, then the result of deploying the strategy in the out-of-sample period will not be affected by the level of price jump activity. Third, the level of price jump activity increases considerably (for all stocks and sampling periods) based on the sampling frequency used. For instance PJD for MSFT increases gradually from 35.60% to 66.16% when the sampling frequency increases from 5 minutes to 1.25 minutes. This is not surprising because, with the increase in sampling frequency, more observations are used per day, leading to higher probability of detection of a price jump for a given day. This feature of the data will be crucial in understanding key patterns in the our predictive accuracy and profitability results.

7 Designing Trading Strategies

The direction-of-change forecasts generated via the interval and point forecasting methods discussed in the previous sections can be translated into trading signals using strategies that we outline below. In particular, we first document the contemporary theories regarding return-volatility association. We then provide empirical evidence of such an association at the daily horizon. Finally, we suggest a feasible long-only day-wise trading strategy which can be easily deployed and evaluated in our context. It should be noted here that the strategy that we develop is different from the “volatility trading” strategy that treats the VIX index as an asset class (see Konstantinidi et al. (2008)) or “volatility timing” strategy that uses covariance forecasts for portfolio construction (see Fleming et al. (2003)).

7.1 Contemporary Theories

There is broad agreement in the literature that rising asset prices are accompanied by declining volatility, and vice versa. There are two documented theories which explain this relationship. The *Leverage Effect* theory (see Black (1976)) posits that a drop in value of a stock increases financial leverage or the debt to equity ratio of the firm, which in turn makes the stock riskier and increases its future volatility. The *Volatility Feedback Effect* theory (see French et al. (1987)) offers an alternative explanation. If volatility is priced, an expected increase

¹⁴The threshold is calculated by taking the maximum increments of log(price) over periods of 20 days, and given the entire sample of maximum increments, we use the 50th percentile of the entire sample of increments as the truncation level, γ .

in volatility raises the required return on equity, leading to an immediate stock price decline to adjust to the change in future expectations¹⁵. Summarizing, the linkage between return and volatility is contemporaneous according to the *Volatility Feedback Effect* (see Bollerslev and Zhou (2006)) while it is inter-temporal according to the *Leverage Effect*.

In our experiments, we utilize the feedback effect based contemporaneous relationship between return and volatility. Bollerslev and Zhou (2006), who find negative estimates of this causal relationship, provide a simple theoretical framework to show that such results are not necessarily inconsistent with the basic ICAPM model¹⁶. The authors conclude that the sign and the magnitude of the relationship depend on parameters of the underlying structural model, the length of return horizon, and the particular volatility measure used. In particular, they find a negative linkage between return and volatility when considering a monthly horizon and RV as the volatility estimator. It remains to be seen how this relationship would change if a daily return-volatility horizon or if other volatility estimators were used.

7.2 Return-Volatility Regression

We consider the following volatility feedback effect based return-volatility regression equation as in Bollerslev and Zhou (2006)

$$R_{t-1,t} = \alpha + \beta RM_{t-1,t} + u_t, \quad (34)$$

where $R_{t-1,t}$ is the daily continuously compounded return for day t and stands for $\log\left(\frac{p_t^c}{p_t^o}\right)$ (p_t^c is the closing price and p_t^o is the opening price) and $RM_{t-1,t}$ stands for estimated volatility in period t . We carry out daily return-volatility regression analysis using the above equation, with the IV estimators discussed in the preceding sections used in place of $RM_{t-1,t}$, across our three different sampling frequencies (5 mins, 2.5 mins, 1.25 mins).

The regression results and correlation coefficients between daily return and volatility are given in Tables (3)-(5), with each table containing results for a given sampling frequency, and for our in-sample period of 1250 days. Inspection of these results indicates that regardless of sampling frequency or the representative stock used, the return-volatility relationship as depicted by estimated β coefficients is significantly negative¹⁷. On the other hand the correlation coefficient between daily returns and volatility is also significantly negative. However, the strength of causality does differ across different volatility estimators, with the jump-robust estimators like TRV and MedRV displaying much stronger causal relationships with daily returns, as opposed to some of the noise robust estimators like TSRV and MSRV. Also, RV, which is neither robust to noise or jumps, displays stronger causal linkages with daily returns than the noise robust volatility estimators. These findings, specifically in terms of the strong overall significant negative association between daily return & volatility, form the foundation of our empirical “volatility direction-of-change” based trading strategy which is now discussed.

7.3 Volatility Direction-of-Change Trading Strategy

The two components that form the basis for our “long-only” trading strategy are: (i) the negative contemporaneous association between daily return and estimated volatility and (ii) forecasts of direction-of-change of volatility. In essence, if a negative linkage can be expected to hold between daily R_t and estimated

¹⁵Using higher frequency five-minute absolute returns to construct a realized volatility proxy over longer horizons, Bollerslev et al. (2006) find the dual presence of a prolonged leverage effect at the intraday level, and an almost instantaneous volatility feedback effect

¹⁶Refer to Proposition 1, pp. 6 in Bollerslev and Zhou (2006) for details.

¹⁷We use a one-sided test with the alternative hypothesis that $\beta < 0$, when checking if our findings match those reported in Bollerslev and Zhou (2006)

IV_t , then forecasts of direction of change of volatility can be utilized to anticipate the direction of change of the market or the asset price movement. This would then motivate either a “buy” signal based on the predicted “upward” movement of the market or the “sell” signal based on the “downward” movement of the market. It should be noted here that a “long-only” strategy means that an investor is allowed to start each set of buy-sell transactions with a buy (long) position only.

Given that PD_{T+1} denotes the one-day ahead directional prediction of future volatility based on either interval forecasts (Section 3.2) or point forecasts (Section ??) and that PD_{T+1} takes a value “up” (down) if the forecasts suggest IV_{T+1} will be greater (lesser) than IV_T , then the following signals can be generated,

$$\text{Buy signal} = \begin{cases} 1, & PD_{T+1} = \text{down} \\ 0, & PD_{T+1} = \text{up} \end{cases}$$

and

$$\text{Sell signal} = \begin{cases} 1, & PD_{T+1} = \text{up} \\ 0, & PD_{T+1} = \text{down} \end{cases}$$

where buy signal represents a “long” position in the stock and sell signal suggests a “short” position in the stock. However a strategy based simply on such trading signals might lead to very frequent trading and hence large transaction costs. Building on Fuertes et al. (2015), we overlay a 5-day simple moving average (SMA) and a 5 day/20 day double crossing average (DCMA) stop-loss rule to the main strategy. The 5-day (weekly) SMA_t can be given as $(\frac{P_{t-5}^C + P_{t-4}^C + P_{t-3}^C + P_{t-2}^C + P_{t-1}^C}{5})$. The DCMA on the other hand combines a weekly and monthly trend; if the weekly SMA falls below the monthly SMA then a “sell” signal is generated. With the help of the SMA, the number of trades can be reduced while potential losses as a result of large price falls can be limited by the DCMA. Such short-term moving averages have been shown to enhance market timing strategies (see Lee et al. (2003)). The final “long-only” strategy that combines these moving averages and previously introduced “buy” and “sell” signals is:

$$\text{Buy stock at the beginning of day (T+1)} = \begin{cases} PD_{T+1} = \text{down} \\ P_{t+1}^O > SMA_{t+1}^{\text{weekly}} \\ SMA_{t+1}^{\text{weekly}} > SMA_{t+1}^{\text{monthly}} \end{cases}$$

and

$$\text{Sell stock at the beginning of day (T+S+1)} = \{ PD_{T+S+1} = \text{up}$$

Thus, the representative stock is held for S days (after it is bought on day T+1) and then sold on the $(T+S+1)^{st}$ day, when PD_{T+S+1} suggests an “upward” movement.

8 Empirical Setup

We use six different representative stocks and three sampling frequencies as described in Section 6 to carry out tests for “directional predictive accuracy” and “trading profitability”. This is done using interval forecasts as well as using our two point-forecasting methodologies (HAR and RNN). For estimation of volatility in each case, we use the ten different volatility estimators discussed in Section 5. In order to evaluate these different modeling approaches, we use the proceeding evaluation methods.

8.1 Directional Predictive Accuracy

The predictive performance of the interval and point forecasts is evaluated on the basis of how accurately the given forecasting procedure predicts *direction of change* of volatility. To that end, the goal is to compute the directional predictive accuracy (DPA) rate. The first step in this procedure is to determine the actual *direction*

of change which is given as

$$D_{T+1}^{actual} = \begin{cases} 1, & RM_{T+1}^{actual} \geq RM_T^{actual} \\ -1, & RM_{T+1}^{actual} < RM_T^{actual}, \end{cases}$$

where a value of 1 (-1) denotes actual volatility for period T+1 that is greater (lesser) than actual volatility of period T. In previous sections, both in the cases of intervals as well as point forecasts, we defined PD_{T+1} to be the predicted *direction of change*. The technique used to construct PD_{T+1} in the case of point forecasts differed from that used for interval forecasts. We now define a new *direction of change* metric, $D_{T+1}^{predicted}$, which can be applied to both interval and point forecasts. Namely, define

$$D_{T+1}^{predicted} = \begin{cases} 1, & PD_{T+1} = up \\ -1, & PD_{T+1} = down, \end{cases}$$

where PD_{T+1} is the predicted *direction of change* according to any given forecasting methodology. If $D_{T+1}^{predicted}$ has value of 1 (-1), it denotes an upward (downward) movement in volatility in period T+1, as compared to period T. We use a type of “confusion matrix” (see Swanson and White (1995)) to classify the predicted and actual directional signals. Namely, define

		D_{T+1}^{actual}	
		1	-1
$D_{T+1}^{predicted}$	1	N_1	N_2
	-1	N_3	N_4

where N_1 (N_4) denotes the number of correct directional predictions in terms of “upward” (downward) volatility movements. Both N_2 and N_3 denote incorrect directional predictions. Thus, we define the directional predictive accuracy rate as follows: $DPA = \frac{N_1 + N_4}{N}$, where $N = (N_1 + N_2 + N_3 + N_4)$. We also conduct the Pesaran and Timmerman (1995) independence tests to evaluate the ability of forecasting models to predict directional change. Under the null hypothesis of these tests, the actual and predicted values of volatility are independently distributed, implying that the forecasting model has no ability to predict the *direction of change*. The statistic for this test is

$$S_N = \frac{\widehat{P} - \widehat{P}_*}{\{\widehat{var}(\widehat{P}) - \widehat{var}(\widehat{P}_*)\}^{0.5}} \rightarrow N(0, 1) \quad (35)$$

where

$$\widehat{P}_* = \{P^A P^P - (1 - P^A)(1 - P^P)\}, \quad \widehat{P} = \frac{N_1 + N_4}{N}, \quad P^P = \frac{N_1 + N_2}{N}, \quad P^A = \frac{N_1 + N_3}{N} \quad (36)$$

and

$$\begin{aligned} \widehat{var}(\widehat{P}) &= \frac{1}{N} \widehat{P}_*(1 - \widehat{P}_*) \\ \widehat{var}(\widehat{P}_*) &= \frac{1}{N} (2\widehat{P} - 1)^2 \widehat{P}_*(1 - \widehat{P}_*) + \frac{1}{N} (2\widehat{P}_* - 1)^2 \widehat{P}(1 - \widehat{P}) + \frac{4}{N^2} \widehat{P}\widehat{P}_*(1 - \widehat{P})(1 - \widehat{P}_*). \end{aligned} \quad (37)$$

If the test statistic, S_N , is greater than the critical value (for a given significance level) of a standard normal random variate, then we have evidence that a model has predictive power for forecasting the *direction of change* of volatility.

8.2 Trading Profitability

Finally, in order to assess the economic value of the forecasts, we employ two metrics - the Sharpe Ratio and the Sortino Ratio. Trading strategies from Section 7.3 are deployed daily, and total percentage returns are calculated at the end of each month. These are in turn used to calculate the above ratios. In particular, the Sharpe Ratio is a measure of the risk-adjusted average rate of return, and can be given as

$$\text{Sharpe Ratio} = \text{ShR} = \frac{R}{SD}, \quad (38)$$

where R is the average monthly rate of return and SD is the standard deviation of the monthly return rate. The Sortino Ratio, on the other hand, is a modified version of the Sharpe Ratio, and penalizes only those returns which fall below a pre-specified target rate of return. Namely, define

$$\text{Sortino Ratio} = \text{SoR} = \frac{R}{DR}, \quad (39)$$

where

$$\text{Downside Risk} = DR = \sqrt{\frac{1}{N} \sum_i^N \min(0, r_i - r_f)^2}, \quad (40)$$

r_i is the monthly rate of return, and the pre-determined target rate r_f is the average 3 month T-bill rate. The higher the value of a Sharpe or Sortino Ratios, the better is the investment strategy.

9 Empirical Findings

In this section we provide our empirical findings based on both “directional predictive accuracy” and “trading profitability”. Additionally based on these findings, we provide our motivation for the follow-up series of Monte-Carlo experiments.

9.1 Directional Predictive Accuracy

Tables (6)-(11) contain *directional predictive accuracy* results for our interval and point forecasting methodologies, based on the use of ten IV estimators and three sampling frequencies, and for our 6 stocks. Refer to Section 8.1 for complete a detailed discussion of our empirical setup. The entries in the tables are DPA rates. In each table, for a given sampling frequency and IV estimator, if a point forecasting methodology (either HAR or RNN) has a lower predictive accuracy rate, relative to the interval method, then that entry is denoted in bold font. Additionally, below the DPA rates S_N test statistic values are given in parentheses. Interestingly, regardless of forecasting methodology, IV estimation technique, sampling frequency or stock, tests based on S_N always lead to rejection of the independence null hypothesis, indicating that all of our models have predictive power for the *direction of change* of volatility. Other findings based upon examination of these tables are as follows. First, HAR point-forecasts display DPA rates which are lower than those associated with interval forecasts, regardless of the sampling frequency or the IV estimator used. This holds true for all the stocks. In particular, predictive intervals outperform HAR forecasts in almost 97% of cases (out of total of 180 cases encompassing our three sampling frequencies, ten volatility estimators and 6 different stocks). Moreover, interval forecasts have 5% higher DPA rates than HAR models, on average.

Second, RNN point-forecasts deliver marginally higher DPA rates than interval forecasts. In particular, interval forecasts display higher DPA rates in only 25% of the cases. However, interval forecast based DPA

rates are only 0.8% lower than the RNN based DPA rates, on average. Interestingly, stocks for which interval forecasts outperform RNN forecasts the most are MSFT and PG. For these stocks the percentage contribution of “large jumps” to the total variation in the price process is higher than for the other stocks in our sample, as evidenced by the large jump variation (LJV) percentages reported in Table (2). This implies that predictive intervals are better at forecasting the direction of change of volatility when there are “larger” rather than “smaller jumps” in the price process.

Tables (6)-(11) also highlight how the different forecasting methodologies perform depending upon the choice of volatility estimation technique and data sampling frequency. To give a more comprehensive overview of our findings in this regard, refer to Tables (12)-(13). The “relative DPA rate” for interval versus point forecasts when averaging across different sampling frequencies and stocks is given in Table (12) while the “relative DPA rate” when averaging across different IV estimators and stocks is given in Table (13). Here, “relative DPA rate” is defined as the ratio of the DPA rates for point forecasts to interval forecasts. Thus, the relative DPA rate for predictive intervals will always be unity. The results collected in Tables (12)-(13) mirror our findings based on Tables (6)-(11). In particular, we find that for any given IV estimator (see Table 12) or sampling frequency (see Table 13), predictive intervals outperform HAR-point forecasts, but RNN-forecasts perform marginally better. Namely, HAR-point forecasts display relative DPA rates of less than 1 while RNN-point forecasts display values which are marginally greater than 1. The overall poor performance of the HAR model is suggestive of the possible existence of a non-linear relationship between estimated IV and lags thereof. Notice also that relative DPA rates for the HAR model in Table (13) drop as the sampling frequency increases.

9.2 Trading Profitability

Tables (14)-(19) contain profitability based results associated with deploying our trading strategies as described in Section 7.3, evaluated using the measures discussed in Section 8.2. We use the same stocks, IV estimators, and sampling frequencies as those used above. The entries in each table (which correspond to a given stock) denote either the monthly Sharpe Ratios (given in Panel A) or Sortino Ratios (given in Panel B) for our interval and point forecasting methodologies. Values in parentheses under Sharpe Ratios denote corresponding return standard deviations and those under Sortino Ratios denote downside risk measures. For a given IV estimator and sampling frequency, the entries in bold denote statistics that have Sharpe (Sortino) ratios that are lesser than those based on interval forecasts. For instance, in Table (14), focusing on the Sharpe Ratio for the HAR model with BPV and sampling frequency 5 minutes (i.e., column 2, row 3), notice that the value is 0.422, and is in bold font, as it is lesser than 0.436 (column 2, row 1), which is the corresponding Sharpe Ratio value based on interval forecasts.

Our key findings based on the results gathered in Tables (14)-(19) can be summarized as follows. First, interval forecasts outperform HAR-point forecasts in around 61% of the cases (out of a total of 180 cases), when considering Sharpe Ratios, and 64% of the cases when considering Sortino Ratios. Second, when compared with RNN-point forecasts, predictive intervals perform better in almost 51% of the cases, when considering Sharpe Ratios, and 61% of the cases when considering Sortino Ratios. Also, we find predictive intervals lead to 29% higher Sharpe Ratios when compared with HAR forecast based ratios and 30% higher Sharpe Ratios when compared with RNN forecast based ratios, on average.

Interval forecasts again dominate HAR-point forecasts, this time in terms of trading profitability (overall). However, the dominance of interval forecasts over HAR forecasts is lessened somewhat when one moves from comparing directional accuracy to comparing economic gains. On the other hand, the strong performance of RNN-point forecasts relative to interval forecasts that was noted in the previous section when comparing directional accuracy is lessened appreciably when comparing trading profitability, as interval forecasts produce

higher (investment based) risk adjusted returns, overall. One contributing factor for such lack of one-to-one mapping between forecast accuracy and profitability could be the choice of volatility estimation techniques at any given sampling frequency (see subsequent discussion based on our Monte Carlo experiments).

To illustrate our profitability results in a different light, Tables (20)-(21) contain “relative Sharpe (Sortino) Ratios” for interval and point forecasting methods. Here, predictive interval based ratios are normalized to have a value of 1, while ratios based point forecasts display are normalized relative to correspondings ratios based on interval forecasts. Table (20) reports the “relative Sharpe (Sortino) Ratio” for our different IV estimators (when averaging across different stocks and sampling frequencies) and Table (21) reports relative ratios for different sampling frequencies (when averaging across different stocks and IV estimators). From these tables, we see that predictive intervals outperform HAR forecasts for any given IV estimator or sampling frequency, with respect to both Sharpe and Sortino Ratios. It is also noteworthy that “relative Sharpe (Sortino) Ratios” are larger for noise-robust IV estimators like TSRV and MSRV while relative ratios are smaller for “truncation based” or “nearest neighbor based” jump-robust IV estimators like TRV, TBPV, and MedRV. Thus, the difference in terms of Sharpe (Sortino) Ratios between the predictive intervals and the HAR model is smaller for these noise-robust IV estimators while it is larger for the jump-robust IV estimators. Interestingly, these same jump-robust IV estimators also displayed stronger causal relationships with daily returns, as evidenced by higher values of the significant estimates of the $-\hat{\beta}$ feedback coefficient (refer to Tables (3)-(5)). This indeed leads to an interesting observation. When the return-volatility relationship is stronger (for given IV estimators), predictive intervals capitalize on their predictive superiority over the HAR model and create larger profit differences. On the other hand, when comparing with RNN forecast based ratios, predictive intervals generate higher (lower) Sharpe ratios but lower (lower) Sortino ratios, overall, for any given IV estimator (sampling frequency). Additionally, the profit difference between interval and RNN-forecasts is lower for Sharpe Ratios than for Sortino Ratios. It should be noted here that Sortino ratios penalize only those returns which fall below a given benchmark or negative returns. Thus, the reason why the RNN model performs much better in terms of the Sortino ratio is because it generates lesser negative returns either in terms of volume or magnitude. This is consistent with our previous observation that predictive intervals (when compared with RNN forecasts) predict the *direction of change* of volatility somewhat more effectively when there are “larger” rather than “smaller” jumps in the price process, thus generating more negative returns.

The empirical analysis given above illustrates a key finding. Namely, although interval forecasts dominate HAR-point forecasts in term of the DPA rate (directional predictive accuracy) in almost all cases (across different IV estimators, sampling frequencies and stocks), this dominance does not persist in all cases when forecasts are translated into economic gains¹⁸. Additionally we found that for noise-robust IV estimators like TSRV and MSRV (which display a weaker causal relationship with daily returns), the difference between interval and HAR-forecast performance in terms of Sharpe (Sortino) Ratios is much smaller than when we used jump-robust IV estimators like TRV and MedRV (which have a stronger causal relationship with daily returns). This suggests that for the translation of predictive improvements into economic gains, the choice of volatility estimation technique is crucial. Specifically, one might argue that if we could determine how well specific IV estimators predict the “true” underlying volatility density (and by extension price movements), we would have important evidence useful for explaining the trade-offs between said estimators when exploiting the return-volatility relationship in order to produce economic gains. Below, we carry out a Monte Carlo investigation to explore this issue.

¹⁸Interval forecasts performed better than HAR forecasts in 97% of cases when comparing directional predictive accuracy and 63% of the cases when comparing trading profitability.

10 Monte Carlo Experiments

The main objective in this experiment is to compare IV estimators on the basis of their interval forecast (conditional confidence interval estimation) accuracy. To this end, we construct IV estimator (or realized measure RM) based confidence intervals, $\widehat{CI}_{(T+1),M}(u_1, u_2 | RM_{T,M})$, and compare them with “pseudo true” IV confidence intervals. A crucial problem is that we do not have a closed form expression for a “pseudo true” IV confidence interval. Indeed, even knowledge of the data generating process of instantaneous volatility is not sufficient to yield the data generating process of integrated volatility. One way to solve this problem is to simulate the instantaneous volatility process, obtain the integrated volatility process from it, and then compute conditional confidence intervals based on simulated IV. This yields the following “pseudo true” IV confidence interval

$$CI_{T+1}(u_1, u_2 | \widehat{IV}_T) = \frac{\frac{1}{T\xi} \sum_{t=0}^{T-1} \mathbf{1}_{(u_1 \leq \widehat{IV}_{t+1} \leq u_2)} K\left(\frac{\widehat{IV}_t - \widehat{IV}_T}{\xi}\right)}{\frac{1}{T\xi} \sum_{t=0}^{T-1} K\left(\frac{\widehat{IV}_t - \widehat{IV}_T}{\xi}\right)}, \quad (41)$$

where \widehat{IV}_t refers to “simulated IV”. If the number of simulations is large enough, $CI_{T+1}(u_1, u_2 | \widehat{IV}_T)$ converges to $\widehat{CI}_{T+1}(u_1, u_2 | RM_{T,M})$, in probability. This leads to a modified version of the asymptotic normality condition given in (11) and to the following test statistic

$$G_{T,M}(u_1, u_2) = \frac{\sqrt{T\xi}(\widehat{CI}_{(T+1),M}(u_1, u_2 | RM_{T,M}) - CI_{T+1}(u_1, u_2 | \widehat{IV}_T))}{\widehat{V}^{1/2}(u_1, u_2)} \sim N(0, 1) \quad (42)$$

where \widehat{V} is the same as in (12). For further discussion, see Corradi et al. (2011).

In our experiments, for each IV estimator, we evaluate the finite sample properties of the test statistic $G_{T,M}(u_1, u_2)$. In particular, we compute the rejection frequency of the test statistic using two sided 10% and 5% nominal level critical values. By rejection frequency, we mean the number of times the test statistic achieves a value greater than the corresponding critical value, out of a total of 100 Monte Carlo iterations. The lower the rejection frequency for the given IV estimator, the better the predictive accuracy of the same estimator in terms of estimating the conditional confidence interval of “pseudo true” IV.

10.1 Data Generation

Our Monte Carlo experiments are based on 12 different data generating processes (DGPs) meant to cover a diverse spectrum of financial market conditions. These DGPs are different versions of the Heston (1993) stochastic volatility model with jumps in both price and volatility. The Heston model (HSV) is

$$\begin{aligned} dY_t &= (m - \sigma_t^2/2)dt + \sigma_t dW_{1t} + J_t^y dN_{1t} \\ d\sigma_t^2 &= \psi(\vartheta - \sigma_t^2)dt + \eta\sigma_t dW_{2t} + J_t^v dN_{2t}. \end{aligned} \quad (43)$$

Here W_{1t} and W_{2t} are two correlated Brownian motions with $\text{corr}(W_{1t}, W_{2t}) = \rho_1$. We consider $m = 0.05$, $\vartheta = 0.04$, $\eta = 0.5$, $\psi = 5$ and $\rho_1 = -0.5$, as in Aït-Sahalia and Mancini (2008). J_t^y refers to the magnitude of price jumps and J_t^v refers to the magnitude of volatility jumps. N_{1t} and N_{2t} are independent poisson processes with the same arrival rates, λ . We assume that jumps occur only once every day and that prices and volatilities jump at the same time. Absence of microstructure noise means $\epsilon_{t+j/M} = 0$, otherwise $\epsilon_{t+j/M} \sim N(0, 10^{-8})$. We consider two different jump sizes, one where the jump magnitude is $J_t^y \sim N(0, 0.02^2)$ and another where the jump magnitude is $J_t^y \sim N(0, 0.06^2)$. Volatility jumps have jump magnitude $J_t^v = z$, where $z = N(-5, 1)$. Table (22) outlines the details of the 12 different DGPs that we consider in our experiments.

In our experiments, data are simulated as follows.

Step 1. We treat one year as a unit of time, so one trading day is $t = 1/252$. N is the number of observations in one day. Using an Milstein scheme we simulate S paths for a DGP with a discrete time interval of $(1/252)N = \Delta_N$. This means there are N observations for both Y and σ^2 in each path. We set $N = 1560$. Once we have the simulated values for instantaneous volatility, σ_i^2 where $i = 1, \dots, N$ we construct integrated volatility (or simulated IV) corresponding to each path from $\sum_i^N \sigma_i^2 \Delta_N$. This simulated IV can then be used to construct infeasible conditional confidence intervals. The interval $[u_1, u_2]$ can be calculated as $[\hat{\mu}_{IV} - \beta\hat{\sigma}_{IV}, \hat{\mu}_{IV} + \beta\hat{\sigma}_{IV}]$ where $\beta = \{0.125, 0.250\}$. Here $\hat{\mu}_{IV}$ and $\hat{\sigma}_{IV}$ are the mean and standard deviation of integrated volatility over the S paths.

Step 2. Using the same simulated data for Y as above, we sample the daily data at a frequency of $1/M$ where M is much smaller than N . If we consider a DGP which does not have microstructure noise then simulated log price is equal to the observed log price, $Y_{t+j/M} = X_{t+j/M}$ (here $t = 1, \dots, T$ and $j = 1, \dots, M$) and microstructure noise $\epsilon_{t+j/M} = 0$. If microstructure noise is present in the DGP then $Y_{t+j/M} + \epsilon_{t+j/M}$ gives the observed log price $X_{t+j/M}$. This observed log price is then used to compute time series of length T days of different IV estimators.

We carry out 100 Monte Carlo iterations for each day, $T = 100$ days and $M = (78, 156, 312)$. This means that $S = 10000$. Finally, IV-estimator based confidence intervals are computed as in (10).

10.2 Results

Tables (33-34) report our results of our tests for conditional confidence interval estimation accuracy. In particular, entries in the table denote rejection frequencies of the test statistic $G_{T,M}(u_1, u_2)$ across ten IV estimators and 3 sampling frequencies, $M = (78, 156, 312)$, for 5% and 10% critical values. Each table contains results for a different data generating process. Entries in bold font denote the lowest rejection frequency achieved by a particular IV estimator, for a given column (i.e., for a given sampling frequency, critical value and interval $[u_1, u_2]$). Additionally, we report the level of price jumps in the data generating processes in Tables (36)-(38) using the Adjusted Jump Ratio test (AJR test, given in Section 6.1). Table (36) reports the percentage of days identified as having jumps, Table (37) reports the relative contribution of the jump component (jump variation) to the total variation and Table (38) reports variation due to large jumps (large jump variation) to the total variation in the price process (in percentage terms). Our main findings from the above tables can be summarized as below.

First, when data-generating processes do not have price jumps (DGPs 1, 2, 7, 8 or Tables (23), (24), (29), (30)), the estimators RV and TRV are the “best performing” measures with overall lowest rejection frequencies. It should be noted here that TRV is a price jump-robust estimator while RV is neither robust to price jumps nor to microstructure noise. This result can be explained with the help of our AJR test results in Tables (36-38). The price-jump free data generating processes were modeled without price jumps but the AJR test picked up on minor discontinuous price movements on almost 13% of the days. However the contribution of price jumps to the total variation in the price process is only 2%, with the contribution coming mostly from “small jumps” rather than large jumps. This indicates that TRV does not overestimate the level of price jumps in the data when there are “small” discontinuous price movements.

Second, we find that none of the noise-robust estimators perform well (in comparison to their jump-robust counterparts) when the DGP has microstructure noise but no price jumps (DGPs 2,8 or Tables 24, 30). In particular noise-robust estimators like TSRV, MSRV, and RK display higher rejection frequencies as compared with jump-robust estimators like MedRV and MinRV, and even when compared with RV. Interestingly, from Tables (36-38) we see that the percentage of jumps, jump variation, and large jump variation is lowered by the

introduction of microstructure noise to any given DGP. For instance, both DGPs 1 and 2 did not have price jumps during the modeling procedure but the AJR test reported some degree of jump activity in both DGPs, with the level of jump activity falling when noise was introduced. This reinforces our previous observation that some of the jump-robust estimators display superior predictive performance even when the DGPs contain “small” jumps.

Third, the “winning” IV estimator for any given DGP with price jumps differs depending upon price jump magnitude (frequency of price jumps, jump variation, or large jump variation) and the sampling frequency. In particular, our “ex ante-threshold based truncated” estimator, TRV, performs the “best” when the DGPs with price jumps have lower jump-magnitude (with 30% of the days having jumps) while the “nearest-neighbor based truncated” estimator, MedRV, is the “winning” estimator when price-jumps have higher magnitude (with around 70% of the days with price jumps). Additionally, MedRV performs better than TRV when the sampling frequency is lower (for eg. 5 minutes).

Finally, Table (35) gives a detailed list of which volatility estimator performs the best, given a particular sampling frequency and DGP. Results are based on averaging the rejection frequencies across different significance levels and intervals, for a particular DGP and sampling frequency. We find that out of the 36 cases listed (3 sampling frequencies for each of the 12 DGPs) TRV is the best performing measure in 23 (more than 63%) of case, including DGPs both with and without jumps. The second and third best are RV (wins in 11 cases, which are all DGPs without price jumps) and MedRV (wins in 7 cases, which are all DGPs with price jumps). As mentioned earlier, none of the “winning” estimators are noise-robust, even when the DGPs contain noise and very small amounts of jump-variation. Additionally, our non-noise, non-jump robust estimator, RV, performs better than some of its jump-robust counterparts, even when there are “small jumps” in the DGP. The ability of RV to more accurately predict the “pseudo true” volatility confidence interval, as compared with all noise robust and some of the jump robust estimators, explains why RV also generates higher Sharpe (Sortino) Ratios in many cases, as discussed in Section 9.2.

11 Concluding Remarks

In this paper, we evaluate the marginal predictive gains associated with interval (density) forecasts versus gains associated with point forecasts, when used to predict the direction-of-change of asset return volatility and when used in investment strategies. We utilize model-free Nadaraya Watson estimators to construct interval forecasts (predictive intervals), and both linear heterogeneous autoregressive (HAR) models and nonlinear recurrent neural network (RNN) models for point predictions. Additionally, a variety of different volatility estimators are used to estimate volatility. Our findings can be summarized as follows. First predictive intervals can be used to improve upon point-forecasts, in terms of directional predictive accuracy and trading profitability. Second, HAR point forecasts perform worse than interval forecasts and nonlinear-model based point forecasts (in terms of both predictive accuracy and risk adjusted returns). This points to the possible existence of a nonlinear relationship between volatility and lags thereof. Third, RNN point-forecasts lead to marginally better predictive accuracy, but lower overall mean-variance return ratios, than predictive intervals. Also, stocks for which interval forecasts outperform RNN forecasts contain a higher percentage of “large jump variation” than other stocks. Finally, Monte Carlo experiments indicate that noise-robust volatility estimators and some jump-robust estimators perform worse than realized volatility, even when the data generating process has noise and “small jumps”. This indicates over-estimation of noise by noise-robust estimators in the presence of “small jumps”. Moreover, the superior performance of RV in such scenarios explains why RV displays a strong causal linkage with daily returns and consequently generates higher risk-adjusted returns in some of the

scenarios examined in this paper. However, an “ex ante-threshold based truncated” estimator, TRV, performs the “best” when the DGPs with price jumps have lower jump-magnitude (with 30% of the days having jumps) while a “nearest-neighbor based truncated” estimator, MedRV, is the “winning” estimator when price-jumps have higher magnitude (with around 70% of the days with price jumps).

References

- AÏT-SAHALIA, Y. AND JACOD, J. 2009. Estimating the degree of activity of jumps in high frequency data. *Annals of Statistics* 37:2202–2244.
- 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* 50:1007–1050.
- AÏT-SAHALIA, Y. AND JACOD, J. 2014. High-frequency financial econometrics. *Princeton University Press: Princeton, NJ, USA*.
- AÏT-SAHALIA, Y. AND MANCINI, L. 2008. Out of sample forecasts of quadratic variation. *Journal of Econometrics* 147:17–33.
- ANDERSEN, T., BOLLERSLEV, T., DIEBOLD, F., AND LABYS, P. 2001. The distribution of realized exchange rate volatility. *Journal of the American statistical association* 96:42–55.
- ANDERSEN, T., DOBREV, D., AND SCHAUMBURG, E. 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169:75–93.
- BARNDORFF-NIELSEN, O., HANSEN, P., LUNDE, A., AND SHEPHARD, N. 2011. Subsampling realised kernels. *Journal of Econometrics* 160:204–219.
- BARNDORFF-NIELSEN, O., HANSEN, P., LUNDE, A., AND SHEPHARD, N. 2008. Designing realised kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76:1481–1536.
- BARNDORFF-NIELSEN, O. AND SHEPHARD, N. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics* 2:1–37.
- BLACK, F. 1976. Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association* pp. 171–181.
- BOLLERSLEV, T., LITINOVA, J., AND TAUCHEN, G. 2006. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics* 4:353–384.
- BOLLERSLEV, T., PATTON, A., AND QUAEDVLIEG, R. 2016. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics* 192:1–18.
- BOLLERSLEV, T. AND ZHOU, H. 2006. Volatility puzzles: a simple framework for gauging return-volatility regressions. *Journal of Econometrics* 131:123–150.
- CAO, C. AND TAY, S. 1992. Nonlinear time-series analysis of stock volatilities. *Journal of Applied Econometrics* 7:165–185.
- CHRISTOFFERSEN, P., DIEBOLD, F., MARIANO, R., TAY, A., AND TSE, Y. 2007. Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. *Journal of Financial Forecasting* 1:1–22.
- CHRISTOFFERSEN, P. AND DIEBOLD, F. 2000. How relevant is volatility forecasting for financial risk management? *The Review of Economics and Statistics* 82:12–22.
- CHRISTOFFERSEN, P. AND DIEBOLD, F. 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52:1273–1287.

- CORRADI, V., DISTASO, W., AND SWANSON, N. 2009. Predictive density estimators for daily volatility based on the use of realized measures. *Journal of Econometrics* 150:119–138.
- CORRADI, V., DISTASO, W., AND SWANSON, N. 2011. Predictive inference for integrated volatility. *Journal of the American Statistical Association* 6:1496–1512.
- CORSI, F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7:174–196.
- CORSI, F., PIRINO, D., AND RENO, R. 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159:276–288.
- DUONG, D. AND SWANSON, N. 2011. Volatility in discrete and continuous time models: A survey with new evidence on large and small jumps. *Missing Data Methods: Advances in Econometrics* 27:179–233.
- DUONG, D. AND SWANSON, N. 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics* 187:606–621.
- FENG, G., HE, J., AND POLSON, N. 2018. Deep learning for predicting asset returns. *arXiv:1804.09314*.
- FLEMING, J., KIRBY, C., AND OSTIEK, B. 2003. The economic value of volatility timing using realised volatility. *Journal of Financial Economics* 67:473–409.
- FRENCH, K., SCHWERT, G., AND STAMBAUGH, R. 1987. Expected stock returns and volatility. *Journal of Financial Economics* p. 3–29.
- FUERTES, A., KALOTYCHOU, E., AND TODOROVIC, N. 2015. Daily volume, intraday and overnight returns for volatility prediction: profitability or accuracy? *Review of Quantitative Finance and Accounting* 45:251–278.
- GALLANT, A. AND WHITE, H. 1992. On learning the derivatives of an unknown mapping with multilayer feedforward networks. *Neural Networks* 5:129–138.
- GRANGER, C. AND PESARAN, H. 2000. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 19:537–560.
- GU, S., KELLY, B., AND XIU, D. 2019. Empirical asset pricing via machine learning. *Chicago Booth Research Paper* 18-04.
- HEATON, J., POLSON., N., AND WITTE, J. 2016. Deep learning in finance. *ArXiv:1602.06561*.
- HUANG, X. AND TAUCHEN, G. 2005. The relative contribution of jumps to total price variance. *Journal of financial econometrics* 3:456–499.
- JACOD, J. 2018. Limit of random measures associated with the increments of a brownian semimartingale. *Journal of Financial Econometrics* 16:526–569.
- KONSTANTINIDI, E., SKIADOPoulos, G., AND TZAKKARAKI, E. 2008. Can the evolution of implied volatility be forecasted? evidence from european and us implied volatility indices. *Journal of Bank Finance* 32:2401–2411.
- LE, Q. ET AL. 2015. A tutorial on deep learning part 1: Nonlinear classifiers and the backpropagation algorithm.

- LEE, D., CHAN, H., FAFF, R., AND KALEV, P. 2003. Short-term contrarian investing-it is profitable-yes and no. *Journal of Multinational Financial Management* 13:385–404.
- LEITCH, G. AND TANNER, J. 1991. Economic forecast evaluation: Profits versus the conventional error measures. 81:580–590.
- LIU, L., PATTON, A., AND SHEPPARD, K. 2015. Does anything beat 5-minute rv? A comparison of realized measures across multiple asset classes. *Journal of Econometrics* 187:293–311.
- MUKHERJEE, A., PENG, W., SWANSON, N., AND YANG, X. 2020. Financial econometrics and big data: A survey of volatility estimators and tests for the presence of jumps and co-jumps. *Handbook of Statistics* 42:3–59.
- PATTON, A. AND SHEPPARD, K. 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics* 97:683–697.
- PESARAN, M. AND TIMMERMANN, A. 1995. A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* 10:561–565.
- SATCHELL, S. AND TIMMERMANN, A. 1995. An assessment of the economic value of non-linear foreign exchange rate forecasts. *Journal of Forecasting* 14:477–497.
- SWANSON, N. R. AND WHITE, H. 1995. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics* 13:265–275.
- TAY, A. AND WALLIS, K. 2000. Density forecasting: A survey. *Journal of Forecasting* 19:235–254.
- ZHANG, L. ET AL. 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12:1019–1043.
- ZHANG, L., MYKLAND, P. A., AND AÏT-SAHLIA, Y. 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100:1394–1411.

Table 1: Estimators of Integrated Volatility

1. Realized Volatility (RV)	Robust to neither price jumps nor microstructure noise
2. Realized Bipower Variation (BPV)	Robust to price jumps
3. Tripower Variation (TPV)	Robust to price jumps
4. Two Scale Realized Volatility (TSRV)	Robust to microstructure noise
5. Multi Scale Realized Volatility (MSRV)	Robust to microstructure noise
6. Realized Kernel (RK)	Robust to microstructure noise
7. Truncated Realized Variance (TRV)	Robust to price jumps
8. Threshold Bipower Variation (TBPV)	Robust to price jumps
9. MedRV	Robust to both price jumps and microstructure noise
10. MinRV	Robust to both price jumps and microstructure noise

*Notes: This table summarizes the different realized measures of integrated volatility used in this study.

Table 2: Price Jump Activity

Panel A. Sampling frequency 5 minutes

	BA	GE	IBM	JPM	MSFT	PG
<i>In-sample period, 2010-2014</i>						
Jump Days to Total Days (PJD)	29.36	33.44	31.28	27.84	35.60	35.36
Jump Variation to Quadratic Variation (JV)	9.59	11.00	10.32	8.72	12.00	11.90
Large Jump Variation to Quadratic Variation (LJV)	0.98	0.89	1.16	0.68	1.58	1.46
<i>Out-of-sample period, 2015-2017</i>						
Jump Days to Total Days (PJD)	29.88	35.72	30.15	28.95	35.06	34.13
Jump Variation to Quadratic Variation (JV)	10.21	11.84	10.12	9.33	11.53	12.02
Large Jump Variation to Quadratic Variation (LJV)	0.85	0.85	1.12	1.10	1.38	1.45
<i>Entire Sample, 2010-2017</i>						
Jump Days to Total Days (PJD)	29.56	34.30	30.85	28.26	35.40	34.90
Jump Variation to Quadratic Variation (JV)	9.82	11.31	10.25	8.95	11.82	11.95
Large Jump Variation to Quadratic Variation (LJV)	1.01	0.89	1.10	0.80	1.34	1.42

Panel B. Sampling frequency: 2.5 minutes

	BA	GE	IBM	JPM	MSFT	PG
<i>In-sample period, 2010-2014</i>						
Jump Days to Total Days (PJD)	36.08	49.04	35.76	34.00	44.72	40.00
Jump Variation to Quadratic Variation (JV)	9.63	13.14	9.15	8.45	12.00	11.29
Large Jump Variation to Quadratic Variation (LJV)	0.81	0.77	0.91	0.66	0.89	1.31
<i>Out-of-sample period, 2015-2017</i>						
Jump Days to Total Days (PJD)	32.01	45.68	34.93	31.74	40.72	37.05
Jump Variation to Quadratic Variation (JV)	8.75	12.37	9.74	7.79	10.62	10.77
Large Jump Variation to Quadratic Variation (LJV)	0.84	1.00	1.14	0.54	0.79	1.37
<i>Entire Sample, 2010-2017</i>						
Jump Days to Total Days (PJD)	34.55	47.78	35.45	33.15	43.24	38.89
Jump Variation to Quadratic Variation (JV)	9.30	12.85	9.37	8.20	11.48	11.10
Large Jump Variation to Quadratic Variation (LJV)	0.87	0.82	1.04	0.72	0.89	1.32

Panel C. Sampling frequency, 1.25 minutes

	BA	GE	IBM	JPM	MSFT	PG
<i>In-sample period, 2010-2014</i>						
Jump Days to Total Days (PJD)	46.48	71.2	40	41.92	66.16	60.48
Jump Variation to Quadratic Variation (JV)	10.72	16.56	8.52	8.84	15.62	14.71
Large Jump Variation to Quadratic Variation (LJV)	0.77	0.81	0.91	0.83	1.15	1.25
<i>Out-of-sample period, 2015-2017</i>						
Jump Days to Total Days (PJD)	44.83	66.67	43.03	43.69	58.43	55.38
Jump Variation to Quadratic Variation (JV)	10.51	16.42	9.79	9.28	13.55	13.67
Large Jump Variation to Quadratic Variation (LJV)	1.02	0.9	1.23	1.04	1.15	1.4
<i>Entire Sample, 2010-2017</i>						
Jump Days to Total Days (PJD)	45.83	69.5	41.14	42.59	63.26	58.56
Jump Variation to Quadratic Variation (JV)	10.63	16.51	9	9	14.84	14.32
Large Jump Variation to Quadratic Variation (LJV)	0.89	0.89	1.03	0.93	1.05	1.44

*Notes: Each panel (A-C) of this table displays the level of price jump activity in terms of percentage of jump days (PJD), contribution of jump variation (JV) and contribution of large jump variation (LJV). This is done across six representative stocks for our: (i) in-sample period, (ii) out-of-sample period, (iii) entire dataset. There are three panels for the three sampling frequencies. For more details on PJD, JV, and LJV refer to section Section (6.1).

Table 3: Volatility Feedback Effect based Return-Volatility Regressions (Sampling frequency 5 minutes)

	Volatility Estimators									
	RV	BPV	TPV	TSRV	MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>BA</i>										
<i>Feedback</i> $\hat{\beta}$	-11.36*** (-3.88)	-10.58*** (-4.00)	-10.89*** (-4.19)	-0.86*** (-6.09)	-2.47*** (-5.67)	-7.56*** (-3.16)	-10.62*** (-3.59)	-11.24*** (-2.97)	-10.69*** (-3.73)	-9.47*** (-3.87)
<i>Regression R</i> ²	0.025	0.023	0.024	0.023	0.024	0.019	0.022	0.016	0.023	0.022
$\rho(R_t, \widehat{IV}_t)$	-0.15***	-0.15***	-0.16***	-0.15***	-0.15***	-0.14***	-0.15***	-0.13***	-0.15***	-0.15***
<i>PG</i>										
<i>Feedback</i> $\hat{\beta}$	-6.01*** (-4.33)	-5.07*** (-4.40)	-5.32*** (-3.99)	-0.18 (-0.27)	-1.24*** (-2.41)	-3.94** (-2.00)	-6.28*** (-4.52)	-6.59*** (-2.38)	-5.31*** (-3.95)	-4.35*** (-4.60)
<i>Regression R</i> ²	0.018	0.015	0.014	0.001	0.009	0.009	0.019	0.014	0.014	0.013
$\rho(R_t, \widehat{IV}_t)$	-0.13***	-0.12***	-0.12***	-0.023	-0.10***	-0.10***	-0.14***	-0.12***	-0.12***	-0.12***
<i>IBM</i>										
<i>Feedback</i> $\hat{\beta}$	-5.59** (-1.79)	-7.62** (-2.00)	-8.33** (-2.25)	-0.34 (-0.61)	-0.88 (-0.99)	-1.73 (-0.46)	-7.01** (-1.88)	-8.48** (-2.05)	-7.45** (-2.10)	-7.20** (-1.95)
<i>Regression R</i> ²	0.004	0.006	0.007	0.001	0.001	0.000	0.005	0.006	0.006	0.006
$\rho(R_t, \widehat{IV}_t)$	-0.06***	-0.07***	-0.08***	-0.03	-0.04	-0.02	-0.07***	-0.08***	-0.08***	-0.07***
<i>JPM</i>										
<i>Feedback</i> $\hat{\beta}$	-5.90** (-2.10)	-5.85** (-2.02)	-6.12** (-2.02)	-0.79* (-1.51)	-1.18 (-1.02)	-1.12 (-0.31)	-5.70** (-1.86)	-5.72** (-1.78)	-5.11** (-1.65)	-4.89* (-1.62)
<i>Regression R</i> ²	0.010	0.009	0.009	0.007	0.004	0.000	0.009	0.007	0.007	0.007
$\rho(R_t, \widehat{IV}_t)$	-0.10***	-0.10***	-0.10***	-0.08***	-0.06**	-0.02	-0.10***	-0.09***	-0.08***	-0.08***
<i>MSFT</i>										
<i>Feedback</i> $\hat{\beta}$	-9.04** (-2.14)	-9.76** (-2.10)	-12.31*** (-2.65)	-1.33** (-1.90)	-2.81** (-2.24)	-7.67** (-1.67)	-11.26*** (-2.52)	-13.31*** (-2.57)	-7.72* (-1.62)	-7.19* (-1.58)
<i>Regression R</i> ²	0.013	0.011	0.016	0.012	0.015	0.010	0.018	0.016	0.008	0.007
$\rho(R_t, \widehat{IV}_t)$	-0.11***	-0.11***	-0.13***	-0.11***	-0.12***	-0.10***	-0.13***	-0.12***	-0.08***	-0.08***
<i>PG</i>										
<i>Feedback</i> $\hat{\beta}$	-1.31*** (-7.03)	-1.22*** (-7.28)	-1.56*** (-5.54)	-0.54* (-1.34)	-0.85*** (-4.72)	-3.36*** (-3.15)	-3.46*** (-4.09)	-9.76** (-1.91)	-1.38*** (-7.06)	-1.22*** (-6.95)
<i>Regression R</i> ²	0.005	0.005	0.006	0.003	0.005	0.005	0.006	0.006	0.005	0.005
$\rho(R_t, \widehat{IV}_t)$	-0.07***	-0.07***	-0.08***	-0.05***	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***

*Notes: This table contains results of the Volatility Feedback Effect regression equation (refer to Section 7.2). The regression coefficient ($\hat{\beta}$) is the feedback estimate, and is given along with t-statistics (in parenthesis). Regression R^2 values are also reported, as well as correlation coefficients between daily R_t and estimated IV_t . Results are given for ten volatility estimators and six representative stocks. ***, **, * denote significance at 1, 5 and 10% levels. The sampling frequency is 5 minutes.

Table 4: Volatility Feedback Effect based Return-Volatility Regressions (Sampling frequency 2.5 minutes)

	Volatility Estimators									
	RV	BPV	TPV	TSRV	MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>BA</i>										
<i>Feedback</i> $\hat{\beta}$	-10.44*** (-4.20)	-10.21*** (-4.14)	-9.99*** (-3.76)	-0.45*** (-5.68)	-2.71*** (-4.78)	-7.86*** (-3.32)	-11.43*** (-3.80)	-11.52*** (-3.19)	-8.89*** (-3.45)	-9.11*** (-3.94)
<i>Regression R</i> ²	0.026	0.023	0.021	0.024	0.026	0.017	0.024	0.020	0.018	0.021
$\rho(R_t, \widehat{IV}_t)$	-0.16***	-0.15***	-0.14***	-0.15***	-0.16***	-0.13***	-0.15***	-0.14***	-0.13***	-0.14***
<i>PG</i>										
<i>Feedback</i> $\hat{\beta}$	-4.74*** (-6.85)	-4.56*** (-6.65)	-4.37*** (-6.52)	-0.12 (-0.36)	-1.16*** (-4.28)	-3.52*** (-5.79)	-5.92*** (-3.74)	-7.19*** (-2.36)	-4.80*** (-5.94)	-4.57*** (-6.17)
<i>Regression R</i> ²	0.020	0.018	0.016	0.001	0.014	0.014	0.018	0.015	0.018	0.017
$\rho(R_t, \widehat{IV}_t)$	-0.14***	-0.13***	-0.12***	-0.03	-0.12***	-0.12***	-0.13***	-0.12***	-0.13***	-0.13***
<i>IBM</i>										
<i>Feedback</i> $\hat{\beta}$	-5.38*** (-2.89)	-5.47*** (-3.42)	-5.87*** (-3.27)	-0.25 (-0.89)	-1.34*** (-2.45)	-3.78** (-1.83)	-7.14** (-2.28)	-8.11** (-1.87)	-5.14*** (-3.77)	-4.39*** (-3.88)
<i>Regression R</i> ²	0.007	0.007	0.008	0.001	0.005	0.003	0.007	0.006	0.008	0.008
$\rho(R_t, \widehat{IV}_t)$	-0.08***	-0.09***	-0.08***	-0.04	-0.07***	-0.05***	-0.08***	-0.07***	-0.09***	-0.09***
<i>JPM</i>										
<i>Feedback</i> $\hat{\beta}$	-5.08* (-1.60)	-6.03** (-1.99)	-5.86** (-1.80)	-0.40* (-1.45)	-1.40* (-1.54)	-2.33 (-0.63)	-6.19** (-2.00)	-5.93** (-1.84)	-5.48** (-1.80)	-6.17** (-2.17)
<i>Regression R</i> ²	0.007	0.009	0.008	0.007	0.007	0.001	0.010	0.008	0.008	0.010
$\rho(R_t, \widehat{IV}_t)$	-0.08***	-0.09***	-0.09***	-0.08***	-0.08***	-0.04	-0.10***	-0.08***	-0.08***	-0.10***
<i>MSFT</i>										
<i>Feedback</i> $\hat{\beta}$	-10.64*** (-2.64)	-12.27*** (-3.07)	-13.94*** (-3.16)	-0.71** (-1.97)	-2.87*** (-2.58)	-9.43*** (-2.14)	-13.01*** (-3.06)	-13.55*** (-3.07)	-9.48** (-2.14)	-10.57*** (-2.63)
<i>Regression R</i> ²	0.019	0.024	0.028	0.013	0.018	0.015	0.025	0.025	0.015	0.021
$\rho(R_t, \widehat{IV}_t)$	-0.14***	-0.16***	-0.17***	-0.12***	-0.13***	-0.12***	-0.16***	-0.16***	-0.12***	-0.17***
<i>PG</i>										
<i>Feedback</i> $\hat{\beta}$	-0.24*** (-35.46)	-0.22*** (-39.55)	-0.30*** (-29.82)	-0.18*** (-4.49)	-0.11*** (-29.52)	-0.46*** (-24.42)	-11.18*** (-2.61)	-13.17*** (-2.78)	-0.23*** (-39.10)	-0.20*** (-45.74)
<i>Regression</i>	0.006	0.006	0.006	0.005	0.006	0.006	0.010	0.011	0.006	0.006
$\rho(R_t, \widehat{IV}_t)$	-0.07***	-0.07***	-0.08***	-0.07***	-0.07***	-0.07***	-0.09***	-0.10***	-0.07***	-0.07***

*Notes: See notes to Table 3.

Table 5: Volatility Feedback Effect based Return-Volatility Regressions (Sampling frequency 1.25 minutes)

	Volatility Estimators									
	RV	BPV	TPV	TSRV	MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>BA</i>										
<i>Feedback</i> $\hat{\beta}$	-10.94*** (-4.37)	-11.51*** (-4.44)	-12.20*** (-4.66)	-0.22*** (-5.34)	-2.54*** (-4.76)	-10.19*** (-4.17)	-11.82*** (-3.64)	-11.91*** (-3.23)	-10.88*** (-3.80)	-11.01*** (-3.94)
<i>Regression R</i> ²	0.026	0.026	0.027	0.024	0.026	0.025	0.023	0.020	0.023	0.023
$\rho(R_t, \widehat{IV}_t)$	-0.16***	-0.16***	-0.16***	-0.15***	-0.16***	-0.15***	-0.15***	-0.14***	-0.15***	-0.15***
<i>PG</i>										
<i>Feedback</i> $\hat{\beta}$	-5.69*** (-6.09)	-5.65*** (-6.65)	-6.24*** (-5.74)	-0.08 (-0.47)	-1.13*** (-7.53)	-4.60*** (-6.63)	-6.42*** (-5.20)	-6.52*** (-4.04)	-4.44*** (-8.56)	-4.96*** (-8.11)
<i>Regression R</i> ²	0.022	0.021	0.021	0.002	0.020	0.020	0.022	0.019	0.020	0.021
$\rho(R_t, \widehat{IV}_t)$	-0.14***	-0.14***	-0.14***	-0.04	-0.14***	-0.14***	-0.15***	-0.13***	-0.14***	-0.14***
<i>IBM</i>										
<i>Feedback</i> $\hat{\beta}$	-5.96*** (-3.05)	-5.93*** (-3.09)	-6.48*** (-2.60)	-0.13* (-1.01)	-1.15*** (-3.47)	-4.90*** (-3.14)	-7.68*** (-2.58)	-8.37*** (-2.87)	-5.69*** (-3.44)	-5.68*** (-3.31)
<i>Regression R</i> ²	0.008	0.007	0.007	0.002	0.006	0.006	0.008	0.009	0.007	0.007
$\rho(R_t, \widehat{IV}_t)$	-0.08***	-0.08***	-0.08***	-0.04***	-0.07***	-0.07***	-0.09***	-0.09***	-0.08***	-0.08***
<i>JPM</i>										
<i>Feedback</i> $\hat{\beta}$	-6.03** (-1.93)	-6.05** (-1.80)	-7.12** (-1.95)	-0.21* (-1.54)	-1.49** (-1.89)	-4.49 (-1.20)	-6.57** (-1.94)	-6.31** (-1.68)	-5.85** (-1.94)	-5.59** (-1.72)
<i>Regression R</i> ²	0.010	0.008	0.010	0.007	0.009	0.006	0.010	0.008	0.008	0.007
$\rho(R_t, \widehat{IV}_t)$	-0.09***	-0.09***	-0.10***	-0.08***	-0.09***	-0.07***	-0.10***	-0.08***	-0.08***	-0.08***
<i>MSFT</i>										
<i>Feedback</i> $\hat{\beta}$	-9.90*** (-2.48)	-11.57*** (-2.71)	-14.10*** (-2.80)	-0.369** (-2.03)	-2.95*** (-2.91)	-9.93*** (-2.67)	-13.96*** (-2.80)	-14.58*** (-2.61)	-8.86** (-2.17)	-8.93*** (-2.26)
<i>Regression R</i> ²	0.016	0.019	0.023	0.014	0.020	0.017	0.025	0.022	0.013	0.013
$\rho(R_t, \widehat{IV}_t)$	-0.12***	-0.13***	-0.15***	-0.11***	-0.14***	-0.13***	-0.15***	-0.14***	-0.11***	-0.11***
<i>PG</i>										
<i>Feedback</i> $\hat{\beta}$	-0.34*** (-29.30)	-0.33*** (-30.83)	-0.40*** (-23.31)	-0.12*** (-3.30)	-0.10*** (-25.65)	-1.77*** (-7.20)	-12.66*** (-3.87)	-7.12*** (-3.89)	-0.31*** (-31.17)	-0.33*** (-29.16)
<i>Regression R</i> ²	0.006	0.006	0.007	0.005	0.006	0.007	0.014	0.011	0.006	0.006
$\rho(R_t, \widehat{IV}_t)$	-0.08***	-0.08***	-0.08***	-0.07***	-0.08***	-0.08***	-0.12***	-0.10***	-0.07***	-0.07***

*Notes: See notes to Table 3.

Table 6: Directional Predictive Accuracy (Stock BA)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators					
					MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>5 minute frequency</i>										
Interval (density)	0.680 (9.82)	0.664 (8.97)	0.681 (9.93)	0.704 (11.22)	0.685 (10.18)	0.709 (11.48)	0.664 (8.95)	0.668 (9.22)	0.640 (7.70)	0.691 (10.45)
HAR (point)	0.632 (9.30)	0.633 (9.42)	0.645 (9.79)	0.651 (9.55)	0.645 (9.02)	0.652 (9.88)	0.624 (8.44)	0.652 (9.59)	0.635 (8.74)	0.643 (9.83)
RNN (point)	0.692 (10.82)	0.684 (10.33)	0.681 (10.07)	0.710 (11.57)	0.701 (11.09)	0.713 (11.78)	0.669 (9.52)	0.688 (10.38)	0.677 (9.82)	0.692 (10.62)
<i>2.5 minute frequency</i>										
Interval (density)	0.668 (7.78)	0.642 (7.49)	0.640 (6.97)	0.692 (9.84)	0.676 (7.63)	0.681 (9.56)	0.639 (7.39)	0.656 (7.37)	0.629 (7.62)	0.649 (7.18)
HAR (point)	0.625 (7.77)	0.612 (7.14)	0.617 (7.25)	0.641 (9.31)	0.653 (8.35)	0.669 (10.07)	0.618 (8.90)	0.622 (8.00)	0.601 (6.76)	0.607 (7.24)
RNN (point)	0.673 (9.79)	0.653 (8.83)	0.657 (8.74)	0.709 (11.50)	0.693 (10.81)	0.693 (10.70)	0.656 (8.81)	0.666 (9.39)	0.657 (8.92)	0.656 (8.93)
<i>1.25 minute frequency</i>										
Interval (density)	0.657 (8.54)	0.637 (7.47)	0.636 (7.44)	0.700 (11.01)	0.661 (8.81)	0.707 (11.34)	0.635 (7.39)	0.637 (7.42)	0.632 (7.22)	0.624 (6.77)
HAR (point)	0.595 (7.70)	0.600 (7.54)	0.614 (7.55)	0.643 (9.09)	0.636 (9.89)	0.667 (10.73)	0.604 (7.31)	0.607 (7.47)	0.583 (7.05)	0.598 (7.29)
RNN (point)	0.649 (8.38)	0.648 (8.47)	0.641 (7.87)	0.708 (11.44)	0.673 (9.99)	0.705 (11.47)	0.635 (7.73)	0.649 (8.35)	0.627 (7.40)	0.633 (7.55)

*Notes: Entries are directional predictive accuracy rates (DPRs) of interval and point (HAR model and RNN model) forecasts constructed using ten different volatility estimators and 3 different sampling frequencies. Entries in bold font denotes cases where either the HAR and/or RNN forecasting method is outperformed by using interval forecasts. For all DPA rates, the PT discussed in Section 8 was carried out. Values in parentheses denote PT-test statistics. All of the statistics are sufficiently large in magnitude to indicate rejection of the independence null hypothesis at a 95% level of confidence.

Table 7: Directional Predictive Accuracy (Stock GE)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators					
					MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>5 minute frequency</i>										
Interval (density)	0.657 (8.65)	0.657 (8.61)	0.654 (8.50)	0.672 (9.40)	0.651 (8.23)	0.691 (10.51)	0.661 (8.83)	0.671 (9.45)	0.641 (7.79)	0.659 (8.69)
HAR (point)	0.629 (8.49)	0.624 (8.42)	0.629 (8.55)	0.653 (9.89)	0.623 (8.13)	0.687 (11.30)	0.637 (9.03)	0.636 (8.84)	0.637 (8.90)	0.628 (8.81)
RNN (point)	0.660 (8.89)	0.651 (8.49)	0.645 (8.21)	0.705 (11.40)	0.669 (9.35)	0.696 (10.85)	0.668 (9.30)	0.683 (10.12)	0.652 (8.55)	0.667 (9.41)
<i>2.5 minute frequency</i>										
Interval (density)	0.641 (7.78)	0.637 (7.50)	0.628 (6.98)	0.680 (9.84)	0.641 (7.63)	0.673 (9.56)	0.619 (6.57)	0.635 (7.37)	0.640 (7.63)	0.631 (7.19)
HAR (point)	0.608 (7.77)	0.594 (7.15)	0.599 (7.26)	0.645 (9.31)	0.618 (8.35)	0.660 (10.07)	0.644 (8.90)	0.623 (8.00)	0.590 (6.77)	0.603 (7.25)
RNN (point)	0.652 (8.40)	0.641 (8.04)	0.632 (7.59)	0.693 (10.75)	0.669 (9.31)	0.689 (10.52)	0.668 (9.29)	0.652 (8.56)	0.664 (9.33)	0.660 (8.92)
<i>1.25 minute frequency</i>										
Interval (density)	0.633 (7.31)	0.640 (7.68)	0.640 (7.70)	0.676 (9.63)	0.645 (7.96)	0.644 (7.89)	0.625 (6.84)	0.624 (6.84)	0.628 (7.01)	0.614 (6.23)
HAR (point)	0.598 (6.95)	0.595 (6.88)	0.608 (7.60)	0.653 (9.72)	0.608 (7.60)	0.608 (7.37)	0.618 (7.88)	0.616 (7.77)	0.578 (5.82)	0.583 (6.18)
RNN (point)	0.648 (8.36)	0.645 (8.20)	0.653 (8.83)	0.695 (10.81)	0.668 (9.27)	0.649 (8.27)	0.651 (8.52)	0.653 (8.50)	0.627 (7.41)	0.622 (6.89)

*Notes: See notes of Table 6.

Table 8: Directional Predictive Accuracy (Stock IBM)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators					
					MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>5 minute frequency</i>										
Interval (density)	0.688 (10.32)	0.668 (9.20)	0.666 (9.13)	0.675 (9.62)	0.652 (8.38)	0.709 (11.48)	0.671 (9.38)	0.667 (9.09)	0.680 (9.85)	0.651 (8.25)
HAR (point)	0.652 (9.46)	0.644 (9.15)	0.646 (9.12)	0.693 (11.24)	0.667 (9.72)	0.685 (11.16)	0.657 (9.76)	0.657 (9.96)	0.652 (9.83)	0.649 (9.53)
RNN (point)	0.688 (10.34)	0.692 (10.57)	0.674 (9.59)	0.718 (11.99)	0.676 (9.66)	0.707 (11.39)	0.697 (10.85)	0.687 (10.20)	0.684 (10.19)	0.665 (9.10)
<i>2.5 minute frequency</i>										
Interval (density)	0.676 (9.64)	0.679 (9.82)	0.675 (9.57)	0.679 (9.82)	0.643 (7.82)	0.675 (9.60)	0.668 (9.20)	0.680 (9.90)	0.660 (8.78)	0.672 (9.45)
HAR (point)	0.622 (8.30)	0.628 (8.50)	0.643 (8.90)	0.691 (11.26)	0.645 (9.09)	0.661 (9.56)	0.633 (8.50)	0.651 (9.09)	0.623 (8.06)	0.632 (8.13)
RNN (point)	0.671 (9.65)	0.668 (9.28)	0.681 (9.94)	0.714 (11.78)	0.684 (10.13)	0.693 (10.61)	0.681 (10.01)	0.673 (9.52)	0.659 (8.84)	0.677 (9.74)
<i>1.25 minute frequency</i>										
Interval (density)	0.633 (7.31)	0.652 (8.34)	0.640 (7.67)	0.679 (9.83)	0.644 (7.91)	0.668 (9.20)	0.636 (7.42)	0.639 (7.58)	0.652 (8.35)	0.651 (8.27)
Point (HAR)	0.606 (6.97)	0.615 (7.48)	0.623 (7.62)	0.688 (11.02)	0.628 (7.83)	0.660 (10.00)	0.602 (6.56)	0.635 (8.45)	0.607 (6.92)	0.602 (6.75)
RNN (point)	0.643 (7.95)	0.660 (8.90)	0.679 (9.83)	0.720 (12.07)	0.661 (8.87)	0.681 (9.99)	0.657 (8.65)	0.676 (9.66)	0.648 (8.16)	0.675 (9.63)

*Notes: See notes of Table 6.

Table 9: Directional Predictive Accuracy (Stock JPM)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators					
					MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>5 minute frequency</i>										
Interval (density)	0.651 (8.27)	0.669 (9.33)	0.661 (8.84)	0.683 (10.05)	0.653 (8.45)	0.701 (11.04)	0.648 (8.15)	0.657 (8.69)	0.680 (9.88)	0.693 (10.66)
HAR (point)	0.608 (7.29)	0.637 (8.80)	0.626 (8.62)	0.659 (9.82)	0.629 (8.18)	0.644 (9.34)	0.622 (7.70)	0.647 (9.03)	0.633 (8.68)	0.644 (8.90)
RNN (point)	0.669 (9.57)	0.704 (11.23)	0.707 (11.61)	0.681 (10.18)	0.676 (9.79)	0.708 (11.54)	0.679 (9.88)	0.691 (10.50)	0.695 (10.91)	0.709 (11.49)
<i>2.5 minute frequency</i>										
Interval (density)	0.675 (9.53)	0.664 (8.99)	0.665 (9.05)	0.692 (10.55)	0.664 (8.96)	0.696 (10.72)	0.659 (8.69)	0.655 (8.51)	0.637 (7.55)	0.661 (8.86)
HAR (point)	0.596 (7.87)	0.619 (8.71)	0.623 (8.86)	0.663 (10.18)	0.608 (7.70)	0.625 (8.84)	0.623 (8.36)	0.659 (9.88)	0.623 (8.44)	0.602 (7.37)
RNN (point)	0.680 (10.39)	0.689 (10.56)	0.668 (9.43)	0.685 (10.44)	0.676 (10.02)	0.688 (10.67)	0.683 (10.24)	0.673 (9.55)	0.660 (8.99)	0.684 (10.17)
<i>1.25 minute frequency</i>										
Interval (density)	0.653 (8.39)	0.657 (8.63)	0.641 (7.73)	0.688 (10.33)	0.660 (8.74)	0.691 (10.41)	0.614 (6.09)	0.651 (8.26)	0.675 (9.57)	0.656 (8.55)
HAR (point)	0.563 (5.98)	0.566 (5.94)	0.555 (5.61)	0.660 (10.01)	0.579 (6.76)	0.598 (8.23)	0.598 (7.57)	0.611 (7.69)	0.563 (5.84)	0.559 (5.75)
RNN (point)	0.655 (8.58)	0.649 (8.20)	0.652 (8.37)	0.709 (11.55)	0.657 (8.63)	0.679 (9.76)	0.637 (7.51)	0.651 (8.25)	0.664 (9.00)	0.661 (8.90)

*Notes: See notes of Table 6.

Table 10: Directional Predictive Accuracy (Stock MSFT)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators					
					MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>5 minute frequency</i>										
Interval (density)	0.687 (10.19)	0.664 (8.94)	0.653 (8.39)	0.701 (11.05)	0.676 (9.64)	0.691 (10.48)	0.687 (10.18)	0.652 (8.28)	0.661 (8.83)	0.671 (9.34)
HAR (point)	0.603 (7.59)	0.599 (7.17)	0.621 (8.11)	0.693 (11.54)	0.647 (9.48)	0.676 (10.49)	0.603 (7.47)	0.620 (8.33)	0.599 (7.19)	0.619 (8.24)
RNN (point)	0.656 (9.04)	0.669 (9.77)	0.650 (8.92)	0.736 (12.95)	0.689 (10.54)	0.697 (10.83)	0.675 (9.77)	0.656 (9.06)	0.653 (8.96)	0.657 (9.08)
<i>2.5 minute frequency</i>										
Interval (density)	0.637 (7.54)	0.651 (8.22)	0.647 (8.00)	0.697 (10.83)	0.659 (8.68)	0.693 (10.63)	0.647 (8.01)	0.656 (8.53)	0.681 (9.87)	0.673 (9.49)
HAR (point)	0.606 (7.00)	0.595 (6.91)	0.594 (7.07)	0.691 (11.21)	0.610 (7.69)	0.632 (8.26)	0.592 (6.56)	0.587 (6.33)	0.584 (7.05)	0.592 (6.98)
RNN (point)	0.655 (8.67)	0.631 (7.63)	0.635 (8.10)	0.726 (12.44)	0.663 (9.19)	0.677 (9.79)	0.652 (8.67)	0.631 (7.72)	0.629 (7.86)	0.639 (8.22)
<i>1.25 minute frequency</i>										
Interval (density)	0.639 (7.60)	0.655 (8.44)	0.637 (7.51)	0.701 (11.04)	0.639 (7.57)	0.668 (9.20)	0.639 (7.58)	0.644 (7.84)	0.645 (7.96)	0.645 (7.98)
HAR (point)	0.583 (6.49)	0.598 (7.21)	0.594 (6.69)	0.692 (11.34)	0.600 (7.17)	0.614 (7.90)	0.586 (6.39)	0.578 (6.35)	0.587 (6.41)	0.570 (5.55)
RNN (point)	0.620 (7.12)	0.649 (8.68)	0.627 (7.39)	0.729 (12.58)	0.649 (8.61)	0.648 (8.55)	0.665 (9.47)	0.648 (8.63)	0.641 (8.22)	0.635 (7.79)

*Notes: See notes of Table 6.

Table 11: Directional Predictive Accuracy (Stock PG)

Forecasting Methodology	Volatility Estimators									
	RV	BPV	TPV	TSRV	MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>5 minute frequency</i>										
Interval (density)	0.656 (8.59)	0.649 (8.20)	0.637 (7.55)	0.681 (9.92)	0.668 (9.22)	0.703 (11.12)	0.664 (9.03)	0.629 (7.11)	0.655 (8.50)	0.645 (8.00)
HAR (point)	0.623 (7.42)	0.623 (7.85)	0.632 (8.23)	0.675 (10.80)	0.648 (9.09)	0.664 (10.25)	0.620 (7.26)	0.625 (7.79)	0.620 (7.44)	0.629 (7.80)
RNN (point)	0.671 (9.37)	0.660 (8.85)	0.654 (8.60)	0.699 (10.96)	0.664 (9.03)	0.713 (11.80)	0.665 (9.07)	0.667 (9.16)	0.664 (9.14)	0.652 (8.37)
<i>2.5 minute frequency</i>										
Interval (density)	0.663 (8.93)	0.651 (8.27)	0.651 (8.27)	0.691 (10.44)	0.651 (8.23)	0.673 (9.53)	0.629 (7.10)	0.644 (7.91)	0.652 (8.35)	0.651 (8.29)
HAR (point)	0.616 (7.72)	0.603 (6.86)	0.618 (7.60)	0.669 (10.41)	0.611 (7.66)	0.656 (9.68)	0.606 (6.97)	0.620 (7.51)	0.616 (7.75)	0.607 (7.28)
RNN (point)	0.616 (6.76)	0.616 (7.02)	0.618 (6.49)	0.699 (10.94)	0.604 (5.90)	0.684 (10.11)	0.636 (7.56)	0.659 (8.75)	0.618 (6.74)	0.610 (6.53)
<i>1.25 minute frequency</i>										
Interval (density)	0.633 (7.32)	0.641 (7.77)	0.648 (8.10)	0.695 (10.67)	0.637 (7.53)	0.660 (8.78)	0.599 (5.41)	0.635 (7.41)	0.637 (7.56)	0.639 (7.63)
HAR (point)	0.603 (6.81)	0.615 (7.55)	0.620 (7.96)	0.679 (10.84)	0.598 (6.98)	0.611 (7.30)	0.596 (6.30)	0.607 (6.66)	0.606 (6.962)	0.608 (7.25)
RNN (point)	0.633 (7.38)	0.648 (8.15)	0.607 (5.95)	0.696 (10.78)	0.612 (6.18)	0.667 (9.20)	0.632 (7.29)	0.651 (8.27)	0.622 (7.14)	0.632 (7.48)

*See notes of Table 6.

Table 12: Relative Directional Predictive Accuracy (averaged across stocks and sampling frequencies)

Forecasting Methodology	Volatility Estimators									
	RV	BPV	TPV	TSRV	MSRV	RK	TRV	TBPV	MedRV	MinRV
1.000										
Interval (density)	1.000 (0.65)	1.000 (0.65)	1.000 (0.65)	1.000 (0.68)	1.000 (0.65)	1.000 (0.68)	1.000 (0.64)	1.000 (0.65)	1.000 (0.65)	1.000 (0.65)
HAR (point)	0.927 (0.60)	0.934 (0.61)	0.949 (0.61)	0.972 (0.66)	0.953 (0.62)	0.946 (0.64)	0.955 (0.61)	0.962 (0.62)	0.933 (0.60)	0.932 (0.61)
RNN (point)	1.001 (0.65)	1.007 (0.65)	1.006 (0.65)	1.026 (0.70)	1.015 (0.66)	1.004 (0.68)	1.027 (0.66)	1.022 (0.66)	1.002 (0.65)	1.005 (0.65)

*Notes: See notes to Table 6. This table contains DPA for all models (averaged across all stocks), relative to the DPA rate of the interval forecast models. Values in parentheses are actual un-normalized DPA values, and entries in bold denote cases where a model performs worse than the interval model.

Table 13: Relative Directional Predictive Accuracy (averaged across stocks and volatility estimators)

	Sampling Frequency		
	78	156	312
Interval (density)	1.000 (0.67)	1.000 (0.66)	1.000 (0.65)
HAR (point)	0.956 (0.64)	0.946 (0.62)	0.937 (0.60)
RNN (point)	1.018 (0.68)	1.006 (0.66)	1.011 (0.65)

*Notes: See notes to Table 12.

Table 14: Trading Based Risk-Adjusted Return Performance (Stock BA)

Forecasting Methodology	RV	BPV	TPV	TSRV	MSRV	RK	TRV	TBPV	MedRV	MinRV	Volatility Estimators				
											Panel A: Sharpe Ratio				
											5 minute frequency				
Interval (density)	0.403 (3.63)	0.436 (3.75)	0.378 (3.61)	0.181 (3.97)	0.379 (3.98)	0.314 (3.71)	0.412 (3.65)	0.407 (3.89)	0.430 (3.89)	0.376 (3.39)	0.455 (3.56)	0.455 (3.56)	0.455 (3.56)	0.455 (3.56)	0.455 (3.56)
HAR (point)	0.435 (2.35)	0.422 (2.50)	0.437 (2.62)	0.383 (2.65)	0.224 (2.09)	0.354 (2.25)	0.407 (2.54)	0.330 (2.37)	0.518 (2.37)	0.510 (2.27)	0.427 (2.00)	0.427 (2.04)	0.427 (2.04)	0.427 (2.00)	0.427 (2.00)
RNN (point)	0.695 (4.31)	0.724 (4.91)	0.503 (4.23)	0.062 (5.40)	0.285 (5.16)	0.398 (4.45)	0.396 (5.306)	0.587 (4.53)	0.755 (4.53)	0.630 (4.72)	0.558 (3.67)	0.660 (3.67)	0.660 (3.67)	0.660 (3.67)	0.660 (3.67)
Interval (density)	0.433 (3.56)	0.354 (3.39)	0.334 (3.43)	0.204 (4.11)	0.448 (3.50)	0.413 (3.19)	0.432 (3.39)	0.378 (3.29)	0.455 (3.29)	0.529 (3.56)	0.427 (2.00)	0.427 (2.04)	0.427 (2.04)	0.427 (2.00)	0.427 (2.00)
HAR (point)	0.306 (2.22)	0.476 (2.21)	0.247 (2.67)	0.350 (2.68)	0.325 (2.27)	0.344 (2.27)	0.291 (2.40)	0.336 (2.04)	0.427 (2.04)	0.241 (2.00)	0.558 (3.58)	0.660 (3.67)	0.660 (3.67)	0.660 (3.67)	0.660 (3.67)
RNN (point)	0.850 (3.28)	0.417 (4.10)	0.605 (5.10)	0.123 (5.84)	0.479 (3.97)	0.498 (4.70)	0.544 (4.57)	0.571 (3.58)	0.558 (3.58)	0.660 (3.67)	0.558 (3.67)	0.660 (3.67)	0.660 (3.67)	0.660 (3.67)	0.660 (3.67)
Interval (density)	0.418 (3.30)	0.597 (3.60)	0.637 (3.23)	0.192 (4.00)	0.431 (3.19)	0.361 (3.35)	0.423 (3.04)	0.464 (3.35)	0.546 (3.35)	0.538 (3.54)	0.316 (3.35)	0.357 (3.35)	0.529 (3.54)	0.529 (3.54)	0.529 (3.54)
HAR (point)	0.315 (2.119)	0.274 (1.900)	0.332 (2.731)	0.376 (2.738)	0.509 (2.473)	0.427 (2.486)	0.279 (2.787)	0.316 (1.747)	0.357 (1.747)	0.529 (1.728)	0.558 (1.728)	0.660 (1.728)	0.660 (1.728)	0.660 (1.728)	0.660 (1.728)
Point (RNN)	0.328 (6.56)	0.335 (6.22)	0.697 (5.46)	0.135 (5.89)	0.488 (4.39)	0.633 (4.04)	0.559 (4.91)	0.469 (5.21)	0.605 (5.21)	0.554 (5.21)	0.660 (5.21)	0.605 (5.21)	0.554 (5.21)	0.554 (5.21)	0.554 (5.21)
Panel B: Sortino Ratio															
Interval (density)	1.408 (1.04)	1.529 (1.07)	1.186 (1.14)	0.269 (2.66)	1.128 (1.33)	0.948 (1.23)	1.321 (1.13)	1.284 (1.22)	1.360 (1.22)	1.304 (0.97)	1.792 (0.83)	2.424 (0.83)	2.424 (0.77)	2.424 (0.77)	2.424 (0.77)
Point (HAR)	1.055 (0.97)	1.059 (0.99)	1.151 (0.99)	1.162 (0.87)	0.413 (1.13)	1.042 (0.76)	1.010 (1.02)	0.890 (0.73)	1.670 (0.73)	2.038 (0.56)	1.747 (1.01)	1.728 (1.01)	1.728 (1.01)	1.728 (1.01)	1.728 (1.01)
RNN (point)	1.996 (1.50)	3.032 (1.17)	1.4431 (1.47)	0.079 (4.21)	0.528 (2.78)	0.825 (2.14)	0.784 (2.67)	1.566 (1.16)	2.931 (1.16)	2.126 (1.39)	2.931 (1.39)	2.126 (1.39)	2.126 (1.39)	2.126 (1.39)	2.126 (1.39)
Interval (density)	1.422 (1.08)	1.151 (1.04)	0.958 (1.19)	0.326 (2.57)	1.759 (0.89)	1.400 (0.94)	1.468 (0.99)	1.409 (0.83)	1.792 (0.83)	2.424 (0.77)	1.792 (0.83)	2.424 (0.77)	2.424 (0.77)	2.424 (0.77)	2.424 (0.77)
Point (HAR)	0.574 (1.18)	1.092 (0.96)	0.424 (1.55)	1.018 (0.92)	0.734 (1.01)	0.992 (0.78)	0.706 (0.99)	0.876 (1.01)	0.861 (1.01)	0.441 (1.09)	1.670 (1.01)	2.038 (1.01)	2.038 (1.01)	2.038 (1.01)	2.038 (1.01)
Point (RNN)	2.574 (1.08)	0.721 (2.36)	1.404 (2.20)	0.157 (4.57)	0.946 (2.01)	1.185 (1.97)	1.283 (1.93)	1.336 (1.97)	1.012 (1.97)	1.376 (1.97)	1.747 (1.97)	1.728 (1.97)	1.728 (1.97)	1.728 (1.97)	1.728 (1.97)
Interval (density)	1.580 (0.87)	3.574 (0.60)	3.901 (0.52)	0.295 (2.59)	1.433 (0.96)	1.034 (1.17)	1.498 (0.85)	2.295 (0.55)	3.308 (0.55)	2.205 (0.86)	1.792 (0.83)	2.424 (0.83)	2.424 (0.77)	2.424 (0.77)	2.424 (0.77)
Point (HAR)	0.576 (2.11)	0.461 (1.90)	0.892 (2.73)	1.114 (2.73)	1.312 (2.47)	1.067 (2.48)	0.755 (2.78)	0.833 (1.74)	0.681 (1.74)	1.077 (1.74)	1.670 (1.74)	2.038 (1.74)	2.038 (1.74)	2.038 (1.74)	2.038 (1.74)
Point (RNN)	0.598 (3.60)	0.578 (3.59)	2.146 (1.77)	0.173 (4.57)	1.164 (1.84)	1.649 (1.55)	1.354 (2.02)	0.985 (1.84)	1.710 (1.84)	1.296 (2.31)	1.747 (1.84)	1.728 (1.84)	1.728 (1.84)	1.728 (1.84)	1.728 (1.84)

*Notes: This table contains monthly Sharpe Ratios (Panel A) and Sortino Ratios (Panel B) generated using the trading strategies described in Section 7.3. The values in parentheses below the Sharpe Ratios are corresponding return standard deviations while those below the Sortino Ratios are corresponding return downside risk measures. Entries in bold indicate models that yield inferior performance, relative those the performance based on forecasts constructed using our interval forecasts, for a given volatility estimator and sampling frequency.

Table 15: Trading Based Risk-Adjusted Return Performance (Stock GE)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators					
					MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>Panel A: Sharpe Ratio</i>										
					<i>5 minute frequency</i>					
Interval (density)	-0.160 (2.17)	-0.082 (1.74)	0.008 (1.53)	-0.138 (2.87)	-0.020 (2.08)	-0.111 (2.57)	-0.183 (2.18)	0.030 (1.34)	0.291 (1.34)	-0.052 (2.22)
HAR (point)	0.017 (1.45)	0.108 (1.22)	0.090 (1.24)	-0.099 (1.67)	-0.118 -0.209	-0.005 -0.172	-0.180 -0.150	0.024 (1.37)	0.291 (1.37)	-0.221 (1.05)
RNN (point)	-0.269 (2.14)	-0.049 (1.05)	0.280 (1.10)	-0.417 (2.14)	-0.288 (1.58)	-0.362 (1.72)	-0.059 (1.98)	-0.063 (0.95)	0.291 (0.95)	-0.005 (0.93)
					<i>2.5 minute frequency</i>					
Interval (density)	0.024 (2.12)	0.020 (2.01)	0.067 (2.08)	-0.240 (2.91)	-0.080 (2.49)	-0.159 (1.98)	0.066 (1.72)	0.149 (2.04)	0.029 (2.04)	0.098 (2.01)
HAR (point)	-0.042 (1.52)	-0.023 (1.32)	0.096 (1.46)	-0.052 (1.68)	-0.220 (1.88)	-0.186 (1.68)	0.016 (1.42)	0.039 (1.44)	-0.029 (1.44)	0.103 (1.325)
RNN (point)	-0.342 (1.59)	0.027 (1.07)	0.140 (1.42)	-0.369 (2.24)	0.059 (0.76)	-0.455 (1.29)	0.106 (0.98)	-0.002 (1.32)	0.161 (1.32)	-0.266 (1.41)
					<i>1.25 minute frequency</i>					
Interval (density)	0.067 (2.19)	0.091 (2.24)	0.125 (2.16)	-0.199 (2.85)	-0.012 (2.45)	-0.096 (2.03)	-0.049 (1.83)	0.037 (2.75)	-0.066 (2.75)	0.028 (2.62)
HAR (point)	0.033 (1.30)	0.033 (1.30)	-0.114 (2.01)	-0.057 (1.69)	-0.208 (1.44)	-0.013 (1.74)	-0.107 (1.81)	-0.014 (1.51)	-0.245 (1.51)	-0.081 (1.23)
RNN (point)	-0.335 (1.75)	-0.164 (1.10)	-0.249 (1.68)	-0.422 (2.15)	-0.182 (0.82)	0.195 (1.23)	-0.024 (0.83)	-0.089 (1.32)	-0.106 (1.32)	-0.133 (1.40)
<i>Panel B: Sortino Ratio</i>										
					<i>5 minute frequency</i>					
Interval (density)	-0.185 (1.88)	-0.105 (1.36)	0.012 (1.03)	-0.186 (2.13)	-0.033 (1.30)	-0.149 (1.92)	-0.215 (1.86)	0.050 (0.73)	0.536 (0.73)	-0.062 (1.84)
HAR (point)	0.031 (0.81)	0.268 (0.49)	0.228 (0.49)	-0.119 (1.39)	-0.148 (1.44)	-0.264 (1.16)	-0.008 (0.86)	-0.228 (0.73)	0.044 (0.73)	-0.291 (0.80)
RNN (point)	-0.317 (1.82)	-0.080 (0.64)	0.661 (0.46)	-0.457 (1.96)	-0.286 (1.59)	-0.361 (1.72)	-0.074 (1.57)	-0.091 (0.41)	0.670 (0.41)	-0.008 (0.58)
					<i>2.5 minute frequency</i>					
Interval (density)	0.040 (1.30)	0.033 (1.21)	0.116 (1.19)	-0.292 (2.39)	-0.110 (1.82)	-0.236 (1.34)	0.094 (1.21)	0.243 (1.21)	0.049 (1.21)	0.182 (1.08)
HAR (point)	-0.061 (1.06)	-0.032 (0.92)	0.151 (0.92)	-0.065 (1.35)	-0.253 (1.63)	-0.231 (1.35)	0.025 (0.94)	0.061 (0.92)	-0.046 (0.92)	0.155 (0.88)
RNN (point)	-0.340 (1.60)	0.039 (0.73)	0.257 (0.77)	-0.420 (1.96)	0.088 (0.51)	-0.436 (1.35)	0.146 (0.71)	-0.003 (0.61)	0.344 (0.61)	-0.322 (1.16)
					<i>1.25 minute frequency</i>					
Interval (density)	0.122 (1.20)	0.174 (1.17)	0.234 (1.16)	-0.248 (2.29)	-0.017 (1.74)	-0.123 (1.59)	-0.064 (1.40)	0.061 (2.03)	-0.089 (2.03)	0.040 (1.84)
HAR (point)	0.055 (1.30)	0.055 (1.30)	-0.143 (2.01)	-0.071 (1.69)	-0.272 (1.44)	-0.021 (1.74)	-0.127 (1.81)	-0.018 (1.51)	-0.312 (1.51)	-0.106 (1.23)
RNN (point)	-0.350 (1.68)	-0.215 (0.84)	-0.271 (1.54)	-0.461 (1.96)	-0.242 (0.61)	0.322 (0.75)	-0.034 (0.59)	-0.106 (1.00)	-0.140 (1.00)	-0.190 (0.98)

*Notes: See Notes of Table 14.

Table 16: Trading Based Risk-Adjusted Return Performance (Stock IBM)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators					
					MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>Panel A: Sharpe Ratio</i>										
					<i>5 minute frequency</i>					
Interval (density)	0.101 (2.32)	-0.057 (1.89)	-0.042 (2.35)	-0.044 (2.69)	0.052 (2.63)	0.064 (2.94)	0.090 (2.11)	0.131 (2.09)	-0.037 (2.09)	0.047 (2.09)
HAR (point)	-0.043 (1.74)	0.001 (1.82)	0.074 (2.18)	-0.097 (2.57)	0.050 (1.72)	-0.118 (2.15)	0.055 (1.64)	0.103 (2.23)	0.083 (2.23)	0.019 (1.89)
Point (RNN)	0.092 (3.02)	-0.098 (2.21)	0.133 (3.34)	-0.037 (4.33)	-0.006 (2.79)	-0.272 (3.41)	0.035 (3.08)	0.135 (3.42)	0.154 (3.42)	0.189 (3.65)
					<i>2.5 minute frequency</i>					
Interval (density)	0.170 (2.85)	0.127 (2.14)	0.031 (2.50)	0.074 (2.79)	-0.045 (2.76)	0.065 (2.43)	0.302 (1.59)	0.243 (2.13)	0.166 (2.13)	0.158 (1.83)
HAR (point)	0.195 (1.87)	0.406 (1.37)	0.378 (1.44)	-0.102 (2.52)	0.051 (1.60)	-0.234 (1.61)	0.412 (1.57)	0.263 (1.95)	0.161 (1.95)	0.305 (1.29)
RNN (point)	0.172 (3.88)	0.393 (2.08)	-0.180 (2.72)	0.002 (4.53)	0.037 (3.83)	-0.428 (3.89)	0.381 (2.98)	0.310 (3.53)	0.242 (3.53)	0.293 (3.25)
					<i>1.25 minute frequency</i>					
Interval (density)	0.026 (2.96)	-0.011 (3.11)	0.075 (2.44)	-0.001 (2.93)	-0.108 (3.00)	0.075 (2.56)	0.022 (2.60)	0.020 (2.57)	0.129 (2.57)	0.160 (2.99)
HAR (point)	-0.169 (1.39)	0.024 (1.87)	0.081 (1.97)	-0.099 (2.40)	0.277 (1.58)	-0.031 (2.03)	0.092 (1.99)	0.199 (1.56)	0.202 (1.56)	0.072 (1.48)
RNN (point)	0.093 (3.74)	-0.046 (3.50)	0.013 (3.51)	-0.041 (4.46)	-0.051 (4.04)	-0.143 (4.26)	0.142 (3.43)	-0.047 (3.53)	0.278 (3.53)	-0.143 (3.80)
<i>Panel B: Sortino Ratio</i>										
					<i>5 minute frequency</i>					
Interval (density)	0.154 (1.52)	-0.083 (1.29)	-0.061 (1.59)	-0.060 (1.97)	0.084 (1.63)	0.097 (1.92)	0.148 (1.27)	0.263 (1.38)	-0.056 (1.38)	0.076 (1.30)
HAR (point)	-0.061 (1.23)	0.001 (1.25)	0.129 (1.25)	-0.119 (2.09)	0.078 (1.10)	-0.154 (1.64)	0.089 (1.01)	0.184 (1.19)	0.156 (1.19)	0.028 (1.26)
RNN (point)	0.128 (2.19)	-0.115 (1.88)	0.183 (2.42)	-0.051 (3.17)	-0.009 (1.87)	-0.319 (2.90)	0.049 (2.22)	0.237 (2.09)	0.253 (2.09)	0.371 (1.86)
					<i>2.5 minute frequency</i>					
Interval (density)	0.318 (1.52)	0.225 (1.21)	0.047 (1.65)	0.110 (1.87)	-0.067 (1.85)	0.108 (1.47)	0.715 (0.67)	0.616 (1.21)	0.292 (1.21)	0.274 (1.05)
HAR (point)	0.332 (1.10)	0.955 (0.58)	0.840 (0.65)	-0.124 (2.07)	0.088 (0.93)	-0.276 (1.36)	0.886 (0.73)	0.489 (1.28)	0.246 (1.28)	0.709 (0.55)
RNN (point)	0.313 (2.14)	1.010 (0.81)	-0.216 (2.26)	0.003 (3.08)	0.058 (2.42)	-0.435 (3.83)	1.143 (0.99)	0.714 (1.83)	0.466 (1.83)	0.452 (2.10)
					<i>1.25 minute frequency</i>					
Interval (density)	0.039 (1.94)	-0.017 (2.05)	0.134 (1.35)	-0.002 (2.15)	-0.148 (2.19)	0.127 (1.50)	0.034 (1.73)	0.029 (1.56)	0.212 (1.56)	0.259 (1.85)
HAR (point)	-0.213 (1.39)	0.037 (1.87)	0.129 (1.97)	-0.122 (2.40)	0.591 (1.58)	-0.040 (2.03)	0.143 (1.99)	0.363 (1.56)	0.331 (1.56)	0.103 (1.48)
RNN (point)	0.148 (2.35)	-0.060 (2.66)	0.017 (2.67)	-0.060 (3.06)	-0.066 (3.14)	-0.172 (3.56)	0.243 (2.01)	-0.063 (1.71)	0.574 (1.71)	-0.172 (3.14)

*Notes: See Notes of Table 14.

Table 17: Trading Based Risk-Adjusted Return Performance (Stock JPM)

Forecasting Methodology	RV	BPV	TPV	TSRV	MSRV	RK	TRV	TBPV	MedRV	MinRV	Volatility Estimators								
											Panel A: Sharpe Ratio								
<i>5 minute frequency</i>																			
Interval (density)	0.249 (2.21)	0.084 (1.73)	0.199 (2.82)	0.299 (3.21)	0.127 (2.66)	0.142 (2.87)	0.275 (2.38)	0.146 (1.94)	0.042 (1.94)	0.123 (2.49)									
HAR (point)	0.003 (2.21)	-0.213 (2.16)	0.017 (1.44)	0.319 (2.48)	0.155 (2.28)	0.165 (2.16)	-0.051 (2.03)	-0.022 (1.12)	0.069 (1.12)	-0.141 (1.46)									
RNN (point)	0.291 (2.29)	0.120 (2.76)	0.372 (2.63)	0.267 (4.68)	0.149 (3.84)	0.261 (4.45)	0.081 (2.48)	0.123 (2.98)	0.081 (2.98)	-0.105 (3.08)									
<i>2.5 minute frequency</i>																			
Density (CI)	0.204 (1.97)	0.277 (2.47)	0.260 (1.96)	0.321 (3.13)	0.250 (2.88)	0.361 (2.60)	0.350 (2.97)	0.345 (2.19)	0.261 (2.19)	0.233 (1.85)									
HAR (point)	0.430 (1.23)	0.422 (1.23)	0.147 (1.43)	0.286 (2.38)	0.075 (2.53)	0.229 (1.45)	-0.079 (2.50)	0.049 (1.15)	0.231 (1.15)	0.285 (1.32)									
RNN (point)	0.117 (4.52)	0.236 (3.53)	0.226 (3.13)	0.248 (4.69)	0.254 (3.35)	0.458 (3.55)	0.278 (3.06)	0.293 (3.32)	0.203 (3.32)	0.130 (3.63)									
<i>1.25 minute frequency</i>																			
Interval (density)	0.234 (2.25)	0.219 (2.08)	0.211 (2.17)	0.317 (3.23)	0.232 (1.96)	0.142 (2.46)	0.238 (2.58)	0.284 (2.09)	0.223 (2.09)	0.194 (2.66)									
HAR (point)	0.325 (1.30)	0.379 (1.32)	0.474 (1.31)	0.298 (2.42)	0.342 (1.26)	0.153 (1.81)	-0.079 (1.99)	-0.074 (1.35)	0.331 (1.35)	0.415 (1.27)									
RNN (point)	0.104 (4.73)	-0.113 (4.58)	0.012 (4.22)	0.267 (4.84)	0.291 (4.08)	0.072 (4.21)	-0.051 (4.17)	-0.046 (3.63)	0.199 (3.63)	-0.205 (5.26)									
<i>Panel B: Sortino Ratio</i>																			
<i>5 minute frequency</i>																			
Interval (density)	0.561 (0.98)	0.122 (1.18)	0.445 (1.26)	0.772 (1.24)	0.285 (1.19)	0.260 (1.56)	0.643 (1.01)	0.294 (1.23)	0.066 (1.23)	0.250 (1.22)									
HAR (point)	0.004 (1.80)	-0.229 (2.01)	0.024 (1.01)	0.907 (0.87)	0.224 (1.57)	0.379 (0.93)	-0.058 (1.79)	-0.031 (0.67)	0.115 (0.67)	-0.174 (1.18)									
RNN (point)	0.734 (0.90)	0.215 (1.54)	0.932 (1.05)	0.619 (2.01)	0.294 (1.94)	0.633 (1.83)	0.144 (1.83)	0.277 (1.83)	0.132 (1.83)	-0.140 (2.31)									
<i>2.5 minute frequency</i>																			
Interval (density)	0.421 (0.95)	0.987 (0.69)	0.686 (0.74)	0.826 (1.21)	0.725 (0.99)	1.229 (0.76)	1.509 (0.69)	1.397 (0.94)	0.603 (0.94)	0.542 (0.79)									
HAR (point)	0.916 (0.58)	0.893 (0.58)	0.220 (0.96)	0.781 (0.87)	0.101 (1.88)	0.429 (0.77)	-0.092 (2.13)	0.061 (0.58)	0.461 (0.58)	0.592 (0.63)									
RNN (point)	0.232 (2.28)	0.419 (1.98)	0.534 (1.32)	0.576 (2.02)	0.524 (1.62)	1.639 (0.99)	0.487 (1.75)	0.729 (1.85)	0.365 (1.85)	0.221 (2.14)									
<i>1.25 minute frequency</i>																			
Interval (density)	0.532 (0.99)	0.509 (0.89)	0.473 (0.97)	0.822 (1.24)	0.625 (0.73)	0.322 (1.08)	0.572 (1.07)	0.727 (0.93)	0.498 (0.93)	0.514 (1.00)									
HAR (point)	0.900 (1.30)	1.038 (1.32)	1.412 (1.31)	0.827 (2.42)	0.693 (1.26)	0.237 (1.81)	-0.100 (1.99)	-0.093 (1.34)	0.913 (1.34)	1.139 (1.27)									
RNN (point)	0.189 (2.60)	-0.146 (3.56)	0.017 (2.94)	0.639 (2.02)	0.786 (1.51)	0.126 (2.39)	-0.072 (2.96)	-0.061 (1.86)	0.388 (1.86)	-0.236 (4.55)									

*Notes: See Notes of Table 14.

Table 18: Trading Based Risk-Adjusted Return Performance (Stock MSFT)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators										
					MSRV	RK	TRV	TBPV	MedRV	MinRV					
<i>Panel A: Sharpe Ratio</i>															
<i>5 minute frequency</i>															
Interval (density)	0.361 (3.02)	0.310 (2.95)	0.295 (2.90)	0.368 (3.57)	0.173 (2.81)	0.266 (4.21)	0.335 (2.98)	0.325 (3.12)	0.289 (3.12)	0.346 (3.21)					
HAR (point)	0.347 (3.73)	0.232 (5.68)	0.218 (5.31)	0.182 (3.95)	0.148 (2.38)	0.127 (4.38)	0.381 (3.66)	0.308 (3.56)	0.302 (3.56)	0.311 (3.55)					
RNN (point)	0.384 (5.94)	0.387 (5.68)	0.480 (4.72)	0.274 (6.62)	0.141 (5.98)	0.269 (7.62)	0.385 (5.89)	0.396 (4.17)	0.519 (4.17)	0.394 (2.22)					
<i>2.5 minute frequency</i>															
Interval (density)	0.189 (3.08)	0.217 (3.39)	0.244 (3.27)	0.344 (3.91)	0.169 (2.66)	0.185 (3.15)	0.251 (3.20)	0.077 (3.45)	0.286 (3.45)	0.218 (3.34)					
HAR (point)	0.134 (4.69)	0.198 (3.74)	0.217 (3.51)	0.170 (3.91)	0.183 (2.74)	0.255 (4.87)	0.264 (3.52)	0.238 (5.02)	0.142 (5.02)	0.172 (3.45)					
RNN (point)	0.355 (5.05)	0.485 (4.53)	0.482 (4.47)	0.275 (6.15)	0.160 (6.68)	0.478 (6.91)	0.392 (5.07)	0.459 (4.97)	0.392 (4.97)	0.418 (4.36)					
<i>1.25 minute frequency</i>															
Interval (density)	0.070 (2.69)	0.197 (2.80)	0.202 (2.66)	0.347 (3.62)	0.097 (2.75)	0.167 (3.36)	0.176 (2.93)	0.235 (2.70)	0.305 (2.70)	0.226 (3.01)					
HAR (point)	0.054 (2.93)	-0.013 (2.47)	0.035 (2.51)	0.171 (4.00)	0.150 (2.68)	0.330 (3.31)	0.156 (2.66)	0.142 (2.46)	0.104 (2.46)	0.095 (2.71)					
RNN (point)	0.404 (5.31)	0.434 (5.48)	0.480 (5.91)	0.315 (6.23)	0.344 (5.64)	0.259 (5.94)	0.542 (5.42)	0.468 (4.74)	0.546 (4.74)	0.536 (5.40)					
<i>Panel B: Sortino Ratio</i>															
<i>5 minute frequency</i>															
Interval (density)	0.901 (1.21)	0.749 (1.22)	0.901 (0.94)	1.453 (0.90)	0.326 (1.49)	0.514 (2.18)	0.843 (1.18)	1.008 (1.48)	0.608 (1.48)	0.798 (1.39)					
HAR (point)	1.168 (1.10)	1.095 (1.20)	0.995 (1.21)	0.359 (2.00)	0.298 (1.18)	0.215 (2.58)	1.390 (1.00)	1.075 (1.10)	0.970 (1.10)	1.006 (1.10)					
RNN (point)	2.016 (1.13)	1.929 (1.14)	2.874 (0.79)	0.632 (2.87)	0.226 (3.71)	0.562 (3.65)	2.304 (0.98)	1.463 (0.85)	2.538 (0.85)	1.556 (1.34)					
<i>2.5 minute frequency</i>															
Interval (density)	0.399 (1.46)	0.490 (1.50)	0.562 (1.42)	1.233 (1.09)	0.300 (1.51)	0.362 (1.61)	0.677 (1.19)	0.155 (1.44)	0.683 (1.44)	0.486 (1.50)					
HAR (point)	0.383 (1.64)	0.449 (1.65)	0.568 (1.35)	0.333 (2.00)	0.435 (1.16)	1.219 (1.02)	0.610 (1.52)	0.675 (1.29)	0.556 (1.29)	0.482 (1.23)					
RNN (point)	1.425 (1.26)	2.847 (0.77)	2.286 (0.94)	0.586 (2.90)	0.328 (3.27)	2.740 (1.20)	1.531 (1.30)	2.203 (1.29)	1.508 (1.29)	1.613 (1.13)					
<i>1.25 minute frequency</i>															
Interval (density)	0.114 (1.67)	0.429 (1.29)	0.424 (1.27)	1.258 (1.00)	0.168 (1.60)	0.295 (1.91)	0.335 (1.54)	0.533 (1.06)	0.777 (1.06)	0.444 (1.53)					
HAR (point)	0.117 (2.93)	-0.021 (2.47)	0.057 (2.51)	0.340 (4.00)	0.266 (2.68)	0.961 (3.32)	0.376 (2.66)	0.330 (2.46)	0.213 (2.46)	0.166 (2.71)					
RNN (point)	1.643 (1.30)	1.499 (1.58)	2.652 (1.07)	0.753 (2.61)	1.215 (1.59)	0.603 (2.55)	4.357 (0.67)	1.792 (0.75)	3.453 (0.75)	2.105 (1.37)					

*Notes: See Notes of Table 14.

Table 19: Trading Based Risk-Adjusted Return Performance (Stock PG)

Forecasting Methodology	RV	BPV	TPV	TSRV	Volatility Estimators					
					MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>Panel A: Sharpe Ratio</i>										
Interval (density)	-0.049 (1.92)	-0.163 (2.44)	-0.157 (2.13)	-0.059 (1.65)	-0.092 (1.90)	-0.036 (1.54)	-0.028 (2.14)	-0.101 (2.42)	-0.151 (2.42)	-0.116 (2.01)
HAR (point)	-0.147 (0.92)	-0.212 (1.08)	-0.172 (1.22)	-0.194 (0.74)	0.111 (1.25)	0.101 (1.18)	-0.179 (1.09)	-0.441 (1.28)	-0.305 (1.28)	-0.330 (1.26)
RNN (point)	-0.366 (1.57)	-0.102 (2.36)	-0.192 (1.90)	-0.636 (1.61)	-0.081 (1.88)	-0.108 (1.59)	-0.138 (2.63)	-0.253 (2.61)	0.042 (2.61)	-0.077 (2.34)
<i>5 minute frequency</i>										
Interval (density)	-0.042 (1.98)	-0.036 (2.14)	-0.163 (1.93)	-0.051 (1.52)	-0.037 (1.95)	0.007 (1.79)	-0.139 (2.55)	-0.127 (1.79)	-0.149 (1.79)	-0.087 (1.73)
HAR (point)	0.005 (0.81)	-0.146 (0.96)	-0.316 (1.13)	-0.234 (0.89)	-0.213 (1.07)	0.156 (1.04)	-0.202 (0.81)	-0.268 (0.94)	-0.310 (0.94)	-0.127 (0.99)
RNN (point)	-0.019 (1.99)	-0.219 (2.58)	-0.497 (1.95)	-0.319 (1.59)	-0.004 (1.83)	0.200 (1.59)	-0.039 (2.22)	-0.343 (2.95)	-0.118 (2.95)	-0.097 (3.07)
<i>2.5 minute frequency</i>										
Interval (density)	0.167 (1.04)	0.096 (1.45)	-0.117 (2.17)	-0.129 (1.53)	-0.059 (2.15)	0.221 (1.10)	-0.076 (1.71)	-0.129 (1.68)	-0.022 (1.68)	-0.151 (2.07)
HAR (point)	0.026 (0.77)	0.043 (1.02)	-0.095 (1.10)	-0.146 (0.62)	-0.200 (0.78)	0.085 (0.68)	-0.049 (0.93)	-0.232 (1.18)	-0.230 (1.18)	-0.141 (1.25)
RNN (point)	-0.305 (2.22)	-0.147 (2.64)	-0.400 (1.87)	-0.263 (1.55)	-0.011 (1.98)	-0.047 (2.27)	-0.263 (2.37)	-0.246 (2.60)	-0.089 (2.60)	-0.046 (2.82)
<i>1.25 minute frequency</i>										
Interval (density)	0.167 (1.04)	0.096 (1.45)	-0.117 (2.17)	-0.129 (1.53)	-0.059 (2.15)	0.221 (1.10)	-0.076 (1.71)	-0.129 (1.68)	-0.022 (1.68)	-0.151 (2.07)
HAR (point)	0.026 (0.77)	0.043 (1.02)	-0.095 (1.10)	-0.146 (0.62)	-0.200 (0.78)	0.085 (0.68)	-0.049 (0.93)	-0.232 (1.18)	-0.230 (1.18)	-0.141 (1.25)
RNN (point)	-0.305 (2.22)	-0.147 (2.64)	-0.400 (1.87)	-0.263 (1.55)	-0.011 (1.98)	-0.047 (2.27)	-0.263 (2.37)	-0.246 (2.60)	-0.089 (2.60)	-0.046 (2.82)
<i>Panel B: Sortino Ratio</i>										
<i>5 minute frequency</i>										
Interval (density)	-0.053 (1.75)	-0.174 (2.28)	-0.170 (1.97)	-0.083 (1.17)	-0.108 (1.63)	-0.045 (1.20)	-0.033 (1.84)	-0.110 (2.22)	-0.164 (2.22)	-0.129 (1.81)
HAR (point)	-0.175 (0.77)	-0.224 (1.02)	-0.189 (1.10)	-0.232 (0.62)	0.177 (0.78)	0.175 (0.68)	-0.209 (0.93)	-0.433 (1.18)	-0.330 (1.18)	-0.332 (1.25)
RNN (point)	-0.359 (1.60)	-0.125 (1.92)	-0.214 (1.71)	-0.592 (1.73)	-0.100 (1.51)	-0.123 (1.40)	-0.150 (2.42)	-0.274 (1.91)	0.057 (1.91)	-0.094 (1.82)
<i>2.5 minute frequency</i>										
Interval (density)	-0.047 (1.73)	-0.041 (1.89)	-0.181 (1.74)	-0.066 (1.16)	-0.044 (1.61)	0.007 (1.51)	-0.148 (2.38)	-0.142 (1.60)	-0.166 (1.60)	-0.099 (1.51)
HAR (point)	0.008 (0.52)	-0.184 (0.76)	-0.334 (1.07)	-0.269 (0.77)	-0.245 (0.93)	0.261 (0.62)	-0.234 (0.70)	-0.316 (0.89)	-0.324 (0.89)	-0.156 (0.81)
RNN (point)	-0.023 (1.58)	-0.249 (2.26)	-0.454 (2.13)	-0.343 (1.47)	-0.005 (1.31)	0.260 (1.22)	-0.047 (1.82)	-0.355 (2.40)	-0.144 (2.40)	-0.122 (2.43)
<i>1.25 minute frequency</i>										
Interval (density)	0.262 (0.66)	0.132 (1.06)	-0.129 (1.96)	-0.164 (1.20)	-0.068 (1.88)	0.400 (0.61)	-0.086 (1.50)	-0.146 (1.38)	-0.027 (1.38)	-0.171 (1.83)
HAR (point)	0.040 (0.77)	0.061 (0.64)	-0.111 (0.77)	-0.181 (0.78)	-0.213 (1.11)	0.141 (0.66)	-0.059 (0.84)	-0.249 (0.64)	-0.264 (0.64)	-0.154 (0.97)
RNN (point)	-0.304 (2.22)	-0.165 (2.35)	-0.387 (1.93)	-0.295 (1.38)	-0.014 (1.57)	-0.061 (1.74)	-0.282 (2.21)	-0.271 (2.22)	-0.103 (2.22)	-0.056 (2.27)

*Notes: See Notes of Table 14.

Table 20: Relative Trading Based Risk-Adjusted Return Performance
(averaged across stocks and sampling frequencies)

<i>Forecasting Methodology</i>	Volatility Estimators									
	RV	BPV	TPV	TSRV	MSRV	RK	TRV	TBPV	MedRV	MinRV
<i>Panel A: Sharpe Ratio</i>										
Interval (density)	1.000 (0.15)	1.000 (0.14)	1.000 (0.14)	1.000 (0.09)	1.000 (0.11)	1.000 (0.13)	1.000 (0.16)	1.000 (0.16)	1.000 (0.16)	1.000 (0.16)
HAR (point)	0.776 (0.12)	0.901 (0.13)	0.829 (0.11)	0.815 (0.08)	0.861 (0.09)	0.814 (0.11)	0.553 (0.08)	0.375 (0.06)	0.606 (0.10)	0.703 (0.11)
RNN (point)	0.635 (0.10)	0.742 (0.11)	0.781 (0.11)	-0.237 (-0.02)	0.677 (0.07)	0.361 (0.05)	0.867 (0.13)	0.704 (0.11)	1.333 (0.22)	0.738 (0.12)
<i>Panel A: Sortino Ratio</i>										
Interval (density)	1.000 (0.44)	1.000 (0.538)	1.000 (0.53)	1.000 (0.34)	1.000 (0.34)	1.000 (0.36)	1.000 (0.51)	1.000 (0.55)	1.000 (0.57)	1.000 (0.52)
HAR (point)	0.707 (0.31)	0.699 (0.376)	0.671 (0.35)	0.884 (0.30)	0.686 (0.23)	0.936 (0.34)	0.548 (0.28)	0.449 (0.24)	0.578 (0.33)	0.720 (0.37)
RNN (point)	1.290 (0.57)	1.150 (0.61)	1.454 (0.77)	0.246 (0.08)	0.871 (0.30)	1.318 (0.48)	1.414 (0.73)	1.012 (0.56)	1.557 (0.89)	1.047 (0.54)

*Notes: See notes to Table 14. This table contains monthly relative Sharpe and Sortino Ratios, averaged across stocks and sampling frequencies. The denominator for all ratios is the interval based ratio. Thus entries based on the interval method have a value of 1 throughout. Entries in bold font are those exhibiting inferior performance, relative to the performance of trading strategies based on interval forecasts. Values in parentheses are actual monthly Sharpe or Sortino Ratios value for a given forecasting model.

Table 21: Relative Trading Based Risk-Adjusted Return Performance
(averaged across stocks and volatility estimators)

	Sampling Frequency		
	5 min	2.5 min	1.25 min
<i>Panel A: Sharpe Ratio</i>			
Interval (density)	1.000 (0.12)	1.000 (0.16)	1.000 (0.15)
HAR (point)	0.626 (0.07)	0.765 (0.12)	0.708 (0.11)
RNN (point)	1.056 (0.13)	0.544 (0.08)	0.582 (0.08)
<i>Panel A: Sortino Ratio</i>			
Interval (density)	1.000 (0.37)	1.000 (0.50)	1.000 (0.54)
HAR (point)	0.864 (0.32)	0.656 (0.33)	0.555 (0.29)
RNN (point)	1.484 (0.56)	1.196 (0.60)	0.933 (0.50)

*Notes: See notes to Table 20.

Table 22: Monte Carlo Setup

Panel A. Heston stochastic volatility model with price and volatility jumps (HSV)

$dY_t = (m - \sigma_t^2/2)dt + \sigma_t dW_{1t} + J_t^y dN_{1t}$
$d\sigma_t^2 = \psi(\vartheta - \sigma_t^2)dt + \eta\sigma_t dW_{2t} + J_t^v dN_{2t}$
$m = 0.05, \vartheta = 0.04, \eta = 0.5, \psi = 5, \rho_1 = -0.5$
Microstructure noise $\sim N(0, 10^{-8})$
Volatility jump, $J_t^v = \exp(Z)$ where $Z \sim N(-5, 1)$
Price Jump type 1, $J_t^y \sim N(0, 0.02^2)$, $J_t^v = 0$.
Price Jump type 2, $J_t^y \sim N(0, 0.006^2)$, $J_t^v = 0$.

Panel B. Data Generating Processes

DGP 1 - No microstructure noise or jumps.
DGP 2 - DGP 1 + Microstructure noise
DGP 3 - DGP 1 + Price Jump type 1
DGP 4 - DGP 2 + Price Jump type 1
DGP 5 - DGP 1 + Price Jump type 2
DGP 6 - DGP 2 + Price Jump type 2
DGP 7 - DGP 1 + Volatility jump
DGP 8 - DGP 2 + Volatility jump
DGP 9 - DGP 7 + Price Jump type 1
DGP 10 - DGP 8 + Price Jump type 1
DGP 11 - DGP 7 + Price Jump type 2
DGP 12 - DGP 8 + Price Jump type 2

*Notes: This table contains details about the Monte Carlo experimental setup. Panel A specifies the HSV model used as our baseline DGP. Panel B lists the 12 different DGPs used our experiments.

Table 23: DGP 1 - Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.47	0.38	0.33	0.55	0.45	0.40	0.51	0.42	0.27	0.59	0.49	0.36
BPV	0.51	0.42	0.37	0.58	0.43	0.44	0.60	0.50	0.34	0.64	0.54	0.40
TPV	0.52	0.42	0.37	0.61	0.51	0.46	0.68	0.54	0.37	0.72	0.58	0.43
TSRV	0.77	0.85	0.81	0.78	0.86	0.82	0.93	0.91	0.86	0.94	0.92	0.90
MSRV	0.66	0.56	0.43	0.69	0.61	0.46	0.82	0.72	0.51	0.84	0.78	0.54
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.47	0.38	0.33	0.55	0.45	0.40	0.51	0.42	0.27	0.59	0.49	0.36
TBPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MedRV	0.59	0.42	0.33	0.62	0.46	0.38	0.72	0.53	0.37	0.78	0.59	0.40
MinRV	0.60	0.42	0.42	0.65	0.47	0.50	0.74	0.60	0.44	0.80	0.66	0.51

*Notes: This tables summarizes the main results from our Monte-Carlo experiments, for DGP1. Entries denote rejection frequencies of statistic $G_{T,M}(u_1, u_2)$, at significance levels $\alpha = 5\%$ and 10% . Intervals evaluated include $[u_1, u_2] = \hat{\mu}_{IV} \pm \beta\hat{\sigma}_{IV}$, where $\beta = \{0.125, 0.250\}$, and where $\hat{\mu}_{IV}$ and $\hat{\sigma}_{IV}$ are the mean and standard error of “pseudo true” IV. Results are based on $M = 78$ (5 mins), $M = 156$ (2.5 mins) and $M = 312$ (1.25 mins), $T = 100$, and 100 Monte Carlo iterations, where M is the number of observations per day and T is the number of days. Highlighted values represent the lowest rejection frequencies for a given column of entries See Section 10 for complete details.

Table 24: DGP 2 - Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.45	0.40	0.35	0.50	0.46	0.44	0.52	0.37	0.36	0.55	0.41	0.43
BPV	0.52	0.40	0.30	0.56	0.48	0.42	0.59	0.45	0.32	0.63	0.55	0.38
TPV	0.55	0.43	0.39	0.64	0.49	0.42	0.68	0.54	0.36	0.71	0.54	0.44
TSRV	0.78	0.82	0.80	0.81	0.86	0.83	0.90	0.91	0.86	0.93	0.92	0.91
MSRV	0.66	0.64	0.51	0.70	0.65	0.57	0.79	0.67	0.56	0.84	0.73	0.63
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.45	0.40	0.35	0.50	0.46	0.44	0.52	0.37	0.36	0.55	0.41	0.43
TBPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MedRV	0.56	0.41	0.30	0.63	0.46	0.40	0.74	0.50	0.35	0.81	0.60	0.43
MinRV	0.53	0.41	0.31	0.60	0.46	0.44	0.67	0.56	0.38	0.71	0.62	0.39

*Notes : See notes of Table 23.

Table 25: DGP 3-Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.83	0.79	0.83	0.83	0.81	0.84	0.96	0.96	0.94	0.97	0.96	0.96
BPV	0.66	0.61	0.53	0.69	0.67	0.58	0.83	0.74	0.68	0.84	0.77	0.70
TPV	0.63	0.48	0.41	0.65	0.55	0.49	0.76	0.58	0.44	0.79	0.68	0.53
TSRV	0.91	0.82	0.84	0.92	0.85	0.84	0.98	0.96	0.95	0.98	0.96	0.95
MSRV	0.87	0.87	0.83	0.88	0.89	0.86	0.94	0.96	0.93	0.95	0.96	0.94
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.53	0.39	0.36	0.58	0.45	0.47	0.63	0.50	0.31	0.72	0.53	0.38
MBV	0.90	0.84	0.92	0.90	0.84	0.92	0.97	0.95	0.96	0.97	0.96	0.96
TBPV	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00
MedRV	0.52	0.41	0.25	0.56	0.49	0.34	0.63	0.51	0.30	0.65	0.59	0.36
MinRV	0.58	0.41	0.36	0.61	0.49	0.36	0.69	0.57	0.39	0.71	0.64	0.49

*Notes : See notes of Table 23.

Table 26: DGP 4- Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.82	0.83	0.80	0.83	0.85	0.83	0.95	0.96	0.96	0.96	0.97	0.95
BPV	0.71	0.66	0.56	0.78	0.69	0.60	0.85	0.77	0.74	0.87	0.80	0.81
TPV	0.64	0.53	0.45	0.67	0.60	0.48	0.73	0.68	0.54	0.76	0.73	0.60
TSRV	0.89	0.85	0.86	0.90	0.87	0.87	0.98	0.96	0.95	0.98	0.96	0.95
MSRV	0.93	0.84	0.81	0.93	0.86	0.83	0.97	0.97	0.95	0.97	0.97	0.97
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.54	0.41	0.22	0.57	0.46	0.36	0.62	0.50	0.31	0.70	0.59	0.37
TBPV	0.98	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00
MedRV	0.54	0.45	0.40	0.56	0.48	0.45	0.60	0.52	0.36	0.66	0.56	0.46
MinRV	0.64	0.43	0.42	0.68	0.50	0.53	0.69	0.61	0.45	0.73	0.64	0.54

*Notes : See notes of Table 23.

Table 27: DGP 5-Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.62	0.62	0.51	0.68	0.67	0.59	0.77	0.71	0.67	0.81	0.72	0.72
BPV	0.55	0.43	0.39	0.66	0.53	0.42	0.68	0.52	0.37	0.78	0.59	0.42
TPV	0.56	0.47	0.30	0.60	0.53	0.42	0.65	0.54	0.39	0.70	0.59	0.48
TSRV	0.75	0.79	0.77	0.78	0.82	0.77	0.92	0.89	0.94	0.95	0.90	0.94
MSRV	0.69	0.68	0.59	0.74	0.70	0.64	0.84	0.76	0.73	0.87	0.78	0.78
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.57	0.47	0.31	0.62	0.55	0.41	0.69	0.52	0.25	0.74	0.58	0.40
TBPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MedRV	0.50	0.42	0.30	0.57	0.47	0.34	0.63	0.51	0.33	0.65	0.60	0.41
MinRV	0.57	0.47	0.38	0.66	0.50	0.48	0.70	0.59	0.45	0.74	0.67	0.56

*Notes : See notes of Table 23.

Table 28: DGP 6 - Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.60	0.61	0.57	0.65	0.67	0.65	0.73	0.72	0.72	0.78	0.75	0.77
BPV	0.58	0.54	0.46	0.66	0.60	0.53	0.63	0.54	0.43	0.69	0.60	0.47
TPV	0.54	0.36	0.33	0.55	0.47	0.42	0.67	0.52	0.40	0.72	0.62	0.48
TSRV	0.78	0.75	0.78	0.80	0.78	0.80	0.90	0.90	0.92	0.94	0.91	0.92
MSRV	0.64	0.67	0.56	0.68	0.71	0.64	0.77	0.73	0.72	0.85	0.77	0.73
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.53	0.46	0.23	0.58	0.51	0.30	0.67	0.53	0.29	0.73	0.62	0.39
TBPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SRK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MedRV	0.50	0.44	0.34	0.59	0.46	0.40	0.60	0.57	0.35	0.68	0.60	0.47
MinRV	0.63	0.45	0.45	0.69	0.47	0.52	0.72	0.61	0.46	0.74	0.66	0.52

*Notes : See notes of Table 23.

Table 29: DGP 7-Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.29	0.23	0.14	0.36	0.35	0.24	0.44	0.27	0.17	0.53	0.27	0.24
BPV	0.33	0.29	0.25	0.39	0.37	0.32	0.53	0.40	0.17	0.59	0.46	0.20
TPV	0.42	0.28	0.25	0.48	0.32	0.28	0.61	0.40	0.28	0.67	0.44	0.34
TSRV	0.78	0.77	0.80	0.81	0.81	0.83	0.93	0.95	0.95	0.94	0.95	0.97
MSRV	0.47	0.40	0.34	0.61	0.41	0.43	0.74	0.59	0.45	0.82	0.65	0.53
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.29	0.23	0.14	0.36	0.35	0.24	0.44	0.27	0.17	0.53	0.27	0.24
TBPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MedRV	0.40	0.34	0.22	0.54	0.42	0.25	0.51	0.33	0.25	0.56	0.43	0.31
MinRV	0.35	0.33	0.28	0.40	0.42	0.36	0.54	0.49	0.24	0.63	0.56	0.36

*Notes : See notes of Table 23.

Table 30: DGP 8 - Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.31	0.17	0.15	0.39	0.30	0.19	0.39	0.26	0.10	0.49	0.30	0.20
BPV	0.40	0.26	0.19	0.44	0.38	0.19	0.54	0.26	0.15	0.59	0.37	0.20
TPV	0.44	0.32	0.22	0.49	0.36	0.29	0.60	0.34	0.15	0.66	0.44	0.22
TSRV	0.77	0.77	0.80	0.83	0.78	0.83	0.94	0.94	0.95	0.95	0.94	0.97
MSRV	0.60	0.44	0.34	0.63	0.49	0.39	0.77	0.59	0.40	0.80	0.70	0.47
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.31	0.17	0.15	0.39	0.30	0.19	0.39	0.26	0.10	0.49	0.30	0.20
TBPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MedRV	0.36	0.27	0.21	0.44	0.36	0.31	0.49	0.37	0.14	0.61	0.46	0.22
MinRV	0.43	0.28	0.20	0.48	0.35	0.27	0.55	0.40	0.18	0.65	0.45	0.28

*Notes : See notes of Table 23.

Table 31: DGP 9 - Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.79	0.81	0.81	0.81	0.84	0.85	0.94	0.95	0.96	0.95	0.97	0.96
BPV	0.55	0.44	0.34	0.63	0.56	0.48	0.78	0.63	0.47	0.84	0.70	0.57
TPV	0.52	0.38	0.19	0.58	0.42	0.24	0.72	0.40	0.25	0.73	0.47	0.30
TSRV	0.89	0.90	0.94	0.91	0.91	0.95	0.93	0.96	0.96	0.95	0.97	0.96
MSRV	0.88	0.82	0.83	0.88	0.83	0.85	0.96	0.95	0.97	0.97	0.96	0.97
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.44	0.30	0.20	0.51	0.34	0.25	0.62	0.37	0.24	0.66	0.42	0.36
TBPV	0.95	1.00	1.00	0.95	1.00	1.00	0.98	1.00	1.00	0.98	1.00	1.00
MedRV	0.41	0.28	0.24	0.48	0.32	0.34	0.55	0.42	0.24	0.60	0.47	0.30
MinRV	0.48	0.35	0.16	0.52	0.40	0.25	0.62	0.45	0.28	0.67	0.51	0.35

*Notes : See notes of Table 23.

Table 32: DGP 10 - Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.80	0.85	0.83	0.81	0.86	0.86	0.94	0.94	0.94	0.95	0.96	0.94
BPV	0.62	0.43	0.41	0.66	0.53	0.44	0.79	0.65	0.56	0.81	0.72	0.64
TPV	0.45	0.44	0.20	0.51	0.49	0.30	0.68	0.41	0.28	0.77	0.48	0.36
TSRV	0.88	0.88	0.93	0.90	0.88	0.94	0.94	0.96	0.96	0.96	0.96	0.96
MSRV	0.86	0.82	0.88	0.87	0.82	0.90	0.97	0.95	0.98	0.97	0.96	0.98
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.44	0.23	0.19	0.49	0.29	0.23	0.57	0.37	0.17	0.67	0.43	0.24
TBPV	0.95	1.00	1.00	0.95	1.00	1.00	0.98	1.00	1.00	0.98	1.00	1.00
MedRV	0.41	0.22	0.20	0.45	0.36	0.29	0.57	0.37	0.23	0.61	0.45	0.29
MinRV	0.47	0.35	0.21	0.48	0.38	0.28	0.63	0.48	0.18	0.67	0.58	0.28

*Notes : See notes of Table 23.

Table 33: DGP 11 - Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.49	0.41	0.31	0.51	0.48	0.37	0.63	0.53	0.40	0.66	0.63	0.50
BPV	0.37	0.31	0.23	0.45	0.34	0.23	0.54	0.37	0.19	0.60	0.48	0.28
TPV	0.41	0.30	0.28	0.47	0.37	0.34	0.56	0.34	0.22	0.59	0.44	0.33
TSRV	0.76	0.80	0.77	0.79	0.83	0.79	0.92	0.92	0.88	0.92	0.93	0.88
MSRV	0.59	0.46	0.45	0.66	0.53	0.53	0.74	0.70	0.63	0.81	0.74	0.66
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.46	0.30	0.23	0.50	0.38	0.25	0.59	0.34	0.20	0.65	0.42	0.31
TBPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MedRV	0.44	0.29	0.27	0.48	0.34	0.36	0.51	0.39	0.24	0.60	0.48	0.32
MinRV	0.43	0.32	0.20	0.48	0.40	0.25	0.60	0.45	0.29	0.66	0.54	0.37

*Notes : See notes of Table 23.

Table 34: DGP 12 - Confidence Interval Estimation Accuracy

Measures	Interval = $\hat{\mu}_{IV} \pm 0.125\hat{\sigma}_{IV}$						Interval = $\hat{\mu}_{IV} \pm 0.250\hat{\sigma}_{IV}$					
	5%			10%			5%			10%		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
RV	0.44	0.41	0.38	0.50	0.49	0.43	0.61	0.57	0.50	0.68	0.62	0.56
BPV	0.42	0.33	0.21	0.49	0.40	0.28	0.52	0.37	0.18	0.63	0.44	0.22
TPV	0.40	0.31	0.26	0.49	0.39	0.36	0.57	0.36	0.20	0.63	0.43	0.26
TSRV	0.76	0.77	0.78	0.81	0.81	0.79	0.92	0.91	0.89	0.92	0.94	0.90
MSRV	0.61	0.51	0.48	0.65	0.55	0.54	0.79	0.70	0.62	0.84	0.75	0.69
RK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TRV	0.42	0.27	0.20	0.47	0.34	0.27	0.55	0.34	0.18	0.62	0.43	0.32
TBPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MedRV	0.44	0.26	0.23	0.49	0.38	0.31	0.57	0.31	0.24	0.63	0.44	0.27
MinRV	0.43	0.31	0.20	0.45	0.36	0.28	0.58	0.44	0.18	0.65	0.57	0.26

*Notes : See notes of Table 23.

Table 35: Best Performing IV Estimators in Monte Carlo Experiments

	M_{78}	M_{156}	M_{312}
DGP 1	RV & TRV	RV & TRV	RV & TRV
DGP 2	RV & TRV	RV & TRV	BPV
DGP 3	MedRV	TRV	MedRV
DGP 4	MedRV	TRV	TRV
DGP 5	MedRV	MedRV	TRV
DGP 6	MedRV	TPV	TRV
DGP 7	RV & TRV	RV & TRV	RV & TRV
DGP 8	RV & TRV	RV & TRV	RV & TRV
DGP 9	MedRV	TRV	TPV
DGP 10	MedRV	TRV	TRV
DGP 11	BPV	TRV	BV & TRV
DGP 12	BPV & TRV	TRV	BPV

*Note: See notes to Table 23. Tabulated entries denote the “best” performing IV estimators, based on confidence interval estimation accuracy, as reported in Tables 23-34, for all 12 different DGPs, and for 3 different sampling frequencies, including $M = 78$ (5 mins), $M = 156$ (2.5 mins) and $M = 312$ (1.25 mins).

Table 36: Percentage of Jump Days

DGP	$\alpha = 0.05$			$\alpha = 0.01$		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
DGP 1	13.56	12.74	13.37	4.59	4.76	4.71
DGP 2	13.28	12.87	13.81	4.72	4.78	4.90
DGP 3	73.56	78.84	83.26	67.4	74.6	79.7
DGP 4	73.29	78.66	83.20	67.2	74.3	79.5
DGP 5	31.11	38.91	48.57	19.6	28.9	39.2
DGP 6	30.73	38.66	47.89	19.8	28.5	37.8
DGP 7	13.77	12.84	13.54	4.77	4.75	4.83
DGP 8	13.52	12.98	13.01	4.88	4.22	4.73
DGP 9	72.14	77.67	82.21	65.7	72.9	78.5
DGP 10	71.98	77.51	82.06	65.6	72.9	78.2
DGP 11	29.45	36.82	46.32	17.8	26.6	36.5
DGP 12	29.24	36.59	45.77	17.7	25.9	35.7

*Notes: See notes to Table 35. Tabulated entries denote the percentage of jump days in each of the 12 DGPs and 3 sampling frequencies. Jump tests are carried out using significance levels α at 5% and 10%. Percentage of jump days is calculated as discussed in Section 6.1.

Table 37: Relative Contribution of Jump Component to Total variation

DGP	$\alpha = 0.05$			$\alpha = 0.01$		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
DGP 1	2.149	1.352	0.927	0.916	0.629	0.413
DGP 2	2.103	1.404	0.958	0.951	0.645	0.428
DGP 3	42.76	44.67	46.07	41.54	44.12	45.76
DGP 4	42.58	44.37	45.46	41.37	43.78	45.12
DGP 5	8.891	9.793	10.85	6.823	8.501	10.02
DGP 6	8.801	9.657	10.49	6.810	8.342	9.591
DGP 7	2.020	1.371	0.948	0.960	0.634	0.436
DGP 8	2.169	1.362	0.904	0.992	0.627	0.405
DGP 9	40.88	42.79	44.20	39.59	42.15	43.87
DGP 10	40.74	42.52	43.66	39.45	41.91	43.31
DGP 11	8.167	8.930	9.900	6.066	7.626	9.033
DGP 12	8.126	8.795	9.623	6.020	7.441	8.719

*Notes: See notes to Table 36. Entries denote the average ratio of jump variation to total variation (percentage) for each of the 12 DGPs and 3 sampling frequencies (M), with jump test significance levels of α at 5% and 10%.

Table 38: Relative Contribution of Large Jump Component to Total Variation

DGP	$\alpha = 0.05$			$\alpha = 0.01$		
	M_{78}	M_{156}	M_{312}	M_{78}	M_{156}	M_{312}
DGP 1	0.143	0.072	0.030	0.079	0.039	0.014
DGP 2	0.139	0.062	0.030	0.077	0.029	0.011
DGP 3	3.102	3.112	3.139	3.102	3.112	3.139
DGP 4	3.109	3.100	3.113	3.109	3.100	3.113
DGP 5	1.685	1.677	1.740	1.667	1.677	1.740
DGP 6	1.642	1.663	1.697	1.642	1.663	1.697
DGP 7	0.136	0.072	0.028	0.072	0.026	0.013
DGP 8	0.133	0.073	0.029	0.073	0.033	0.013
DGP 9	3.080	3.075	3.121	3.080	3.075	3.121
DGP 10	3.077	3.116	3.092	3.077	3.116	3.092
DGP 11	1.595	1.607	1.645	1.555	1.645	3.046
DGP 12	1.574	1.578	1.625	1.542	1.578	1.625

*Notes: See notes to Table 37. Entries denote the average ratio of “large” jump variation to total variation (percentage), for each of the 12 DGPs and 3 sampling frequencies (M), with jump test significance levels of α at 5% and 10%.