Volatility estimation and forecasts based on price durations*

Seek Young Hong† Ingmar Nolte‡ Stephen J. Taylor§ Xiaolu Zhao¶

This version: 26 January 2021

Abstract

We investigate price duration variance estimators that have long been neglected in the literature. In particular, we consider simple-to-construct non-parametric duration estimators, and parametric price duration estimators using autoregressive conditional duration specifications. This paper shows i) how price duration estimators can be used for the estimation and forecasting of the integrated variance of an underlying semi-martingale price process and ii) how they are affected by discrete and irregular spacing of observations, market microstructure noise and finite price jumps. Specifically, we contribute to the literature by constructing the asymptotic theory for the non-parametric estimator with and without the presence of bid/ask spread and time discreteness. Further, we provide guidance about how our estimators can best be implemented in practice by appropriately selecting a threshold parameter that defines a price duration event, or by averaging over a range of non-parametric duration estimators. We also provide simulation and forecasting evidence that price duration estimators can extract relevant information from high-frequency data better and produce more accurate forecasts than competing realized volatility and option-implied variance estimators, when considered in isolation or as part of a forecasting combination setting.

Keywords: Price durations; Volatility estimation; High-frequency data; Market microstructure noise; Forecasting.

---

*We are grateful to Federico M. Bandi (the Editor), the Associate Editor and two anonymous referees, whose valuable comments have greatly improved our paper. We would also like to thank Torben Andersen, Dobrislav Dobrev and Oliver Linton, as well as participants of the faculty seminars at Aarhus, Cambridge, Kent, Reading and the Bank of England, the ESEM 2016, Konstanz-Lancaster 2015 and ESRC-NWDTC 2015 workshops, and the SoFiE 2016, IAAE 2016 and Financial Econometrics Lancaster 2016 conferences for helpful comments. This paper was previously circulated under the title "More Accurate Volatility Estimation and Forecasts Using Price Durations". Support from the ESRC-FWF grant 'Bilateral Austria: Order Book Foundations of Price Risks and Liquidity: An Integrated Equity and Derivatives Markets Perspective' (ES/N014588/1) is gratefully acknowledged. The usual disclaimer applies.

†Lancaster University Management School, Lancaster LA1 4YX, UK. email: s.y.hong@lancaster.ac.uk

‡Lancaster University Management School, Lancaster LA1 4YX, UK. email: i.nolte@lancaster.ac.uk.

§Lancaster University Management School, Lancaster LA1 4YX, UK. email: s.taylor@lancaster.ac.uk.

¶Contact author. International Business College, Dongbei University of Finance and Economics, Dalian 116025, China. Phone +86 411847 10248, email: zhaoxiaolu@ibc.dufe-edu.cn.

1 Introduction

Precise volatility estimates are indispensable for many applications in finance. Over the last two decades realized variance (RV) estimators of quadratic variation following Andersen, Bollerslev, Diebold & Ebens (2001) and Barndorff-Nielsen & Shephard (2002) have become the standard tool for the construction of daily variance estimators by exploiting intra-day high-frequency data. In the presence of market microstructure (MMS) noise, apart from some alternative methods (such as for example, those based on Fourier-Malliavin theory, see Malliavin & Mancino (2009) and Mancino & Sanfelici (2012)), we could probably say that there are four major approaches for the estimation of quadratic variation (QV) developed in the literature. First, the sub-sampling method of Zhang, Mykland & Aït-Sahalia (2005) and Aït-Sahalia, Mykland & Zhang (2011) combines realized volatility estimators computed on different return sampling frequencies and gives rise to the two-scale and multi-scale realized variance estimators, related to this approach is the Least Squares based IV estimation framework of Nolte & Voev (2012). Second, Barndorff-Nielsen, Hansen, Lunde & Shephard (2008) develop the class of realized kernel estimators. Third, Podolskij & Vetter (2009), Jacod, Li, Mykland, Podolskij & Vetter (2009) and Christensen, Oomen & Podolskij (2014) introduce the pre-averaging based realized volatility estimators. Fourth, Xiu (2010) develops the class of QML integrated variance estimators, which is later extended to the multivariate case by Shephard & Xiu (2017). Bandi & Russell (2011) investigate finite sample properties of the kernel estimators and their optimal implementation. Liu, Patton & Sheppard (2015) compare the accuracy of these and further estimators across multiple asset classes.

The observation error is also an important issue in volatility estimation with high frequency data. While the underlying price process is assumed to be continuous, prices are in reality observed at discrete times. Furthermore, it is also sensible to think of the time points as random stopping times, or at least unregularly observed and possibly also endogenous to the underlying price process. See Aït-Sahalia & Jacod (2014) for the general overview, and Jacod (2008) and Vetter & Zwingmann (2017) for the detailed econometrics. Recently, Bandi, Pirino & Renò (2017) and Bandi, Kolokolov, Pirino & Renò (2020) introduce the notion of idleness and staleness in the price changes. These have very important implications in volatility measurement under the semimartingale assumption as implied by the fundamental theory of asset pricing.

This paper studies an alternative approach to volatility measurement and forecasting, based on *price duration*. We consider both simple-to-construct non-parametric estimators and parametric price duration estimators with autoregressive conditional duration (ACD) specifications. In contrast to the GARCH, realized variance and option-implied variance estimators, the price duration approach has received very little attention in the literature so far.

Indeed, two hands suffice to count the studies carried out on duration based methods, with the first detailed paper being Cho & Frees (1988). Earlier research focusses on parametric approaches, for example Engle & Russell (1998) and Gerhard & Hautsch (2002), which consider ACD specifications to govern the price duration dynamics. With a parametric assumption for the dynamic price duration process, not only an integrated variance estimator but also a local (intra-day, spot) variance estimator can be obtained, as pointed out by Tse & Yang (2012). All three ACD studies start from a point process concept to construct volatility estimators, but provide little guidance on the practical task of selecting a good price change threshold when MMS noise effects are present, which is important for implementation. Pelletier & Wei (2019) recently propose an intraday spot volatility estimator by specifying stochastic models for the price durations and volatility simultaneously. Li, Nolte & Nolte (2021) study a point process-based approach via Markov-Switching Autoregressive Conditional Intensity models. A notable but neglected working paper by Andersen, Dobrev & Schaumburg (2008) proposes a non-parametric price duration variance estimator similar to ours but without the averaging feature. They show that duration estimators are more efficient than noise-robust realized volatility estimators for price diffusions.

Theoretical and empirical justifications for the duration based methods have been even more scarce than the limited number of papers written. In this paper we aim to fill the gap in the literature, and support the validity of their uses. Specifically, we establish the asymptotic theory for the non-parametric duration based estimator when the underlying process is an Itô semimiartingale. We also investigate how the asymptotics is influenced by the presence of some microstructure noise, observation errors of discrete and irregular forms and finite activity jumps. Further, we discuss practical ways to appropriately choose the threshold parameter, which determines the size of the price change that defines the event times. In addition, we show that the performance of price duration estimators can be further improved by averaging over a range of price duration estimators with different threshold values. Our simulation study shows that, in general, price duration estimators produce lower Root Mean Squared Errors (RMSE) and QLIKE loss values than competing realized volatility-type estimators. This is the case in setups with constant and stochastic volatilities as well as those with noise and observation errors. Within a forecasting analysis we provide evidence for Dow Jones Industrial Average (DJIA) index stocks that price duration variance estimators, especially a parametric price duration estimator and an averaging non-parametric estimator, extract relevant information from (high-frequency) data better, and produce more accurate variance forecasts, than competing realized volatility-type and option-implied variance estimators, when considered either in isolation or as part of a forecasting combination.

Speaking further of the non-parametric price duration approach, we remark that there are some different strands of research papers that can be related to ours. Fukasawa (2010a,b), Fukasawa & Rosenbaum (2012), Li, Zhang & Zheng (2013) and Li, Mykland, Renault, Zhan & Zheng (2014) study the RV estimators with respect to various stochastic sampling times. In some particular special cases (regular grid), the estimators overlap with the non-parametric duration estimator we consider, when there is no bid/ask spread and time discreteness. It is shown in their papers that realized volatility-type estimators on an appropriate stochastic sampling grid are asymptotically more efficient than calendar time analogues using a comparable number of observations. Recently Li, Nolte & Nolte (2019) provide asymptotic results for the general class of renewal process estimators which includes price duration and also range based estimators as examples.

The parametric price duration variance estimator most similar to ours was proposed by Tse & Yang (2012) using a computationally intensive semi-parametric estimation method for an ACD specification. They show through simulation that the semi-parametric estimation method can improve upon maximum-likelihood-estimation (MLE) coupled with an Exponential distribution assumption for the scaled duration, but that the estimates are not sensitive to the choice of the computation method. We thus continue employing MLE which is straightforward to implement. Apart from using a new ACD model that can better accommodate the long-range dependence in price durations, we improve upon their parametric price duration estimator by replacing the Exponential distribution with a Burr distribution which significantly improves the density forecast results. Tse & Yang (2012) select the threshold values by targeting a desired average duration but acknowledge that an optimal choice is important. We address this issue by relating threshold choices to the level of the bid/ask spread, with the underlying assumption that the bid/ask spread can be related to the market's level of volatility, since it's well-known that spread and volatility are positively correlated. We plot using empirical data an upward-sloping then stabilizing plot for the price duration variance estimates against a large range of threshold values, similar to a volatility signature plot. We show that with our threshold selection rule the resulting variance estimates reach the stabilizing region, thus balancing bias against efficiency. A forecasting study confirms our parametric estimator's superiority in predicting future stock volatilities, by comparing it with 10 established RV-type estimators and one option-implied volatility estimator.

The paper is organized in the following way. Section 2 lays out the theoretical foundations for the duration based variance estimators, and derive the asymptotic properties thereof. In particular, we establish the central limit theorems, and show how they are affected by the presence of market microstructure noise, observation errors (time discreteness) and finite activity jumps. Section 3 describes the high-frequency data used subsequently and provides descriptive results that motivate the simulation study. Section 4 contains the simulation study that assesses the effects of market microstructure noise components on our duration based variance estimators, provides guidance on the choice of a preferred price change threshold value, and compares the accuracy and efficiency of the duration based estimator with competing estimators for both constant and stochastic volatility models. Section 5 contains the empirical analysis of our estimators including a discussion on the construction of the parametric duration based variance estimators and empirical evidence on the choice of a preferred price change threshold value. Section 6 contains the forecasting study and Section 7 concludes.

2 Theoretical foundations

In Section 2.1 we provide the theoretical foundations for parametric and non-parametric duration based variance estimators. Section 2.2 studies the asymptotic properties of our non-parametric estimator in the presence of time discreteness, market microsturcture noise such as the bid/ask spread and finite activity jumps.

2.1 Duration based integrated variance estimation

Suppose the efficient log-price process X_t is a continuous Itô semimartingale defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ represented by

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{1}$$

where W_t is a standard Brownian motion and μ_t and σ_t are $(\mathcal{F}_t)$-adapted and locally-bounded. These assumptions are sufficient for suppressing the drift term using standard methods via Girsanov's theorem, see Mykland & Zhang (2009). In this subsection we assume continuous time observations on the price process; the issue of time discreteness is discussed later in Section 2.2.

Let n be the parameter that defines the observation frequency and derives the asymptotics. For each trading day and a selected threshold δ, a set of event times $\{\tau_{n,j}\}_{j\in\mathbb{Z}_+\cup\{0\}}$ is defined in terms of absolute cumulative price changes exceeding δ. For asymptotic derivations, we suppose that the sequence of thresholds $\delta = \delta_n \to 0$ as $n \to \infty$. To elaborate, we are considering a random sampling scheme of hitting times defined as $\tau_{n,0} = 0$ and

$$\tau_{n,j+1} := \inf\Big\{t > \tau_{n,j};\ \Big|X_t - X_{\tau_{n,j}}\Big| \geq \delta_n\Big\}; \quad j \in \mathbb{Z}_+ \cup \{0\}. \tag{2}$$

Note that the resulting times form a sequence of strictly increasing stopping times. The total number of price duration events, that is the number of 'hits' up to time t, is defined by

$$N_{n,t} := \max\{j \geq 0;\ \tau_{j,n} \leq t\}. \tag{3}$$

Note that in the case of deterministic regular time sampling (calendar time sampling) for example: $\tau_{n,j} = j/n;\ j = 1, \dots, n$ over $[0, 1]$, it follows that $N_{n,t} \equiv n$ and the rate of convergence $\delta_n = n^{-1/2}$. The object of interest is the quadratic variation $[X, X]_t$ which is equal to the integrated variance a.s. (Jacod & Shiryaev (2003)):

$$[X, X]_t = \int_0^t \sigma_s^2\, ds \tag{4}$$

Note that, from hereafter, subscripts will sometimes be omitted where no confusion is likely, in particular the dependence of the processes upon the parameter n.

2.1.1 Non-parametric estimation

Below we sketch the underlying idea of our non-parametric duration based estimator and motivate its use. Let $x_{n,j} = \tau_{n,j} - \tau_{n,j-1}$ denote the time duration between consecutive events. For the conditional distribution of $x_{n,j}|\mathcal{F}_{\tau_{n,j-1}}$, we denote the density function by $f(x_{n,j}|\mathcal{F}_{\tau_{n,j-1}})$, the cumulative distribution function by $F(x_{n,j}|\mathcal{F}_{\tau_{n,j-1}})$ and the intensity (or hazard) function by $\lambda(x_{n,j}|\mathcal{F}_{\tau_{n,j-1}}) = f(x_{n,j}|\mathcal{F}_{\tau_{n,j-1}})/(1 - F(x_{n,j}|\mathcal{F}_{\tau_{n,j-1}}))$.

Following Engle & Russell (1998) and Tse & Yang (2012), duration based variance estimators rely on a relationship between the conditional intensity function and the conditional instantaneous variance of a step process. The step process $\{\widetilde{X}_t, t \geq 0\}$ is defined by $\widetilde{X}_t = X_t$ when $t \in \{\tau_{n,j}, j \geq 0\}$ and by $\widetilde{X}_t = \widetilde{X}_{\tau_{n,j-1}}$ whenever $\tau_{n,j-1} < t < \tau_{n,j}$. The conditional instantaneous variance of $\widetilde{X}_t$ equals

$$\sigma^2_{\widetilde{X},t} = \lim_{\Delta \to 0} \frac{1}{\Delta} \operatorname{var}(\widetilde{X}_{t+\Delta} - \widetilde{X}_t | \mathcal{F}_{n,j-1}), \quad \tau_{n,j-1} < t < \tau_{n,j}. \tag{5}$$

As Δ approaches zero we may ignore the possibility of two or more events between times t and $t+\Delta$, so that the only possible outcomes for $\widetilde{X}_{t+\Delta} - \widetilde{X}_t$ can be assumed to be 0, δ and $-\delta$. The probability of a non-zero outcome is determined by $\lambda(x_{n,j}|\mathcal{F}_{\tau_{n,j-1}})$ and consequently

$$\sigma^2_{\widetilde{X},t} = \delta^2 \lambda(t - \tau_{n,j-1}|\mathcal{F}_{\tau_{n,j-1}}), \quad \tau_{n,j-1} < t < \tau_{n,j}. \tag{6}$$

The integral of $\sigma^2_{\widetilde{X},t}$ over a fixed time interval provides an approximation to the integral of $\sigma^2_{X,t}$ over the same time interval, and the approximation error disappears as $\delta \to 0$.

The general duration based estimator of the integrated variance, IV, is given by

$$\begin{aligned} \widetilde{IV} &= \int_0^{\tau_{n,N}} \sigma^2_{\widetilde{X},t} dt = \sum_{j=1}^{N} \delta^2 \int_{\tau_{n,j-1}}^{\tau_{n,j}} \lambda\Big(t - \tau_{n,j-1}|\mathcal{F}_{\tau_{n,j-1}}\Big) dt \\ &= -\delta^2 \sum_{j=1}^{N} \ln\Big(1 - F\Big(x_{n,j}|\mathcal{F}_{\tau_{n,j-1}}\Big)\Big). \end{aligned} \tag{7}$$

In fact, the above estimator is ignoring price variation over the interval between the last price event of the day at time $\tau_{n,N}$ and the end of the day, $\tau_{n,eod}$, which is expected to be of minor importance when δ is relatively small. A natural bias corrected general duration based variance estimator is therefore

$$\widetilde{IV}_+ = \int_0^{\tau_{n,eod}} \sigma^2_{\widetilde{X},t} dt = -\delta^2 \sum_{j=1}^{N} \ln(1 - F(x_{n,j}|\mathcal{F}_{\tau_{n,j-1}})) + \delta^2 \int_{\tau_{n,N}}^{\tau_{n,eod}} \lambda(t - \tau_{n,N}|\mathcal{F}_{\tau_{n,N}}) dt. \tag{8}$$

In practice, we do not know the true intensity function. We must therefore either estimate the functions $\lambda(.|.)$ or we can replace the summed integrals in (7) by their expectations. Noting that these expectations are always one, we can define the following estimator which will be one of the main objects of study in this paper:

Definition 1. *The non-parametric duration based variance estimator (NPDV) over the interval $[0, t]$ is defined by*

$$NPDV_t = NPDV_t(\delta) := N_{n,t} \cdot \delta^2_n, \tag{9}$$

where $N_{n,t}$ is as defined in (3).

This equals the quadratic variation of the approximating step process over a single day, which we may hope is a good estimate of the quadratic variation of the price process over the same time interval.[1] Below we show it is indeed the case. An equation like (9), for the special case of constant volatility, can be found in the early investigations of duration based methods by Cho & Frees (1988) and Marsh & Rosenfeld (1986). We note that in Andersen et al. (2008), their approach is to estimate local volatility at each single time point, resulting in a different form of the estimator. Also, they consider the case of constant volatility in the first instance. A further discussion on the differences between our estimator and theirs is provided in Section 2.2.1.

As aforementioned in the introduction and also briefly discussed in Fukasawa (2010a), the $NPDV$ estimator overlaps with the RV in some special cases. This happens when the RV is stochastically sampled at times at which the price hits the regular (i.e. symmetric and equidistant) grid of size δ. It is however no longer the case when there is a bias due to time discretization and/or microstructure noise; Section 2.2 details how the asymptotics are affected. As for detailed discussions for the RV with respect to stochastic sampling, the interested reader is referred to in Fukasawa (2010b), Fukasawa & Rosenbaum (2012), Li, Zhang & Zheng (2013), Li et al. (2014) and references therein.

For the limiting theory, we shall impose the following condition within this section for the mesh of the sampling interval.

Assumption A. *The mesh of the sampling points satisfy the following:*

$$\max_j |\tau_{n,j+1} - \tau_{n,j}| = o_p(1). \tag{10}$$

Note that this assumption above makes the sequence of stopping times $\{\tau_{n,j}\}$ an adapted subdivision of a Riemann sequence. We therefore obtain the Law of Large Numbers for (9) in view of Jacod & Shiryaev (2003, Theorem 1.4.47), implying consistency of our estimator:

$$NPDV_t \xrightarrow{P} [X, X]_t \tag{11}$$

as $n \to \infty$ (so that $\delta_n \to 0$).

We now move on to the limiting distribution.

[1] The asymptotic downward bias introduced by ignoring end of day effects is equal to $\delta^2/6$ from Li et al. (2019). A bias corrected version of (9) is therefore given by $NPDV_+ = (N + 1/6)\delta^2$.

Theorem 1. *Suppose Assumption A holds. For all t we have the following convergence in law to a mixed normal distribution for the estimator defined in (9):*

$$\delta_n^{-1}\Big(NPDV_t - [X,X]_t\Big) \xrightarrow{\mathcal{L}} \mathcal{MN}\left(0,\ \frac{2}{3}\int_0^t \sigma_s^2\ ds\right). \tag{12}$$

Proof. See Web-Appendix C. □

Remark. The 'symmetric nature' of our sampling scheme (2) is worth noting, since without it the bias may not asymptotically vanish, as pointed out by Fukasawa & Rosenbaum (2012). We note that (12) is consistent with the limiting distribution of the realized variance and renewal estimators in the special cases where all three estimators overlap, see Fukasawa (2010b, Theorem 3.10), and Li, Nolte & Nolte (2019, Remark 4.3).

The NPDV has a lower limiting variance than that of the RV estimator sampled in business time, or in conditionally independent time (e.g. Poisson type), or in equidistant calendar time. See Hansen & Lunde (2006) and Aït-Sahalia & Jacod (2014, Chapter 9).

2.1.2 Parametric estimation

We now introduce some parametric approaches; detailed derivations of their asymptotic theories are omitted. A parametric implementation of (7) requires the selection of appropriate hazard functions $\lambda(.|.)$. As first suggested by Engle & Russell (1998), we assume the durations $x_{n,j} = \tau_{n,j} - \tau_{n,j-1}$ have conditional expectations ψ_j determined by $\mathcal{F}_{\tau_{n,j-1}}$ and that scaled durations are independent variables. More precisely,

$$x_{n,j} = \psi_{n,j}\varepsilon_{n,j},\ \text{with } \psi_{n,j} = \mathrm{E}[x_{n,j}|\mathcal{F}_{\tau_{n,j-1}}], \tag{13}$$

and the scaled durations ε_j are i.i.d., positive random variables which are stochastically independent of the expected durations ψ_j.

Autoregressive specifications for ψ_j are standard choices, such as the autoregressive conditional duration (ACD) model of Engle & Russell (1998), the logarithmic ACD model of Bauwens & Giot (2000), the augmented ACD model of Fernandes & Grammig (2006) and others reviewed by Pacurar (2008). These specifications do not accommodate the long-range dependence present in our durations data. As a practical alternative to the fractionally integrated ACD model of Jasiak (1999), we develop the heterogenous autoregressive conditional duration (HACD) model in the spirit of the HAR model for volatility introduced by Corsi (2009). Short, medium and long range effects are arbitrarily associated with 1, 5 and 20 durations, and our HACD specification is then

$$\psi_{n,j} = \omega + \alpha x_{n,j-1} + \beta_1 \psi_{n,j-1} + \beta_2(\psi_{n,j-5} + \ldots + \psi_{n,j-1}) + \beta_3(\psi_{n,j-20} + \ldots + \psi_{n,j-1}). \quad (14)$$

A flexible shape for the hazard function can be obtained by assuming the scaled durations have a Burr distribution, as in Grammig & Maurer (2000) and Bauwens, Giot, Grammig & Veredas (2004). Recently, in Pelletier & Wei (2019) a Gamma distribution is assumed for the scaled durations. Note that the general Burr density and cumulative distribution functions, as parameterized by Lancaster (1997) and Hautsch (2004), are given by

$$f(y|\xi,\eta,\gamma) = \frac{\gamma}{\xi}\left(\frac{y}{\xi}\right)^{\gamma-1}[1+\eta(y/\xi)^{\gamma}]^{-(1+(1/\eta))}, \quad y > 0, \quad (15)$$

and

$$F(y|\xi,\eta,\gamma) = 1 - [1+\eta(y/\xi)^{\gamma}]^{-1/\eta}, \quad y > 0, \quad (16)$$

with three positive parameters (ξ, η, γ). The Weibull special case is obtained when $\eta \to 0$ and its special case of an exponential distribution is given by also requiring $\gamma = 1$. The mean μ of the general Burr distribution is

$$\mu = \xi c(\eta,\gamma), \text{ with } c(\eta,\gamma) = \mathrm{B}\Big(1+\gamma^{-1}, \eta^{-1} - \gamma^{-1}\Big)/\eta^{1+(1/\gamma)}, \quad (17)$$

with $\mathrm{B}(\cdot,\cdot)$ denoting the Beta function. For each scaled duration the mean is 1 so that ξ is replaced by $1/c(\eta,\gamma)$. For each duration $x_{n,j}$ (having conditional mean ψ_j) we replace ξ by $\psi_j/c(\eta,\gamma)$. From (7) our parametric, duration based variance estimator, PDV over the period $[0,t]$, is therefore

$$PDV_t = PDV_t(\delta) = \frac{\delta^2}{\eta}\sum_{j=1}^{N_t}\log\left(1+\eta\left[c(\eta,\gamma)\frac{x_j}{\psi_j}\right]^{\gamma}\right). \quad (18)$$

When we implement (18), we take account of the intraday pattern in the durations data. The durations x_j in (13) and (14) are replaced by the scaled quantities $x_j^* = x_j/s_j$ and each expected duration ψ_{j-u} is replaced by the scaled quantity $\psi_{j-u}^* = \psi_{j-u}/s_{j-u}$, with s_{j-u} the estimated average time between events at the time-of-day corresponding to time $t_{d-\tau}$; each term s_{j-u} is obtained from a Nadaraya-Watson kernel regression, with Gaussian kernel and a data-based automatic bandwidth as proposed by Silverman (1986), of price durations against time-of-day using one month of durations data. Then ψ_j is replaced by s_j/ψ_j^*, so the scaled duration x_j/ψ_j in (18) is simply x_j^*/ψ_j^*. End of day bias correction is obtained by adding

$1/6 \cdot \delta^2$ as above. The associated log-likelihood function for the Burr case is given by

$$\log \mathcal{L}(\theta) = \sum_j \log \left(\gamma c(\eta,\gamma)^\gamma \left(\varepsilon_j^*\right)^{\gamma-1} [1 + \eta(\varepsilon_j^* c(\eta,\gamma))^\gamma]^{-(1+(1/\eta))} \right) \tag{19}$$

where $\varepsilon_j^* = x_j^*/\psi_j^*$, ψ_j^* follows the specification in (14), $c(\eta,\gamma)$ is given by (17), initial values for ψ_j^* are set to the the unconditional mean of x_j^* at the beginning of each day, and $\theta = (\omega, \alpha, \beta_1, \beta_2, \beta_3, \gamma, \eta)$ is the corresponding parameter vector.

The theoretical framework above is for the logarithms of prices. It is much easier to set the threshold to be a dollar quantity related to the magnitude of the bid/ask spread. We then replace the log-price X_t in (2) by the price $P_t = \exp(X_t)$. As a small change δ in the price is equivalent to a change δ/P_t in the log-price, we can define the following (asymptotically equivalent) alternative definitions for our estimators:

$$NPDV_t = \sum_{j=0}^{N_t} \frac{\delta_n^2}{P_{\tau_{n,j}}^2} \tag{20}$$

and

$$PDV_t = \frac{\delta_n^2}{\eta} \cdot \sum_{j=0}^{N_t} \log \left(1 + \eta \left[c(\eta,\gamma) \frac{x_j}{\psi_j} \right]^\gamma \right) / P_{\tau_{n,j}}^2. \tag{21}$$

These alternative definitions will be used in Sections 4 and beyond for computational practicality. Obviously, their asymptotic equivalence with the previous definitions can be easily seen via Taylor series expansion for logarithms.

Speaking from a practical point of view, we see that the non-parametric estimator can easily be constructed with a reasonable number of events $N \equiv N_t$ over the interval $[0,t]$, which represents a day for example. On the other hand, the additional parametric form assumption of the parametric estimator also guarantees a volatility estimator for small N and yields for example a local (intraday) volatility estimator. The end of day bias correction is now to add $\delta^2/(6P_N^2)$ to these estimators.

2.2 Time discreteness and market microstructure noise

2.2.1 Time discretization

To investigate the effect of time discretization on our estimation theory, we first suppose that the observations are first sampled according to a Poisson process, after which the hitting time sampling we reviewed in the previous section is considered. Poisson sampling is a widely implemented random sampling scheme within the high frequency framework. It is a continuous-time version of Bernoulli process modelling whose inter-arrival time is geometric. For example, Campbell, Lo & MacKinlay (1997) state that their non-synchronous trading model converges to the continuous time Poisson process (under suitable normalisation). However, due to its exogenous nature, Poisson sampling does not take the information of the price process into account. This form of limitation has been discussed in the literature, for example in Aït-Sahalia & Jacod (2014) and Li et al. (2014). The setup we consider in this subsection can be viewed as "integrating" the exogenous and endogenous aspects in the sampling procedure. This allows us to cover a wide range of plausible situations from a practical point of view, but as a cost to pay the rate of convergence slows down.

Suppose the stopping times $\mathcal{I}_n = \{\tau_{n,0}, \tau_{n,1}, \dots, \}$ follow a Poisson process with intensity Δ. For the asymptotics, the sequence of positive numbers $\Delta = \Delta_n$ is set to approach to zero, since it represents the time to next "arrival", which should decrease as the number of sample increases. We then consider time points chosen according to the hitting time procedure we defined above. That is, the sampling times under consideration are given by

$$\tau^*_{n,j+1} := \inf \left\{ t \in \mathcal{I}_n > \tau^*_{n,j};\ \left| X_t - X_{\tau^*_{n,j}} \right| \geq \delta_n \right\}; \quad j \in \mathbb{Z}_+ \cup \{0\}. \tag{22}$$

Therefore, the total number of hitting times become $N^*_{n,t} := \max\{j \geq 0; \tau^*_{j,n} \leq t\}$, and our non-parametric estimator becomes

$$NPDV_t = N^*_{n,t} \cdot \delta_n^2. \tag{23}$$

Since we are essentially selecting those times which satisfy the "criteria" out of the sample set $\mathcal{I}_n$, the generalized thinning argument of Poisson processes applies. Therefore, it follows that (conditionally on $\mathcal{F}_{\tau_{n,j}}$) we have

$$\tau^*_{n,j+1} - \tau^*_{n,j} \sim \Delta_n^2 v_{\tau_{n,j}} e, \tag{24}$$

where $e \sim \exp(1)$ and

$$v_{\tau_{n,j}} = P\left(\left| X_{\tau_{n,j+1}} - X_{\tau_{n,j}} \right| > \delta_n \right). \tag{25}$$

The sequence of positive $(\mathcal{F}_t)$-measurable processes (v_n) can be understood as representing the "likelihood" of the Poisson time points being selected.

Lemma 1. *Suppose σ_t in (1) is bounded above by some positive constant σ_* for all t. Then* *it follows that* $v_{\tau_{n,j}} = O_P(\Delta_n^3/\delta_n)$, *and the sampling points* $\{\tau^*_{n,j}\}_j$ *(24) satisfy Assumption A when* $\Delta_n = o(\delta_n^{1/3})$.

Proof. This is proved along with Theorem 2 in Web-Appendix C. Note that the boundedness of σ is a often improsed in the literature as an innocuous assumption, see for example, Li et al. (2014). $\square$

Furthermore, with standard renewal arguments, it is rather straightforward to see that the following structural assumption below (Assumption B) holds for the sequence of stopping times $\{\tau^*_{n,j}\}_j$. The implication is that $\tau^*_{n,j}$ is now $\mathcal{F}'_{t-1}$ measurable while still being defined on the original space Ω.

Assumption B. The hitting time $\tau^*_{n,j+1}$ defined in (22) is an $(\mathcal{F}'_t)$-stopping time where the σ-field $\mathcal{F}'$ is bigger than or equal to $\mathcal{F}$ defined in (1), and the probability measure P is defined on $\mathcal{F}'$.

The limiting distribution now follows. Note that (26) can be broadly seen as a generalized version of the CLT of Andersen et al. (2008, page 17), where a directional local volatility estimator is considered on a pre-defined fixed time grid.

Theorem 2. *Suppose* σ_t *in (1) is bounded above by some positive constant* σ_* *for all* t, *and* $\Delta_n = O(\delta_n^{3/5})$. *Assume also that there exists a non-vanishing càdlàg process* v, *which is the probability limit of the* $(\mathcal{F}_t)$*-progressively measurable process* v_n *defined in (25). Then under Assumption B, for all* t *we have the following convergence in law to a mixed normal distribution for the estimator (23):*

$$\Delta_n^{-1}\Big(NPDV_t - [X,X]_t - \mathcal{B}_{n,t}\Big) \xrightarrow{\mathcal{L}} \mathcal{MN}\left(0,\ 2\int_0^t \sigma_s^4 v_s\ ds\right), \tag{26}$$

where $\mathcal{B}_{n,t}$ *is the bias term due to time descretization, which is of order* $O_P(\Delta_n^5/\delta_n^2)$.

Proof. See Web-Appendix C. $\square$

2.2.2 Market microstructure noise

Next we consider the bid/ask spread, which is arguably one of the most important market microstructure noise components for transaction price datasets. We discuss how it affects the limiting theory of the non-parametric duration based volatility estimator.

Specifically, assume that at general times t we observe a contaminated noisy log price

$$Y_t = \log P_t + \frac{1}{2}\mathcal{D}_t \cdot \varsigma \quad \left(\equiv X_t + \frac{\varsigma}{2}\mathcal{D}_t\right), \tag{27}$$

where ς denotes2 the size of the bid/ask spread and P_t is the unobserved true price. Further, $\mathcal{D}_t$ is the binary variable that takes the value of 1 when Y_t represents the log of an ask price at time t, and -1 when Y_t represents the log of a bid price at time t.

Microstructure noise as a whole is often modeled as some additive random error ϵ_t, see Zhang et al. (2005), Hansen & Lunde (2006), Bandi & Russell (2008). Our assumption (27) can be understood as a detailed specification thereof for addressing the bid/ask spread, i.e. $\epsilon_t = (\varsigma/2)\mathcal{D}_t$. In the context of stochastic sampling for the realized volatility, Fukasawa (2010a) deals with the bid-ask spread component of a different specification.

We assume that ς is constant throughout the day and does not depend on n, which is consistent with the standard practice of modelling $\mathrm{E}(\epsilon_t)$ as fixed. As Zhang (2011) and Aït-Sahalia and Jacod (2014, 7.1.1) discuss, a shrinking noise asymptotics is sometimes, albeit relatively uncommonly, considered e.g. to examine the bias-variance tradeoff in detail. As such we also consider the case where ς depend on n and shrinks asymptotically, i.e. $\varsigma = \varsigma_n (\to 0)$, see below.

We suppose that Y_t takes prices on the ask side with probability p_a and on the bid side with probability $p_b(= 1-p_a)$. For example, we may set them to be 0.5 so that both situations would happen with equal probability. Note that we do not consider the time discretization issue for expositional simplicity in this subsection.

We set the price events to occur at $\{\tau'_{n,j}\}_{j=1}^{N'_{n,t}}$, where $N'_{n,t}$ denotes the total number of hitting times over the interval $[0, t]$. In line with our previous definition, these hitting times are now alternatively defined as follows:

$$\left|Y_{\tau'_{n,j+1}} - Y_{\tau'_{n,j}}\right| = \left|\left(X_{\tau'_{n,j+1}} - X_{\tau'_{n,j}}\right) + \frac{1}{2}\varsigma\left\{\mathcal{D}_{\tau'_{n,j+1}} - \mathcal{D}_{\tau'_{n,j}}\right\}\right| \geq \delta. \tag{28}$$

The setting suggests that an event is triggered by the combined magnitudes of the unobserved efficient price change component $(X_{\tau'_{n,j+1}} - X_{\tau'_{n,j}})$ and the bid/ask spread component $1/2 \cdot \{\mathcal{D}_{\tau'_{n,j+1}} - \mathcal{D}_{\tau'_{n,j}}\}$. The bid/ask spread component can therefore take three values: either

$-1, 0$, or 1, which together with an upward (downward) move of the log price component constitutes the following three possible scenarios:

I. Bid-Bid or **Ask-Ask**

When both the first price and the last price of the price duration lie on the same side of the limit order book, i.e. bid-bid or ask-ask, we note that the contribution from the bid-ask spread becomes zero. This means the price component itself solely triggers a price event in our construction, and the noise induced by the presence of bid-ask spread cancels out automatically:

$$\begin{aligned} \mathrm{NPDV}_t \equiv N'_{n,t}\delta_n^2 &= N'_{n,t}\left[\left(Y_{\tau'_{n,j+1}} - Y_{\tau'_{n,j}}\right)\right]^2 \\ &= N'_{n,t}\left[\left(X_{\tau'_{n,j+1}} - X_{\tau'_{n,j}}\right)\right]^2. \end{aligned} \tag{29}$$

So, we can see the robustness of our estimator to the noise in the modelling procedure. Asymptotic theory follows straightforwardly.

II. Bid-Ask

In this case, the preceeding binary variable $\mathcal{D}.$ in (28) takes the value of -1 while the other is 1. It turns out that the magnitude of the noise ς from the spread plays a key role as we sketch now. We see that

$$\mathrm{NPDV}_t \equiv N'_{n,t}\delta_n^2 = N'_{n,t}\left[\left(X_{\tau'_{n,j+1}} - X_{\tau'_{n,j}}\right) + \varsigma\right]^2 \tag{30}$$

by construction. In the meantime, applying the triangle inequality on (28) we see that

$$\left|X_{\tau'_{n,j+1}} - X_{\tau'_{n,j}}\right| \geq \delta - \varsigma, \tag{31}$$

with an implicit assumption that $\delta \geq \varsigma^2$. This suggests that the noise induced in this situation is equivalent to requiring a "lower bar" (of $\delta - \varsigma$) for the unobserved price process in triggering a price event.

Standard arguments relating to the limiting behavior of the conditional first moment of price change imply that the leading bias term comes from the term $\varsigma^2 \cdot N'_{n,t}$ which is positive. This highlights the possible opposite roles that ς and δ play to the

[2] In practice δ will always be chosen to be larger than ς. We discuss the case $\delta < \varsigma$ in the context of the simulation study in Section 4.

magnitude of bias. Specifically, the bias increases in ς, while it decreases in δ (as an increase in δ reduces N).

As for the asymptotics where each term is set to depend upon n, we recall that $N'_{n,t} = O_p(\delta_n^{-2})$, so we require at least $\varsigma = \varsigma_n = o(\delta_n)$ in order to eliminate the bias, see proof of Theorem 3 for details. Note that in Theorem 3 below, we impose a slightly weaker condition of $O(\delta_n)$ so that the explicit form of the limiting bias can be specified. Obviously, this can be smoothed away upon imposing a stronger condition $\varsigma_n = o(\delta_n)$, for example.

III. Ask-Bid

This case is similar to the second situation. The preceeding binary variable $\mathcal{D}.$ in (28) takes the value of $+1$ while the other is -1. So it follows that

$$\text{NPDV}_t \equiv N'_{n,t}\delta_n^2 = N'_{n,t}\left[\left(X_{\tau'_{n,j+1}} - X_{\tau'_{n,j}}\right) - \varsigma\right]^2, \tag{32}$$

and similarly as before,

$$\left|X_{\tau'_{n,j+1}} - X_{\tau'_{n,j}}\right| \geq \delta + \varsigma. \tag{33}$$

The leading bias term is still $\varsigma^2 \cdot N'_{n,t}$ (and the same arguments for the asymptotics therefore applies). So we expect an upward bias in the estimation, although the downward bias contribution from the cross term in (32) suggests that the magnitude of the bias can be slightly lower than the "bid-ask" case in the finite sample. Roughly speaking, this is in line with the "higher bar" we end up requiring in (33).

Note that even in the second and third cases, we can see that the noise induced by bid-ask spreads can be easily tracked down.

Remark. As aforementioned, for the limiting theory below we let ς depend upon n. This is to control the asymptotic bias in relation to the threshold parameter δ_n. We note that in the asymptotic distribution below, if $\varsigma = \varsigma_n$ is chosen to be of order $o(\delta_n)$ the asymptotic bias in (34) vanishes in the limit, implying certain degree of robustness of our estimator to bid/ask spread (of this particular specification).

Theorem 3. *Suppose the NPDV estimator is defined according to the hitting time scheme (28), and suppose $\varsigma = \varsigma_n = C_\varsigma \delta_n$ for some positive constant C_ς. Then, for all t we have the* *following convergence in law to a mixed normal distribution for the estimator (23):*

$$\delta_n^{-1}\Big(NPDV_t - [X,X]_t\Big) \xrightarrow{\mathcal{L}} \mathcal{MN}\left(\frac{1}{2}C_\varsigma^2 \cdot \Big\{1-(p_a-p_b)^2\Big\},\ \frac{2}{3}\int_0^t \sigma_s^2\ ds\right) \tag{34}$$

as $n \to \infty$.

Proof. See Web-Appendix C. □

Lastly, let us further consider the possibility of having finite number of jumps, and consider the case where a jump of size κ occurs over the interval $[\tau'_{n,}, \tau'_{n,j+1}]$; i.e.

$$\left|Y_{\tau'_{n,j+1}} - Y_{\tau'_{n,j}}\right| = \left|\Big(X_{\tau'_{n,j+1}} - X_{\tau'_{n,j}}\Big) + \frac{1}{2}\varsigma\left\{\mathcal{D}_{\tau'_{n,j+1}} - \mathcal{D}_{\tau'_{n,j}}\right\} + \kappa\right| \tag{35}$$

As we expect $|\kappa| \gg \delta$, a price jump would most likely trigger an immediate price event. Yet its impact on the integrated variance estimator is substantially mitigated as only one duration event is caused and thus κ is effectively truncated. In addition, as the occurrences of large jumps are rare, we expect them to have very limited influence on the duration based variance estimator. However, when we have a finite number of many small jumps of size similar to δ, they are likely to affect the estimator. In the asymptotics, it is straightforward to see that they would appear in the form of the squared sum of κ's.

In the simulation study in Section 4, we further evaluate the performance of our duration based variance estimators under different market microstructure noise scenarios. To obtain some representative input parameters for this study we first report a descriptive analysis of our high-frequency data.

3 Data properties

In the empirical analysis we use 20 of the 30 stocks of the Dow Jones Industrial Average (DJIA) index. The tick-by-tick trades and quotes data spanning 11 years (2769 trading days) from January 2002 to December 2012 are obtained from the New York Stock Exchange (NYSE) TAQ database and are time-stamped to a second. All 20 stocks have their primary listing at NYSE.[3]

[3]From the list of 30 DJIA stocks as of December 2012, BAC, CSCO, CVX, HPQ, PFE, TRV, UNH, and VZ are excluded because of incomplete data samples. INTC and MSFT are excluded because their primary listing is at NASDAQ.

On 1 August, 2006, NASDAQ started to operate as an independent registered national securities exchange for NASDAQ listed securities, separate from the National Association of Securities Dealer (NASD) which has

The raw data is cleaned using the methods of Barndorff-Nielsen, Hansen, Lunde & Shephard (2009). Data entries, trades and quotes, that meet one or more of the following conditions are deleted: 1) entries outside of the normal 9:30am to 4pm daily trading session; 2) entries with either bid, ask or transaction price equal to zero; 3) transaction prices that are above the ask price plus the bid/ask spread or below the bid price minus the bid/ask spread; 4) entries with negative bid/ask spread; 5) entries with spread larger than 50 times the median spread of the day; 6) entries for which the mid-quote deviates by more than 10 mean-absolute-deviations from a rolling centered median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after). When multiple transaction, bid or ask prices have the same time stamp, the median price is used. We match trades with corresponding bid and ask quotes using a refined Lee and Ready algorithm as outlined in Nolte (2008), which yields the bid/ask spreads.

The list of stocks and their descriptive statistics for the whole sample period are presented in Table 1. Table 1 shows means and medians for bid/ask spreads and inter-trade times, as well as means for the prices and volatilities, sorted in ascending order of their mean spread level in the first column. The mean values of daily average bid/ask spreads range from 1.3 to 3.7 cents, and from 4.15 to 9.08 seconds for average inter-trade times. The corresponding medians range from 1 to 3 cents, and 3.83 to 7.98 seconds, respectively, implying right-skewed distributions for both variables. Table 1 also presents means and medians for a simple measure of a jump frequency. A jump is recorded when the absolute value of a price change exceeds five times the average bid/ask spread for a given day. The mean is 0.3 to 1.89 while the median is 0 to 1 jump per day. We also observe that the average level of volatility across the whole sample period lies between 16% and 32%, while the average price level ranges from \$26 to \$99. We observe that the average bid/ask spread is roughly increasing with the average price level. In our empirical analysis we select four reference stocks on the basis of their bid/ask spread levels: Home Depot (HD), McDonald's (MCD), American Express (AXP), and International Business Machines (IBM).

To obtain an idea of the time variation of the key variables, we plot (log) bid/ask spread, (log) inter-trade time, and (log) annualized volatility calculated using (20) for AXP from 2002 to 2012 in Figure 1. We observe that periods of higher volatility coincide with periods of wider bid/ask spreads and more frequent trades. We observe very much the same pattern

a different set of trading and reporting rules. This break is also reflected in the TAQ data, which started to record trades and quotes with exchange code "Q" instead of "T". Before 1 August, 2006, the average numbers of jumps (recorded when the absolute value of a price change is larger than 5 times the day's average spread) per day for INTC and MSFT are 8.08 and 10.37, and after that date the averages are 0.97 and 1.29 jumps per day. Given this structural break for both stocks, we decided to exclude them from the data sample.

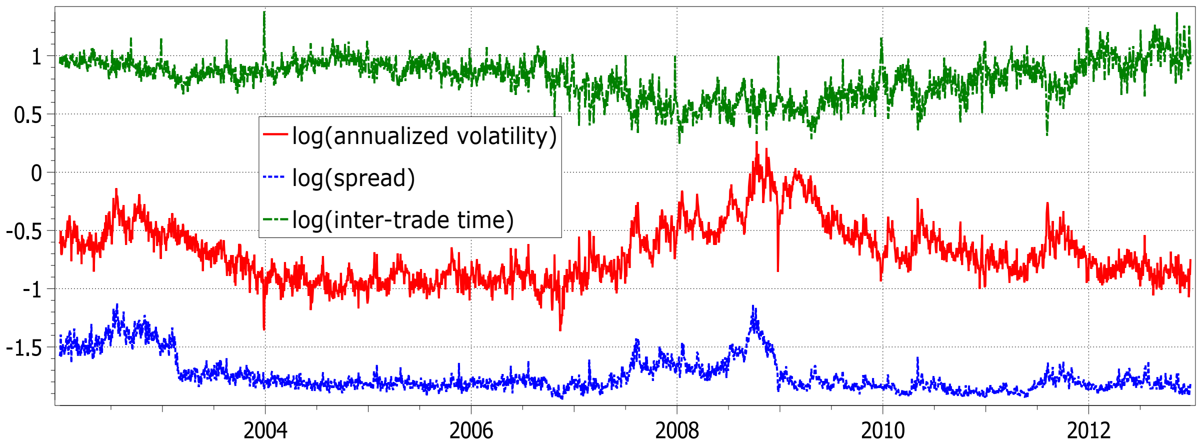
Table 1: Descriptive statistics for 20 DJIA stocks

Stock	bid/ask spread		inter-trade times		number of jumps		price	volatility
	mean	median	mean	median	mean	median	mean	mean
T	0.013	0.01	7.74	7.98	0.56	0.00	26.49	0.23
GE	0.014	0.01	5.66	4.91	0.30	0.00	28.93	0.26
DIS	0.015	0.01	6.60	6.52	0.84	0.00	29.16	0.25
HD	0.016	0.01	6.15	5.91	0.77	0.00	35.49	0.25
AA	0.016	0.01	9.08	7.60	0.62	0.00	26.54	0.32
KO	0.017	0.01	6.74	6.56	0.96	0.00	49.53	0.16
JPM	0.017	0.01	4.97	5.02	0.97	0.00	38.68	0.32
MRK	0.017	0.01	6.68	6.23	1.26	0.00	38.18	0.21
MCD	0.018	0.02	7.19	6.94	1.01	0.00	50.58	0.20
WMT	0.018	0.01	5.84	5.17	1.04	0.00	52.88	0.18
XOM	0.019	0.02	4.15	3.83	1.27	0.00	67.30	0.21
JNJ	0.019	0.01	6.12	5.87	1.23	0.00	62.04	0.16
DD	0.019	0.02	7.66	7.19	1.01	0.00	43.93	0.23
AXP	0.019	0.02	6.84	6.85	1.22	0.00	48.01	0.29
PG	0.020	0.02	6.24	6.07	1.35	1.00	63.57	0.16
BA	0.025	0.02	7.57	7.09	1.65	1.00	64.33	0.23
UTX	0.027	0.02	7.96	7.47	1.89	1.00	70.65	0.20
CAT	0.028	0.03	7.27	6.42	1.18	1.00	71.24	0.25
MMM	0.030	0.02	7.83	7.36	1.65	1.00	82.36	0.18
IBM	0.037	0.03	5.95	5.44	1.51	1.00	99.10	0.19

Notes: Descriptive statistics for the daily average bid/ask spread (in dollars), the daily average time between consecutive transactions (in seconds), the number of large price jumps per day, the transaction price, and the annualized volatility. A "large jump" is recorded when the absolute value of a price change exceeds 5 times the average bid/ask spread of the day. "Volatility" is calculated using (20) and then converted to an annualized standard deviation.

for all other NYSE stocks in our sample.

Figure 1: Bid/ask spread, inter-trade times and volatility for American Express (AXP)



Notes: Time series of logs of inter-trade time, volatility, and bid/ask spread from 2002 to 2012. Bid/ask spread is the average spread in dollars per day and inter-trade time is the average duration per day (in seconds). The annualized volatility is calculated using (20).

In Section 4, we carry out a comprehensive simulation study to analyze the properties of the duration based variance estimators. We will consider as a benchmark a scenario with 25% annualized volatility and 6 seconds average inter-trade time, which corresponds approximately to the average levels in Table 1. Likewise, to assess the effect of bid/ask spreads, we will consider scenarios with varying spreads from 1 to 4 cents.

4 Simulation results

We first assess the effects of market microstructure (MMS) as well as price jumps on the non-parametric duration estimator assuming constant volatility. Then we compare the bias and accuracy of duration and RV-type estimators for a variety of well-known stochastic volatility processes.

4.1 Constant volatility case

We separate the MMS noise into time-discretization (Δ), bid/ask spread (ς), and price-discretization components. We investigate the separate and combined effects of the noise components as well as jumps on the non-parametric duration based volatility estimator, $NPDV$ (hereafter NP for convenience), in a Monte Carlo study with 10000 replications. The performance of NP depends on the selection of the threshold value δ. Following the discussion of the two main sources of noise, bid/ask spread and time-discretization, we discuss the tradeoff between efficiency and bias in the context of choosing a preferred threshold value δ^*.

4.1.1 Time-discretization

We first look at the effect of time discretization we studied in Section 2.2.1. For practical implementation, let us consider a discrete-time setting in which we estimate the integrated variance over one trading day. We first discretize the diffusion process on a half-second interval so that there are 46800 efficient returns from a normal distribution in a 6.5-hour daily trading session. Upon this foundation process, we sample trade points according to random Bernoulli distributions with probabilities 1/2, 1/6, and 1/12, resulting in three further time-discretized processes with average inter-trade times equal to 1, 3, and 6 seconds respectively. Note that this is in line with the previously-discussed time discretization via Poisson sampling in the continuous case.

For any time-discretized process we now let Δ denote its expected inter-trade time. Specifically, we employ the following simulation setup to obtain the discretized versions of the log-price X_t in (1) and its corresponding price P_t:

$$\begin{aligned}
X_0 &= \ln(P_0), && (36)\\
X_s &= X_{s-1} + \sigma_X \sqrt{1/46800} Z_s, \quad \text{for } s = 1, \ldots, 46800, && (37)\\
B_s &\sim Bernoulli(1/(2 \cdot \Delta)), && (38)\\
N_s^B &= \sum_{j=1}^{s} B_j, && (39)\\
X_{t_i} &= X_{\inf\{s | N_s^B = i\}}, && (40)\\
P_{t_i} &= \exp(X_{t_i}), && (41)
\end{aligned}$$

where t_i for $i = 1, \ldots, I$ denotes the time stamps of the discretized log-price process with expected inter-trade time equal to Δ, I the number of observations on a given day, Z_s a standard normally distributed random variable, σ_X the constant value of the daily volatility and P_0 the initial price. We denote by $\mathcal{I} := \{t_1, t_2, \ldots, t_I\}$.

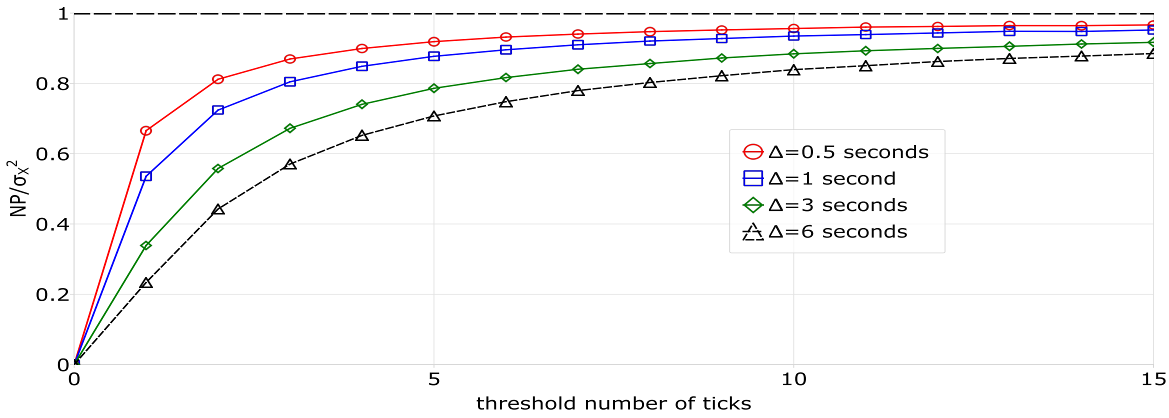
The estimator NP can now be defined as in (22) with respect to $\mathcal{I}$:

$$NP_t = N_t^* \delta^2, \tag{42}$$

where N^* is the number of duration event (within I) chosen according to the hitting time scheme (so that $N_s^* \leq N_s^B$ for all s). In Figure 2, the average values of NP divided by the true integrated variance $\int_0^1 \sigma_s^2 ds \equiv \sigma_X^2$ are plotted against the threshold value δ, for Δ ranging within 0.5, 1, 3, and 6 seconds. The annualized volatility of efficient log-prices is 25% throughout Section 4.1, using 252 trading days per year, although the magnitude of volatility is irrelevant for Figure 2.

Time-discretization decreases the number of duration events observed, due to the absence of prices that might have defined price events. As Δ decreases, the number of trades N^* increases and NP approaches its "true value" (which occurs when prices are observed continuously), see (25). Thus, given δ, a smaller Δ leads to more accurate estimates of the integrated variance represented by the unit line in Figure 2. On the other hand, increasing δ for a given Δ reduces the bias introduced by time-discretization, see (24). Note the tradeoff between δ and N^*, and the asymptotic convergence of δ to the average of the difference in selected log prices.

Figure 2: The time-discretization bias



Notes: NP variance estimates divided by σ_X^2. Average inter-trade times Δ are 6, 3, 1, and 0.5 seconds from the bottom to the top. $\sigma_X = 0.25$ per year. Thresholds δ are from 0 to 15 ticks. $P_0 = 50$, tick size = 0.01.

4.1.2 Bid/ask spread and time-discretization

As discussed in Section 2.2.2, when time is not discretized, the introduction of a bid/ask spread and corresponding bid and ask transaction prices bias the duration variance estimates upwards. Also, as we remarked before, the bias increases with the size of the spread ς, and decreases with the threshold value δ (assuming $\delta > \varsigma$).

Throughout the remainder of Section 4 we consider Δ equal to 6 seconds, with bid and ask transaction prices generated by

$$Y_t = \log P_t + \frac{1}{2}\mathcal{D}_t \cdot \varsigma. \tag{43}$$

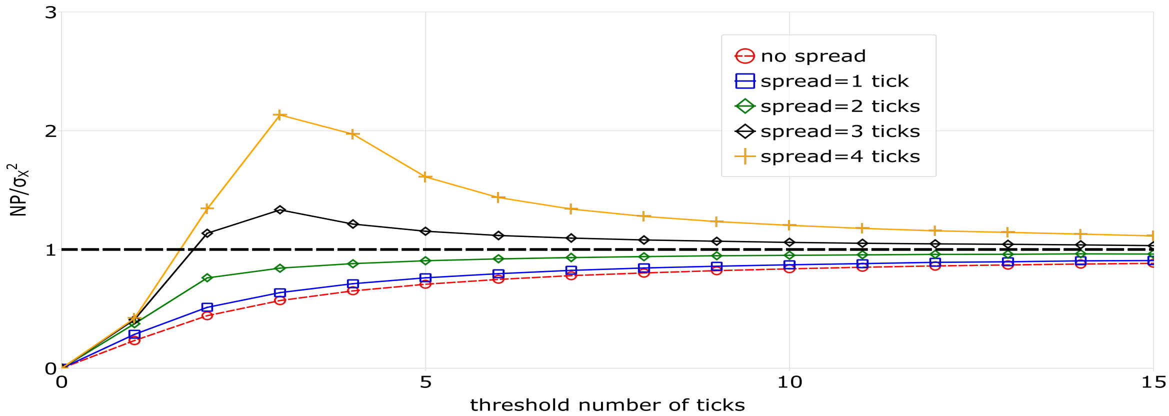
The transaction price takes either the bid or the ask side with probability 0.5 (i.e. $p_A = 1/2 = p_B$) and the variables $\mathcal{D}.$'s are i.i.d. over the time index. Note here the subtle difference between ς, a proxy for the bid/ask spread component of noise, and the real bid/ask spread which is measured in ticks. ς represents a difference between observed and efficient log-prices, and thus is essentially a return measure. But in our simulation setting, the relation between ς and spread is quite straightforward. Given an initial price $P_0 = 50$, and no drift, a rough relationship follows: $\varsigma \approx spread/50$, where $spread = 0.01, 0.02, etc.$[4] All graphs are plotted with the bid/ask spread measured in cents, matching the threshold values (expressed in ticks as well).

Figure 3 shows ratios of the average NP variance estimates over the true integrated variance. A deviation from the unit line indicates a bias. The hump-shaped curves occur

[4] ς is sometimes used inter-changably with "spread" to explain ideas in the text.

as a result of the bid/ask spread component bias when the spread is relatively large. When $\delta < \varsigma$, one bid/ask bounce is enough to trigger a price event and N is inflated in comparison to the case when $\varsigma \to 0$ (dotted line). N does not decrease much as δ increases as long as $\delta < \varsigma$, causing the NP estimate, $N\delta^2$, to increase rapidly. When δ further increases, to $\delta > \varsigma$, the influence of bid/ask bounces is mitigated by the price changes from the efficient price component as a price event is now increasingly caused by the cumulative efficient price changes rather than by the bid/ask spread component. The bid/ask spread has the largest influence at or near the point where $\delta = \varsigma$.

Figure 3: Combined effects of spread and time-discretization biases



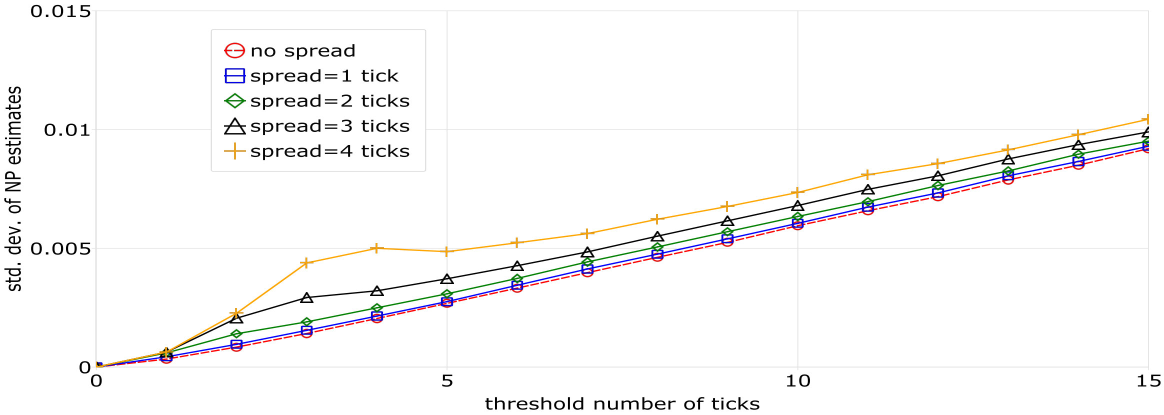
Notes: NP variance estimates divided by σ_X^2, with the range of thresholds from 0 to 15 ticks. Bid/ask spreads from bottom to the top are 0 to 4 ticks. $\sigma_X = 0.25$ per year. Δ is 6 seconds. $P_0 = 50$, tick size=0.01.

As δ increases past ς, the NP estimates start to stabilize, since both the time-discretization and the bid/ask spread biases are reduced by larger threshold values of δ. We observe two scenarios: 1) for smaller bid/ask spread levels (here 1 and 2 ticks) the negative bias contribution of the time-discretization is partially off-set by the positive contribution of the bid/spread components and the curves in Figure 3 for these cases tend to the unit line from below; 2) for larger bid/ask spread levels (here 3 and 4 ticks) the negative bias contribution of the time-discretization is, as discussed above, clearly dominated by the positive contribution of the bid/ask spread component and the curves in Figure 3 for these cases tend after the initial hump to approach the unit line from above.

4.1.3 Bias versus efficiency: the preferred threshold value

We must choose a threshold level δ for the implementation of our estimators. Figure 3 shows that the bias of the NP estimator decreases for a large enough threshold value, regardless of the bid/ask spread level. But, increasing the threshold level will inevitably result in a decreasing number of price events over the course of a day, rendering the NP estimates more dispersed and hence less efficient. Figure 4 shows this effect, as the standard deviation of the NP variance estimates is seen to increase over the range of δ from 0 to 15 ticks.

Figure 4: Standard deviations of the NP variance estimator



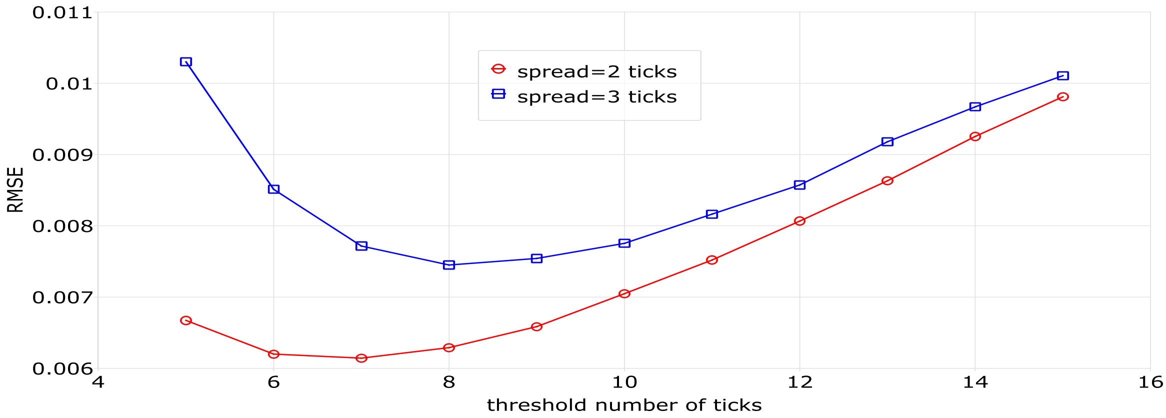
Notes: Standard deviations of the NP variance estimates over the range of thresholds from 0 to 15 ticks. Bid/ask spreads from bottom to the top are 0 to 4 ticks. $\sigma_X = 0.25$ per year. Δ is 6 seconds. $P_0 = 50$, tick size=0.01.

To illustrate this trade-off we present in Figure 5 root mean squared error (RMSE) statistics for the NP estimator over the range of δ from 5 to 15 ticks, for 2-tick and 3-tick bid/ask spread levels. These are realistic bid/ask spread levels as shown in Table 1. For the 2-tick bid/ask spread case, the minimum RMSE lies at $\delta^* = 7$ ticks, while for the 3-tick spread case, the minimum is given for $\delta^* = 8$ ticks. As these minimum RMSE values increase with the size of the bid/ask spread, we suggest for practical implementations to choose a preferred threshold δ^* equal to 2.5 to 3.5 times the bid/ask spread. A threshold in the range of 3 to 6 times the log-spread is recommended in Andersen et al. (2008) for a different duration based estimator. Further guidance about the choice of δ^* for real data on the basis of bias-type curves, similar to those in Figure 3, is presented in Section 5.2.

4.1.4 Price-discretization

Transaction prices are recorded as multiples of a minimum tick size, usually one cent. To account for this additional price-discretization component of market microstructure noise in our simulation study we now consider a setup in which, in addition to the above, bid and ask prices and consequently transaction prices are recorded discretely as multiples of 0.01 (one tick). First we obtain mid-quote prices M_{t_i} by rounding the efficient price P_{t_i} to the nearest half-cent price (50.005, 50.015, etc.) when $\varsigma/0.01$ is an odd number and to the nearest cent when $\varsigma/0.01$ is an even number. The resulting transaction prices are then given by replacing

Figure 5: Plot of RMSE as a function of the threshold value



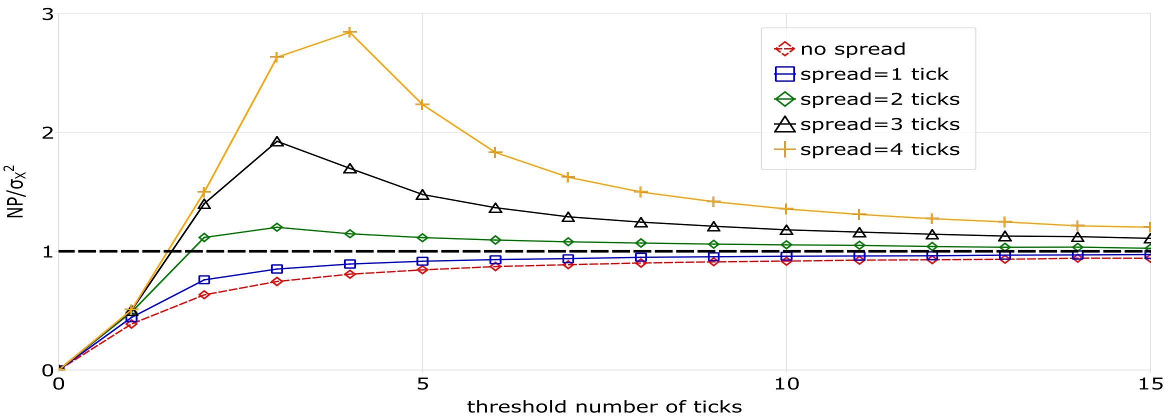
Notes: RMSE of the NP variance estimates over the range of threshold from 5 to 15 ticks. Bid/ask spreads are 2 and 3 ticks. $\sigma_X = 0.25$ per year. Δ is 6 seconds. $P_0 = 50$, tick size=0.01.

(43) by

$$
\begin{aligned}
M_{t_i} &= \begin{cases} [100P_{t_i}]/100 & \text{if } 100\varsigma \text{ is even,} \\ 1/200 + [100P_{t_i} - 0.5]/100 & \text{if } 100\varsigma \text{ is odd,} \end{cases} && (44) \\
Y_{t_i} &= M_{t_i} + 0.5 1_{t_i}\varsigma, && (45)
\end{aligned}
$$

where $[x]$ is the integer nearest to x. Figure 6 shows that price-discretization further increases the NP estimates compared to Figure 3. The general effects of bid/ask spreads and time-discretization are, however, unchanged and the estimates still tend to the unit line as δ increases beyond ς.

Figure 6: Including price-discretization noise

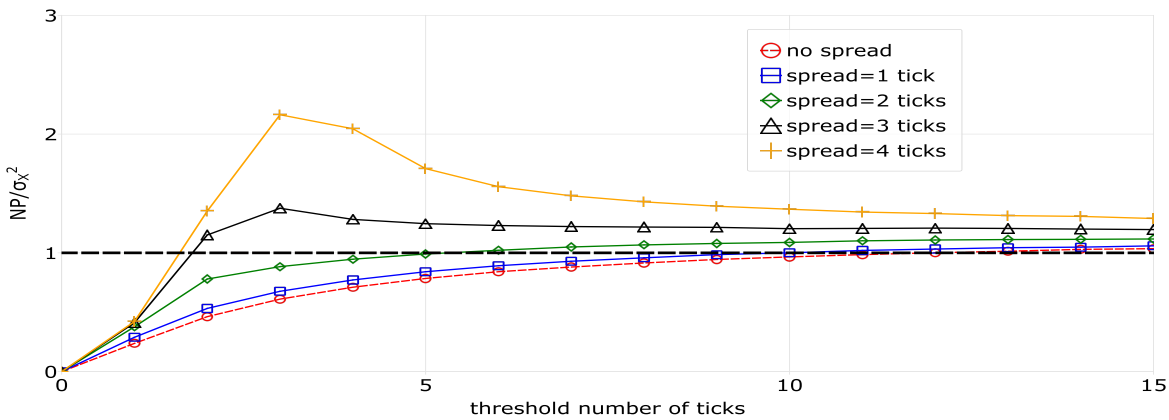


Notes: NP variance estimates divided by σ_X^2. Prices are multiples of one tick. Bid/ask spreads from bottom to the top are 0 to 4 ticks. Δ is 6 seconds. Thresholds are from 0 to 15 ticks. $\sigma_X = 0.25$ per year. $P_0 = 50$, tick size=0.01.

4.1.5 Jumps

To investigate how jumps affect our duration based variance estimators, we adapt the simulation setup of Section 4.1.2. The jumps are normally distributed with mean zero and expected jump variation equal to 20% of the quadratic variation. Jumps are simulated to arrive according to a Poisson process. We consider two scenarios: 1) one large jump on average and 2) 100 small jumps on average during a day.

Figure 7: 100 small jumps a day



Notes: NP variance estimates divided by σ_X^2. The discretization interval is 6 seconds on average. There are on average 100 small jumps a day, with a total variance of 20% of the integrated variance. Bid/ask spreads from bottom to the top are 0 to 4 ticks. The discretization interval is 6 seconds on average. Thresholds are from 0 to 15 ticks. $\sigma = 0.25$ per year. $P_0 = 50$, tick size=0.01.

As discussed in Section 2.2, due to a truncation of price changes at δ, rare large jumps are expected to have little influence on the duration based variance estimates and indeed in scenario 1) there is no visible impact[5] as N is large and an increase of one potential additional price event, triggered by an expected single large jump, results only in a tiny upward bias of the NP estimator in the order of $1/N$. In scenario 2) the standard deviation of the jump size is 3.5 ticks. Here, on the contrary, we do observe in Figure 7 that small jumps increase the integrated variance estimates by around 16.3% in comparison to the no jump case. In this case estimates are inflated considerably as small jumps are mixed with the diffusion price changes and effectively increase the number of price events by a non-trivial amount.

In reality we expect there to be less than one large jump per day, to which the duration based estimator is robust, and at most only a small number of detectable smaller jumps per day. Studies focussing on the detection of large jumps find on average less than one jump per week (e.g. Andersen, Bollerslev & Dobrev (2007)). Lee & Hannig (2010) investigate

[5] We omit the graph for brevity.

the occurrence of big and small jumps and find approximately 0.3 big jumps and 0.6 small jumps per day for individual stocks. Nonetheless, if the number of jumps is known (or can be estimated) a bias correction for jumps can readily be obtained.

4.2 Stochastic volatility processes

We also consider three stochastic processes that are commonly used to incorporate stochastic volatility (SV) into high-frequency simulations, for example as in Huang & Tauchen (2005) and Barndorff-Nielsen et al. (2008). These processes are special cases of a general jump-diffusion process introduced in Chernov, Gallant, Ghysels & Tauchen (2003). All the simulated processes have expected annualized quadratic variation equal to 0.0625.

The first is a one-factor SV model, without jumps (SV1F):

$$dX_t = \sigma_t dW_t, \tag{46}$$

$$\sigma_t = \exp(\beta_0 + \beta_1 \tau_t), \tag{47}$$

$$d\tau_t = \alpha \tau_t dt + dB_t. \tag{48}$$

The SV parameters are selected to give a standard deviation of log-volatility equal to 0.4 and a half-life for log-volatility equal to 63 trading days (3 months). With t measured in trading days, we obtain $\beta_0 = -4.311$, $\beta_1 = 0.05934$, and $\alpha = -0.011$. Like Huang & Tauchen (2005), we set $corr(dW_t, dB_t) = -0.3$. Each day, the initial value of τ_t is drawn from its unconditional distribution, which is $N(0, -0.5/\alpha)$.

The second model is SV1FJ, which is SV1F augmented by a Poisson jump process. We select an intensity of one jump per day and suppose the jumps are Gaussian with mean zero and with expected jump variation equal to 20% of the quadratic variation.

The third model is the two-factor SV model of Chernov et al. (2003), referred to as SV2F:

$$dX_t = \sigma_t dW_t, \tag{49}$$

$$\sigma_t = \text{s-exp}(\beta_0 + \beta_1 \tau_{1t} + \beta_2 \tau_{2t}), \tag{50}$$

$$d\tau_{1t} = \alpha_1 \tau_{1t} dt + dB_{1t}, \tag{51}$$

$$d\tau_{2t} = \alpha_2 \tau_{2t} dt + (1 + \phi \tau_{2t}) dB_{2t}, \tag{52}$$

The spline-exponential function in (50) is the usual exponential function with an appropriate polynomial function splined in at a very high value of its argument. The knot point for the spline implies a 150% annualized volatility, which is very unlikely to occur. We select some parameters by firstly supposing the two log-volatility components in (51) and (52) have approximately equal variance and respective half-lives equal to 126 and 0.5 days and others by following Huang & Tauchen (2005) and Barndorff-Nielsen et al. (2008). Our choices for the SV parameters are $\beta_0 = -4.442$, $\beta_1 = 0.04$, $\beta_2 = 0.635$, $\alpha_1 = -0.005501$, $\alpha_2 = -1.3863$, and $\phi = 0.25$, and the correlations between the increments of the Wiener processes are $corr(dW_t, dB_{1t}) = corr(dW_t, dB_{2t}) = -0.3$ and $corr(dB_{1t}, dB_{2t}) = 0$.

The persistent first factor is initialized each day by drawing from its unconditional distribution while the strongly mean-reverting second factor is simply started at zero. It is well-known (e.g. Nolte (2008)) that the bid/ask spread tends to increase as volatility increases. We report results when the bid/ask spread is the following deterministic function of the annualized volatility σ_A:

$$\varsigma = (1 + \lfloor 8\sigma_A \rfloor)/100. \tag{53}$$

The bid/ask spread is then one tick when the annualized volatility is less than 12.5%, two ticks for annualized volatilities between 12.5% and 25%, and three ticks for annualized volatilities between 25% and 37.5%. This formula is motivated by the empirical evidence in Table 1 and ensures that the minimum bid/ask spread is equal to one tick, i.e. one cent.

To implement the non-parametric duration estimator NP, we calculate the average bid/ask spread during each day and then set the threshold δ for a selected simulated day equal to a multiple of the average value of ς for that day.

We simulate 100000 days and incorporate time-discretization, price-discretization and bid/ask spreads as described in Section 4.1 and by (36) to (41), (43) and (45); we retain an average time between trades equal to six seconds.

We now consider two loss functions when evaluating the accuracy of a set of estimates of the integrated variance. These are RMSE, as before, and QLIKE from Patton (2011) given by averaging across days the values of

$$L(e, \hat{e}) = e/\hat{e} - log(e/\hat{e}) - 1, \tag{54}$$

where e is the true value of the integrated variance and $\hat{e}$ is its estimate.

Figures 8 and 9 show the values of RMSE and QLIKE for the estimator NP, for the processes SV1F and SV2F, over a range of thresholds from 1 to 10 times the average bid/ask spread. We can see that RMSE and QLIKE are minimized when the threshold is around 3 times the spread, for both processes. Also, the loss function values are near their corresponding minimum levels for the threshold range from 2 to 4 times the spread. This observation motivates the introduction of an average version of the NP estimator, called ANP, which is simply the average of a range of NP estimators, given by

$$ANP = \frac{1}{\#\mathcal{D}} \sum_{\delta \in \mathcal{D}} NPDV(\delta),$$

where $\mathcal{D}$ denotes the set of δ multipliers and $\#\mathcal{D}$ the number of elements in $\mathcal{D}$. We anticipate that ANP is more accurate than NP. We compare three non-parametric duration estimators (NP, ANP_1 and ANP_2) with five established RV-type estimators plus the standard 5-minute realized variance. For NP the threshold multiplier is 3, while ANP_1 is the average across 21 multipliers ranging from 2 to 4 with increment equal to 0.1; ANP_2 is the average across multipliers from 2 to 8 again with increment 0.1.

The five RV estimators are designed to be robust against microstructure noise, time discreteness and/or price jumps, and are all calculated from the complete record of trade prices with parameter values selected as recommended by the authors.. PAV_1 and PAV_2 are values of RV calculated from pre-averaged prices, using the equations in Christensen et al. (2014) which give very similar results to the formulae of Jacod et al. (2009). The size of the pre-averaging window is $\theta\sqrt{n}$ when there are n trades during a day. Our PAV_2 follows the cited papers by adopting the recommendation $\theta = 1$. As we find narrower windows provide more accurate estimates[6], we define PAV_1 by choosing $\theta = 0.25$.

The estimators RK and $RKNP$ are realized kernel values of RV, based upon the methods of Barndorff-Nielsen et al. (2008), computed using tick-by-tick returns. The Parzen kernel is used and two bandwidths are compared. For RK we use the optimal bandwidth of Barndorff-Nielsen et al. (2008), which requires estimates of the noise variance and the integrated quarticity. The noise variance is estimated using a sub-sampled tick-by-tick RV estimator on a dense grid of on average 5 observations, which corresponds to 30 seconds on average in calendar time, and the integrated quarticity is obtained correspondingly on a sparse equidistant grid of 50 observations, which corresponds to 5 minutes on average in calendar time. In contrast, $RKNP$ equates each day's bandwidth with the day's number of duration events as counted by NP. The final RV-type estimator is the two-scale RV of Zhang et al. (2005), denoted $TSRV$, with the fast scale using a 5 observations grid corresponding to 30 seconds on average in calender time and the sparse scale using a 50 observations grid corresponding to 5 minutes, which are the recommended sampling frequencies.

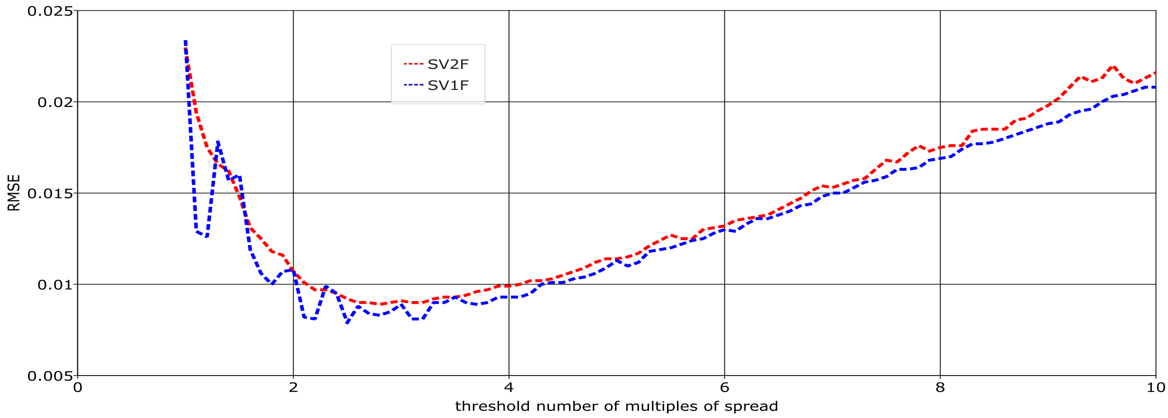
We evaluate two more volatility estimators, selected because they are robust to large price jumps. The first is the 5-minute realized bipower variation subsampled at 30-second

[6] Such simulation results are available from the authors upon request.

grids, denoted SBV, from Barndorff-Nielsen & Shephard (2006), and the second is the noise robust pre-averaged bipower variation, $PABV$, from Christensen et al. (2014), computed using tick-by-tick returns. Like the pre-averaged RV estimators, we compute $PABV_1$ with $\theta = 0.25$ and $PABV_2$ with $\theta = 1$.

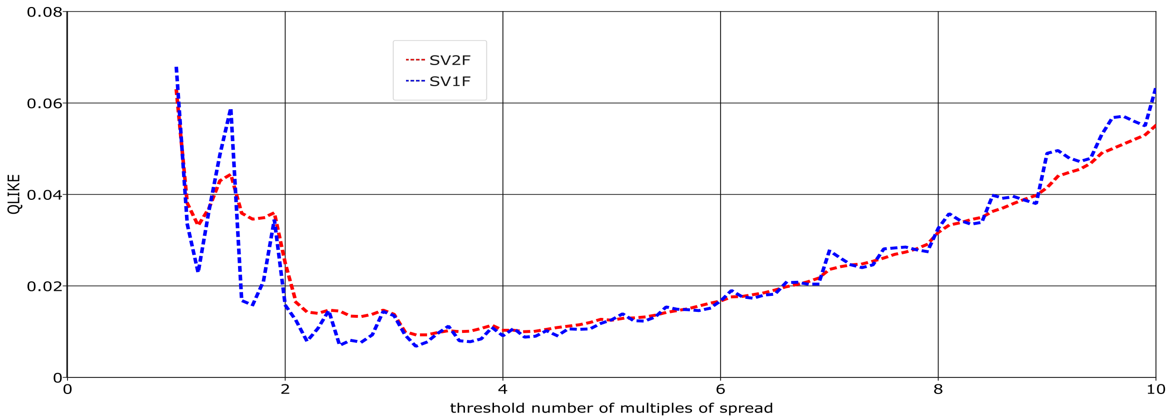
Finally, we add the simple 5-minute variance, RV_5, and its subsampled version, SRV_5, as they are popular candidates for a volatility forecasting study (see Liu et al. (2015)) which will be our major empirical application.

Figure 8: RMSE for the estimator NP when volatility is stochastic



Notes: RMSE of the NP variance estimates over thresholds that are multiples of the day's average bid/ask spread. SV1F and SV2F respectively refer to the one and two-factor stochastic volatility models defined by equations (46) to (52). The bid/ask spread is related to annualized volatility by $\varsigma = (1 + \lfloor 8\sigma_X \rfloor)/100$. Δ is 6 seconds. $P_0 = 50$, tick size=0.01.

Figure 9: QLIKE for the estimator NP when volatility is stochastic



Notes: QLIKE is the average of the loss values defined by (54). SV1F and SV2F respectively refer to the one and two-factor stochastic volatility models defined by equations (46) to (52). The bid/ask spread is related to annualized volatility by $\varsigma = (1 + \lfloor 8\sigma_X \rfloor)/100$. Δ is 6 seconds. $P_0 = 50$, tick size=0.01.

Table 2 summarizes the results from simulating 100000 days, initially for a constant annualized volatility of 25% and then for the three stochastic volatility models, referred to as SV1F, SV1FJ and SV2F. First we note that NP minimizes both RMSE and QLIKE when volatility is constant. Next we focus on comparing estimators of integrated variance using four numbers, namely RMSE and QLIKE evaluated for SV1F and SV2F. We see that ANP_1 always outperforms ANP_2 and that ANP_2 always outperforms NP. Likewise, PAV_1 is superior to PAV_2 and RK is superior to $RKNP$. $TSRV$ performs better comparably when there are jumps but is otherwise less successful. Then we note that ANP_1 and PAV_1 are each superior to RK, with ANP_1 minimizing RMSE while PAV_1 minimizes QLIKE. We conclude that the best duration estimator compares well with the best RV-type estimator.

The results for SV1FJ show the impact of occasional large jumps, occurring once a day on average and representing 20% of the quadratic variation. As expected, the duration estimators remain almost unbiased for the integrated variance but PAV_1, PAV_2, RK and $RKNP$ are upward biased by about 0.0125 which represents the expected value of the jump variation. Both $PABV_1$ and $PABV_2$ are robust to both large jumps and MMS noise, thus they perform better than all other RV-type estimators. Similar to the no jump scenarios, ANP_1 is minimizing RMSE while $PABV_1$ minimizes QLIKE.

5 Empirical analysis

We first discuss the estimation framework for the parametric duration estimator and then the choice of the threshold value for the NP estimator from an empirical point of view.

5.1 Parametric duration based variance estimator

The parametric duration based variance estimator, PDV, is implemented by choosing the Burr distribution specification described in Section 2.1. Parameter estimates are obtained by maximizing the log-likelihood function in (19) on a monthly basis. We also consider less flexible Weibull and Exponential distribution specifications for ε_d in (13), which can be obtained as special (limiting) cases from the Burr distribution. Tse & Yang (2012) use an Exponential specification for their parametric estimator. Parameter estimates are obtained for a range of threshold values δ yielding different price durations.

We perform likelihood ratio (LR), Ljung-Box (LB), and density forecast (DF) tests to assess the goodness-of-fit of the models. The LR test compares the overall model fit between two nested models on the basis of their likelihood values. The LB test has the null hypothesis of i.i.d. distributed ε_d. The DF test of Diebold, Gunther & Tay (1998) tests the null

Table 2: Simulation Results

	NP	ANP_1	ANP_2	PAV_1	PAV_2	RK	$RKNP$	$TSRV$	SBV	$PABV_1$	$PABV_2$	RV_5	SRV_5
							Constant Volatility						
Bias	.0000	-.0044	-.0033	-.0001	-.0001	-.0002	.0000	-.0075	.0007	-.0016	-.0005	.0008	.0008
StD	.0040	.0031	.0046	.0045	.0060	.0060	.0145	.0081	.0089	.0046	.0091	.0102	.0081
RMSE	.0040	.0054	.0057	.0045	.0060	.0060	.0145	.0111	.0089	.0049	.0091	.0103	.0082
QLIKE	.0021	.0044	.0049	.0026	.0047	.0048	.0288	.0228	.0098	.0034	.0112	.0135	.0083
							SV1F						
Bias	.0020	.0001	-.0001	-.0002	-.0000	-.0002	.0000	-.0076	.0007	-.0017	-.0005	.0009	.0008
StD	.0087	.0059	.0079	.0061	.0083	.0083	.0171	.0132	.0120	.0065	.0124	.0140	.0110
RMSE	.0089	.0059	.0079	.0061	.0083	.0083	.0171	.0152	.0120	.0067	.0124	.0140	.0110
QLIKE	.0135	.0037	.0056	.0026	.0048	.0051	.0187	.0232	.0100	.0035	.0114	.0136	.0085
							SV1FJ						
Bias	.0071	.0020	.0020	.0124	.0125	.0123	.0125	.0037	.0058	.0011	.0043	.0134	.0134
StD	.0087	.0057	.0080	.0330	.0339	.0335	.0381	.0310	.0174	.0085	.0172	.0377	.0363
RMSE	.0112	.0060	.0082	.0353	.0361	.0357	.0401	.0312	.0183	.0085	.0178	.0400	.0387
QLIKE	.0230	.0055	.0069	.0325	.0342	.0342	.0449	.0337	.0158	.0044	.0153	.0417	.0377
							SV2F						
Bias	.0004	.0004	.0002	-.0002	-.0000	-.0002	.0001	-.0071	.0005	-.0016	-.0005	.0009	.0008
StD	.0091	.0061	.0081	.0078	.0107	.0105	.0210	.0168	.0182	.0094	.0190	.0199	.0162
RMSE	.0091	.0061	.0081	.0078	.0107	.0105	.0210	.0182	.0183	.0095	.0190	.0200	.0163
QLIKE	.0138	.0042	.0056	.0031	.0056	.0060	.0187	.0253	.0114	.0040	.0130	.0157	.0096

Notes: Bias and standard deviation (StD) are calculated from the estimation errors, which are the estimate of the annualized variance minus the true annualized variance. RMSE is the associated root mean squared error, and QLIKE is calculated as in (54). For the constant volatility model the bid/ask spread is equal to 2 ticks. For the stochastic volatility models the bid/ask spread is linearly related to annualized volatility through $\varsigma = (1 + \lfloor 8\sigma_X \rfloor)/100$. Δ is 6 seconds on average. $P_0 = 50$, tick size=0.01.

hypothesis that the assumed distribution for ε_d is actually the true distribution and relies on a probability integral transformation of ε_d, namely the c.d.f. $F(\varepsilon_d)$, which under the null is i.i.d. $U(0,1)$ distributed. Provided that the HACD specification in (14) accommodates long-range dependence of the price durations data appropriately, and the assumed distribution for ε_d reflects the true distribution of the scaled duration, neither the LB nor the DF test should reject its null hypothesis.

All tests are performed, for each of the 132 months from January 2002 to December 2012, over a selected range of δ threshold values (between 2 to up to 20 ticks) for four reference stocks: HD, MCD, AXP and IBM. In the interest of brevity, all tests results are relegated to Web-Appendix A. The conclusion from the LR tests is unequivocal: conditional Burr distributions fit the price durations data best. As an illustration, Table 3 presents the parameter values for the Burr-HACD model for AXP in 2008, with δ equal to 12 ticks, together with LB and DF test results. As expected, we observe that, although there is some variation over the months, generally price durations are very persistent with an average β_1 equal to 0.64 and an average α equal to 0.22. The parameters η and γ have values that are significantly different from 0 and 1, respectively, which shows that the Burr specification provides a better fit than the Weibull or Exponential specifications. The LB test's p-values at lag 50 for the generalized model residuals indicate that the null hypothesis can only be rejected in 2 out of 12 cases at a 5% significance level and shows that generally the HACD specification provides a satisfactory fit. The density forecasting test's p-values reveal that the null hypothesis can be rejected in 5 out of 12 cases at the 5% level and indicates that there is scope to further improve, especially through the choice of a more flexible density function for ε_d, upon the Burr-HACD specification. The selection of more flexible densities, such as a stochastic model for durations as in Pelletier & Wei (2019), than the Burr density will probably come at the cost of losing some computational tractability and we refrain from considering them in this paper. Taken together, the fit provided by the Burr-HACD specification is good, and confirmed in Section 6 which focuses on out-of-sample forecasting comparisons.

5.2 The preferred threshold value

As discussed in Section 4.1.3, the selection of δ^* needs to take into account the tradeoff between improving efficiency and reducing bias: a larger δ reduces bias while a smaller δ improves efficiency. In the simulation study we know the true values of the integrated variance, and their RMSE statistics for appropriate simulation setups suggest that a threshold

Table 3: Illustrative parameter values and tests results: AXP, year 2008, with threshold value equal to 12 ticks

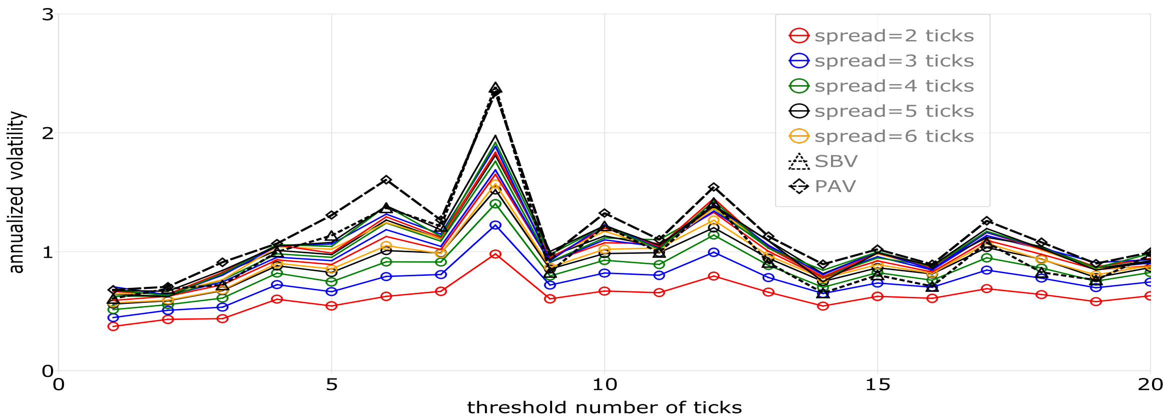
Month	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
ω	0.053	0.173	0.035	0.077	0.161	0.167	0.020	0.085	0.008	0.031	0.025	0.067
	(0.015)	(0.059)	(0.011)	(0.044)	(0.083)	(0.043)	(0.008)	(0.028)	(0.003)	(0.007)	(0.010)	(0.028)
α	0.223	0.261	0.179	0.234	0.197	0.177	0.137	0.229	0.242	0.294	0.208	0.234
	(0.027)	(0.042)	(0.027)	(0.053)	(0.067)	(0.028)	(0.019)	(0.044)	(0.026)	(0.029)	(0.033)	(0.044)
β_1	0.727	0.462	0.835	0.643	0.683	0.652	0.866	0.346	0.675	0.436	0.644	0.676
	(0.078)	(0.154)	(0.093)	(0.161)	(0.317)	(0.075)	(0.074)	(0.167)	(0.059)	(0.103)	(0.135)	(0.139)
β_2	-0.014	0.032	-0.015	0.008	0.004	0.017	-0.031	0.086	0.002	0.050	0.015	0.006
	(0.021)	(0.030)	(0.017)	(0.031)	(0.051)	(0.015)	(0.014)	(0.037)	(0.011)	(0.018)	(0.024)	(0.023)
β_3	0.004	-0.002	0.002	0.001	-0.003	-0.004	0.007	-0.004	0.004	0.000	0.003	0.000
	(0.003)	(0.003)	(0.001)	(0.005)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)	(0.001)	(0.002)	(0.003)
γ	1.396	1.344	1.449	1.377	1.238	1.432	1.536	1.391	1.274	1.316	1.554	1.532
	(0.035)	(0.053)	(0.042)	(0.063)	(0.048)	(0.048)	(0.038)	(0.058)	(0.030)	(0.027)	(0.052)	(0.068)
η	0.518	0.421	0.480	0.475	0.187	0.424	0.533	0.383	0.478	0.481	0.571	0.552
	(0.050)	(0.074)	(0.056)	(0.089)	(0.059)	(0.062)	(0.051)	(0.077)	(0.043)	(0.038)	(0.067)	(0.087)
LL	-0.834	-0.907	-0.842	-0.900	-0.917	-0.908	-0.889	-0.896	-0.755	-0.810	-0.857	-0.770
LB50	0.051	0.951	0.643	0.200	0.097	0.057	0.349	0.291	0.044	0.016	0.678	0.761
DF	0.000	0.749	0.193	0.515	0.307	0.623	0.000	0.030	0.586	0.034	0.002	0.432
obs.	2848	1416	2298	1287	1237	1668	3099	1336	3689	4884	2237	1303

Notes: The first 14 rows are the parameter estimates and robust standard errors in parentheses for the Burr-HACD model in (13), (14) and (15). LL are the average log-likelihood values (over the number of duration observations), LB50 and DF are the p-values for LB statistics (at 50 lags) and DF tests, respectively; and the last row contains the number of duration observations for each month.

value δ^* should preferably be chosen to lie within the range of 2.5 to 3.5 times the bid/ask spread. In this section we provide a number of selective empirical results that support the conclusions of the simulation study and provide further guidance on how to select a preferred threshold δ^*. But it is important to note that the threshold selection rule promoted in this study should be regarded as an empirically appropriate data-driven preference rather than a theoretically optimized value. The results presented in this section focus on the reference stock AXP.[7]

We start by considering NP variance estimates in October 2008, when volatility peaked during the financial crisis. This month is governed by high uncertainty and the average bid/ask spread level of 4.6 ticks in this month is amongst the highest in our sample period. Figure 10 plots the NP variance estimates for the first 20 trading days of October 2008 for stock AXP, over the range of threshold values from 2 ticks to 15 ticks. SBV and PAV_1 are added to the plot as benchmarks, with the former being robust to price jumps and the latter robust to MMS noise. We observe that, even during this high bid/ask spread level regime, duration based variance estimates first increase with the chosen threshold value and then stabilize, which is a stabilizing pattern that is similar to the one shown in Figure 3 for the simulation setting. We also observe that SBV estimates are very close to NP estimates, given the threshold values chosen within the stabilizing region, while PAV_1 generates higher integrated variance estimates over this volatile time period, because it is not robust to large price jumps.

Figure 10: Daily NP, SBV, and PAV_1 estimates for AXP: October 2008



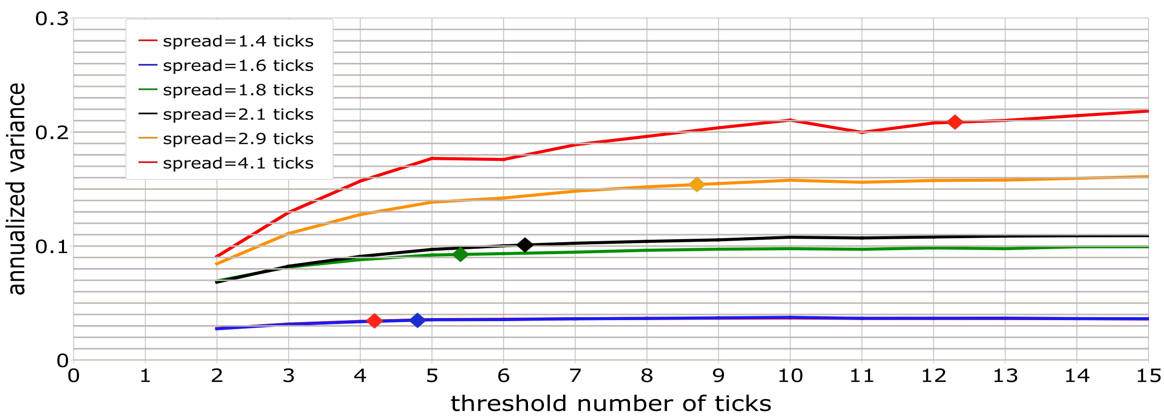
Notes: Daily NP estimates for the first 20 trading days of October 2008 for stock AXP, over the range of threshold values from 2 to 15 ticks (ordered generally from bottom to top), together with SBV and PAV_1 estimates over the same period.

The results of the simulation study suggest that estimates are less biased once stabilization

[7]Results for the other stocks are available from the authors upon request.

has been achieved and pinpointing the lower bound of this stabilizing region would provide a good trade-off between bias and efficiency and a good choice for the preferred threshold value δ^*. To obtain a better picture of this stabilizing behavior, and its relationship to the level of the bid/ask spread in reality, we consider the full data sample for AXP. We divide the 132 months into 6 groups based on their average spread levels and obtain for each group daily NP variance estimates (annualized) for δ between 2 and 15 ticks and show their averages across days in Figure 11. The six groups are in ascending average spread level order. The first two are the bottom third and the middle third of the spread distribution. Groups 3 to 6 represent the upper third, subdivided into 4 ascending groups (1/12 each of the data). Table 4 shows the distribution of the 6 groups across the 132 months in the data sample. It should be noted that many of the high bid/ask spread level months, besides those during the financial crisis of 2007/8, are in the early years of the data sample when trading was less liquid, and consequently many of the low bid/ask spread level months are concentrated at the end of the data sample.

Figure 11: Duration based variance signature plot, six levels of spread, AXP



Notes: The average spreads of groups 1 to 6 for AXP are 1.4, 1.6, 1.8, 2.1, 2.9, and 4.1 ticks. One tick equals one cent. Diamonds indicate three times the respective average spread.

Figure 11 shows the stabilizing behavior of the duration based variance estimates very clearly and, upon visual inspection, we observe that the threshold value at the point where the estimates start to stabilize, δ^*, is roughly three times the average bid/ask spread which is in line with the guidance obtained from the simulation study. We will use the "three-times-bid/ask-spread" rule henceforth as guidance to select δ^* for the computation of the PDV and NP estimators in the subsequent forecasting study. Let us add a note at this point that this threshold selection rule is more important to PDV than to ANP, since the latter uses a wide range of threshold values that do not necessarily center at 3-times-spread. It is possible to further improve PDV by adding an averaging procedure, but since the "local

Table 4: Bid/ask spread level groups, AXP

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Jan.	5	6	2	2	2	1	5	3	1	1	1
Feb.	6	5	2	2	2	1	4	2	1	1	1
Mar.	5	2	2	1	1	2	4	1	1	1	1
Apr.	5	3	2	2	2	1	3	2	1	1	2
May	5	3	2	2	2	2	4	3	2	1	2
Jun.	6	4	1	1	2	2	4	1	1	1	2
Jul.	6	4	2	1	2	4	5	2	2	2	2
Aug.	6	3	1	1	2	5	5	2	1	3	2
Sep.	6	3	1	1	2	4	6	1	1	3	1
Oct.	6	3	2	2	1	4	6	1	1	3	2
Nov.	6	2	2	2	1	5	5	2	1	2	1
Dec.	6	2	2	1	1	4	4	1	1	1	1

variance" within each price event has been adjusted by the intensity function for its longer or shorter price duration (compared with the expected duration), the improvement on PDV from averaging may be limited. Andersen et al. (2008) recommend a threshold range of 3 to 6 times the *log-spread* for their non-parametric duration based variance estimator. Since their thresholds are set in log-scales, the resulting estimates can be different.

Table 22 in Web-Appendix A presents goodness-of-fit results (LB and DF tests) of the Burr-HACD model for all 20 stocks, with the price durations obtained by setting the threshold value to be δ^*. It confirms that, when the threshold value is set to be three times the average bid/ask-spread, the Burr-HACD fits the price durations data well.

It should be noted that in the absence of high-frequency bid/ask spread data a price duration volatility estimator can still be obtained by, for example, selecting a threshold so that on average price durations have a length, say, 5 minutes, as is done in Tse & Yang (2012). An ad hoc choice of δ, equal to 10 ticks say, is also always possible and can be adjusted with some general notion about the liquidity of the assets under consideration.

It should also be noted that this threshold selection rule is based on the average spread of the day, thus ignoring the time-variation of spreads within the day. A possible improvement on NP can be made by taking into account local movements in spreads which translate into time-varying thresholds, based on the same "3-times-spread" rule. However, it brings an issue of choosing the size of the local window that defines the moving-average local spreads. We would like to leave this for future research.

6 Volatility forecasting

Our objective in this section is to forecast the integrated variance, so we choose a jump-robust variance estimator, namely the subsampled 5-minute realized bipower, SBV, as the target.[8] We include a total of 15 volatility estimators in the forecasting comparisons, 13 of which come from Section 4.2 (the other 2 are PDV and ATM). All estimators are obtained on a daily basis and their tuning parameters are obtained using only historical data to rule out any forward looking bias. In Table 5 we present the means and standard deviations of the 15 volatility estimators including the forecast object SBV for the 4 representative stocks, and relegate statistics for all 20 stocks to Web-Appendix B. Note that we have included the forecast target itself into the race as a benchmark. All estimators are annualized and converted from variance estimates into volatility (i.e. standard deviation) estimates. We consider four duration variance estimators: PDV, NP, ANP_1 and ANP_2. PDV is obtained using a threshold equal to 3 times the average spread of the last month and maximum likelihood parameters estimates are obtained on a monthly basis. NP, ANP_1 and ANP_2 follow the specifications detailed in Section 4.2 with thresholds based on the average spread of the corresponding day. PAV_1, PAV_2, $PABV_1$, $PABV_2$, RK, $RKNP$, $TSRV$, SBV, RV_5, and SRV_5 are also obtained following the setups detailed in Section 4.2. In addition, we include ATM which is an at-the-money option implied volatility estimator obtained from daily OptionMetrics data. From Table 5 we can see that the duration volatility estimates are on average slightly smaller than RV-type estimates and also exhibit lower standard deviations. First- and fifth-order autocorrelations of the above estimators across 20 stocks are presented in Table 24 in Web-Appendix B as well, which shows the four price duration estimators have higher autocorrelations than the RV-type estimators.

6.1 Individual forecasts

We use a HAR-type forecasting specification,

$$SBV_{n:n+h-1} = c + b_1 Z_{n-1:n-1} + b_2 Z_{n-5:n-1} + b_3 Z_{n-22:n-1} + \epsilon_{n:n+h-1} \tag{55}$$

to obtain h-step ahead forecasts. Here Z_n represents the day-n volatility estimate from one of the 15 estimators discussed above. Both $Z_{n-h:n-1}$ and $SBV_{n:n+h-1}$ aggregate h terms and are in their logarithmic forms: $Z_{n-h:n-1} = 0.5 \log(\sum_{s=n-h}^{n-1} Z_s^2)$, similarly for $SBV_{n:n+h-1}$,

[8]Results with standard 5-minute realized volatility (RV) and its subsampled version as targets, recommended by Andersen, Bollerslev & Meddahi (2011), are similar and available from the authors upon request.

Table 5: Volatility estimator means and standard deviations

		HD	MCD	AXP	IBM	avg.
				Mean		
1	PDV	.238	.189	.248	.174	.210
2	ANP_1	.216	.173	.230	.167	.193
3	ANP_2	.225	.180	.241	.172	.200
4	NP	.217	.174	.231	.167	.193
5	PAV_1	.246	.197	.270	.191	.222
6	PAV_2	.244	.196	.263	.185	.218
7	$PABV_1$	.238	.189	.261	.185	.214
8	$PABV_2$	.239	.190	.257	.182	.213
9	RK	.261	.237	.321	.232	.239
10	$RKNP$	.245	.198	.265	.185	.219
11	$TSRV$	.230	.219	.278	.209	.212
12	SBV	.230	.184	.248	.175	.205
13	RV_5	.238	.192	.257	.180	.213
14	SRV_5	.241	.194	.259	.182	.216
15	ATM	.244	.194	.272	.202	.223
			Standard Deviation			
1	PDV	.123	.090	.182	.092	.113
2	ANP_1	.121	.092	.178	.096	.112
3	ANP_2	.127	.096	.187	.100	.117
4	NP	.122	.093	.180	.097	.113
5	PAV_1	.139	.107	.206	.112	.131
6	PAV_2	.140	.111	.206	.110	.132
7	$PABV_1$	.136	.101	.202	.108	.128
8	$PABV_2$	.139	.107	.204	.109	.130
9	RK	.139	.176	.189	.125	.144
10	$RKNP$	.138	.110	.206	.108	.131
11	$TSRV$	.129	.166	.176	.118	.134
12	SBV	.132	.104	.193	.105	.125
13	RV_5	.135	.108	.201	.108	.129
14	SRV_5	.137	.108	.201	.107	.130
15	ATM	.104	.073	.175	.088	.100

Notes: Mean and standard deviation statistics for 15 daily volatility estimators for 4 representative stocks using data from January 2002 to December 2012. The "avg." values are averages of all 20 stocks. The means and standard deviations are for annualized volatilities.

h=1,5, or 22, with SBV_n the day-n sub-sampled bipower variation estimate.

For one day ($h = 1$) ahead forecasts the in-sample estimation period for the HAR model ranges from 1 February 2002 to 29 January 2010 (2013 trading days) and the first out-of-sample forecast is obtained for 1 February 2010. For one week ($h = 5$) and one month ($h = 22$) horizons forecasts are constructed similarly and a total of 735, 731 and 714 out-of-sample predictions are obtained for $h = 1, 5$ and 22, respectively, with the final predictions made in December 2012. All forecasts are constructed using a rolling window of explanatory variables with a fixed length of 2013 trading days. Setting $\epsilon_{n:n+h-1} = 0$ defines the forecast of $SBV_{n:n+h-1}$ made at time $n-1$, denoted $\hat{S}BV_{n:n+h-1}$. The forecast error obtained at time $n+h-1$ is then $SBV_{n:n+h-1} - \hat{S}BV_{n:n+h-1}$.

We evaluate 15 individual forecasts, defined by the selected 15 estimators, for the 20 stocks over 3 horizons. Table 6 lists the average values across stocks of RMSE (root-mean-squared-errors) and QLIKE following Patton (2011) as in (54). The RMSE and QLIKE values by stock are provided in Web-Appendix B as Tables 25 and 27. The average values of RMSE and QLIKE are similar across the 15 sets of forecasts. Nevertheless, the parametric duration estimator, PDV, has the lowest average values in most cases.

Table 6: Average RMSE and QLIKE values for individual forecasts

		RMSE	QLIKE	RMSE	QLIKE	RMSE	QLIKE
		one day ahead		one week ahead		one month ahead	
1	PDV	.0028	.0300	.0111	.0230	.0501	.0285
2	ANP_1	.0029	.0308	.0117	.0244	.0515	.0295
3	ANP_2	.0029	.0303	.0116	.0241	.0516	.0293
4	NP	.0029	.0314	.0118	.0249	.0519	.0298
5	PAV_1	.0029	.0307	.0122	.0249	.0537	.0301
6	PAV_2	.0029	.0306	.0122	.0244	.0534	.0294
7	$PABV_1$	.0029	.0309	.0122	.0250	.0539	.0300
8	$PABV_2$	.0029	.0306	.0122	.0244	.0536	.0294
9	RK	.0038	.0466	.0157	.0383	.0668	.0402
10	$RKNP$	.0029	.0310	.0121	.0245	.0527	.0293
11	$TSRV$	.0038	.0474	.0159	.0387	.0669	.0399
12	SBV	.0029	.0304	.0121	.0242	.0531	.0292
13	RV_5	.0029	.0309	.0121	.0246	.0530	.0295
14	SRV_5	.0029	.0307	.0123	.0244	.0537	.0294
15	ATM	.0031	.0334	.0123	.0252	.0552	.0296

Notes: Average RMSE and QLIKE values across 20 stocks using trade data for stated forecasting horizons for 15 individual volatility estimators. Forecasts are obtained from (55).

We also evaluate the accuracy of the forecasts using the Model Confidence Set (MCS) method of Hansen, Lunde & Nason (2011). Table 7 shows the counts, across stocks, of how often each forecasting method is included in the MCS. These counts are provided for all six combinations of loss functions and forecast horizons for a 20% significance level.[9] The MCS p-values are shown in Web-Appendix B in Tables 26 and 28, respectively for RMSE and QLIKE. In these tables, under each stock on the left we report the MCS p-value and on the right we list the estimator number, ranked by the p-value; the estimator numbers are as in Table 23. We see that very often estimator 1 (PDV) has the highest confidence, which is consistent with it having the least average values of RMSE and QLIKE. The MCS results summarized by Table 7 show that PDV provides the best forecasts, for any choices of loss

[9]Hansen et al. (2011) use the 10% and 25% levels for their examples.

function, forecast horizon and significance level. Adding the counts across the combinations of loss function and horizon gives a total of 104 at the 20% significance level for PDV. At the same level, the second-best methods are SBV and ATM striking a tie at 37, followed by SRV_5 with a total of 30, RV_5 with a total of 28, and ANP_2 with a total of 26. Both ANP_1 and ANP_2 clearly outscore NP, confirming that averaging across thresholds improves the non-parametric duration estimator. Though SBV performs best in the RV group, we need to keep in mind that it being the forecast target might be an advantage. Apart from SBV, the two simple 5-minute RV estimators outperform all other RV estimators in this group, confirming the conclusion of Liu et al. (2015) that 5-minute RV is difficult to beat when it comes to forecasting future stock volatilities.

Table 7: MCS summary results for 15 individual forecasts using trade data

		RMSE	QLIKE	RMSE	QLIKE	RMSE	QLIKE	Total
		one day ahead		one week ahead		one month ahead		
1	PDV	15	20	19	16	17	17	104
2	ANP_1	4	3	4	1	2	4	18
3	ANP_2	7	3	4	7	2	3	26
4	NP	0	1	2	0	0	2	5
5	PAV_1	5	0	2	5	0	3	15
6	PAV_2	4	1	3	6	6	2	22
7	$PABV_1$	3	1	2	4	1	4	15
8	$PABV_2$	3	1	4	6	4	5	23
9	RK	0	3	5	0	2	3	13
10	$RKNP$	2	3	6	2	5	6	24
11	$TSRV$	0	1	7	0	1	6	15
12	SBV	8	4	4	8	5	8	37
13	RV_5	3	4	5	5	5	6	28
14	SRV_5	5	2	5	6	5	7	30
15	ATM	6	7	6	4	7	7	37

Notes: Across 20 stocks, each number represents the number of times a given estimator is included in the model confidence set with 20% significance level

6.2 Combination forecasts

The same HAR-type forecasting specification, namely (55), is evaluated again with the quantity Z_n now denoting the average of all volatility estimators included in some set of estimators. We want to find out firstly whether combining information from multiple volatility estimates improves upon the individual forecasts. Therefore we include PDV, SBV and ATM as benchmarks in this combination setting. Secondly, we investigate which group of estimators provides the best forecasts and whether combining different groups improves the accuracy of our forecasts.

Fourteen combinations are studied. Three of these are the individual forecasts: PDV, SBV and ATM. Combining any two benchmarks gives 3 distinct combinations: $PDV + SBV$, $PDV + ATM$, and $SBV + ATM$. $Duration4$ is the average of all four duration estimators (PDV, ANP_1, ANP_2 and NP), and $RV10$ is the average of all ten RV-type estimators (PAV_1, PAV_2, $PABV_1$, $PABV_2$, RK, $RKNP$, $TSRV$, SBV, RV_5 and SRV_5). Mixing price duration, RV, and option-implied volatility estimators gives $Duration4 + SBV$, $Duration4 + ATM$, $RV10 + PDV$, $RV10 + ATM$, and $Duration4 + RV10$. Finally, taking all 15 estimators in, we have All, i.e. it is $Duration4 + RV10 + ATM$.

Table 8: Average RMSE and QLIKE values for combination forecasts

		RMSE	QLIKE	RMSE	QLIKE	RMSE	QLIKE
		one day ahead		one week ahead		one month ahead	
1	PDV	.0028	.0300	.0111	.0230	.0501	.0300
2	SBV	.0029	.0304	.0121	.0242	.0531	.0308
3	ATM	.0031	.0334	.0123	.0252	.0552	.0311
4	$PDV + SBV$	.0028	.0294	.0114	.0231	.0508	.0300
5	$PDV + ATM$	.0026	.0275	.0106	.0206	.0501	.0286
6	$SBV + ATM$	.0027	.0281	.0112	.0219	.0515	.0294
7	$Duration4$	.0028	.0303	.0115	.0239	.0511	.0308
8	$SBV + Duration4$	.0028	.0300	.0115	.0237	.0511	.0306
9	$ATM + Duration4$	.0027	.0288	.0111	.0225	.0505	.0300
10	$RV10$	.0028	.0302	.0120	.0241	.0535	.0311
11	$PDV + RV10$	.0028	.0299	.0118	.0238	.0529	.0308
12	$ATM + RV10$	.0028	.0295	.0118	.0235	.0529	.0307
13	$Duration4 + RV10$	.0028	.0297	.0117	.0236	.0522	.0306
14	All	.0028	.0292	.0116	.0232	.0519	.0304

Notes: Average RMSE and QLIKE values across 20 stocks using trade data for stated forecasting horizons for 14 combination volatility estimators. Forecasts are obtained from (55).

Table 8 shows the average values across stocks of RMSE and QLIKE for the 14 combinations. The RMSE and QLIKE values by stock are included in Web-Appendix B as Tables 29 and 31, and their corresponding MCS p-values are in Tables 30 and 32. We now find that $PDV + ATM$ has the least average values in Table 8 for both RMSE and QLIKE loss functions at all three horizons.

Table 9 summarizes the MCS results when the candidate set of forecasts is defined by the fourteen combinations. When we rank the combinations by counting membership of the MCS across loss functions and horizons at the 20% significance level, as before, the best combination is $PDV + ATM$ with a count of 106, followed by PDV at 46, $SBV + ATM$ at 35 and $ATM + Duration4$ at 28. It is noteworthy that PDV by itself outperforms all but one of the combination forecasts. In contrast, simply combining either all duration estimators or all RV estimators scores poorly, with respective counts equal to 5 and 10. For our study, the most successful combinations of more than one forecast are averages across different data sources, namely duration measures from high-frequency stock prices and implied volatilities from daily option prices.

Table 9: MCS summary results for 14 combination forecasts using trade data

		RMSE	QLIKE	RMSE	QLIKE	RMSE	QLIKE	Total
		one day ahead		one week ahead		one month ahead		
1	PDV	2	9	17	1	5	12	46
2	SBV	0	0	3	0	0	4	7
3	ATM	0	1	3	0	1	9	14
4	$PDV + SBV$	0	2	8	1	3	11	25
5	$PDV + ATM$	19	18	15	18	19	17	106
6	$SBV + ATM$	6	2	6	7	4	10	35
7	$Duration4$	0	0	2	0	0	3	5
8	$SBV + Duration4$	0	0	3	0	0	4	7
9	$ATM + Duration4$	6	3	6	4	2	7	28
10	$RV10$	0	0	3	0	0	7	10
11	$PDV + RV10$	0	0	5	0	0	7	12
12	$ATM + RV10$	1	1	5	1	2	8	18
13	$Duration4 + RV10$	0	0	3	0	0	6	9
14	*All*	3	2	5	3	3	7	23

Notes: The combinations are defined in Section 6.2. Across 20 stocks, each number represents the number of times a given estimator is included in the confidence set with 20% significance level.

7 Conclusions

Duration based variance estimators are calculated by using the times of price change events; an event occurs when the magnitude of the price change since the previous event first equals or exceeds some threshold value. These estimators have been neglected in previous research, despite them being very simple to use and Andersen et al. (2008) documenting their nice performance compared to realized variance estimators. Market microstructure noise prevents comprehensive theoretical comparisons for realistic data generating processes and, furthermore, requires careful consideration to be given to the selection of the threshold value. In this paper, we establish limiting theories for our non-parametric duration based estimator, with and without the presence of microstructure noise, supporting the validity of its use. Specifically, we elaborate on how different types of noise specifically may affect the results.

We use Monte Carlo methods for a selection of volatility processes and also tick price data for U.S. stocks to recommend that an appropriate choice of the threshold is three times a measure of the average bid/ask spread. We introduce average non-parametric duration estimators, which average across thresholds from two to four or more bid/ask spreads, to further reduce bias and improve efficiency. We evaluate both parametric and non-parametric duration based estimators and find the parametric specification forecasts more accurately than RV-type estimators calculated from tick prices, for our sample period; the non-parametric duration estimators have the same accuracy as the best RV-type estimators. Applying the model confidence set methodology of Hansen et al. (2011) shows that the parametric duration estimator significantly outperforms all RV-type estimators and at-the-money implied volatility, for one-day, one-week and one-month forecast horizons, for both the RMSE and the QLIKE loss functions.

Calculation of the non-parametric duration estimator from a complete record of transaction prices is very easy. Bid and ask prices can be used to select our preferred threshold value, alternatively an appropriate multiple of the tick size can be chosen. The parametric estimator is more accurate but does require the estimation of a parametric model for price events, which requires specifying intensity functions for durations whose conditional expectations are functions of previous durations. We recommend considering duration based estimators of integrated variance whenever transaction prices are available because of their potential to provide more accurate estimates and forecasts.

Future research might also evaluate additional duration estimators, following Andersen et al. (2008). Semiparametric RV-type estimators, as motivated for example by Becker, Clements & Hurn (2011) and Žikeš & Baruník (2016), may have the potential to achieve some of the efficiency gained by the parametric duration estimator so they too may deserve further attention. Finally, it is possible to extend the duration methodology to estimate the integrated covariance between the returns from two assets and we are investigating multivariate methods in ongoing research.

References

Aït-Sahalia, Y. & J. Jacod (2014): *High-Frequency Financial Econometrics*, Princeton University Press.

Aït-Sahalia, Y., P. A. Mykland, & L. Zhang (2011): "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise," *Journal of Econometrics*, 160, 160–175.

Andersen, T., T. Bollerslev, F. Diebold, & H. Ebens (2001): “The Distribution of Realized Stock Return Volatility,” *Journal of Financial Economics*, 61, 43–76.

Andersen, T., T. Bollerslev, & D. Dobrev (2007): “No-arbitrage Semi-martingale Restrictions for Continuous-time Volatility Models Subject to Leverage Effects, Jumps and i.i.d. Noise: Theory and Testable Distributional Implications,” *Journal of Econometrics*, 138 (1), 125 – 180.

Andersen, T., D. Dobrev, & E. Schaumburg (2008): “Duration-Based Volatility Estimation,” Working Paper, Northwestern University.

Andersen, T., T. Bollerslev, & N. Meddahi (2011): “Realized Volatility Forecasting and Market Microstructure Noise,” *Journal of Econometrics*, 160, 220–234.

Bandi, F. M. & J. R. Russell (2008): “Microstructure Noise, Realized Variance, and Optimal Sampling,” *The Review of Economic Studies*, 75, 339–369.

—— (2011): “Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations,” *Journal of Econometrics*, 160, 145–159.

Bandi, F. M., D. Pirino, & R. Renò (2017): “EXcess Idle Time,” *Econometrica*, 85(6), 1793–1846.

Bandi, F. M., A. Kolokolov, D. Pirino, & R. Renò (2020): “Zeros,” *Management Science*, 66(8), 3466–3479.

Barndorff-Nielsen, O. & N. Shephard (2002): “Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models,” *Journal of the Royal Statistical Society*, 64, 253–280.

—— (2006): “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation,” *Journal of Financial Econometrics*, 4, 1–30.

Barndorff-Nielsen, O., P. R. Hansen, A. Lunde, & N. Shephard (2008): “Designing Realized Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76, 1481–1536.

—— (2009): “Realized Kernels in Practice: Trades and Quotes,” *Econometrics Journal*, 12, C1–C32.

Bauwens, L. & P. Giot (2000): "The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks," *Annals of Economics and Statistics*, 117–149.

Bauwens, L., P. Giot, J. Grammig, & D. Veredas (2004): "A Comparison of Financial Duration Models Via Density Forecasts," *International Journal of Forecasting*, 20, 589–609.

Becker, R., A. Clements, & S. Hurn (2011): "Semi-parametric Forecasting of Realized Volatility," *Studies in Nonlinear Dynamics and Econometrics*, 15 (3), 1–21.

Campbell, J., A. W. Lo, & A. C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.

Chernov, M., A. R. Gallant, E. Ghysels, & G. Tauchen (2003): "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116, 225–257.

Cho, D. & E. Frees (1988): "Estimating the Volatility of Discrete Stock Prices," *Journal of Finance*, 43, 451–466.

Christensen, K., R. Oomen, & M. Podolskij (2014): "Fact or Friction: Jumps at Ultra High Frequency," *Journal of Financial Economics*, 114, 576–599.

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7, 174–196.

Diebold, F. X., T. Gunther, & A. Tay (1998): "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

Engle, R. & J. R. Russell (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Fernandes, M. & J. Grammig (2006): "A Family of Autoregressive Conditional Duration Models," *Journal of Econometrics*, 130, 1–23.

Fukasawa, M. (2010a): "Central Limit Theorem for the Realized Volatility based on Tick Time Sampling," *Finance and Stochastics*, 14 (2), 209–233.

—— (2010b): "Realized Volatility with Stochastic Sampling," *Stochastic Processes and their Applications*, 120 (6), 829 – 852.

Fukasawa, M. & M. Rosenbaum (2012): "Central Limit Theorems for Realized Volatility under Hitting Times of an Irregular Grid," *Stochastic Processes and their Applications*, 122 (12), 3901 – 3920.

Gerhard, F. & N. Hautsch (2002): "Volatility Estimation on the Basis of Price Intensities," *Journal of Empirical Finance*, 9, 57–89.

Grammig, J. & K. Maurer (2000): "Non-monotonic Hazard Functions and the Autoregressive Conditional Duration Model," *The Econometrics Journal*, 3, 16–38.

Hansen, P. R. & A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business & Economic Statistics*, 24, 127–161.

Hansen, P. R., A. Lunde, & J. Nason (2011): "The Model Confidence Set," *Econometrica*, 79, 453–497.

Hautsch, N. (2004): *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*, Springer Science and Business Media.

Huang, X. & G. Tauchen (2005): "The Relative Contribution of Jumps to Total Price Variation," *Journal of Financial Econometrics*, 3, 456–499.

Jacod, J. (2008): "Asymptotic properties of realized power variations and related functionals of semimartingales," *Stochastic processes and their applications*, 118, 517–559.

Jacod, J. & A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*, Springer-Verlag.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, & M. Vetter (2009): "Microstructure Noise in the Continuous Case: The Pre-averaging Approach," *Stochastic Processes and their Applications*, 119, 2249–2276.

Jasiak, J. (1999): "Persistence in Intertrade Durations," *Finance*, 19, 166–195.

Lancaster, T. (1997): *The Econometric Analysis of Transition Data*, Cambridge University Press.

Lee, S. & J. Hannig (2010): "Detecting Jumps from Lévy Jump Diffusion Processes," *Journal of Financial Economics*, 96 (2), 271–290.

Li, Y., Z. Zhang, & X. Zheng (2013): "Volatility inference in the presence of both endogenous time and microstructure noise," *Stochastic Processes and their Applications*, 123, 2696–2727.

Li, Y., P. A. Mykland, E. Renault, L. Zhan, & X. Zheng (2014): "Realized volatility when sampling times are possibly endogenous," *Econometric Theory*, 30, 580–605.

Li, Y., I. Nolte, & S. Nolte (2019): "Renewal Based Volatility Estimation," Working Paper.

—— (2021): "High-Frequency Volatility Modelling: A Markov-Switching Autoregressive Conditional Intensity Model," *Journal of Economic Dynamics and Control*, Forthcoming.

Liu, L., A. Patton, & K. Sheppard (2015): "Does Anything Beat 5-minute RV? A Comparison of Realized Measures Across Multiple Asset Classes," *Journal of Econometrics*, 187 (1), 293–311.

Malliavin, P. & M. E. Mancino (2009): "Estimating Covariation: Epps Effect, Microstructure Noise," *Annals of Statistics*, 37, 1983–2010.

Mancino, M. E. & S. Sanfelici (2012): "Estimation of Quarticity with High Frequency Data," *Quantitative Finance*, 12, 607–622.

Marsh, T. A. & E. R. Rosenfeld (1986): "Non-trading, market making, and estimates of stock price volatility," *Journal of Financial Economics*, 15, 359–372.

Mykland, P. A. & L. Zhang (2009): "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1445.

Nolte, I. (2008): "Modeling a Multivariate Transaction Process," *Journal of Financial Econometrics*, 6, 143–170.

Nolte, I. & V. Voev (2012): "Least Squares Inference on Integrated Volatility and the Relationship Between Efficient Prices and Noise," *Journal of Business & Economic Statistics*, 30, 94–108.

Pacurar, M. (2008): "Autoregressive Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature," *Journal of Economic Surveys*, 22, 711–751.

Patton, A. (2011): "Data-based Ranking of Realised Volatility Estimators," *Journal of Econometrics*, 161, 284–303.

Pelletier, D. & W. Wei (2019): "A Stochastic Price Duration Model for Estimating High-Frequency Volatility," Working Paper.

Podolskij, M. & M. Vetter (2009): "Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps," *Bernoulli*, 15(3), 634–658.

Shephard, N. & D. Xiu (2017): "Econometric analysis of multivariate realised QML: Estimation of the covariation of equity prices under asynchronous trading," *Journal of Econometrics*, 201 (1), 19–42.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

Tse, Y. & T. T. Yang (2012): "Estimation of High-Frequency Volatility: An Autoregressive Conditional Duration Approach," *Journal of Business & Economic Statistics*, 30, 533–545.

Vetter, M. & T. Zwingmann (2017): "A note on central limit theorems for quadratic variation in case of endogenous observation times," *Electronic Journal of Statistics*, 11, 963–980.

Žikeš, F. & J. Baruník (2016): "Semi-parametric Conditional Quantile Models for Financial Returns and Realized Volatility," *Journal of Financial Econometrics*, 14 (1), 185–226.

Xiu, D. (2010): "Quasi-maximum Likelihood Estimation of Volatility with High Frequency Data," *Journal of Econometrics*, 159, 235–250.

Zhang, L. (2011): "Estimating Covariation: Epps Effect, Microstructure Noise," *Journal of Econometrics*, 160, 33–47.

Zhang, L., P. A. Mykland, & Y. Aït-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

Web-Appendix A: Comparison of density functions

For the choice of a suitable density function for the scaled price durations we first consider LR tests for the four reference stocks: HD, MCD, AXP and IBM. The results in Tables 10, 13, 16 and 19 show that the Burr density is preferred over the Weibull and Exponential densities most of the time over a wide range of price change threshold values δ.

Corresponding LB test results for LB statistics with lags 30 and 50 are presented in Tables 11, 14, 17 and 20. For the majority of the months the null hypothesis of i.i.d. distributed generalized residuals cannot be rejected at the 1% and 5% significance levels, which indicates that the price duration dynamics are well captured by the HACD specification.

The associated density forecast (DF) test results in Tables 12, 15, 18 and 21 show that the Burr density clearly outperforms the other two distributional assumptions, by giving the highest percentages of months in which the null is not rejected at either the 1% or 5% significance level. From the three densities considered the Burr density provides the best fit for the scaled price durations.

Overall, the test results for the four reference stocks indicate that the HACD-Burr combination fits the price duration data best.

Finally, we present in Table 22 the LB and DF tests results for all 19 stocks, when the price change threshold δ is selected using the "3-times-spread" rule. We observe that the HACD-Burr model fits the price durations data well.

Table 10: LR test results, HD

δ(ticks)	2	3	4	5	6	7	8	9	10
Wei. vs. Burr	505.77	260.88	155.93	100.55	67.86	45.56	35.09	24.14	21.30
Exp. vs. Burr	574.24	307.80	189.89	127.16	87.73	63.65	51.02	38.30	34.75
Exp. vs. Wei.	68.47	46.92	33.96	26.32	19.72	18.34	16.15	13.91	13.85
Wei. vs. Burr	1.00	1.00	1.00	1.00	1.00	0.95	0.93	0.74	0.68
Exp. vs. Burr	1.00	1.00	1.00	1.00	1.00	0.99	0.96	0.89	0.84
Exp. vs. Wei.	0.78	0.75	0.69	0.69	0.67	0.69	0.63	0.61	0.61

Notes: The first three rows are the LR test statistics (averaged over 132 months), and the last three rows are LR test results presented as proportions of the months in which the null is rejected at the 1% significance level. The assumed density under the null is stated first in the 1st column.

Table 11: LB test results for 30 and 50 lags, HD

δ(ticks)	2	3	4	5	6	7	8	9	10
				30 lags 1% significance level					
Exp.	0.98	0.95	0.97	0.98	0.97	0.98	0.94	0.92	0.89
Weibull	0.97	0.94	0.95	0.98	0.95	0.98	0.93	0.92	0.92
Burr	0.87	0.86	0.90	0.92	0.95	0.95	0.94	0.86	0.89
				30 lags 5% significance level					
Exp.	0.86	0.89	0.91	0.90	0.93	0.93	0.89	0.87	0.85
Weibull	0.82	0.85	0.88	0.86	0.89	0.92	0.87	0.87	0.86
Burr	0.66	0.70	0.78	0.76	0.80	0.83	0.81	0.77	0.80
				50 lags 1% significance level					
Exp.	0.94	0.96	0.96	0.96	0.98	0.99	0.96	0.92	0.89
Weibull	0.93	0.95	0.96	0.96	0.96	0.98	0.95	0.93	0.92
Burr	0.87	0.91	0.93	0.90	0.96	0.99	0.96	0.87	0.90
				50 lags 5% significance level					
Exp.	0.82	0.86	0.91	0.89	0.92	0.95	0.90	0.86	0.86
Weibull	0.79	0.83	0.90	0.86	0.90	0.95	0.88	0.86	0.88
Burr	0.67	0.73	0.81	0.80	0.88	0.89	0.83	0.80	0.86

Notes: The upper part of the table are LB test results for 30 lags, and the lower part are the results for 50 lags. Significance levels of 1% and 5% are considered. Each figure is the proportion of months in which the null is not rejected.

Table 12: DF test results, HD

δ(ticks)	2	3	4	5	6	7	8	9	10
				1% significance level					
Exp.	0.00	0.00	0.01	0.03	0.11	0.34	0.31	0.52	0.53
Weibull	0.00	0.02	0.02	0.08	0.21	0.36	0.49	0.60	0.67
Burr	0.21	0.57	0.69	0.80	0.86	0.95	0.92	0.88	0.89
				5% significance level					
Exp.	0.00	0.00	0.00	0.01	0.03	0.20	0.23	0.32	0.44
Weibull	0.00	0.00	0.01	0.04	0.11	0.25	0.30	0.45	0.53
Burr	0.14	0.43	0.56	0.67	0.76	0.85	0.80	0.81	0.83

Notes: DF test results for significance levels of 1% and 5% are presented. Each figure is the proportion of months in which the null that the assumed density is the true density is not rejected.

Table 13: LR test results, MCD

δ(ticks)	2	3	4	5	6	7	8	9	10
Wei. vs. Burr	460.17	268.57	181.59	129.51	91.43	68.76	52.41	40.05	32.77
Exp. vs. Burr	577.22	328.81	219.52	156.81	113.24	86.20	62.74	53.14	46.29
Exp. vs. Wei.	117.05	60.24	37.93	27.09	21.33	17.38	10.87	12.03	12.24
Wei. vs. Burr	1.00	1.00	0.99	0.98	0.95	0.88	0.84	0.78	0.73
Exp. vs. Burr	1.00	1.00	0.99	0.99	0.97	0.93	0.84	0.88	0.84
Exp. vs. Wei.	0.87	0.64	0.57	0.55	0.50	0.52	0.45	0.46	0.45

Notes: The first three rows are the LR test statistics (averaged over 132 months), and the last three rows are LR test results presented as proportions of the months in which the null is rejected at the 1% significance level. The assumed density under the null is stated first in the 1st column.

Table 14: LB test results for 30 and 50 lags, MCD

δ(ticks)	2	3	4	5	6	7	8	9	10
				30 lags 1% significance level					
Exp.	0.93	0.96	0.98	0.95	0.98	0.99	0.93	0.93	0.89
Weibull	0.92	0.96	0.96	0.95	0.98	0.98	0.93	0.94	0.92
Burr	0.90	0.87	0.89	0.86	0.94	0.96	0.91	0.90	0.86
				30 lags 5% significance level					
Exp.	0.83	0.86	0.88	0.86	0.87	0.96	0.88	0.90	0.83
Weibull	0.82	0.83	0.86	0.83	0.86	0.95	0.84	0.89	0.85
Burr	0.73	0.67	0.74	0.76	0.82	0.89	0.77	0.80	0.77
				50 lags 1% significance level					
Exp.	0.90	0.92	0.97	0.95	0.98	0.99	0.90	0.92	0.88
Weibull	0.90	0.92	0.97	0.95	0.98	0.98	0.89	0.93	0.90
Burr	0.89	0.89	0.92	0.92	0.96	0.98	0.89	0.89	0.86
				50 lags 5% significance level					
Exp.	0.85	0.87	0.88	0.89	0.96	0.98	0.86	0.86	0.86
Weibull	0.84	0.86	0.87	0.88	0.92	0.97	0.85	0.86	0.88
Burr	0.76	0.73	0.77	0.80	0.86	0.91	0.79	0.81	0.80

Notes: The upper part of the table are LB test results for 30 lags, and the lower part are the results for 50 lags. Significance levels of 1% and 5% are considered. Each figure is the proportion of months in which the null is not rejected.

Table 15: DF test results, MCD

δ(ticks)	2	3	4	5	6	7	8	9	10
				1% significance level					
Exp.	0.00	0.04	0.11	0.15	0.27	0.39	0.51	0.51	0.48
Weibull	0.01	0.07	0.18	0.21	0.34	0.43	0.57	0.63	0.61
Burr	0.24	0.55	0.75	0.80	0.83	0.92	0.88	0.85	0.84
				5% significance level					
Exp.	0.00	0.00	0.07	0.08	0.13	0.27	0.43	0.38	0.36
Weibull	0.01	0.03	0.10	0.13	0.23	0.32	0.40	0.47	0.48
Burr	0.14	0.45	0.61	0.70	0.72	0.83	0.83	0.80	0.76

Notes: DF test results for significance levels of 1% and 5% are presented. Each figure is the proportion of months in which the null that the assumed density is the true density is not rejected.

Table 16: LR test results, AXP

δ(ticks)	2	3	4	5	6	7	8	9	10	11	12
Wei. vs. Burr	678.13	382.69	253.40	172.94	128.72	98.31	74.79	59.54	44.10	35.64	28.78
Exp. vs. Burr	759.60	435.54	292.96	206.03	155.16	121.43	94.94	75.91	59.25	52.71	42.91
Exp. vs. Wei.	81.47	52.85	39.56	29.77	26.70	22.26	19.39	18.16	15.46	15.49	14.66
Wei. vs. Burr	1.00	1.00	1.00	1.00	0.99	0.99	0.96	0.95	0.89	0.77	0.65
Exp. vs. Burr	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.95	0.95	0.89	0.84
Exp. vs. Wei.	0.64	0.71	0.72	0.72	0.66	0.63	0.63	0.64	0.65	0.60	0.63

Notes: The first three rows are the LR test statistics (averaged over 132 months), and the last three rows are LR test results presented as proportions of the months in which the null is rejected at the 1% significance level. The assumed density under the null is stated first in the 1st column.

Table 17: LB test results for 30 and 50 lags, AXP

δ(ticks)	2	3	4	5	6	7	8	9	10	11	12
					30 lags 1% significance level						
Exp.	0.93	0.93	0.95	0.98	0.98	0.97	0.96	0.97	0.95	0.87	0.92
Weibull	0.91	0.93	0.95	0.97	0.95	0.96	0.96	0.96	0.95	0.88	0.92
Burr	0.86	0.86	0.82	0.92	0.90	0.92	0.89	0.89	0.92	0.89	0.90
					30 lags 5% significance level						
Exp.	0.79	0.89	0.86	0.90	0.92	0.91	0.83	0.91	0.92	0.83	0.90
Weibull	0.73	0.88	0.85	0.88	0.89	0.90	0.83	0.92	0.91	0.82	0.90
Burr	0.60	0.69	0.67	0.75	0.73	0.82	0.77	0.81	0.83	0.77	0.82
					50 lags 1% significance level						
Exp.	0.89	0.95	0.98	0.97	0.98	0.96	0.94	0.98	0.95	0.87	0.91
Weibull	0.89	0.95	0.97	0.95	0.96	0.97	0.95	0.98	0.95	0.89	0.92
Burr	0.85	0.92	0.88	0.89	0.92	0.94	0.93	0.96	0.92	0.90	0.90
					50 lags 5% significance level						
Exp.	0.74	0.89	0.86	0.90	0.92	0.95	0.89	0.96	0.92	0.83	0.89
Weibull	0.73	0.88	0.83	0.88	0.89	0.95	0.89	0.94	0.91	0.85	0.89
Burr	0.65	0.75	0.77	0.79	0.80	0.88	0.85	0.83	0.86	0.80	0.85

Notes: The upper part of the table are LB test results for 30 lags, and the lower part are the results for 50 lags. Significance levels of 1% and 5% are considered. Each figure is the proportion of months in which the null is not rejected.

Table 18: DF test results, AXP

δ(ticks)	2	3	4	5	6	7	8	9	10	11	12
					1% significance level						
Exp.	0.00	0.00	0.00	0.02	0.08	0.13	0.27	0.35	0.45	0.46	0.52
Weibull	0.00	0.00	0.00	0.04	0.12	0.16	0.34	0.45	0.54	0.55	0.64
Burr	0.14	0.45	0.57	0.70	0.74	0.82	0.83	0.86	0.86	0.85	0.86
					5% significance level						
Exp.	0.00	0.00	0.00	0.01	0.02	0.06	0.16	0.20	0.30	0.30	0.40
Weibull	0.00	0.00	0.00	0.02	0.05	0.08	0.22	0.27	0.36	0.48	0.51
Burr	0.11	0.35	0.42	0.51	0.66	0.64	0.74	0.76	0.78	0.74	0.80

Notes: DF test results for significance levels of 1% and 5% are presented. Each figure is the proportion of months in which the null that the assumed density is the true density is not rejected.

Table 19: LR test results, IBM

δ(ticks)	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Wei. vs. Burr	1226.36	756.87	519.93	386.78	289.61	237.13	192.94	162.79	138.47	113.99	98.94	86.18	77.35	68.19	62.71	52.73	47.47	40.66	35.94
Exp. vs. Burr	1456.15	904.80	613.80	457.87	340.02	280.01	234.20	195.56	165.85	142.20	126.07	108.08	101.79	89.13	79.64	69.18	66.39	60.73	52.39
Exp. vs. Wei.	286.91	147.92	93.87	71.09	52.99	44.79	40.66	34.32	31.24	30.33	27.14	24.45	24.41	22.38	21.90	19.86	17.90	20.12	15.73
Wei. vs. Burr	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.99	0.98	0.93	0.92	0.84	0.85	0.75
Exp. vs. Burr	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	0.98	0.94	0.89	0.92	0.93	0.88
Exp. vs. Wei.	0.86	0.83	0.77	0.71	0.69	0.69	0.66	0.70	0.70	0.69	0.66	0.63	0.65	0.68	0.65	0.64	0.64	0.65	0.63

Notes: The first three rows are the LR test statistics (averaged over 132 months), and the last three rows are LR test results presented as proportions of the months in which the null is rejected at the 1% significance level. The assumed density under the null is stated first in the 1st column.

Table 20: LB test results for 30 and 50 lags, IBM

δ(ticks)	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
30 lags 1% significance level																			
Exp.	0.76	0.88	0.92	0.94	0.98	0.98	0.96	0.97	0.96	0.98	0.97	0.91	0.98	0.99	0.95	0.93	0.91	0.95	0.91
Wei.	0.72	0.87	0.89	0.90	0.97	0.98	0.94	0.95	0.98	0.99	0.97	0.93	0.97	0.95	0.87	0.89	0.91	0.89	0.86
Burr	0.68	0.80	0.83	0.81	0.89	0.92	0.89	0.89	0.92	0.92	0.92	0.94	0.94	0.91	0.95	0.92	0.91	0.91	0.88
30 lags 5% significance level																			
Exp.	0.60	0.77	0.81	0.83	0.90	0.92	0.89	0.87	0.90	0.92	0.87	0.87	0.92	0.92	0.90	0.90	0.83	0.89	0.89
Wei.	0.55	0.76	0.78	0.81	0.89	0.91	0.89	0.86	0.92	0.91	0.84	0.88	0.91	0.89	0.83	0.86	0.83	0.82	0.83
Burr	0.43	0.59	0.65	0.67	0.72	0.75	0.75	0.75	0.75	0.82	0.72	0.82	0.82	0.83	0.83	0.80	0.80	0.78	0.80
50 lags 1% significance level																			
Exp.	0.72	0.89	0.89	0.96	0.95	0.98	0.98	0.98	0.98	0.99	0.97	0.89	0.98	0.99	0.95	0.92	0.92	0.93	0.90
Wei.	0.70	0.86	0.88	0.94	0.96	0.98	0.98	0.98	0.99	1.00	0.98	0.91	0.96	0.96	0.89	0.88	0.92	0.88	0.86
Burr	0.69	0.83	0.84	0.91	0.92	0.95	0.94	0.93	0.96	0.97	0.92	0.92	0.94	0.93	0.95	0.91	0.95	0.90	0.87
50 lags 5% significance level																			
Exp.	0.58	0.71	0.80	0.86	0.91	0.92	0.91	0.90	0.93	0.96	0.86	0.87	0.95	0.95	0.90	0.90	0.89	0.89	0.87
Wei.	0.53	0.69	0.77	0.83	0.88	0.90	0.89	0.89	0.95	0.95	0.86	0.89	0.93	0.89	0.82	0.85	0.87	0.83	0.83
Burr	0.49	0.60	0.65	0.67	0.81	0.80	0.79	0.80	0.82	0.86	0.79	0.83	0.84	0.85	0.81	0.82	0.85	0.83	0.83

Notes: The upper part of the table are LB test results for 30 lags, and the lower part are the results for 50 lags. Significance levels of 1% and 5% are considered. Each figure is the proportion of months in which the null is not rejected.

Table 21: DF test results, IBM

δ(ticks)	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1% significance level																			
Exp.	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.06	0.08	0.18	0.17	0.23	0.27	0.30	0.35	0.37	0.45	0.47
Wei.	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.06	0.10	0.18	0.22	0.25	0.35	0.33	0.41	0.41	0.48	0.54
Burr	0.00	0.00	0.11	0.31	0.45	0.54	0.56	0.64	0.75	0.78	0.83	0.83	0.84	0.87	0.88	0.85	0.87	0.87	0.84
5% significance level																			
Exp.	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.02	0.03	0.07	0.11	0.13	0.17	0.23	0.25	0.25	0.36	0.33
Wei.	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.02	0.02	0.04	0.12	0.16	0.15	0.20	0.20	0.27	0.33	0.39	0.37
Burr	0.00	0.00	0.07	0.20	0.30	0.40	0.39	0.48	0.57	0.67	0.70	0.76	0.78	0.75	0.77	0.77	0.80	0.77	0.75

Notes: DF test results for significance levels of 1% and 5% are presented. Each figure is the proportion of months in which the null that the assumed density is the true density is not rejected.

Table 22: Diagnostic test results for 22 DJIA stocks

	LB30(1%)	LB30(5%)	LB50(1%)	LB50(5%)	DF(1%)	DF(5%)
HD	0.93	0.84	0.95	0.83	0.80	0.68
MCD	0.91	0.75	0.94	0.80	0.82	0.73
AXP	0.88	0.69	0.91	0.77	0.73	0.55
IBM	0.94	0.80	0.95	0.83	0.73	0.57
AA	0.90	0.75	0.92	0.80	0.80	0.70
BA	0.87	0.73	0.92	0.82	0.87	0.79
CAT	0.95	0.84	0.95	0.86	0.67	0.51
DD	0.91	0.82	0.96	0.86	0.82	0.67
DIS	0.92	0.78	0.98	0.84	0.92	0.78
GE	0.96	0.80	0.93	0.85	0.82	0.62
JNJ	0.91	0.72	0.91	0.77	0.80	0.68
JPM	0.89	0.70	0.89	0.77	0.58	0.42
KO	0.90	0.73	0.94	0.81	0.83	0.73
MMM	0.96	0.83	0.97	0.89	0.79	0.69
MRK	0.90	0.77	0.92	0.86	0.77	0.61
PG	0.92	0.73	0.94	0.80	0.77	0.63
T	0.92	0.81	0.93	0.84	0.81	0.70
UTX	0.92	0.81	0.96	0.88	0.86	0.69
WMT	0.95	0.78	0.92	0.80	0.79	0.61
XOM	0.91	0.77	0.94	0.83	0.44	0.28
INTC	0.81	0.64	0.85	0.70	0.73	0.67
MSFT	0.85	0.67	0.90	0.77	0.73	0.58
Avg.	0.91	0.76	0.93	0.81	0.77	0.63

Notes: LB and DF test results from the MLE of the HACD-Burr model in equations (13), (14) and (15). The price durations are obtained with δ^* given by the "3-times-spread" rule. Each figure in the table is the proportion of months in which the null is not rejected at the stated significance level.

Web-Appendix B: Forecasting results

Table 23: Volatility estimator means and standard deviations

		AA	AXP	BA	CAT	DD	DIS	GE	HD	IBM	JNJ	JPM	KO	MCD	MMM	MRK	PG	T	UTX	WMT	XOM	avg.
												Mean										
1	PDV	.313	.248	.219	.231	.218	.240	.241	.238	.174	.149	.277	.161	.189	.168	.204	.150	.221	.190	.174	.195	.210
2	ANP_1	.281	.230	.206	.218	.201	.216	.214	.216	.167	.136	.254	.146	.173	.159	.186	.138	.198	.178	.159	.179	.193
3	ANP_2	.292	.241	.215	.231	.209	.220	.219	.225	.172	.141	.265	.151	.180	.167	.195	.143	.205	.187	.166	.185	.200
4	NP	.279	.231	.207	.219	.202	.218	.213	.217	.167	.137	.256	.147	.174	.160	.187	.139	.198	.179	.160	.180	.193
5	PAV_1	.318	.270	.240	.259	.232	.240	.236	.246	.191	.157	.294	.168	.197	.190	.218	.159	.219	.208	.184	.207	.222
6	PAV_2	.318	.263	.235	.258	.227	.232	.234	.244	.185	.153	.289	.162	.196	.185	.216	.155	.220	.202	.182	.200	.218
7	$PABV_1$	.304	.261	.232	.253	.225	.231	.227	.238	.185	.151	.285	.161	.189	.184	.208	.153	.209	.201	.177	.201	.214
8	$PABV_2$	.312	.257	.230	.254	.223	.228	.230	.239	.182	.148	.284	.159	.190	.180	.210	.152	.214	.198	.177	.197	.213
9	RK	.321	.321	.270	.268	.240	.239	.262	.261	.232	.231	.239	.237	.237	.237	.248	.248	.191	.189	.158	.156	.239
10	$RKNP$	.322	.265	.236	.260	.228	.237	.236	.245	.185	.154	.289	.163	.198	.184	.218	.155	.223	.203	.183	.199	.219
11	$TSRV$	.292	.278	.244	.234	.215	.203	.239	.230	.209	.199	.214	.202	.219	.212	.227	.218	.172	.164	.141	.133	.212
12	SBV	.300	.248	.222	.245	.216	.219	.220	.230	.175	.142	.271	.152	.184	.175	.201	.145	.206	.191	.170	.190	.205
13	RV_5	.312	.257	.229	.252	.224	.227	.228	.238	.180	.149	.281	.158	.192	.181	.210	.151	.217	.198	.177	.195	.213
14	SRV_5	.317	.259	.233	.256	.226	.231	.231	.241	.182	.151	.284	.160	.194	.184	.213	.153	.219	.201	.179	.197	.216
15	ATM	.402	.272	.240	.267	.215	.243	.237	.244	.202	.153	.293	.159	.194	.189	.212	.150	.207	.209	.181	.193	.223
												Standard Deviation										
1	PDV	.176	.182	.108	.120	.107	.122	.161	.123	.092	.074	.188	.072	.090	.083	.095	.067	.115	.092	.081	.102	.113
2	ANP_1	.165	.178	.109	.121	.107	.119	.150	.121	.096	.075	.186	.073	.092	.085	.097	.069	.113	.094	.082	.104	.112
3	ANP_2	.173	.187	.111	.126	.112	.123	.157	.127	.100	.080	.198	.078	.096	.089	.104	.073	.121	.099	.087	.107	.117
4	NP	.166	.180	.110	.122	.108	.121	.151	.122	.097	.076	.188	.074	.093	.085	.098	.069	.115	.095	.083	.104	.113
5	PAV_1	.189	.206	.122	.139	.123	.134	.175	.139	.112	.090	.226	.088	.107	.105	.125	.085	.132	.109	.096	.120	.131
6	PAV_2	.188	.206	.120	.138	.125	.136	.176	.140	.110	.095	.226	.091	.111	.105	.126	.085	.138	.112	.099	.117	.132
7	$PABV_1$	.187	.202	.119	.139	.121	.130	.173	.136	.108	.086	.221	.085	.101	.101	.117	.081	.125	.107	.093	.118	.128
8	$PABV_2$	.189	.204	.120	.138	.125	.133	.174	.139	.109	.092	.223	.089	.107	.102	.122	.083	.134	.112	.097	.116	.130
9	RK	.189	.189	.207	.208	.122	.122	.139	.139	.125	.125	.135	.135	.176	.176	.141	.142	.113	.112	.093	.094	.144
10	$RKNP$	.188	.206	.121	.139	.125	.137	.174	.138	.108	.095	.222	.089	.110	.098	.125	.084	.137	.111	.099	.115	.131
11	$TSRV$	.178	.176	.194	.189	.113	.111	.130	.129	.118	.115	.126	.123	.166	.164	.132	.128	.105	.102	.088	.086	.134
12	SBV	.180	.193	.117	.134	.120	.127	.164	.132	.105	.088	.213	.085	.104	.100	.115	.079	.130	.106	.093	.114	.125
13	RV_5	.185	.201	.118	.136	.122	.132	.167	.135	.108	.092	.221	.087	.108	.101	.126	.082	.136	.106	.097	.119	.129
14	SRV_5	.185	.201	.120	.137	.123	.134	.170	.137	.107	.093	.222	.088	.108	.103	.125	.086	.135	.109	.097	.117	.130
15	ATM	.192	.175	.089	.101	.089	.100	.140	.104	.088	.064	.172	.065	.073	.070	.080	.059	.102	.086	.069	.075	.100

Notes: Mean and standard deviation statistics for 15 daily volatility estimators for 20 stocks using data from January 2002 to December 2012. The means and standard deviations are for annualized volatilities.

Table 24: Volatility estimator autocorrelations

		AA	AXP	BA	CAT	DD	DIS	GE	HD	IBM	JNJ	JPM	KO	MCD	MMM	MRK	PG	T	UTX	WMT	XOM	avg.
												First order										
1	PDV	.850	.888	.834	.825	.786	.828	.845	.831	.845	.812	.878	.775	.801	.748	.773	.768	.840	.789	.777	.722	.811
2	ANP_1	.791	.817	.763	.758	.705	.755	.776	.728	.760	.711	.824	.712	.660	.687	.672	.639	.751	.706	.710	.636	.728
3	ANP_2	.788	.790	.778	.772	.693	.721	.754	.701	.741	.673	.810	.665	.586	.675	.641	.604	.726	.703	.674	.598	.705
4	NP	.793	.815	.763	.741	.693	.744	.775	.724	.768	.706	.823	.711	.632	.694	.680	.639	.756	.708	.728	.645	.727
5	PAV_1	.772	.774	.754	.743	.695	.646	.734	.649	.704	.622	.760	.613	.500	.266	.470	.434	.672	.657	.633	.557	.633
6	PAV_2	.759	.684	.769	.783	.666	.564	.693	.608	.720	.511	.726	.583	.428	.333	.448	.453	.613	.626	.493	.527	.599
7	$PABV_1$	.774	.794	.746	.740	.680	.650	.726	.650	.720	.613	.756	.596	.486	.377	.511	.501	.684	.661	.585	.545	.640
8	$PABV_2$	.780	.677	.776	.777	.655	.584	.689	.550	.734	.472	.721	.560	.415	.482	.444	.417	.616	.616	.517	.520	.600
9	RK	.764	.760	.761	.746	.751	.752	.749	.760	.685	.678	.634	.624	.733	.729	.641	.630	.708	.714	.623	.601	.702
10	$RKNP$	.711	.640	.760	.747	.661	.633	.681	.617	.709	.583	.713	.671	.440	.637	.522	.564	.650	.614	.517	.628	.635
11	$TSRV$	.760	.729	.695	.636	.748	.721	.777	.738	.661	.611	.567	.484	.703	.669	.604	.556	.705	.673	.511	.433	.649
12	SBV	.783	.683	.776	.785	.732	.760	.708	.696	.726	.627	.741	.609	.438	.307	.578	.504	.656	.647	.636	.524	.646
13	RV_5	.738	.612	.764	.782	.737	.700	.713	.687	.666	.661	.721	.690	.514	.451	.343	.628	.664	.709	.674	.439	.645
14	SRV_5	.742	.666	.768	.763	.687	.609	.681	.588	.697	.553	.727	.613	.435	.279	.470	.333	.628	.612	.535	.460	.592
15	ATM	.941	.936	.926	.947	.933	.943	.936	.955	.948	.891	.925	.943	.939	.936	.908	.932	.904	.943	.942	.855	.929
												Fifth order										
1	PDV	.723	.823	.725	.730	.704	.741	.715	.747	.763	.707	.694	.666	.717	.670	.659	.667	.734	.688	.679	.694	.712
2	ANP_1	.651	.749	.652	.650	.613	.649	.648	.636	.696	.597	.585	.597	.564	.604	.544	.545	.640	.594	.595	.602	.621
3	ANP_2	.635	.711	.660	.664	.596	.602	.620	.609	.682	.545	.534	.551	.508	.590	.507	.503	.601	.585	.547	.561	.591
4	NP	.640	.750	.657	.634	.605	.640	.651	.619	.697	.587	.582	.586	.545	.601	.550	.549	.643	.604	.601	.608	.617
5	PAV_1	.623	.695	.642	.640	.588	.530	.604	.558	.651	.496	.495	.504	.430	.211	.355	.358	.548	.555	.520	.523	.526
6	PAV_2	.590	.591	.637	.664	.542	.451	.545	.530	.649	.393	.425	.464	.356	.268	.312	.360	.492	.520	.384	.469	.482
7	$PABV_1$	.625	.719	.633	.635	.584	.527	.597	.554	.664	.496	.501	.495	.425	.303	.395	.411	.575	.549	.480	.518	.534
8	$PABV_2$	.595	.590	.641	.651	.524	.467	.539	.485	.653	.372	.425	.456	.351	.404	.321	.336	.495	.507	.402	.465	.484
9	RK	.608	.601	.669	.648	.633	.629	.646	.653	.573	.562	.516	.501	.599	.590	.557	.550	.654	.656	.492	.468	.590
10	$RKNP$	.553	.546	.634	.628	.542	.514	.527	.533	.645	.462	.424	.515	.360	.553	.362	.445	.486	.511	.389	.540	.508
11	$TSRV$	.585	.549	.600	.544	.617	.583	.660	.628	.535	.504	.449	.393	.559	.518	.527	.483	.641	.608	.392	.334	.535
12	SBV	.616	.595	.649	.659	.610	.651	.544	.603	.652	.505	.441	.497	.367	.245	.415	.424	.501	.554	.500	.476	.525
13	RV_5	.545	.543	.619	.666	.588	.591	.537	.603	.622	.527	.417	.570	.405	.388	.232	.477	.491	.596	.512	.355	.514
14	SRV_5	.566	.584	.627	.638	.554	.492	.525	.509	.629	.440	.430	.488	.345	.218	.319	.256	.476	.511	.395	.391	.470
15	ATM	.836	.852	.838	.860	.833	.872	.820	.871	.814	.758	.801	.846	.845	.812	.822	.841	.809	.816	.843	.749	.827

Notes: First-order and fifth-order autocorrelation statistics for 15 daily volatility estimators for 20 stocks using data from January 2002 to December 2012.

Table 25: RMSE values, 15 individual volatility estimators, three horizons, trade data

		AA	AXP	BA	CAT	DD	DIS	GE	HD	IBM	JNJ	JPM	KO	MCD	MMM	MRK	PG	T	UTX	WMT	XOM
											one day ahead										
1	PDV	.0048	.0031	.0028	.0039	.0030	.0029	.0049	.0021	.0021	.0013	.0041	.0013	.0012	.0067	.0023	.0021	.0014	.0028	.0010	.0026
2	ANP_1	.0049	.0032	.0029	.0041	.0031	.0029	.0051	.0021	.0021	.0013	.0042	.0013	.0012	.0068	.0023	.0021	.0014	.0028	.0011	.0027
3	ANP_2	.0049	.0031	.0029	.0041	.0031	.0028	.0051	.0021	.0021	.0013	.0041	.0012	.0012	.0068	.0023	.0021	.0014	.0027	.0011	.0027
4	NP	.0050	.0033	.0029	.0041	.0032	.0029	.0053	.0021	.0021	.0013	.0043	.0013	.0012	.0069	.0023	.0021	.0015	.0028	.0011	.0028
5	PAV_1	.0051	.0031	.0029	.0042	.0030	.0027	.0051	.0021	.0021	.0013	.0041	.0012	.0012	.0075	.0024	.0022	.0015	.0028	.0011	.0027
6	PAV_2	.0050	.0031	.0030	.0042	.0030	.0027	.0052	.0021	.0021	.0014	.0041	.0012	.0012	.0074	.0024	.0022	.0015	.0028	.0011	.0028
7	$PABV_1$	.0051	.0031	.0029	.0042	.0031	.0027	.0051	.0022	.0021	.0013	.0041	.0012	.0012	.0073	.0024	.0022	.0015	.0028	.0011	.0027
8	$PABV_2$	.0049	.0031	.0030	.0041	.0030	.0027	.0052	.0021	.0021	.0014	.0041	.0012	.0012	.0071	.0025	.0021	.0015	.0028	.0011	.0028
9	RK	.0050	.0045	.0036	.0059	.0035	.0029	.0065	.0035	.0025	.0017	.0069	.0017	.0038	.0073	.0032	.0024	.0024	.0029	.0015	.0034
10	$RKNP$	.0050	.0031	.0030	.0042	.0030	.0027	.0052	.0022	.0021	.0014	.0042	.0012	.0012	.0069	.0024	.0022	.0015	.0028	.0011	.0029
11	$TSRV$	.0050	.0043	.0036	.0060	.0035	.0031	.0064	.0036	.0026	.0018	.0070	.0018	.0039	.0073	.0032	.0024	.0025	.0031	.0016	.0036
12	SBV	.0049	.0031	.0029	.0041	.0030	.0026	.0052	.0021	.0021	.0013	.0041	.0012	.0012	.0074	.0024	.0022	.0015	.0027	.0011	.0028
13	RV_5	.0051	.0032	.0030	.0042	.0030	.0027	.0052	.0022	.0021	.0014	.0042	.0012	.0012	.0072	.0024	.0022	.0015	.0028	.0011	.0029
14	SRV_5	.0050	.0031	.0029	.0041	.0030	.0027	.0052	.0022	.0021	.0014	.0041	.0012	.0012	.0075	.0024	.0025	.0015	.0028	.0011	.0029
15	ATM	.0060	.0031	.0028	.0042	.0032	.0024	.0054	.0021	.0026	.0016	.0049	.0014	.0013	.0068	.0024	.0023	.0017	.0028	.0013	.0027
											one week ahead										
1	PDV	.0188	.0121	.0105	.0148	.0117	.0109	.0201	.0075	.0090	.0051	.0152	.0051	.0041	.0285	.0089	.0083	.0051	.0113	.0037	.0110
2	ANP_1	.0195	.0130	.0110	.0157	.0127	.0113	.0214	.0079	.0094	.0055	.0163	.0053	.0043	.0290	.0094	.0086	.0055	.0119	.0041	.0115
3	ANP_2	.0194	.0130	.0110	.0159	.0124	.0109	.0213	.0080	.0093	.0055	.0164	.0052	.0043	.0290	.0095	.0086	.0056	.0117	.0041	.0117
4	NP	.0196	.0132	.0110	.0158	.0129	.0113	.0222	.0079	.0095	.0056	.0165	.0053	.0043	.0291	.0095	.0086	.0057	.0121	.0041	.0118
5	PAV_1	.0196	.0132	.0112	.0163	.0128	.0111	.0218	.0083	.0095	.0056	.0164	.0052	.0045	.0337	.0101	.0097	.0061	.0120	.0045	.0122
6	PAV_2	.0191	.0134	.0113	.0161	.0125	.0117	.0227	.0082	.0095	.0060	.0170	.0052	.0044	.0330	.0102	.0089	.0059	.0118	.0042	.0127
7	$PABV_1$	.0194	.0133	.0112	.0164	.0128	.0110	.0218	.0085	.0096	.0056	.0165	.0053	.0047	.0322	.0105	.0095	.0062	.0120	.0046	.0122
8	$PABV_2$	.0191	.0135	.0113	.0159	.0125	.0115	.0228	.0084	.0098	.0061	.0172	.0053	.0046	.0310	.0107	.0088	.0063	.0120	.0043	.0127
9	RK	.0194	.0191	.0132	.0224	.0130	.0105	.0284	.0152	.0110	.0072	.0290	.0073	.0191	.0301	.0138	.0096	.0120	.0119	.0067	.0142
10	$RKNP$	.0193	.0136	.0113	.0166	.0122	.0115	.0228	.0084	.0095	.0060	.0177	.0053	.0044	.0295	.0098	.0085	.0059	.0118	.0043	.0126
11	$TSRV$	.0191	.0181	.0129	.0222	.0130	.0120	.0279	.0160	.0111	.0074	.0299	.0082	.0192	.0303	.0137	.0097	.0123	.0126	.0072	.0147
12	SBV	.0192	.0133	.0112	.0161	.0121	.0108	.0227	.0083	.0095	.0058	.0173	.0051	.0044	.0334	.0098	.0095	.0057	.0116	.0042	.0125
13	RV_5	.0197	.0135	.0111	.0166	.0117	.0114	.0221	.0081	.0093	.0060	.0175	.0053	.0044	.0316	.0097	.0092	.0059	.0114	.0042	.0127
14	SRV_5	.0193	.0132	.0112	.0163	.0123	.0115	.0227	.0082	.0097	.0059	.0173	.0052	.0044	.0338	.0099	.0109	.0058	.0120	.0042	.0130
15	ATM	.0228	.0141	.0104	.0167	.0134	.0093	.0225	.0068	.0110	.0063	.0197	.0054	.0046	.0291	.0092	.0088	.0063	.0116	.0043	.0134
											one month ahead										
1	PDV	.0841	.0547	.0486	.0713	.0520	.0472	.0861	.0356	.0428	.0259	.0696	.0231	.0172	.1144	.0434	.0358	.0254	.0521	.0180	.0557
2	ANP_1	.0831	.0566	.0493	.0741	.0552	.0489	.0893	.0371	.0437	.0267	.0720	.0234	.0180	.1157	.0448	.0369	.0265	.0535	.0189	.0562
3	ANP_2	.0837	.0571	.0490	.0742	.0547	.0478	.0886	.0380	.0434	.0268	.0722	.0234	.0187	.1160	.0452	.0375	.0269	.0528	.0194	.0574
4	NP	.0824	.0570	.0499	.0744	.0543	.0494	.0928	.0376	.0438	.0267	.0725	.0229	.0180	.1158	.0455	.0369	.0271	.0540	.0193	.0569
5	PAV_1	.0820	.0587	.0504	.0750	.0576	.0497	.0904	.0392	.0445	.0272	.0713	.0237	.0201	.1310	.0480	.0412	.0308	.0541	.0208	.0593
6	PAV_2	.0813	.0599	.0491	.0724	.0549	.0538	.0937	.0387	.0445	.0285	.0739	.0233	.0200	.1267	.0471	.0387	.0294	.0524	.0201	.0603
7	$PABV_1$	.0806	.0584	.0506	.0753	.0579	.0493	.0912	.0414	.0449	.0273	.0720	.0238	.0216	.1263	.0499	.0407	.0315	.0537	.0218	.0597
8	$PABV_2$	.0808	.0603	.0495	.0708	.0548	.0530	.0943	.0403	.0454	.0293	.0750	.0239	.0211	.1202	.0492	.0385	.0316	.0529	.0207	.0606
9	RK	.0813	.0901	.0503	.0868	.0546	.0439	.1253	.0666	.0442	.0299	.1223	.0345	.0858	.1090	.0653	.0384	.0626	.0526	.0320	.0601
10	$RKNP$	.0817	.0589	.0488	.0739	.0526	.0531	.0947	.0386	.0435	.0282	.0775	.0230	.0204	.1164	.0439	.0369	.0285	.0524	.0197	.0611
11	$TSRV$	.0788	.0849	.0484	.0833	.0528	.0484	.1217	.0694	.0438	.0300	.1286	.0383	.0857	.1080	.0651	.0394	.0633	.0531	.0340	.0616
12	SBV	.0815	.0590	.0492	.0726	.0535	.0502	.0943	.0390	.0439	.0273	.0740	.0230	.0202	.1271	.0461	.0412	.0279	.0521	.0198	.0603
13	RV_5	.0838	.0609	.0489	.0757	.0521	.0521	.0914	.0377	.0435	.0281	.0753	.0231	.0204	.1232	.0454	.0397	.0275	.0507	.0195	.0619
14	SRV_5	.0826	.0587	.0489	.0731	.0539	.0530	.0933	.0382	.0448	.0278	.0749	.0229	.0199	.1288	.0455	.0458	.0275	.0532	.0197	.0615
15	ATM	.0916	.0693	.0507	.0799	.0603	.0473	.0945	.0408	.0459	.0299	.0931	.0253	.0201	.1154	.0418	.0347	.0263	.0542	.0219	.0617

Table 26: MCS p-values, RMSE loss function, 15 individual forecasts, three horizons, trade data

AA		AXP		BA		CAT		DD		DIS		GE		HD		IBM		JNJ	
one day ahead																			
.00	15	.00	11	.02	11	.00	9	.00	9	.00	11	.02	11	.00	11	.00	11	.00	11
.01	5	.00	9	.02	9	.00	11	.00	11	.00	1	.02	9	.00	9	.00	15	.00	9
.01	7	.00	13	.02	6	.00	5	.00	4	.00	9	.02	4	.01	10	.00	9	.00	15
.02	13	.04	4	.05	8	.00	7	.00	2	.01	2	.03	15	.03	14	.02	4	.00	13
.02	9	.04	2	.06	10	.02	13	.02	15	.01	4	.03	14	.03	7	.04	10	.00	12
.09	11	.04	3	.06	5	.02	10	.02	3	.01	6	.03	6	.03	4	.04	1	.00	6
.16	4	.04	7	.06	12	.02	6	.02	5	.01	3	.03	10	.14	13	.04	14	.00	14
.22	6	.04	8	.06	14	.02	14	.02	7	.01	13	.03	12	.14	5	.04	2	.00	8
.34	10	.04	10	.06	7	.02	15	.06	6	.02	8	.03	3	.21	8	.09	13	.00	10
.47	8	.10	6	.06	13	.02	8	.06	8	.02	14	.03	13	.21	12	.09	6	.00	4
.47	14	.36	5	.06	4	.02	4	.13	10	.06	5	.06	2	.21	6	.09	8	.01	2
.58	2	.64	1	.09	2	.02	3	.13	14	.11	10	.06	8	.21	15	.39	5	.13	5
.66	12	.87	15	.17	3	.02	12	.37	1	.11	7	.06	5	.21	2	.61	7	.13	3
.66	3	.88	12	.45	1	.02	2	.60	12	.13	12	.06	7	.22	3	.68	12	.13	7
1	1	1	14	1	15	1	1	1	13	1	15	1	1	1	1	1	3	1	1
one week ahead																			
.03	15	.00	9	.01	9	.00	9	.00	5	.02	6	.00	9	.00	11	.00	11	.00	11
.13	5	.00	11	.01	11	.00	11	.00	7	.02	10	.00	11	.00	9	.00	15	.00	9
.15	4	.00	8	.01	8	.00	5	.01	4	.02	14	.00	10	.00	8	.00	9	.00	10
.35	2	.00	10	.01	6	.00	13	.01	9	.02	8	.00	14	.00	7	.00	7	.00	6
.35	13	.02	13	.02	12	.00	7	.01	15	.02	11	.00	8	.00	10	.01	14	.00	12
.35	3	.02	6	.02	5	.00	10	.01	11	.02	4	.00	13	.00	5	.01	8	.00	14
.35	7	.02	12	.02	14	.00	14	.01	2	.02	7	.00	15	.02	6	.01	5	.00	15
.35	14	.02	15	.02	7	.00	6	.01	14	.02	13	.00	12	.02	12	.01	6	.00	13
.35	9	.02	5	.05	13	.00	12	.01	8	.02	5	.00	4	.03	14	.01	10	.00	8
.35	10	.02	7	.05	10	.00	15	.01	6	.02	2	.00	5	.03	3	.04	4	.00	7
.35	12	.02	4	.05	3	.00	3	.01	3	.52	1	.00	6	.13	4	.12	2	.00	5
.73	8	.02	2	.05	2	.00	2	.01	12	.52	12	.00	2	.27	13	.12	12	.05	2
.73	6	.02	3	.05	4	.00	4	.14	10	.52	3	.00	7	.27	2	.12	3	.05	3
.73	11	.02	14	.77	1	.02	8	.91	13	.52	9	.00	3	.33	1	.13	13	.05	4
1	1	1	1	1	15	1	1	1	1	1	15	1	1	1	15	1	1	1	1
one month ahead																			
.11	3	.01	9	.03	7	.00	9	.00	5	.01	10	.02	9	.00	11	.35	7	.00	15
.11	15	.01	11	.03	5	.00	15	.00	7	.01	6	.02	11	.00	9	.39	5	.00	8
.11	13	.01	15	.11	15	.00	13	.00	15	.01	8	.02	10	.00	8	.39	8	.00	11
.11	14	.01	8	.11	4	.01	11	.00	2	.01	12	.02	8	.00	7	.39	15	.00	6
.11	2	.01	13	.11	8	.01	5	.00	6	.01	14	.02	12	.00	5	.39	6	.02	9
.11	5	.01	6	.11	9	.03	3	.00	3	.01	5	.02	6	.00	12	.39	14	.02	10
.11	1	.03	12	.69	6	.03	7	.00	8	.01	13	.02	14	.00	15	.39	12	.02	13
.11	4	.03	14	.69	12	.03	2	.00	9	.01	7	.02	15	.00	3	.39	9	.02	5
.11	12	.07	5	.78	2	.03	10	.03	14	.01	4	.02	4	.00	6	.39	4	.02	7
.13	6	.07	10	.85	3	.03	4	.03	4	.01	2	.02	13	.00	10	.52	2	.02	14
.13	10	.19	3	.86	14	.06	12	.03	12	.22	11	.02	7	.00	14	.72	13	.02	3
.37	9	.19	7	.86	13	.07	14	.81	11	.56	3	.03	5	.00	4	.72	10	.05	12
.37	8	.19	2	.88	10	.11	6	.81	10	.60	1	.07	2	.03	13	.72	11	.05	4
.37	7	.19	4	.88	1	.68	1	.82	13	.60	15	.07	3	.03	2	.72	3	.05	2
1	11	1	1	1	11	1	8	1	1	1	9	1	1	1	1	1	1	1	1

JPM		KO		MCD		MMM		MRK		PG		T		UTX		WMT		XOM	
one day ahead																			
.00	11	.00	11	.00	9	.00	5	.00	9	.00	11	.00	11	.00	11	.00	11	.00	11
.00	9	.00	9	.00	11	.00	14	.00	11	.00	9	.00	9	.01	10	.00	9	.00	9
.00	15	.00	15	.00	15	.00	11	.01	6	.02	15	.00	15	.01	4	.00	15	.00	13
.00	4	.00	1	.00	7	.00	13	.01	8	.02	12	.00	8	.04	9	.00	14	.00	10
.00	2	.01	4	.00	1	.00	6	.01	10	.02	14	.01	13	.04	8	.00	7	.00	14
.02	12	.01	10	.00	4	.00	7	.01	15	.02	13	.01	4	.04	6	.00	13	.00	8
.02	13	.02	13	.00	8	.00	9	.01	13	.02	6	.13	6	.05	14	.00	6	.00	6
.02	10	.02	2	.34	13	.00	8	.01	14	.02	10	.13	12	.06	5	.00	10	.00	4
.56	8	.06	6	.34	2	.00	12	.01	5	.02	5	.13	10	.06	1	.00	5	.02	7
.56	3	.06	3	.57	3	.00	10	.01	7	.02	8	.13	14	.06	2	.00	12	.02	12
.70	6	.06	7	.61	5	.00	4	.01	4	.02	7	.13	7	.29	15	.00	8	.02	5
.71	14	.06	8	.84	10	.00	2	.06	3	.02	2	.13	5	.29	7	.01	4	.02	15
.71	5	.42	5	.84	14	.01	3	.06	12	.02	4	.13	2	.56	13	.85	2	.02	2
.71	7	.42	14	.84	6	.35	15	.08	2	.02	3	.13	3	.56	3	.85	3	.02	3
1	1	1	12	1	12	1	1	1	1	1	1	1	1	1	12	1	1	1	1
one week ahead																			
.00	9	.00	9	.00	9	.02	5	.00	9	.02	11	.00	11	.00	5	.00	11	.00	11
.00	11	.00	11	.00	11	.02	7	.00	11	.02	7	.00	9	.00	4	.00	9	.00	15
.00	13	.02	10	.00	7	.02	14	.00	7	.02	9	.00	5	.00	7	.00	5	.00	9
.00	15	.02	2	.00	8	.02	6	.00	8	.02	12	.00	7	.00	11	.00	7	.00	8
.00	8	.02	8	.00	15	.02	12	.00	6	.02	13	.00	8	.03	2	.00	8	.00	13
.00	14	.02	4	.00	5	.02	13	.00	10	.03	5	.00	13	.14	8	.00	6	.00	6
.00	6	.02	15	.00	13	.02	8	.00	12	.03	14	.00	15	.14	14	.00	14	.00	12
.00	12	.02	13	.00	14	.02	11	.00	5	.03	6	.00	6	.14	6	.00	10	.00	14
.00	10	.02	5	.00	10	.02	9	.00	14	.07	8	.00	10	.14	3	.00	13	.00	10
.00	5	.02	6	.00	12	.02	10	.00	4	.30	15	.00	12	.32	9	.00	15	.00	5
.00	4	.02	7	.00	3	.05	4	.00	3	.30	4	.00	4	.32	10	.00	3	.00	7
.00	7	.11	3	.00	4	.05	3	.01	13	.30	2	.00	14	.35	12	.00	4	.00	4
.01	2	.53	14	.00	6	.18	2	.01	2	.30	3	.00	3	.80	15	.00	2	.00	3
.02	3	.83	12	.00	2	.53	15	.57	15	.56	10	.00	2	.80	13	.01	12	.01	2
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
one month ahead																			
.00	9	.00	9	.00	9	.00	7	.00	11	.01	7	.00	11	.03	5	.00	11	.00	13
.01	11	.00	11	.00	11	.00	5	.00	9	.01	5	.00	9	.04	7	.00	9	.00	14
.01	15	.01	5	.00	12	.00	6	.03	7	.01	14	.00	7	.04	4	.00	7	.00	10
.01	8	.01	8	.00	8	.00	14	.04	8	.01	13	.00	5	.04	2	.00	8	.00	8
.03	14	.01	15	.00	13	.00	12	.04	5	.01	12	.00	8	.63	15	.00	5	.00	7
.03	10	.01	7	.00	10	.00	13	.04	12	.01	8	.01	6	.63	3	.00	6	.00	15
.03	6	.01	3	.00	7	.00	8	.04	4	.01	6	.01	10	.66	14	.00	15	.00	6
.03	13	.01	2	.00	14	.00	10	.09	6	.01	11	.01	12	.66	8	.00	12	.00	12
.03	12	.01	6	.00	6	.02	3	.10	3	.01	3	.01	14	.68	11	.00	3	.00	5
.03	4	.40	13	.00	5	.05	15	.17	13	.01	9	.01	13	.68	6	.00	14	.00	11
.06	3	.73	1	.00	15	.05	4	.29	2	.01	10	.01	3	.68	10	.00	13	.00	9
.15	7	.73	10	.00	3	.08	2	.29	14	.01	4	.01	4	.68	9	.00	4	.00	3
.15	2	.73	12	.04	2	.54	1	.67	10	.01	2	.01	2	.68	12	.00	10	.00	4
.20	5	.82	4	.06	4	.54	9	.67	1	.57	1	.49	15	.68	1	.00	2	.31	2
1	1	1	14	1	1	1	11	1	15	1	15	1	1	1	13	1	1	1	1

Notes: MCS p-values for the RMSE loss function. In each stock sub-panel we report the corresponding p-values on the left and on the right we list the estimator numbers ranked in the ascending orders of the p-values. Estimators 1 to 15 are: PDV, ANP_1, ANP_2, NP, PAV_1, PAV_2, $PABV_1$, $PABV_2$, RK, RKNP, TSRV, SBV, RV_5, SRV_5 and ATM.

Table 27: QLIKE values, 15 individual volatility estimators, three horizons, trade data

		AA	AXP	BA	CAT	DD	DIS	GE	HD	IBM	JNJ	JPM	KO	MCD	MMM	MRK	PG	T	UTX	WMT	XOM
											one day ahead										
1	PDV	.0275	.0278	.0288	.0279	.0293	.0345	.0317	.0257	.0326	.0295	.0306	.0321	.0323	.0367	.0269	.0304	.0267	.0336	.0273	.0283
2	ANP_1	.0278	.0288	.0297	.0287	.0308	.0344	.0337	.0262	.0325	.0307	.0316	.0321	.0318	.0386	.0279	.0315	.0276	.0341	.0278	.0293
3	ANP_2	.0280	.0281	.0298	.0287	.0298	.0319	.0337	.0258	.0316	.0297	.0309	.0306	.0315	.0385	.0281	.0311	.0276	.0334	.0278	.0289
4	NP	.0284	.0294	.0297	.0289	.0315	.0346	.0359	.0265	.0331	.0309	.0319	.0326	.0320	.0394	.0285	.0316	.0291	.0348	.0283	.0305
5	PAV_1	.0286	.0283	.0302	.0295	.0295	.0310	.0329	.0263	.0321	.0300	.0304	.0302	.0317	.0414	.0284	.0322	.0284	.0340	.0289	.0299
6	PAV_2	.0282	.0282	.0309	.0294	.0290	.0311	.0336	.0260	.0313	.0311	.0306	.0301	.0313	.0410	.0286	.0318	.0283	.0334	.0286	.0297
7	$PABV_1$	.0289	.0286	.0301	.0294	.0298	.0309	.0331	.0268	.0319	.0297	.0304	.0305	.0326	.0410	.0289	.0325	.0284	.0336	.0295	.0304
8	$PABV_2$	.0279	.0282	.0308	.0290	.0290	.0308	.0334	.0261	.0312	.0309	.0307	.0300	.0317	.0405	.0290	.0318	.0288	.0333	.0289	.0297
9	RK	.0285	.0462	.0388	.0425	.0360	.0342	.0601	.0438	.0417	.0411	.0761	.0487	.0917	.0486	.0446	.0400	.0492	.0350	.0412	.0442
10	$RKNP$	.0282	.0290	.0315	.0297	.0295	.0309	.0334	.0269	.0323	.0315	.0315	.0308	.0316	.0398	.0292	.0329	.0286	.0335	.0287	.0305
11	$TSRV$	.0284	.0451	.0385	.0431	.0363	.0374	.0593	.0451	.0428	.0435	.0756	.0512	.0912	.0485	.0440	.0405	.0491	.0383	.0429	.0467
12	SBV	.0277	.0281	.0304	.0286	.0289	.0300	.0336	.0262	.0309	.0308	.0311	.0294	.0311	.0415	.0282	.0321	.0281	.0327	.0284	.0294
13	RV_5	.0288	.0291	.0307	.0294	.0289	.0310	.0337	.0263	.0316	.0316	.0313	.0306	.0314	.0411	.0289	.0331	.0289	.0327	.0287	.0304
14	SRV_5	.0280	.0279	.0304	.0292	.0291	.0306	.0338	.0264	.0314	.0311	.0309	.0297	.0313	.0415	.0285	.0333	.0285	.0331	.0286	.0298
15	ATM	.0337	.0277	.0286	.0307	.0334	.0276	.0371	.0278	.0417	.0358	.0360	.0360	.0337	.0380	.0305	.0369	.0335	.0343	.0333	.0325
											one week ahead										
1	PDV	.0199	.0194	.0197	.0192	.0223	.0257	.0240	.0173	.0286	.0235	.0217	.0248	.0206	.0379	.0199	.0276	.0194	.0273	.0186	.0234
2	ANP_1	.0206	.0211	.0207	.0204	.0246	.0268	.0269	.0184	.0293	.0254	.0236	.0256	.0210	.0392	.0216	.0291	.0210	.0286	.0204	.0244
3	ANP_2	.0207	.0208	.0206	.0206	.0235	.0248	.0266	.0186	.0285	.0246	.0233	.0247	.0210	.0393	.0221	.0286	.0210	.0276	.0204	.0245
4	NP	.0209	.0214	.0207	.0204	.0251	.0269	.0293	.0185	.0297	.0256	.0236	.0254	.0214	.0395	.0219	.0292	.0223	.0293	.0209	.0255
5	PAV_1	.0206	.0211	.0211	.0213	.0240	.0248	.0267	.0193	.0291	.0251	.0227	.0242	.0219	.0433	.0228	.0303	.0227	.0285	.0220	.0258
6	PAV_2	.0201	.0214	.0210	.0210	.0229	.0253	.0270	.0187	.0280	.0261	.0235	.0239	.0212	.0426	.0226	.0284	.0216	.0264	.0208	.0257
7	$PABV_1$	.0204	.0212	.0212	.0213	.0241	.0246	.0268	.0201	.0291	.0249	.0225	.0244	.0230	.0427	.0237	.0307	.0229	.0281	.0228	.0262
8	$PABV_2$	.0198	.0216	.0209	.0204	.0227	.0248	.0267	.0191	.0284	.0261	.0237	.0242	.0218	.0413	.0232	.0284	.0224	.0263	.0211	.0257
9	RK	.0204	.0366	.0271	.0310	.0264	.0227	.0532	.0355	.0348	.0346	.0629	.0395	.0819	.0454	.0392	.0360	.0427	.0274	.0315	.0373
10	$RKNP$	.0204	.0218	.0208	.0216	.0228	.0249	.0270	.0191	.0287	.0260	.0248	.0246	.0213	.0402	.0226	.0283	.0215	.0257	.0209	.0263
11	$TSRV$	.0201	.0350	.0263	.0309	.0262	.0257	.0518	.0369	.0357	.0363	.0625	.0420	.0814	.0450	.0384	.0364	.0421	.0299	.0332	.0388
12	SBV	.0200	.0214	.0207	.0208	.0228	.0238	.0271	.0189	.0279	.0256	.0243	.0235	.0211	.0426	.0221	.0292	.0211	.0260	.0205	.0255
13	RV_5	.0209	.0221	.0206	.0217	.0222	.0253	.0273	.0185	.0279	.0260	.0247	.0241	.0214	.0421	.0224	.0303	.0219	.0255	.0204	.0265
14	SRV_5	.0202	.0210	.0206	.0211	.0230	.0248	.0270	.0188	.0281	.0256	.0241	.0235	.0210	.0422	.0221	.0301	.0212	.0262	.0207	.0259
15	ATM	.0239	.0202	.0183	.0218	.0257	.0187	.0302	.0163	.0354	.0285	.0247	.0261	.0227	.0382	.0215	.0310	.0238	.0258	.0209	.0298
											one month ahead										
1	PDV	.0212	.0212	.0230	.0237	.0257	.0292	.0298	.0202	.0380	.0330	.0235	.0275	.0203	.0544	.0271	.0378	.0235	.0335	.0205	.0366
2	ANP_1	.0210	.0225	.0234	.0246	.0276	.0303	.0317	.0214	.0378	.0344	.0250	.0276	.0209	.0558	.0283	.0397	.0247	.0344	.0215	.0373
3	ANP_2	.0212	.0224	.0230	.0245	.0269	.0287	.0310	.0218	.0367	.0342	.0245	.0269	.0214	.0559	.0285	.0400	.0246	.0335	.0219	.0380
4	NP	.0211	.0226	.0238	.0247	.0271	.0308	.0341	.0217	.0380	.0341	.0250	.0268	.0208	.0558	.0287	.0395	.0259	.0350	.0220	.0378
5	PAV_1	.0207	.0228	.0238	.0249	.0278	.0287	.0313	.0219	.0374	.0343	.0232	.0268	.0226	.0611	.0299	.0410	.0274	.0341	.0226	.0389
6	PAV_2	.0206	.0232	.0228	.0241	.0262	.0302	.0316	.0218	.0361	.0349	.0243	.0256	.0226	.0578	.0290	.0398	.0254	.0315	.0224	.0392
7	$PABV_1$	.0198	.0224	.0236	.0246	.0277	.0281	.0315	.0229	.0373	.0340	.0228	.0266	.0237	.0593	.0309	.0408	.0278	.0331	.0232	.0389
8	$PABV_2$	.0201	.0229	.0226	.0231	.0258	.0294	.0311	.0223	.0361	.0350	.0242	.0259	.0234	.0559	.0298	.0400	.0267	.0312	.0228	.0389
9	RK	.0206	.0408	.0247	.0280	.0274	.0236	.0616	.0334	.0342	.0359	.0616	.0424	.0774	.0509	.0447	.0409	.0481	.0328	.0314	.0432
10	$RKNP$	.0208	.0226	.0224	.0244	.0251	.0301	.0320	.0216	.0359	.0344	.0256	.0255	.0232	.0558	.0278	.0394	.0255	.0308	.0222	.0411
11	$TSRV$	.0201	.0387	.0236	.0268	.0259	.0247	.0594	.0345	.0335	.0357	.0618	.0442	.0764	.0493	.0446	.0421	.0470	.0328	.0326	.0441
12	SBV	.0205	.0222	.0225	.0239	.0256	.0288	.0315	.0217	.0354	.0337	.0245	.0251	.0226	.0573	.0287	.0415	.0249	.0317	.0220	.0393
13	RV_5	.0215	.0238	.0226	.0253	.0251	.0298	.0315	.0210	.0360	.0343	.0250	.0254	.0231	.0578	.0292	.0413	.0244	.0305	.0218	.0410
14	SRV_5	.0210	.0223	.0224	.0243	.0259	.0298	.0313	.0215	.0361	.0340	.0247	.0250	.0224	.0581	.0284	.0423	.0246	.0314	.0220	.0401
15	ATM	.0221	.0248	.0226	.0246	.0272	.0254	.0336	.0227	.0379	.0393	.0261	.0288	.0234	.0529	.0276	.0360	.0239	.0303	.0250	.0386

Table 28: MCS p-values, QLIKE loss function, 15 individual forecasts, three horizons, trade data

AA		AXP		BA		CAT		DD		DIS		GE		HD		IBM		JNJ		JPM		KO		MCD		MMM		MRK		PG		T		UTX		WMT		XOM	
one day ahead																																							
1	11	1	1	1	11	1	8	1	1	1	9	1	1	1	1	1	1	1	1	1																			
.00	15	.00	11	.00	11	.00	11	.00	11	.00	1	.00	11	.00	11	.00	11	.00	11	.00	11	.00	11	.00	9	.00	11	.00	9	.00	9	.00	11	.00	11	.00	11	.00	11
.00	7	.00	9	.00	9	.00	9	.00	9	.00	11	.00	9	.00	9	.00	9	.00	9	.00	9	.00	9	.00	11	.00	9	.00	11	.00	11	.00	15	.01	4	.00	9	.00	9
.00	13	.00	4	.00	6	.00	10	.00	4	.00	2	.00	4	.01	10	.00	15	.00	15	.00	15	.00	15	.04	15	.00	7	.00	10	.00	15	.00	9	.01	5	.00	15	.00	15
.00	5	.00	10	.00	10	.00	5	.00	15	.00	4	.00	15	.02	15	.00	4	.00	13	.00	2	.00	4	.04	7	.00	5	.00	8	.00	13	.00	8	.01	9	.00	7	.00	4
.00	9	.00	13	.00	13	.00	6	.00	2	.00	9	.00	14	.02	7	.00	2	.00	10	.00	4	.00	1	.04	1	.00	13	.00	15	.00	10	.01	13	.01	2	.00	14	.00	7
.06	4	.00	2	.00	8	.00	7	.04	5	.00	3	.00	3	.02	4	.00	10	.00	14	.02	12	.00	2	.04	4	.00	8	.00	13	.00	14	.02	4	.01	15	.00	6	.00	13
.15	11	.25	7	.00	14	.00	13	.04	7	.00	6	.00	2	.03	14	.22	1	.00	4	.02	10	.00	10	.04	2	.00	10	.00	7	.00	7	.02	6	.01	7	.00	8	.00	10
.15	10	.52	8	.00	12	.00	15	.31	3	.01	13	.00	6	.03	5	.22	5	.00	6	.18	13	.00	13	.04	8	.00	6	.00	6	.00	5	.02	10	.01	6	.00	5	.00	5
.15	6	.52	5	.00	5	.00	14	.42	10	.01	5	.00	12	.03	2	.22	3	.00	2	.18	14	.03	7	.60	5	.00	14	.00	4	.00	6	.02	14	.01	10	.01	10	.00	8
.29	14	.52	12	.00	7	.02	8	.71	14	.01	8	.00	8	.32	13	.36	7	.08	8	.18	3	.12	3	.60	10	.00	4	.00	14	.00	12	.02	7	.01	1	.01	13	.00	14
.75	8	.52	3	.00	3	.07	4	.85	1	.07	7	.00	13	.32	12	.36	14	.08	12	.50	8	.12	6	.75	3	.01	12	.00	5	.00	2	.02	12	.01	8	.01	4	.00	2
.77	2	.52	6	.00	2	.07	3	.85	6	.07	10	.01	10	.37	6	.36	13	.43	5	.50	6	.12	5	.77	13	.01	2	.01	3	.00	8	.03	5	.02	3	.03	12	.02	6
.77	3	.83	14	.00	4	.08	2	.85	8	.07	14	.06	7	.40	8	.36	6	.57	7	.70	1	.12	8	.77	14	.03	3	.02	12	.00	4	.03	2	.02	14	.13	2	.03	12
.77	12	.83	1	.74	1	.08	12	.85	12	.17	12	.06	5	.69	3	.36	8	.57	3	.70	7	.33	14	.77	6	.44	15	.02	2	.00	3	.03	3	.83	13	.13	3	.10	3
1	1	1	15	1	15	1	1	1	13	1	15	1	1	1	1	1	12	1	1	1	5	1	12	1	12	1	1	1	1	1	1	1	1	1	12	1	1	1	1
one week ahead																																							
.01	15	.00	11	.00	9	.00	9	.00	9	.00	4	.00	9	.00	11	.00	11	.00	11	.00	11	.00	11	.00	9	.00	7	.00	9	.00	9	.00	11	.00	4	.00	11	.00	9
.06	4	.00	9	.00	11	.00	11	.00	4	.00	2	.00	11	.00	9	.00	15	.00	9	.00	9	.00	9	.00	11	.00	5	.00	11	.00	11	.00	9	.00	5	.00	9	.00	11
.06	5	.00	13	.00	6	.00	5	.00	5	.00	6	.00	4	.00	7	.01	9	.00	6	.00	13	.01	4	.07	7	.00	11	.00	8	.00	13	.00	7	.00	2	.00	5	.00	15
.15	13	.00	10	.07	8	.00	10	.00	7	.00	1	.00	15	.00	8	.01	4	.00	15	.00	14	.01	2	.08	15	.00	9	.00	7	.00	7	.00	5	.00	11	.00	7	.00	6
.15	3	.00	8	.07	5	.00	13	.00	11	.00	13	.00	13	.00	5	.01	2	.00	10	.00	12	.01	15	.08	8	.00	13	.00	6	.00	5	.00	8	.00	7	.00	8	.00	13
.20	2	.00	6	.07	14	.00	7	.00	2	.00	8	.00	14	.06	10	.01	10	.00	14	.00	10	.01	10	.08	5	.00	8	.00	5	.02	14	.00	15	.05	3	.00	6	.00	8
.20	7	.00	4	.07	7	.00	6	.14	15	.00	11	.00	10	.06	6	.05	7	.00	13	.00	8	.01	3	.08	4	.00	6	.00	10	.10	15	.00	13	.10	1	.00	4	.00	14
.20	9	.00	12	.07	12	.00	14	.14	14	.02	7	.00	12	.06	12	.05	5	.00	12	.00	2	.01	7	.59	10	.00	12	.00	13	.10	4	.00	4	.39	9	.00	14	.00	5
.44	10	.00	5	.07	13	.00	12	.14	3	.02	5	.00	2	.10	4	.12	3	.00	8	.01	15	.03	1	.59	13	.00	14	.00	3	.10	2	.00	6	.39	8	.00	10	.00	7
.60	14	.00	2	.07	3	.00	15	.14	12	.07	10	.00	8	.10	3	.12	8	.00	4	.01	4	.03	5	.59	6	.00	10	.00	12	.12	12	.00	10	.39	6	.00	2	.00	12
.60	11	.01	7	.07	2	.00	3	.24	6	.07	14	.00	6	.10	14	.77	1	.00	5	.01	3	.03	8	.59	12	.00	4	.00	4	.48	3	.00	2	.55	14	.00	3	.00	4
.60	6	.01	14	.07	10	.00	2	.60	10	.16	3	.00	7	.19	2	.77	14	.02	2	.01	6	.14	13	.59	3	.00	2	.00	14	.62	6	.00	12	.55	12	.09	15	.00	10
.76	12	.01	3	.07	4	.00	4	.60	8	.16	12	.00	3	.28	13	.77	6	.02	7	.01	5	.14	6	.59	2	.00	3	.00	2	.67	8	.00	14	.55	15	.09	13	.00	2
.86	1	.55	15	.33	1	.01	8	.80	1	.16	9	.00	5	.51	1	.85	12	.06	3	.01	7	.83	14	.59	14	.82	15	.30	15	.67	10	.00	3	.55	10	.09	12	.00	3
1	8	1	1	1	15	1	1	1	13	1	15	1	1	1	15	1	13	1	1	1	1	1	12	1	1	1	1	1	1	1	1	1	1	1	13	1	1	1	1
one month ahead																																							
.11	13	.01	9	.00	7	.01	9	.00	5	.01	6	.01	9	.01	11	.05	5	.00	15	.00	11	.00	9	.00	9	.00	7	.00	11	.03	14	.00	11	.03	5	.00	11	.00	13
.11	3	.01	11	.00	5	.01	13	.00	7	.01	10	.01	11	.01	9	.05	7	.00	8	.00	9	.00	11	.00	11	.00	5	.00	9	.03	5	.00	9	.03	7	.00	9	.00	10
.11	2	.02	13	.05	4	.02	5	.00	2	.01	8	.01	4	.01	8	.10	4	.00	6	.01	10	.00	5	.03	8	.00	14	.01	7	.06	13	.00	7	.03	4	.00	8	.01	11
.11	14	.02	6	.05	9	.02	11	.00	3	.01	5	.01	10	.01	7	.10	15	.03	9	.01	14	.02	15	.03	10	.00	6	.01	8	.06	12	.00	5	.29	2	.00	15	.01	14
.37	4	.02	15	.05	2	.02	4	.00	9	.01	13	.01	15	.01	15	.10	2	.03	11	.01	13	.02	3	.03	12	.00	13	.01	6	.06	7	.00	8	.59	3	.00	7	.01	9
.37	15	.02	8	.05	3	.02	7	.01	4	.01	12	.01	7	.01	3	.10	1	.03	5	.01	4	.02	2	.03	13	.00	12	.02	5	.06	11	.00	6	.69	1	.00	6	.01	6
.37	5	.02	5	.05	11	.04	3	.01	6	.01	14	.07	12	.01	6	.10	6	.03	10	.02	2	.02	7	.03	15	.00	10	.02	13	.06	8	.00	10	.69	9	.00	5	.01	12
.37	10	.02	4	.05	6	.04	10	.11	15	.01	4	.07	6	.01	5	.10	3	.03	3	.02	15	.06	1	.03	6	.00	3	.02	4	.06	6	.06	4	.69	11	.00	12	.01	5
.53	1	.02	10	.66	1	.04	2	.11	14	.01	7	.07	13	.01	12	.10	14	.03	2	.02	6	.06	4	.03	7	.01	8	.27	3	.06	3	.06	12	.69	12	.00	3	.01	8
.53	6	.02	3	.66	8	.04	14	.38	8	.01	2	.07	2	.01	4	.10	13	.03	13	.02	8	.06	8	.03	14	.01	2	.27	12	.08	9	.06	14	.69	6	.00	4	.01	7
.54	12	.02	12	.74	13	.04	6	.38	11	.32	3	.10	5	.01	10	.10	8	.03	4	.02	3	.06	6	.03	3	.01	4	.30	14	.08	10	.06	2	.69	14	.03	14	.01	15
.54	9	.02	2	.75	15	.04	15	.38	12	.36	1	.10	14	.01	2	.10	10	.61	7	.35	12	.37	13	.24	5	.14	15	.33	2	.08	4	.18	3	.69	8	.03	13	.01	3
.77	11	.39	14	.75	12	.04	12	.71	1	.36	15	.13	3	.23	14	.10	12	.61	14	.52	1	.37	10	.33	2	.14	1	.49	10	.08	2	.63	13	.69	10	.03	10	.01	4
.77	8	.41	7	.87	10	.47	1	.86	10	.36	11	.13	8	.23	13	.10	9	.61	12	.52	5	.63	12	.33	4	.14	9	.66	15	.55	1	.72	15	.84	13	.03	2	.32	2
1	7	1	1	1	14	1	8	1	13	1	9	1	1	1	1	1	11	1	1	1	7	1	14	1	1	1	11	1	1	1	15	1	1	1	15	1	1	1	1

Notes: MCS p-values for the QLIKE loss function. In each stock sub-panel we report the corresponding p-values on the left and on the right we list the estimator numbers ranked in the ascending orders of the p-values. Estimators 1 to 15 are: PDV, ANP_1, ANP_2, NP, PAV_1, PAV_2, $PABV_1$, $PABV_2$, RK, RKNP, TSRV, SBV, RV_5, SRV_5 and ATM.

Table 29: RMSE values, 14 combination volatility estimators, three horizons, trade data

		AA	AXP	BA	CAT	DD	DIS	GE	HD	IBM	JNJ	JPM	KO	MCD	MMM	MRK	PG	T	UTX	WMT	XOM
											one day ahead										
1	PDV	.0048	.0031	.0028	.0039	.0030	.0029	.0049	.0021	.0021	.0013	.0041	.0013	.0012	.0067	.0023	.0021	.0014	.0028	.0010	.0026
2	SBV	.0049	.0031	.0029	.0041	.0030	.0026	.0052	.0021	.0021	.0013	.0041	.0012	.0012	.0074	.0024	.0022	.0015	.0027	.0011	.0028
3	ATM	.0060	.0031	.0028	.0042	.0032	.0024	.0054	.0021	.0026	.0016	.0049	.0014	.0013	.0068	.0024	.0023	.0017	.0028	.0013	.0027
4	$PDV + SBV$	.0047	.0030	.0028	.0039	.0029	.0027	.0050	.0021	.0020	.0013	.0040	.0012	.0012	.0069	.0023	.0021	.0014	.0027	.0010	.0026
5	$PDV + ATM$	.0046	.0027	.0025	.0036	.0027	.0022	.0047	.0018	.0021	.0013	.0038	.0011	.0011	.0065	.0021	.0020	.0014	.0024	.0010	.0024
6	$SBV + ATM$	.0046	.0028	.0026	.0037	.0027	.0022	.0048	.0018	.0020	.0013	.0039	.0011	.0011	.0068	.0022	.0020	.0014	.0025	.0010	.0025
7	$Duration4$	.0048	.0031	.0028	.0040	.0031	.0028	.0051	.0021	.0021	.0013	.0041	.0012	.0012	.0068	.0023	.0021	.0014	.0028	.0010	.0027
8	$SBV + Duration4$	.0048	.0031	.0028	.0040	.0030	.0027	.0051	.0021	.0021	.0013	.0041	.0012	.0012	.0068	.0023	.0021	.0014	.0027	.0010	.0027
9	$ATM + Duration4$	.0046	.0030	.0027	.0039	.0029	.0026	.0049	.0020	.0020	.0013	.0039	.0012	.0012	.0067	.0022	.0020	.0014	.0026	.0010	.0026
10	$RV10$	.0049	.0030	.0028	.0041	.0029	.0026	.0051	.0021	.0021	.0013	.0041	.0012	.0013	.0070	.0023	.0021	.0015	.0027	.0011	.0028
11	$PDV + RV10$	.0049	.0030	.0028	.0041	.0029	.0026	.0050	.0021	.0020	.0013	.0041	.0012	.0013	.0069	.0023	.0021	.0014	.0027	.0010	.0028
12	$ATM + RV10$	.0048	.0029	.0028	.0041	.0029	.0025	.0050	.0020	.0020	.0013	.0040	.0012	.0013	.0069	.0023	.0021	.0014	.0026	.0010	.0028
13	$Duration4 + RV10$	.0048	.0030	.0028	.0041	.0029	.0026	.0050	.0021	.0020	.0013	.0040	.0012	.0012	.0069	.0023	.0021	.0014	.0027	.0010	.0027
14	All	.0048	.0029	.0028	.0040	.0029	.0025	.0050	.0020	.0020	.0013	.0040	.0011	.0012	.0068	.0023	.0021	.0014	.0026	.0010	.0027
											one week ahead										
1	PDV	.0188	.0121	.0105	.0148	.0117	.0109	.0201	.0075	.0090	.0051	.0152	.0051	.0041	.0285	.0089	.0083	.0051	.0113	.0037	.0110
2	SBV	.0192	.0133	.0112	.0161	.0121	.0108	.0227	.0083	.0095	.0058	.0173	.0051	.0044	.0334	.0098	.0095	.0057	.0116	.0042	.0125
3	ATM	.0228	.0141	.0104	.0167	.0134	.0093	.0225	.0068	.0110	.0063	.0197	.0054	.0046	.0291	.0092	.0088	.0063	.0116	.0043	.0134
4	$PDV + SBV$	.0186	.0126	.0107	.0154	.0117	.0103	.0211	.0078	.0090	.0054	.0162	.0050	.0041	.0297	.0093	.0087	.0053	.0112	.0039	.0116
5	$PDV + ATM$	.0186	.0121	.0095	.0147	.0110	.0084	.0198	.0060	.0090	.0053	.0154	.0044	.0040	.0279	.0082	.0079	.0052	.0103	.0033	.0115
6	$SBV + ATM$	.0185	.0130	.0103	.0155	.0115	.0092	.0208	.0069	.0090	.0056	.0167	.0048	.0041	.0297	.0088	.0083	.0054	.0109	.0037	.0119
7	$Duration4$	.0190	.0128	.0108	.0155	.0123	.0110	.0211	.0077	.0092	.0054	.0160	.0052	.0042	.0289	.0093	.0085	.0054	.0117	.0040	.0115
8	$SBV + Duration4$	.0189	.0128	.0108	.0156	.0122	.0108	.0213	.0078	.0092	.0054	.0162	.0051	.0042	.0291	.0094	.0086	.0055	.0115	.0040	.0116
9	$ATM + Duration4$	.0185	.0125	.0104	.0153	.0117	.0100	.0206	.0071	.0091	.0054	.0154	.0049	.0041	.0285	.0089	.0083	.0054	.0112	.0038	.0114
10	$RV10$	.0190	.0132	.0107	.0161	.0120	.0108	.0224	.0084	.0095	.0059	.0174	.0053	.0054	.0306	.0099	.0089	.0062	.0115	.0043	.0128
11	$PDV + RV10$	.0189	.0131	.0107	.0159	.0119	.0106	.0221	.0082	.0093	.0058	.0172	.0052	.0051	.0302	.0098	.0088	.0060	.0114	.0042	.0125
12	$ATM + RV10$	.0187	.0131	.0106	.0158	.0118	.0105	.0220	.0081	.0093	.0058	.0171	.0052	.0051	.0302	.0097	.0087	.0060	.0113	.0042	.0126
13	$Duration4 + RV10$	.0189	.0130	.0106	.0158	.0120	.0105	.0219	.0080	.0092	.0056	.0169	.0051	.0047	.0296	.0097	.0087	.0058	.0114	.0041	.0122
14	All	.0187	.0129	.0105	.0156	.0118	.0103	.0216	.0079	.0091	.0056	.0166	.0051	.0046	.0294	.0095	.0086	.0058	.0113	.0040	.0121
											one month ahead										
1	PDV	.0841	.0547	.0486	.0713	.0520	.0472	.0861	.0356	.0428	.0259	.0696	.0231	.0172	.1144	.0434	.0358	.0254	.0521	.0180	.0557
2	SBV	.0815	.0590	.0492	.0726	.0535	.0502	.0943	.0390	.0439	.0273	.0740	.0230	.0202	.1271	.0461	.0412	.0279	.0521	.0198	.0603
3	ATM	.0916	.0693	.0507	.0799	.0603	.0473	.0945	.0408	.0459	.0299	.0931	.0253	.0201	.1154	.0418	.0347	.0263	.0542	.0219	.0617
4	$PDV + SBV$	.0812	.0561	.0483	.0718	.0520	.0469	.0887	.0370	.0424	.0264	.0713	.0225	.0180	.1169	.0445	.0378	.0263	.0511	.0187	.0575
5	$PDV + ATM$	.0851	.0586	.0473	.0722	.0517	.0421	.0878	.0355	.0420	.0272	.0735	.0219	.0174	.1130	.0407	.0337	.0247	.0508	.0184	.0577
6	$SBV + ATM$	.0803	.0616	.0485	.0732	.0537	.0469	.0894	.0387	.0418	.0275	.0765	.0232	.0189	.1162	.0420	.0356	.0256	.0516	.0197	.0587
7	$Duration4$	.0823	.0562	.0491	.0733	.0539	.0481	.0887	.0369	.0433	.0265	.0714	.0231	.0178	.1154	.0446	.0367	.0263	.0530	.0188	.0565
8	$SBV + Duration4$	.0818	.0564	.0489	.0731	.0535	.0477	.0891	.0372	.0431	.0265	.0716	.0229	.0181	.1158	.0447	.0372	.0265	.0525	.0189	.0570
9	$ATM + Duration4$	.0816	.0569	.0483	.0729	.0528	.0460	.0881	.0367	.0427	.0268	.0708	.0226	.0177	.1146	.0432	.0357	.0258	.0521	.0188	.0567
10	$RV10$	.0808	.0615	.0467	.0713	.0536	.0496	.0956	.0408	.0437	.0281	.0772	.0247	.0263	.1175	.0476	.0386	.0322	.0520	.0217	.0606
11	$PDV + RV10$	.0806	.0606	.0466	.0711	.0533	.0486	.0943	.0400	.0433	.0277	.0763	.0243	.0246	.1165	.0472	.0382	.0313	.0517	.0212	.0599
12	$ATM + RV10$	.0799	.0616	.0466	.0710	.0533	.0489	.0942	.0406	.0431	.0279	.0764	.0245	.0250	.1164	.0467	.0379	.0312	.0516	.0214	.0602
13	$Duration4 + RV10$	.0810	.0595	.0469	.0714	.0533	.0478	.0929	.0390	.0431	.0273	.0750	.0238	.0223	.1155	.0466	.0378	.0299	.0519	.0206	.0590
14	All	.0804	.0596	.0468	.0712	.0531	.0474	.0922	.0390	.0428	.0272	.0744	.0237	.0218	.1150	.0460	.0374	.0294	.0516	.0205	.0589

Table 30: MCS p-values, RMSE loss function, 14 combination forecasts, three horizons, trade data

AA		AXP		BA		CAT		DD		DIS		GE		HD		IBM		JNJ		JPM		KO		MCD		MMM		MRK		PG		T		UTX		WMT		XOM	
one day ahead																																							
.00	3	.00	2	.00	2	.00	2	.00	7	.00	1	.00	3	.00	10	.00	3	.02	2	.00	3	.00	3	.00	10	.01	11	.00	3	.00	3	.00	3	.00	7	.00	3	.00	2
.00	10	.00	7	.00	8	.00	3	.00	3	.00	7	.00	13	.00	11	.00	1	.02	3	.00	2	.00	1	.00	3	.01	10	.00	2	.01	10	.00	10	.00	8	.00	2	.00	13
.00	2	.00	8	.00	7	.00	13	.01	13	.00	8	.00	2	.00	2	.00	7	.02	10	.00	13	.00	7	.00	11	.01	12	.00	10	.01	11	.00	2	.00	1	.00	10	.00	10
.00	11	.00	1	.00	4	.00	10	.01	2	.00	10	.00	11	.00	13	.00	2	.02	11	.00	10	.00	10	.00	1	.01	13	.00	7	.01	2	.00	11	.00	2	.00	1	.00	11
.00	7	.00	10	.00	10	.00	7	.01	8	.00	11	.00	7	.00	7	.00	8	.02	12	.00	11	.00	13	.00	12	.01	2	.00	8	.01	13	.00	8	.00	3	.00	11	.00	3
.00	13	.00	13	.00	13	.00	11	.01	11	.00	13	.00	10	.00	1	.00	10	.02	4	.02	7	.00	11	.00	2	.01	7	.00	11	.03	12	.00	6	.01	4	.00	4	.01	14
.00	1	.00	11	.00	11	.00	8	.01	10	.00	2	.01	8	.00	3	.00	5	.02	6	.02	8	.00	4	.00	7	.01	8	.00	4	.04	4	.00	7	.01	10	.01	13	.01	12
.06	12	.00	4	.00	1	.01	4	.02	4	.00	4	.01	14	.00	12	.00	11	.02	8	.02	4	.00	8	.00	13	.01	14	.00	13	.04	14	.00	12	.01	13	.01	8	.01	7
.06	8	.00	3	.00	12	.01	1	.02	14	.00	12	.01	12	.00	14	.00	4	.02	7	.05	14	.00	14	.00	8	.01	4	.01	12	.04	8	.00	4	.01	11	.02	12	.01	8
.07	14	.00	12	.00	3	.01	14	.02	1	.00	14	.01	4	.00	8	.00	9	.02	13	.05	1	.00	12	.00	4	.01	3	.01	14	.04	7	.00	13	.01	9	.02	7	.07	9
.07	4	.00	14	.00	14	.01	12	.02	12	.00	9	.01	9	.01	4	.00	13	.34	14	.06	12	.00	2	.02	9	.02	1	.01	1	.05	6	.78	14	.01	14	.06	6	.07	4
.85	6	.00	9	.00	9	.01	9	.02	9	.01	3	.02	1	.01	9	.20	6	.73	1	.45	9	.01	9	.05	14	.03	9	.02	9	.14	1	.78	9	.01	12	.14	14	.07	1
.85	5	.00	6	.00	6	.03	6	.15	6	.67	5	.02	6	.02	6	.20	12	.73	9	.45	6	.23	6	.24	5	.03	6	.02	6	.20	9	.78	5	.01	6	.51	9	.07	6
1	9	1	5	1	5	1	5	1	5	1	6	1	5	1	5	1	14	1	5	1	5	1	5	1	6	1	5	1	5	1	5	1	1	1	5	1	5	1	5
one week ahead																																							
.00	3	.00	10	.00	2	.01	2	.00	3	.00	2	.00	3	.00	10	.00	3	.00	2	.00	3	.00	10	.00	10	.02	10	.00	2	.00	10	.00	10	.00	7	.00	10	.00	3
.01	7	.00	11	.00	8	.01	3	.00	2	.00	1	.00	13	.00	12	.00	10	.00	10	.00	10	.00	11	.00	12	.02	11	.00	10	.00	11	.00	13	.00	8	.00	11	.04	2
.01	2	.00	2	.00	7	.01	10	.01	13	.00	7	.00	14	.00	11	.00	11	.00	3	.00	11	.00	12	.00	11	.02	2	.00	11	.00	2	.00	11	.00	2	.00	12	.04	12
.01	8	.00	12	.00	13	.01	13	.01	7	.00	10	.00	10	.00	13	.00	12	.00	13	.00	12	.00	7	.00	13	.02	13	.00	12	.00	12	.00	12	.00	13	.00	13	.04	14
.01	10	.00	13	.00	10	.01	11	.01	10	.00	13	.00	2	.00	14	.00	2	.00	14	.00	2	.00	13	.00	2	.02	12	.00	13	.00	13	.00	14	.00	10	.00	14	.04	10
.01	13	.00	14	.00	11	.01	8	.01	8	.00	8	.00	11	.00	2	.00	13	.01	12	.00	13	.00	3	.00	3	.02	6	.00	8	.00	4	.00	2	.00	1	.00	2	.04	13
.02	11	.00	3	.00	4	.01	12	.01	14	.00	11	.00	8	.00	8	.03	8	.01	11	.00	14	.00	2	.00	14	.02	14	.00	14	.00	14	.00	3	.00	11	.00	3	.04	11
.80	14	.00	8	.00	14	.01	7	.01	11	.00	12	.00	12	.00	7	.03	7	.01	6	.01	4	.00	8	.00	8	.02	8	.00	7	.00	3	.00	8	.00	14	.00	7	.04	6
.82	1	.00	7	.00	12	.01	14	.01	12	.00	4	.00	7	.00	4	.46	14	.01	8	.01	8	.00	14	.00	7	.02	4	.02	4	.02	8	.00	7	.00	3	.00	8	.04	8
.82	12	.03	6	.00	6	.01	9	.01	9	.00	14	.01	4	.00	1	.75	9	.01	4	.01	6	.00	1	.00	9	.02	7	.02	3	.02	7	.01	9	.00	9	.01	4	.04	9
.88	5	.05	4	.00	9	.01	6	.01	4	.00	9	.01	6	.00	9	.79	1	.09	7	.01	7	.00	4	.02	1	.05	3	.02	9	.11	1	.01	6	.00	12	.01	9	.04	4
.88	4	.05	9	.00	1	.01	4	.01	1	.00	6	.01	9	.00	6	.84	6	.12	9	.40	9	.00	9	.09	4	.05	9	.02	6	.11	6	.01	4	.00	4	.01	1	.04	5
.88	9	.75	1	.03	3	.69	1	.01	6	.22	3	.31	1	.06	3	.84	5	.16	5	.61	5	.00	6	.09	6	.13	1	.06	1	.20	9	.43	5	.00	6	.01	6	.04	7
1	6	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	4	1	1	1	1	1	5	1	5	1	5	1	5	1	5	1	1	1	5	1	5	1	1
one month ahead																																							
.01	3	.01	10	.02	2	.01	3	.03	3	.00	2	.11	10	.00	12	.04	3	.02	3	.00	3	.00	12	.00	10	.00	2	.07	10	.01	12	.00	13	.09	7	.00	12	.04	2
.01	7	.01	12	.09	3	.01	9	.03	2	.00	10	.11	12	.00	10	.21	2	.02	12	.00	12	.00	10	.00	12	.00	10	.08	12	.01	10	.00	10	.17	3	.00	10	.04	12
.01	1	.01	3	.09	8	.01	7	.03	10	.00	12	.11	2	.00	11	.21	10	.02	10	.00	10	.00	11	.00	11	.00	6	.08	11	.01	2	.00	11	.17	8	.00	11	.04	3
.01	8	.01	11	.09	7	.01	8	.03	7	.00	11	.11	13	.00	14	.23	13	.02	14	.00	11	.03	14	.00	2	.00	12	.08	2	.01	11	.00	14	.17	10	.00	14	.04	10
.01	5	.01	13	.09	9	.55	6	.03	13	.00	13	.11	11	.00	13	.23	7	.03	6	.00	2	.03	13	.02	13	.00	11	.17	13	.01	14	.00	12	.17	9	.00	13	.04	14
.01	2	.01	6	.09	6	.55	2	.07	8	.00	7	.11	14	.00	6	.23	8	.03	11	.00	14	.14	3	.02	14	.00	4	.19	14	.01	13	.00	2	.17	13	.00	3	.04	11
.01	4	.01	2	.60	4	.70	5	.07	12	.00	8	.11	3	.00	2	.35	11	.03	13	.01	13	.14	7	.02	3	.00	8	.22	8	.01	4	.00	8	.28	1	.00	2	.04	6
.01	9	.02	14	.69	1	.70	13	.07	14	.00	14	.11	8	.00	3	.35	12	.03	5	.08	6	.14	1	.05	6	.09	13	.22	7	.01	8	.00	4	.28	2	.00	6	.04	13
.01	13	.15	5	.69	5	.70	4	.24	11	.00	6	.11	6	.00	8	.35	9	.03	2	.08	5	.14	2	.05	8	.09	7	.22	4	.01	7	.00	7	.28	14	.00	9	.04	4
.16	10	.15	8	.69	13	.70	10	.24	6	.04	4	.11	4	.00	9	.55	14	.03	8	.08	8	.14	6	.11	4	.49	14	.22	1	.01	6	.00	3	.28	11	.00	8	.04	5
.16	11	.15	9	.69	14	.83	14	.60	9	.05	1	.13	7	.00	7	.66	1	.05	9	.08	4	.44	8	.11	9	.49	3	.22	9	.13	9	.00	9	.65	12	.00	7	.04	8
.18	14	.15	4	.69	10	.83	1	.84	4	.05	3	.13	9	.02	4	.69	4	.05	7	.08	7	.53	9	.11	7	.49	9	.22	6	.21	1	.00	6	.65	6	.33	4	.04	9
.79	6	.15	7	.85	11	.83	11	.84	1	.05	9	.26	5	.89	1	.75	5	.13	4	.30	9	.53	4	.55	5	.49	1	.40	3	.40	3	.27	1	.73	4	.48	5	.04	7
1	12	1	1	1	12	1	12	1	5	1	5	1	1	1	5	1	6	1	1	1	1	1	5	1	1	1	5	1	5	1	5	1	5	1	5	1	1	1	1

Notes: MCS p-values for the RMSE loss function. In each stock sub-panel we report the corresponding p-values on the left and on the right we list the estimator combination numbers ranked in the ascending orders of the p-values. Combinations 1 to 14 are: PDV; SBV; ATM; PDV+SBV; PDV+ATM; SBV+ATM; four duration-based estimators, estimators 1 to 4, called Duration4; SBV+Duration4; ATM+Duration4; ten RV-type estimators, estimators 5 to 14, called RV10; PDV+RV10; ATM+RV10; Duration4+RV10; All.

Table 31: QLIKE values, 14 combination volatility estimators, three horizons, trade data

		AA	AXP	BA	CAT	DD	DIS	GE	HD	IBM	JNJ	JPM	KO	MCD	MMM	MRK	PG	T	UTX	WMT	XOM
											one day ahead										
1	PDV	.0275	.0278	.0288	.0279	.0293	.0345	.0317	.0257	.0326	.0295	.0306	.0321	.0323	.0367	.0269	.0304	.0267	.0336	.0273	.0283
2	SBV	.0277	.0281	.0304	.0286	.0289	.0300	.0336	.0262	.0309	.0308	.0311	.0294	.0311	.0415	.0282	.0321	.0281	.0327	.0284	.0294
3	ATM	.0337	.0277	.0286	.0307	.0334	.0276	.0371	.0278	.0417	.0358	.0360	.0360	.0337	.0380	.0305	.0369	.0335	.0343	.0333	.0325
4	$PDV+SBV$	.0268	.0274	.0292	.0279	.0283	.0305	.0322	.0254	.0305	.0295	.0303	.0294	.0307	.0384	.0272	.0305	.0269	.0322	.0274	.0283
5	$PDV+ATM$	.0268	.0246	.0253	.0257	.0264	.0252	.0300	.0228	.0320	.0294	.0290	.0265	.0299	.0335	.0256	.0288	.0264	.0287	.0262	.0271
6	$SBV+ATM$	.0263	.0252	.0273	.0263	.0269	.0254	.0311	.0234	.0301	.0295	.0291	.0277	.0294	.0366	.0262	.0297	.0270	.0299	.0271	.0276
7	$Duration4$	.0272	.0283	.0293	.0284	.0300	.0334	.0332	.0257	.0321	.0299	.0310	.0314	.0316	.0381	.0276	.0308	.0274	.0337	.0275	.0290
8	$SBV+Duration4$	.0270	.0279	.0293	.0282	.0294	.0320	.0331	.0256	.0314	.0297	.0307	.0305	.0312	.0383	.0276	.0307	.0273	.0331	.0275	.0288
9	$ATM+Duration4$	.0262	.0267	.0277	.0272	.0281	.0295	.0315	.0241	.0310	.0294	.0294	.0288	.0307	.0364	.0266	.0295	.0267	.0314	.0266	.0282
10	$RV10$	.0277	.0277	.0295	.0290	.0284	.0296	.0337	.0259	.0306	.0298	.0316	.0302	.0343	.0387	.0282	.0309	.0281	.0319	.0281	.0298
11	$PDV+RV10$	.0275	.0275	.0292	.0287	.0283	.0295	.0333	.0256	.0304	.0295	.0314	.0298	.0330	.0383	.0279	.0307	.0276	.0318	.0278	.0295
12	$ATM+RV10$	.0271	.0271	.0290	.0284	.0279	.0286	.0328	.0253	.0299	.0294	.0307	.0297	.0330	.0380	.0276	.0304	.0276	.0312	.0277	.0294
13	$Duration4+RV10$	.0274	.0275	.0289	.0285	.0286	.0298	.0333	.0253	.0304	.0293	.0311	.0296	.0315	.0380	.0277	.0305	.0272	.0321	.0275	.0293
14	All	.0270	.0271	.0285	.0281	.0281	.0290	.0326	.0249	.0300	.0291	.0305	.0291	.0309	.0375	.0273	.0301	.0270	.0315	.0272	.0290
											one week ahead										
1	PDV	.0199	.0194	.0197	.0192	.0223	.0257	.0240	.0173	.0286	.0235	.0217	.0248	.0206	.0379	.0199	.0276	.0194	.0273	.0186	.0234
2	SBV	.0200	.0214	.0207	.0208	.0228	.0238	.0271	.0189	.0279	.0256	.0243	.0235	.0211	.0426	.0221	.0292	.0211	.0260	.0205	.0255
3	ATM	.0239	.0202	.0183	.0218	.0257	.0187	.0302	.0163	.0354	.0285	.0247	.0261	.0227	.0382	.0215	.0310	.0238	.0258	.0209	.0298
4	$PDV+SBV$	.0193	.0201	.0199	.0198	.0219	.0234	.0252	.0178	.0273	.0241	.0229	.0232	.0200	.0398	.0208	.0277	.0199	.0259	.0192	.0240
5	$PDV+ATM$	.0194	.0177	.0166	.0176	.0198	.0172	.0236	.0133	.0278	.0231	.0196	.0191	.0192	.0350	.0178	.0251	.0187	.0223	.0162	.0237
6	$SBV+ATM$	.0190	.0194	.0183	.0190	.0209	.0187	.0252	.0153	.0269	.0241	.0215	.0215	.0196	.0382	.0194	.0258	.0196	.0234	.0181	.0245
7	$Duration4$	.0200	.0206	.0203	.0201	.0237	.0258	.0263	.0180	.0288	.0245	.0230	.0249	.0208	.0389	.0212	.0284	.0207	.0280	.0199	.0243
8	$SBV+Duration4$	.0199	.0205	.0202	.0201	.0232	.0248	.0263	.0180	.0283	.0245	.0231	.0242	.0205	.0393	.0213	.0282	.0206	.0273	.0198	.0243
9	$ATM+Duration4$	.0193	.0195	.0190	.0191	.0219	.0222	.0251	.0162	.0277	.0240	.0213	.0225	.0202	.0374	.0199	.0267	.0201	.0258	.0186	.0239
10	$RV10$	.0199	.0210	.0198	.0206	.0221	.0230	.0276	.0190	.0273	.0248	.0246	.0243	.0254	.0392	.0226	.0280	.0221	.0255	.0205	.0254
11	$PDV+RV10$	.0197	.0207	.0195	.0204	.0220	.0228	.0271	.0186	.0271	.0244	.0243	.0238	.0240	.0389	.0223	.0278	.0216	.0254	.0201	.0251
12	$ATM+RV10$	.0195	.0205	.0194	.0202	.0215	.0220	.0269	.0183	.0266	.0244	.0237	.0237	.0241	.0386	.0219	.0274	.0215	.0248	.0199	.0251
13	$Duration4+RV10$	.0198	.0206	.0195	.0202	.0223	.0230	.0270	.0182	.0272	.0243	.0240	.0236	.0222	.0387	.0220	.0278	.0211	.0259	.0200	.0248
14	All	.0195	.0203	.0192	.0199	.0218	.0222	.0265	.0177	.0268	.0240	.0233	.0232	.0216	.0383	.0215	.0273	.0208	.0253	.0196	.0246
											one month ahead										
1	PDV	.0221	.0224	.0243	.0250	.0272	.0306	.0315	.0211	.0402	.0349	.0248	.0290	.0214	.0575	.0286	.0399	.0247	.0354	.0216	.0386
2	SBV	.0216	.0235	.0238	.0253	.0271	.0304	.0334	.0229	.0374	.0356	.0259	.0266	.0239	.0607	.0303	.0439	.0263	.0335	.0232	.0416
3	ATM	.0232	.0261	.0237	.0260	.0285	.0264	.0355	.0239	.0398	.0404	.0275	.0303	.0244	.0558	.0285	.0380	.0252	.0319	.0260	.0406
4	$PDV+SBV$	.0214	.0227	.0237	.0251	.0267	.0295	.0321	.0218	.0381	.0349	.0252	.0270	.0219	.0584	.0292	.0416	.0253	.0339	.0222	.0400
5	$PDV+ATM$	.0219	.0229	.0225	.0237	.0253	.0249	.0316	.0203	.0379	.0367	.0238	.0254	.0209	.0552	.0268	.0362	.0235	.0313	.0222	.0385
6	$SBV+ATM$	.0208	.0239	.0229	.0243	.0262	.0272	.0323	.0225	.0362	.0363	.0246	.0265	.0224	.0567	.0279	.0384	.0243	.0311	.0236	.0398
7	$Duration4$	.0219	.0233	.0246	.0257	.0283	.0312	.0332	.0222	.0396	.0357	.0258	.0286	.0219	.0586	.0296	.0414	.0259	.0360	.0225	.0395
8	$SBV+Duration4$	.0217	.0232	.0243	.0256	.0278	.0306	.0331	.0223	.0390	.0356	.0257	.0279	.0221	.0586	.0297	.0417	.0259	.0353	.0226	.0398
9	$ATM+Duration4$	.0214	.0232	.0238	.0250	.0271	.0289	.0325	.0218	.0387	.0363	.0248	.0271	.0216	.0576	.0287	.0396	.0252	.0342	.0226	.0391
10	$RV10$	.0215	.0248	.0226	.0243	.0270	.0289	.0353	.0230	.0366	.0352	.0276	.0290	.0286	.0558	.0317	.0420	.0291	.0331	.0239	.0416
11	$PDV+RV10$	.0214	.0245	.0225	.0243	.0269	.0287	.0348	.0226	.0367	.0350	.0273	.0286	.0270	.0557	.0313	.0417	.0285	.0332	.0236	.0413
12	$ATM+RV10$	.0211	.0247	.0225	.0241	.0267	.0283	.0347	.0229	.0363	.0352	.0269	.0287	.0274	.0555	.0310	.0411	.0284	.0326	.0238	.0413
13	$Duration4+RV10$	.0215	.0242	.0228	.0245	.0272	.0289	.0345	.0224	.0372	.0350	.0270	.0282	.0251	.0562	.0310	.0416	.0277	.0337	.0233	.0408
14	All	.0213	.0241	.0227	.0243	.0269	.0285	.0341	.0223	.0369	.0351	.0265	.0280	.0245	.0560	.0305	.0410	.0273	.0333	.0233	.0406

Table 32: MCS p-values, QLIKE loss function, 14 combination forecasts, three horizons, trade data

AA		AXP		BA		CAT		DD		DIS		GE		HD		IBM		JNJ		JPM		KO		MCD		MMM		MRK		PG		T		UTX		WMT		XOM	
																				one day ahead																			
.00	3	.00	7	.00	2	.00	3	.00	3	.00	1	.00	3	.00	10	.00	3	.00	3	.00	3	.00	3	.00	10	.00	10	.00	3	.00	3	.00	3	.00	7	.00	3	.00	3
.00	10	.00	1	.00	10	.00	13	.00	8	.00	7	.00	2	.00	3	.00	7	.00	2	.00	10	.00	1	.00	3	.00	11	.00	10	.00	2	.00	2	.00	1	.00	10	.00	10
.00	2	.00	2	.00	11	.00	2	.00	7	.00	13	.00	13	.00	11	.00	1	.00	8	.00	13	.00	7	.00	11	.00	12	.00	2	.00	10	.00	10	.00	8	.00	2	.00	11
.00	11	.00	8	.00	8	.00	10	.00	13	.00	8	.00	8	.00	7	.00	8	.00	7	.00	11	.00	10	.00	1	.00	13	.00	7	.00	11	.00	11	.00	4	.00	11	.00	7
.00	13	.00	10	.00	13	.00	11	.00	2	.00	4	.00	7	.00	13	.00	5	.00	10	.00	2	.00	13	.00	12	.00	7	.00	11	.00	13	.00	7	.00	3	.00	12	.00	2
.00	7	.00	13	.00	7	.00	7	.00	10	.00	11	.00	10	.00	8	.00	2	.00	4	.00	7	.00	11	.00	7	.00	2	.00	8	.00	8	.00	12	.00	2	.00	7	.00	13
.00	1	.00	11	.00	12	.00	14	.00	4	.00	2	.00	11	.00	2	.00	9	.00	11	.00	12	.00	8	.00	2	.00	8	.00	13	.00	7	.00	8	.00	13	.00	8	.00	
.00	12	.00	4	.00	4	.00	1	.00	1	.00	10	.00	14	.00	12	.00	10	.00	6	.00	1	.00	12	.00	13	.00	14	.00	12	.04	12	.00	13	.00	11	.00	13	.00	
.00	8	.00	12	.00	14	.00	8	.00	11	.00	14	.00	4	.00	1	.00	13	.00	1	.00	14	.00	14	.00	8	.00	4	.00	4	.06	4	.50	6	.00	10	.00	1	.01	14
.03	14	.00	14	.00	1	.00	12	.02	14	.00	12	.00	12	.00	14	.00	4	.00	5	.00	4	.00	4	.00	4	.00	3	.00	14	.07	14	.64	4	.00	14	.00	4	.04	1
.07	4	.00	3	.00	3	.00	4	.02	9	.00	9	.01	9	.00	4	.00	11	.00	12	.00	8	.00	2	.01	9	.05	1	.01	1	.07	1	.64	14	.00	9	.00	14	.05	9
.47	5	.00	9	.00	9	.00	9	.07	12	.05	3	.01	1	.00	9	.72	6	.00	9	.52	9	.00	9	.03	14	.05	9	.04	9	.07	6	.64	9	.00	12	.00	6	.09	4
.70	6	.00	6	.00	6	.03	6	.28	6	.66	6	.01	6	.00	6	.72	14	.00	13	.67	6	.02	6	.22	5	.05	6	.17	6	.13	9	.64	1	.00	6	.29	9	.14	6
1	9	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	12	1	14	1	5	1	5	1	6	1	5	1	5	1	5	1	5	1	5	1	5	1	5
																				one week ahead																			
.00	3	.00	10	.00	2	.00	2	.00	3	.00	1	.00	3	.00	10	.00	3	.00	2	.00	10	.00	7	.00	10	.00	10	.00	10	.00	2	.00	3	.00	7	.00	10	.00	3
.00	7	.00	11	.00	7	.00	10	.00	2	.00	2	.00	2	.00	11	.00	8	.00	3	.00	11	.00	10	.00	12	.00	2	.00	11	.00	3	.00	10	.00	1	.00	11	.00	10
.00	8	.00	13	.00	8	.00	3	.00	8	.00	7	.00	13	.00	13	.00	7	.00	8	.00	13	.00	1	.00	11	.00	11	.00	12	.00	10	.00	13	.00	8	.00	7	.00	12
.00	13	.00	2	.00	13	.00	13	.00	13	.00	8	.00	10	.00	12	.00	2	.00	7	.00	2	.00	11	.00	3	.00	13	.00	2	.00	11	.00	2	.00	9	.00	13	.00	11
.00	10	.00	12	.00	10	.00	11	.00	7	.00	10	.00	11	.00	14	.00	1	.00	10	.00	12	.00	3	.00	13	.00	12	.00	13	.00	7	.00	11	.00	2	.00	8	.00	13
.00	2	.00	14	.00	4	.00	8	.00	10	.00	13	.00	14	.00	2	.00	9	.01	11	.00	14	.00	13	.00	2	.00	7	.00	14	.00	13	.00	12	.00	13	.00	12	.00	2
.00	11	.00	8	.00	11	.00	14	.03	11	.00	4	.00	8	.00	7	.00	5	.01	12	.00	3	.00	2	.01	7	.00	8	.00	8	.02	8	.00	7	.00	4	.00	2	.00	14
.48	1	.00	7	.00	14	.00	7	.04	4	.00	11	.00	7	.00	8	.00	13	.01	4	.00	8	.00	14	.01	14	.00	14	.00	7	.05	4	.00	8	.00	10	.00	14	.00	6
.48	14	.00	4	.00	12	.00	12	.04	14	.00	14	.00	12	.00	4	.00	10	.01	13	.00	4	.00	8	.01	8	.04	4	.00	4	.05	12	.02	14	.00	11	.00	3	.00	7
.56	12	.00	3	.00	1	.00	4	.04	1	.00	12	.00	4	.00	1	.00	4	.05	9	.00	7	.00	12	.04	1	.04	6	.00	3	.12	1	.02	9	.00	3	.00	9	.02	8
.72	5	.00	9	.00	9	.00	9	.04	12	.00	9	.00	9	.00	9	.00	11	.05	6	.00	1	.00	4	.04	9	.04	3	.00	9	.12	14	.02	4	.00	14	.00	4	.10	9
.72	9	.00	1	.00	6	.00	6	.04	9	.00	6	.00	6	.00	3	.60	6	.47	14	.00	6	.00	9	.39	4	.04	9	.00	1	.33	9	.02	6	.00	12	.00	1	.25	4
.72	4	.00	6	.04	3	.02	1	.04	6	.41	3	.48	1	.00	6	.60	14	.47	1	.01	9	.00	6	.44	6	.05	1	.00	6	.35	6	.23	1	.00	6	.00	6	.55	5
1	6	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	12	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	1
																				one month ahead																			
.00	7	.00	10	.03	2	.08	7	.02	7	.00	2	.03	10	.00	10	.02	7	.04	3	.00	10	.00	10	.00	10	.01	2	.05	10	.04	2	.00	13	.07	7	.01	12	.19	2
.00	1	.00	12	.15	7	.08	8	.11	8	.00	7	.03	13	.00	3	.02	8	.04	6	.00	11	.00	11	.00	12	.01	8	.05	12	.04	10	.00	14	.07	8	.01	10	.37	10
.00	3	.00	11	.15	8	.55	2	.37	3	.00	8	.03	11	.00	2	.02	3	.04	5	.00	12	.03	12	.01	11	.02	7	.10	11	.04	11	.00	10	.18	1	.03	3	.37	12
.00	8	.01	14	.31	9	.55	9	.37	13	.00	1	.03	12	.00	12	.12	1	.04	9	.00	13	.03	13	.03	2	.02	9	.14	13	.04	13	.00	11	.18	9	.03	11	.37	11
.00	5	.01	13	.47	1	.55	3	.46	2	.00	13	.03	14	.00	13	.12	9	.06	7	.00	14	.04	3	.03	3	.02	6	.16	14	.04	12	.00	12	.18	13	.03	14	.37	13
.00	13	.04	3	.47	4	.55	4	.46	14	.00	10	.03	3	.00	6	.12	4	.13	8	.01	3	.04	1	.05	13	.14	4	.17	2	.04	14	.00	2	.41	4	.03	13	.37	14
.11	2	.04	6	.58	3	.55	1	.46	9	.00	4	.03	2	.00	8	.12	13	.13	2	.01	2	.04	14	.33	14	.52	1	.21	7	.04	8	.00	8	.41	14	.03	2	.37	4
.11	4	.04	2	.58	6	.55	13	.46	10	.00	11	.04	8	.00	14	.19	2	.13	14	.01	8	.08	7	.33	8	.52	13	.23	8	.04	7	.03	7	.41	11	.03	9	.37	3
.11	10	.04	9	.58	13	.55	14	.51	11	.00	14	.05	7	.00	7	.19	5	.13	12	.01	7	.24	8	.33	6	.52	14	.23	4	.04	4	.04	4	.41	10	.03	6	.37	8
.11	9	.04	8	.58	14	.55	10	.51	1	.00	12	.09	9	.01	11	.19	14	.34	10	.14	4	.32	9	.42	7	.52	10	.23	1	.04	9	.06	9	.58	2	.03	8	.37	6
.11	14	.04	7	.58	10	.55	11	.51	12	.01	9	.09	6	.04	9	.19	11	.86	13	.14	9	.49	4	.42	4	.65	11	.23	9	.04	1	.06	3	.58	12	.03	7	.37	7
.11	11	.53	5	.81	11	.55	6	.51	4	.01	6	.09	4	.16	4	.19	10	.87	11	.14	1	.53	6	.42	9	.76	3	.23	3	.04	6	.06	1	.74	3	.45	4	.55	9
.62	12	.53	4	.90	5	.55	12	.51	6	.52	3	.81	5	.29	1	.81	12	.87	4	.14	6	.53	2	.42	1	.76	12	.23	6	.37	3	.06	6	.74	5	.47	5	.78	1
1	6	1	1	1	12	1	5	1	5	1	5	1	1	1	5	1	6	1	1	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	6	1	1	1	5

Notes: MCS p-values for the QLIKE loss function. In each stock sub-panel we report the corresponding p-values on the left and on the right we list the estimator combination numbers ranked in the ascending orders of the p-values. Combinations 1 to 14 are: PDV; SBV; ATM; PDV+SBV; PDV+ATM; SBV+ATM; four duration-based estimators, estimators 1 to 4, called Duration4; SBV+Duration4; ATM+Duration4; ten RV-type estimators, estimators 5 to 14, called RV10; PDV+RV10; ATM+RV10; Duration4+RV10; All.

Web-Appendix C: Proofs of Main Results

Proof of Theorem 1. In the 'continuous case', our non-parametric estimator $NPDV_t$ can be viewed as the realized volatility estimator with respect to a particular stochastic sampling on regular grids, where the "barriers" are always equidistant and symmetric.

We shall derive the quadratic variation of the process U_n defined as

$$U_n := \delta_n^{-1}\Big(NPDV_t - [X,X]_t\Big) \equiv \delta_n^{-1}\left\{\sum_{j=0}^{N_t}\Big(X_{\tau_{n,j+1}} - X_{\tau_{n,j}}\Big)^2 - [X,X]_t\right\}.$$

We first note that the limit of the sample third moment, the *tricity*:

$$\delta_n{}^{-1}\sum_{j=0}^{N_{n,t}}\Big(X_{\tau_{n,j+1}} - X_{\tau_{n,j}}\Big)^3, \tag{56}$$

is crucial in the determination of the asymptotic bias in any stochastic sampling type framework, see for example Li et al. (2014, Section 1). In view of the definition of our sampling points (2) and straightforward applications of Doob's optional sampling theorem, it follows that the probability limit of (56) is zero.

Now, since by Itô's lemma we have

$$d(X_t - X_{\tau_{n,j}\wedge t})^4 = 4(X_t - X_{\tau_{n,j}\wedge t})^3 dX_t + 6(X_t - X_{\tau_{n,j}\wedge t})^2 d\,[X,X]_t\,,$$

and using standard measure change arguments through Girsanov's theorem

$$dU_t = 2\delta_n^{-1}\Big[\Big(X_t - X_{\tau_{n,j}\wedge t}\Big)dX_t\Big], \tag{57}$$

it follows that, along with weak consistency of NPDV (11) we have

$$\begin{aligned}
[U,U]_t = {} & \frac{4}{\delta_n^2}\sum_{j=0}^{N_{n,t}}\int_{\tau_{n,j}}^{\tau_{n,j+1}}\Big(X_s - X_{\tau_{n,j}}\Big)^2\, d[X,X]_s \\
= {} & \frac{2}{3}\frac{1}{\delta_n^2}\sum_{j=0}^{N_{n,t}}\Big(X_{\tau_{n,j+1}} - X_{\tau_{n,j}}\Big)^4 - \frac{8}{3}\frac{1}{\delta_n^2}\sum_{j=0}^{N_{n,t}}\int_{\tau_{n,j}}^{\tau_{n,j+1}}\Big(X_s - X_{\tau_{n,j}}\Big)^3\, dX_s && (58)\\
= {} & \frac{2}{3}N_{n,t}\delta_n^2 + o_p(1) \xrightarrow{P} \frac{2}{3}[X,X]_t \equiv \frac{2}{3}\int_0^t \sigma_s^2\, ds. \quad \text{a.s.} && (59)
\end{aligned}$$

The proof is now complete upon employing the stable limit theorem of Jacod & Shiryaev (2003, Theorem 9.7.3) applied to the stochastic sequence $\{U_n\}$. □

Proof of Lemma 1 and Theorem 2.

By Itô isometry and boundedness of σ we can see that

$$\mathrm{E}\Big(X_{\tau_{n,j+1}} - X_{\tau_{n,j}}\Big)^2 = \mathrm{E}\left(\int\limits_{\tau_j}^{\tau_{j+1}} \sigma_s \, dW_s\right)^2 = \mathrm{E}\left(\int\limits_{\tau_j}^{\tau_{j+1}} \sigma_s^2 \, ds\right) \leq (\sigma_*^2) \cdot \mathrm{E}(\tau_{j+1} - \tau_j),$$

from which it follows by Markov's inequality that

$$v_{\tau_{n,j}} = P\left[\Big|X_{\tau_{n,j+1}} - X_{\tau_{n,j}}\Big| > \delta_n\right] = O_P\left(\frac{\Delta_n}{\delta_n}\right).$$

Therefore, we have $\tau^*_{n,j+1} - \tau^*_{n,j} = O_P(\Delta_n^3/\delta_n)$ and also, since $\Delta_n = o(\delta_n^{1/3})$ it is clear that the thinned Poisson sample $\{\tau_n^*\}$ satisfies Assumption A. This proves Lemma 1.

We note that we can write

$$\Delta_n^{-1}\left(N_t^*\delta_n^2 - [X,X]_t\right) = \Delta_n^{-1}\left(\sum_{j=0}^{N_t^*}(X_{\tau^*_{n,j+1}} - X_{\tau^*_{n,j}})^2 - [X,X]_t - \mathcal{B}_t\right)$$

where $\mathcal{B}_t = \mathcal{B}_{n,t} = [N_t^*\delta_n^2 - \sum_j (X_{\tau^*_{j+1}} - X_{\tau^*_j})^2]$. Now, we see that

$$\begin{aligned}
\mathrm{E}\left(\left[X_{\tau^*_{n,j+1}} - X_{\tau^*_{n,j}}\right]^2 \Big| \mathcal{F}^*_{n,j}\right) - \delta_n^2 &= \sigma^2_{\tau_{n,j}} \mathrm{E}\left(\left[W_{\tau^*_{n,j+1}} - W_{\tau^*_{n,j}}\right]^2 \Big| \mathcal{F}^*_{n,j}\right) - \delta_n^2 \\
&= \sigma^2_{\tau_{n,j}}(2-1)!!(\tau^*_{n,j+1} - \tau^*_{n,j}) - \delta_n^2 \\
&\sim 1 \cdot \sigma^2_{\tau_{n,j}}\Delta_n^2 v_n - \delta_n^2 \qquad (60)
\end{aligned}$$

where !! means double factorial. But since $\Delta_n = O(\delta_n^{3/5})$ and $v = O(\Delta_n^3/\delta_n)$, it straightfowardly follows that the last term $(60) = O(\Delta_n^5/\delta_n - \delta_n^2) = O(\delta_n^2)$, implying that the bias term $\mathcal{B}_{n,t} = O_P(\Delta_n^5\delta_n^{-1}\delta_n^{-1})$.

Therefore, we see that the bias contribution from the time discretization asymptotically tends to zero, since $\delta_n^1/\delta_n^{3/5} \to 0$.

Now it suffices to derive the limiting law of $\Delta_n^{-1}(\sum_j (X_{\tau^*_{n,j+1}} - X_{\tau^*_{n,j}})^2 - [X,X]_t)$. Following the same argument leading to (58) in Theorem 1 and applying Lemma 9.1. of Aït-Sahalia & Jacod (2014), we see that the desired CLT holds with asymptotic variance

$$\frac{2}{3} \cdot 3 \cdot \int\limits_0^t \sigma_s^4 v_s \, ds$$

as required. □

Proof of Theorem 3. From (31) and the property of conditional expectation of the unobserved price change, it is straightforward to see that the leading bias term comes from

$$N'_{n,t} \cdot \frac{1}{4}\varsigma^2 \left\{\mathcal{D}_{\tau'_{n,j+1}} - \mathcal{D}_{\tau'_{n,j}}\right\}^2. \tag{61}$$

Now, on noting that both binary variables $\mathcal{D}.$ take values of 1 with probability p_a and -1 with probability $p_q(= 1 - p_a)$, we readily see that the expected value of (61) is given by $2 - 2(p_a - p_b)^2$.

Therefore, since $\varsigma_n = C_\varsigma \delta_n$, the asymptotic bias of (37) then follows from Theorem 1, completing the proof. $\square$