# Realized Variance in the presence of non-IID Microstructure noise: A Structural Approach

Bart Frijns*
Auckland University of Technology

Thorsten Lehnert†
Limburg Institute of Financial Economics

October 12, 2004

**Abstract**

It is a well known fact that at high sampling frequencies, the contamination of microstructure noise causes the Realized Variance to be a biased measure of the Integrated Variance. Recent developments in this field propose sampling on lower frequencies, sub-sampling techniques, or bias corrections using the autocorrelation patterns in the data. In this paper we propose a structural decomposition of the efficient price process and the microstructure noise. At the highest sampling frequency, we allow for potential correlation between the efficient price and the microstructure noise. For 20 actively traded stocks at Nasdaq, we find that the method provides a lower bound on Realized Variance. Applying a recently introduced bias correction reveals a very long persistence in transaction by transaction returns corrects the Realized Variance upwards to a level equal to low frequency Realized Variance. It remains questionable, however, whether this long persistence should be seen as microstructure noise, or is an inherent feature of the price process.

Keywords: Realized Variance, Ultra-High Frequency data, Tick Time models, Nasdaq, Microstructure noise.

JEL Classifications: C22, G15.

*Corresponding author: Bart Frijns, Department of Finance, Auckland University of Technology, Private Bag 92006, 1020 Auckland, New Zealand. Phone: +64-9-917 9999 extn 5713, Fax: +64-9-917 9629, E-mail: bart.frijns@aut.ac.nz

†Thorsten Lehnert, LIFE, Maastricht University, E-mail: t.lehnert@berfin.unimaas.nl

# 1 Introduction

Probably one of the most rapidly developing fields in finance is the area of realized variance ($RV$). Introduced by Andersen, Bollerslev, Diebold, and Labys (2001), $RV$, defined as the sum of squared intraday returns, converges to the integrated variance ($IV$) when the sampling frequency goes to infinity (see Barndorff-Nielsen and Shephard (2002) and Meddahi (2002)). This consistency relies on the assumption that a market operates without frictions and that there are no microstructure effects present. But when sampling frequencies are very high, prices are contaminated by microstructure noise and $RV$ is biased and an inconsistent estimate for $IV$ (see e.g. Andreou and Ghysels (2002) and Oomen (2002)). Sampling at lower frequencies reduces the bias of the microstructure noise, but increases the variance of the volatility estimate. Finding the optimal sampling frequency results in finding the best trade-off between bias and variance and is the approach followed by e.g. Bandi and Russell (2003). Zhang, Mykland, and Aït-Sahalia (2003) propose a measure for $RV$ based on sub-sampling techniques to reduce the loss of information inherent in data aggregation, which in the presence of microstructure noise provides a consistent estimate for $IV$. However, all these approaches rely on data aggregation to mitigate impact of the microstructure noise. In this paper, we propose a structural model for the price process at the highest sampling frequency. We decompose the microstructure noise from the "efficient" price process and propose a measure for $RV$.

A well established notion in market microstructure literature is that at high frequencies the contaminated price process can be decomposed into two components, an efficient price component and microstructure noise. Both components, however, are unobserved. A common assumption adhered to by previous studies (e.g. Bandi and Russell (2003), Hansen and Lunde (2004a) and Zhang, Mykland, and Aït-Sahalia (2003)) is that the microstructure noise is IID and uncorrelated with the innovation in the efficient price. More recently, Hansen and Lunde (2004b) have relaxed this assumption and allow for potential correlation between microstructure noise and the efficient price. Incorporating this correlation they find that the best unbiased estimator for $RV$ is achieved at the highest possible sampling frequency.

We contribute to the literature on $RV$ in several ways.

Inspired by previous research which focuses on the autocorrelation structure in high frequency returns to correct for the bias in $RV$, we take a structural approach for the decomposition of the price process. Motivated by Hansen and Lunde (2004b), we allow for potential correlation between the innovation in the efficient price and the microstructure noise. Additionally, the microstructure noise can depend on the time of the day. Given the parameter estimates of the model we can explicitly express the bias in $RV$ as derived by Hansen and Lunde (2004b). In line with their reasoning, we sample at the highest possible frequency, at every transaction. As such it can be seen as an empirical application of Engle (2000). Although Hansen and Lunde (2004b) prefer sampling in calendar time at every second using a previous-tick rule, [1] this is essentially the same as sampling in transaction time (see section 2). However, the impact of microstructure noise on the autocorrelations in observed returns are very different.

As we sample in transaction time, we observe the durations between transactions. This allows us to test for the impact of time between transactions on the innovation in the efficient price. We additionally test for the impact of time between trades on the correlation between the microstructure noise and the innovation in the efficient price. At the highest sampling frequency these durations have a negative impact, but this impact becomes positive at lower frequencies.

The explicit modelling of the price process and the microstructure noise gives us an insight into their properties. We find high correlations between the efficient price and the microstructure noise when sampling at the highest frequency. At lower frequencies this correlation diminishes, although the idiosyncratic noise persists. In line with Madhavan, Richardson, and Roomans (1997), we find a larger impact of the microstructure noise near the open and the close and a lower impact in the middle of the trading day.

At the highest possible frequency, we achieve an unbiased measure for $RV$ that is substantially lower than the traditional $RV$. Furthermore, our results suggest that data aggregation has a positive impact on the average level of $RV$. This increase in $RV$ seems to be related to a decrease in the correlation between the efficient price and the microstructure noise. We additionally find that this increase in volatility is more pronounced for the liquid stocks than for the illiquid stocks.

We compare our results with the bias correction proposed by Hansen and Lunde (2004b). For all of the stocks we find the lowest value for $rv$ at a one or five tick correction on the data. These values for $rv$ closely correspond to the values for $rv$ found with

[1] The previous-tick rule (Wasserfallen and Zimmermann (1985)) is defined as recording the last observation before a pre-determined time interval.

our structural model. Additionally, we find the standard deviations for average $rv$ to be almost always lowest when $rv$ is lowest. Moreover, there seems to be persistence in tick returns of up to about 30 transactions. Correcting for those long persistence would result in an increase in $rv$ up to a level that corresponds to the 30min. calendar time sampling frequency, which was recently proposed by Hansen and Lunde (2004b). However, we wonder whether this is microstructure noise or whether this is an inherent persistency in the price process.

The remainder of this paper is structured as follows. In the next section we discuss the concept of realized volatility. We discuss some relevant issues for the realized volatility when sampling is done in different time scales. Subsequently, section 3 addresses microstructure noise and the effects it has on observed prices. In this section we address a structural approach to estimating realized variance and we consider a model-free correction on data sampled in tick time and data sampled in calendar time. Section 4 discusses the data used in this study. In section 5 we present the results for realized variance applying different ways of sampling and different corrections. Finally, section 6 concludes.

# 2 Realized Variance

In this section we introduce the concept of Realized Variance. We start with the case where no microstructure noise is present and assume the price process evolves in calendar time. Consecutively, we show that Realized Variance is the same whether sampling is done in calendar time or tick time.

Realized Variance was introduced by Andersen, Bollerslev, Diebold, and Labys (2001) and is the discrete time approximation to the Integrated Variance ($IV$). This approximation holds under the assumption that the continuous price process is a semimartingale. We first define $IV$ and then $RV$.

Let $p_t$ be the log price of an asset at time $t$, then the return process is assumed to follow a stochastic differential equation of the form

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

where $\mu_t$ is a drift term, $\sigma_t$ the spot volatility and $W_t$ is a standard Brownian process. If we assume that $t$ is measured in days, then the return from $t-1$ to $t$ is given by the integral

$$p_t - p_{t-1} = \int_{t-1}^{t} \mu_u du + \int_{t-1}^{t} \sigma_u dW_u. \tag{2}$$

The Integrated Variance of this day is then given by

$$\begin{aligned} IV_t &= \left\{ \int_{t-1}^{t} \mu_u du + \int_{t-1}^{t} \sigma_u dW_u \right\}^2 \\ &= \left\{ \int_{t-1}^{t} \sigma_u dW_u \right\}^2 \\ &= \int_{t-1}^{t} \sigma_u^2 du, \end{aligned} \tag{3}$$

where the last line holds by Itô isometry.

Realized Variance is the discrete time approximation to the Integrated Variance and is defined as the sum of squared intraday returns. In calendar time these returns are sampled at equidistant intervals. On day $t-1$ we define the $i^{th}$ intraday calendar time return as

$$r_{C,t-1+ih}^{(h)} = p_{t-1+ih} - p_{t-1+(i-1)h}, \tag{4}$$

where $i = 1, \ldots, h^{-1}$ and $h^{-1}$ is a positive integer referring to the intraday sampling frequency. The superscript $(h)$ is used to indicate the frequency sampled at. The subscript $C$ refers to the fact that sampling is done in calendar time.

The Realized Variance in calendar time sampling on day $t$ is defined by

$$RV_{C,t}^{(h)} \equiv \sum_{i=1}^{h^{-1}} \left\{ r_{C,t-1+ih}^{(h)} \right\}^2. \tag{5}$$

By the theory of quadratic variation it holds that $RV$ is a consistent measure $IV$(see e.g. Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2002)), i.e. when $h \to 0$,

$$\text{plim} \sum_{i=1}^{h^{-1}} \left\{ r_{C,t-1+ih}^{(h)} \right\}^2 \to \int_{t-1}^{t} \sigma_u^2 du. \tag{6}$$

This convergence, however, only holds when returns are calculated in a correct manner. As sampling is done in calendar time, not every point sampled at will have a corresponding observation. The way these "missing observations" are filled in determine whether $RV$ converges to $IV$. Hansen and Lunde (2004b), for example, propose two different methods to fill in these observations. The first method is referred to as the previous tick method (introduced by Wasserfallen and Zimmermann (1985)). This method records the previous price at every sampling point until a new price appears. For very high frequency returns this means that many returns are equal to zero, with some nonzero returns when prices change. Sampling in this manner provides a consistent estimate for $IV$. A second method often employed is linear interpolation (e.g. Andersen, Bollerslev, Diebold, and Labys (2001)). In this case, if no new price is observed at a sampling point, the observation is interpolated from the previous and the next price. Hansen and Lunde (2004b) show that in this case $RV$ converges to zero in probability and thus does not provide a consistent measure for $IV$. Moreover, we note that from a theoretical point of view, calculating $RV$ with the linear interpolation method cannot be a consistent measure of $IV$ as it does not satisfy the martingale assumption, which is the assumption made for the price process. The correct manner to calculate $RV$ is thus by calculating returns using the previous tick method.

Additional to sampling in calendar time, one can also sample in tick time. When sampling in tick time we basically ignore the in-between zero returns and only observe the price when a transaction has taken place. Returns in transaction time are defined as follows. On day $t-1$ a total of $k(t)^{-1}$ transactions are observed, where $j = 1, \ldots, k(t)^{-1}$. The $j^{th}$ intraday return at sampling frequency $(a)$ is defined as

$$r^{(a)}_{T,t-1+jak(t)} = p_{t-1+jak(t)} - p_{t-1+(j-1)ak(t)}. \tag{7}$$

The subscript $T$ in this case refers to sampling in tick time. If $a = 1$ sampling is done at every tick, for $a = 2$ at two ticks, etcetera. Notice that these returns are spaced equidistantly in tick time, but are spaced irregularly in calendar time. The Realized Variance for sampling frequency $(a)$ in tick time is given by

$$RV^{(a)}_{T,t} \equiv \sum_{j=1}^{k(t)^{-1}} \left\{ r^{(a)}_{t-1+jak(t)} \right\}^2. \tag{8}$$

An interesting feature of $RV_{T,t}$ is that at the highest sampling frequency, it is the same as $RV_{C,t}$ at the highest sampling frequency, when the previous tick rule is applied for the calendar time observations. If, for example, we record prices at every second using a previous tick rule, zero returns are recorded if prices do not change. When summing the squared returns, only the non-zero returns determine $RV$. These non-zero returns are exactly the same as the returns that we sample in tick time, by sampling at every tick. This entails that both methods provide the same consistent measure for $IV$.

# 3 Microstructure Noise

In the previous section we addressed $RV$ when no microstructure noise is present and have shown that $RV$ is the same whether it is determined in calendar time or in tick time. In this section we discuss microstructure noise and its impact on $RV$. We further address the impact of microstructure noise on the time series properties of observed returns in both time scales. Consecutively, we propose a structural approach to filter out the microstructure noise. Finally, we address a newly introduced method to correct for the microstructure noise.

## 3.1 The Effects of Microstructure Noise on Realized Variance

Although RV converges to $IV$ when the sampling frequency goes to infinity, the presence of microstructure noise at these frequencies seriously contaminates the observed price process. This causes $RV$ to be a biased measure for $IV$. Without making any reference to a particular time scale the observed price at a given time $d$ on day $t$ can be decomposed as

$$p_{t+d} = p^*_{t+d} + u_{t+d}, \tag{9}$$

where $p^*_{t+d}$ refers to the latent *true* price and $u_{t+d}$ refers to the microstructure noise, which is also latent. The true price is often referred to as the efficient price and as such is often assumed to follow a random walk. The microstructure noise, on the other hand is considered transitory and is often assumed to be a covariance stationary process. Moreover, the microstructure noise is often assumed to be i.i.d. (see e.g. Bandi and Russell (2004), Hansen and Lunde (2004a) and Zhang, Mykland, and Aït-Sahalia (2003)).

One of the most trivial ways to circumvent the issue of microstructure noise is by aggregating data. Frictions, inherent to trading mechanisms have no permanent impact on prices. Therefore, if prices are observed at low frequencies the impact of microstructure noise on observed returns is negligible. This has been the first approach to circumvent the problems of dealing with microstructure noise (see e.g. Andersen, Bollerslev, Diebold, and Labys (2001) and Andersen (2004)), where sampling was done at five minute intervals.

However, the sampling interval of five minutes is chosen rather arbitrarily and several other studies have addressed the issue of finding the correct sampling frequency. The correct frequency here entails finding the best trade-off between the minimizing the bias introduced by the microstructure noise and minimizing the measurement error for $RV$ introduced by data aggregation (see e.g. Bandi and Russell (2004)). The microstructure noise is minimized at the lowest possible frequency, whereas the measurement error of $RV$ is minimized at the highest possible frequency.

An alternative approach is adopted by Zhang, Mykland, and Aït-Sahalia (2003). Their approach, based on sub-sampling techniques,leads to a lower measurement error in $RV$ without introducing a bias due to microstructure noise. The idea of this approach is to sample at lower frequencies (e.g. five minutes), while moving the sampling window forward with small steps. This allows all the information to be incorporated in the estimate of volatility without suffering from the effects of microstructure noise.

These approaches all circumvent dealing with microstructure noise by relying on data aggregation in some way, without explicitly addressing the properties of the microstructure noise and its impact on observed returns. To address the impact for observed returns, assume, for example that sampling is done in calendar time. The price observed then can be written as

$$p_{t-1+ih} = p^*_{t-1+ih} + u_{t-1+ih}, \tag{10}$$

and returns in calendar time are given by

$$r^{(h)}_{C,t-1+ih} = (p^*_{t-1+ih} - p^*_{t-1+(i-1)h}) + (u_{t-1+ih} - u_{t-1+(i-1)h}). \tag{11}$$

Assuming that the microstructure noise is i.i.d. and independent of the innovation in the efficient price, the variance of returns is given by

$$E[\{r^{(h)}_{C,t-1+ih}\})^2] = \sigma^2 + 2\omega^2, \tag{12}$$

where $\sigma^2$ is the variance of the true price and $\omega^2$ is the variance of the microstructure noise. If we further assume that the innovations in the efficient price are independent, then the microstructure noise introduces an MA(1) term in observed returns,

$$E[\{r^{(h)}_{C,t-1+ih}\}\{r^{(h)}_{C,t-1+(i-1)h}\}] = -\omega^2. \tag{13}$$

All higher order autocorrelations are equal to zero. The last two equations (12) and (13) reveal that the first order autocorrelation in returns has a lower bound of $-0.5$.

The example shown above relies on the assumption that at each observation the price changes, and we expect to observe a MA(1) structure in returns if the sampling frequency is low enough to ensure this. But if calendar time sampling is done using the previous tick rule, then when no price change is observed the previous price is recorded. This has no impact for the variance of returns in (12), but does affect the first order autocorrelation in (13). When zero returns are recorded the first order autocorrelation will be higher than that of a pure MA(1) structure. As the observed price in this way of sampling changes after a longer time period, sampling using the previous tick rule will induce higher order negative autocorrelations in returns. Moreover, the induced higher order autocorrelation will be a function of the duration between observations.

On the other hand, if sampling is done in tick time, we do not observe these zero return observations. As we observe only prices when they change, sampling in this time scale will not induce higher order negative autocorrelations, assuming i.i.d.-ness of the microstructure noise. This time scale should reveal the MA(1) structure and therefore, in our opinion, is the more appropriate sampling time scale.

Recent research (Hansen and Lunde (2004b)), however, has questioned the i.i.d.-ness property of the microstructure noise. Firstly, the microstructure noise can be correlated with the efficient price process. Further, the size of the microstructure noise can depend on the time of the day (see e.g. Madhavan, Richardson, and Roomans (1997)). In the next subsection we introduce a model, for tick time data, that decomposes the microstructure noise from the efficient price process. We assume that the microstructure noise is independently distributed, but allow for some time of the day variation. We further allow the microstructure noise to correlate with the innovations in the efficient price.

## 3.2 A structural model for the price process

In this section we propose a structural model for the price process of a security that uses ultra-high frequency data. In essence the model is very simple in the sense that we decompose the stock price into two components, one referring to the efficient price, the other to microstructure noise (see e.g. Hasbrouck (1993), Aït-Sahalia, Mykland, and Zhang (2003) and Zhang, Mykland, and Aït-Sahalia (2003) among others). However, instead of assuming that the microstructure noise is i.i.d. we allow for time of the day effects and correlation between the microstructure noise and the innovation in the efficient price. In this model we consider all informative transactions (the next section elaborates on what we refer to as informative transactions) at the time they are reported to the trading system. We thus sample each transaction and do not aggregate the data to some equidistant sampling interval. For sake of clarity we omit the $(a)$ superscript for the sampling frequency and assume that sampling is done at every tick. We will later reintroduce the superscript when needed.

Consider the price process observed in tick time. This price process is decomposed in a latent component referring to the efficient price and microstructure noise

$$p_{t+jk(t)} = p^*_{t+jk(t)} + u_{t+jk(t)}, \tag{14}$$

where the first component $(p^*_{t+jk(t)})$ refers to the true, or efficient price of an asset and the second component $(u_{t+jk(t)})$ measures the microstructure noise in each transaction. The efficient price is assumed to follow a random walk. However, we let the model determine which time scale the price process evolves in. This is done by including a duration function in front of the innovation term of the random walk,

$$p^*_{t+jk(t)} = p^*_{t+(j-1)k(t)} + \tau^{\delta_{1,t}}_{t+jk(t)} \varepsilon_{t+jk(t)}, \tag{15}$$

where

$$\tau_{t+jk(t)} \equiv \frac{\ell_{t+jk(t)} - \ell_{t+(j-1)k(t)}}{\frac{1}{N}\sum_t k(t) \sum_j (\ell_{t+jk(t)} - \ell_{t+(j-1)k(t)})}$$

measures the deviation from the average duration over the sample, where $N$ is the total number of days evaluated. The innovation term $\varepsilon_{t+jk(t)}$ measures the innovation in the random walk at average durations and has a variance of $\sigma^2_t$. The parameter $\delta_{1,t}$ measures the impact duration has on the innovation in the random walk. When $\delta_{1,t} = \frac{1}{2}$ the variance of the random walk innovation grows proportional to the duration between observation. In this case the price process evolves in calendar time, where the variance of the innovation in the efficient price grows proportional to the time length between observation. This process would be the discrete time equivalent of the continuous price process in (1). On the other hand when $\delta_{1,t} = 0$, the time between observations does not influence the innovation in the random walk. In this case the variance of the random walk grows proportional to the number of transactions. This is the time scale that has been promoted more recently by Ané and Geman (2000).

We further want to specify the microstructure noise $u_{t+jk(t)}$. One factor that we consider is whether the innovation in the efficient price is correlated with the microstructure noise.

Hansen and Lunde (2004b) stress the importance of the correlation that exists between microstructure noise and the innovation in the efficient price. Moreover, by including a parameter that measures the impact of time between transactions on these correlations we allow for richer dynamics than when sampling at fixed time intervals. We include a parameter that measures this correlation and we consider the effects of the opening and closing of the market. The amount of microstructure noise might differ substantially around the open and close of the market (see e.g. Madhavan, Richardson, and Roomans (1997)). Combining these two factors we specify the noise as

$$\begin{aligned}
&u_{t+jk(t)} = \alpha_t \tau_{t+jk(t)}^{\delta_{2,t}} \varepsilon_{t+jk(t)} + D_{1,t+jk(t)} e_{1,t+jk(t)} + D_{2,t+jk(t)} e_{2,t+jk(t)} + D_{3,t+jk(t)} e_{3,t+jk(t)}, \\
&Var(e_{1,t+jk(t)}) = \omega_{1,t}, \\
&Var(e_{2,t+jk(t)}) = \omega_{2,t}, \\
&Var(e_{3,t+jk(t)}) = \omega_{3,t},
\end{aligned} \tag{16}$$

where $\alpha_t$ measures the extent to which the innovation in the efficient price and the microstructure noise co-vary at average durations. The parameter $\delta_{2,t}$ measures the impact that time has on this correlation. When $\delta_{2,t} > 0$ the dependence increases with an increase in durations and vice versa. The three dummy variables $D_{1,t+jk(t)}$, $D_{2,t+jk(t)}$ and $D_{3,t+jk(t)}$ are equal to zero or one, such that the three idiosyncratic noise terms ($e_{1,t+jk(t)}$, $e_{2,t+jk(t)}$, $e_{3,t+jk(t)}$) capture the microstructure noise at different times of the day. The dummy variable $D_{1,t+jk(t)}$ equals 1 near the opening of the market (between 9.30 and 11.00) and zero otherwise, $D_{2,t+jk(t)}$ equals 1 over the normal part of the trading day (11.00 - 14.30) and zero otherwise, and $D_{3,t+jk(t)}$ equals 1 near the close (14.30 -16.00) and zero otherwise.

The model presented above is can be put into a state space model and estimated by QML using a Kalman Filter[2]

$$\begin{aligned}
&p_{t+jk(t)} = p^*_{t+jk(t)} + u_{t+jk(t)}, \\
&p^*_{t+jk(t)} = p^*_{t+(j-1)k(t)} + \tau_{t+jk(t)}^{\delta_1} \varepsilon_{t+jk(t)}, \\
&u_{t+jk(t)} = \alpha \tau_{t+jk(t)}^{\delta_2} \varepsilon_{t+jk(t)} + D_{1,t+jk(t)} e_{1t+jk(t)} + \\
&D_{2,t+jk(t)} e_{2,t+jk(t)} + D_{3,t+jk(t)} e_{3,t+jk(t)}.
\end{aligned} \tag{17}$$

[2]For more information on state space models and Kalman Filter techniques, we refer to Harvey (1989) and Durbin and Koopman (2001).

As the latent variable in the state equation is a random walk the model is initialized using a diffuse prior. This entails that the initial prediction error variance is set at a very large number. For the prediction error variance to converge to normal levels we leave out the first 50 observations for the calculation of the likelihood function. As the innovation in the random walk depends on the time length between observations, the model does not converge to a steady state. This increases the computational efforts of this filter as the prediction error variance needs to be computed at every single recursion. All parameters in the model are identified, as long as $\delta_2 \neq 0$ and $\delta_1 + \delta_2 \neq 0$ and the variance of durations is not equal to zero (See the Appendix for a full derivation of the moment conditions).

In a first step, we estimate the model assuming constant parameters over all trading days. In this case, the subscript $t$ can be omitted for the parameters. Every day, we re-initialize the system by increasing the prediction error variance and leaving out the first 50 observations on each day in the calculation of the likelihood function. This procedure allows us to analyze the properties of the microstructure noise effectively. As a second step, the model is re-estimated on a daily basis. In this case, the re-initialization procedure does not allow us to calibrate the model at lower than the highest frequency. In addition to the full calibration, this is a more correct approach to compute $RV$.

Given the parameterization in the model above we can define the realized variance for the price process $p_{t+jk(t)}$,

$$\begin{aligned} RV_{T,t} &= \sum_{j=1}^{k(t)^{-1}} (\Delta p^*_{t+jk(t)} + u_{t+jk(t)} - u_{t+(j-1)k(t)})^2 \\ &= \sum_{j=1}^{k(t)^{-1}} \tau_{t+jk(t)}^{2\delta_{1,t}} \sigma_t^2 + 2\alpha_t^2 \sigma_t^2 \tau_{t+jk(t)}^{2\delta_{2,t}} + 2\omega_{\cdot,t} + \alpha_t \sigma_t^2 \tau_{t+jk(t)}^{(\delta_{1,t}+\delta_{2,t})}, \end{aligned} \tag{18}$$

where the first part

$$RV_{T,p^*,t} = \sum_{j=1}^{k(t)^{-1}} \tau_{t+jk(t)}^{2\delta_{1,t}} \sigma_t^2, \tag{19}$$

is an unbiased measure for the $RV$ of the efficient price. The remaining part in (18) is the bias in $RV$ introduced by the microstructure noise.

The model proposed above can be estimated at different frequencies and therefore we can estimate $RV$ at different sampling frequencies

$$RV^{(a)}_{T,p^*,t} = \sum_{j=1}^{k(t)^{-1}} \tau_{t+jak(t)}^{2\delta_{1,t}} \{\sigma_t^{(a)}\}^2, \tag{20}$$

## 3.3 A Model Free Correction

The previous subsection introduced a structural model, explicitly designed for sampling in tick time, to correct for the microstructure noise. This section discusses a recently introduced approach by Hansen and Lunde (2004b) that relies on a model free correction of the microstructure noise. The correction is based on the autocorrelation induced by the microstructure noise. Here we define their measure for $RV$ and compare the results of this measure to the results of our structural approach in section 5.

Hansen and Lunde (2004b) start with the general notion that prices are contaminated by noise, which introduces autocorrelations in observed returns. The correction they propose assumes that $RV$ is biased similar to the bias shown in (18). However, their approach relies on calendar time sampling instead of tick time sampling. When sampling in calendar time using the previous tick rule, Hansen and Lunde (2004b) prove that this bias can be corrected for by correcting for all induced higher order autocorrelations in the observed returns. The correction takes the form of a Newey-West type correction

$$RV_{C,q,t}^{(h)} = \sum_{i=1}^{h^{-1}} \{r_{C,t-1+ih}^{(h)}\}^2 + 2\sum_{s=1}^{q} \frac{h^{-1}}{h^{-1}-s} \sum_{i=1}^{h^{-1}-s} r_{C,t-1+ih}^{(h)} r_{C,t-1+(i+s)h}^{(h)}, \tag{21}$$

where $q$ is the maximum number of lags, which is depends on the sampling frequency $h$..

There are several important finding in their paper. Firstly, the best unbiased measure for $RV$ is achieved when sampling at the highest frequency (in their case 1 second). Second, their correction is not done over a fixed number of lags but depends on a certain amount of time that has to be corrected for. By this we mean that autocorrelation has to be corrected for e.g. 60 seconds, independent of the sampling frequency adopted.

The correction over longer lags is in contrast with the notion that the microstructure noise induces an MA(1) in observed returns, but can be explained by the time scale that the process evolves in. As mentioned in section 2 sampling calendar time at the highest possible frequency using the pre-tick rule, leads to many zero observations. Although the price process can be assumed to evolve continuously, the microstructure noise is only observed when a new transaction occurs. Therefore, if the average time between transactions is 10 seconds, most of the negative autocorrelation will be observed at the $10^{th}$ lag. The autocorrelation function will therefore display negative autocorrelations up to longer lags and will be influenced by the distribution of durations between trades.

If only the time scale explain this autocorrelation function for observed calendar time returns, then we can also address the Hansen & Lunde correction in tick time. In a similar fashion we define $RV$ in tick time as

$$RV^{(a)}_{T,q,t} = \sum_{j=1}^{k(t)^{-1}} \{r^{(a)}_{T,t-1+jak(t)}\}^2 + 2\sum_{s=1}^{q} \frac{k(t)^{-1}}{k(t)^{-1}-s} \sum_{j=1}^{k(t)^{-1}-s} r^{(a)}_{T,t-1+jak(t)} r^{(a)}_{T,t-1+(j-s)ak(t)}, \quad (22)$$

where $q$ is the maximum number of lags in tick time, which depends on the sampling frequency $(a)$. When sampling is done at every transaction, we expect only a first order negative autocorrelation, induced by the microstructure noise. Applying this correction in tick time should therefore reveal whether there are other effects present in observed returns.

# 4 Data

In this study we examine transaction data of 20 actively traded stocks at Nasdaq. The selected stocks vary in liquidity, but were all included in the Nasdaq-100 index during 1999. Data were provided by Nastraq. The sample period extends from February $1^{st}$ until July $30^{th}$, spanning a total of 124 trading days and contains all transactions within normal trading hours (9.30 - 16.00) of the Nasdaq National Market. The transaction data contains the reported time of the trade, the executed time of the trade, the price at which the trade took place and the volume traded. Some indicators were added to trades that were reported late. We use transactions data in contrast to midquote data (as used by e.g. Bandi and Russell (2004) and Hansen and Lunde (2004b)) as Nasdaq inside quote are updated much less frequently then the occurrence of transaction (see e.g. Frijns (2004)).

As we are interested in evaluating informative transactions, we remove all trades with a late reported indicator. In a similar way we remove all trades that were executed before, but reported at the same time or after the next executed trade. If two trades were executed at the same time, the trade reported first is considered and the other one is removed. This procedure removes a substantial amount of the transaction data. If these transactions are not removed from the data then tick data would be noisy as we constantly observe past information. The data set we construct contains all trades where the discrepancy between executed and reported time is minimized and thus has the highest information content.

From this data set we remove all outliers. Observations are considered outliers when there are either two consecutive observations that deviate more than 5 standard deviations from the mean, or a single observation that deviates more than 10 standard deviations.

The information used from this data set are the log price at which the transaction takes place ($p_{t+jk(t)}$) and the reported time of the $j^{th}$ transaction $\ell_{t+jk(t)}$.

INSERT TABLE 1 HERE

Table 1 reports some summary statistics of the data. In our filtered sample, the most frequently traded stock, Dell computer corporation, had 1,218,852 transactions over 124 trading days. This averages to 13,842 trades per day. In contrast the least liquid stock in the sample had 213,749 trades over the whole period, resulting in 2,051 trades per day on average. More interesting are the large differences between the minimum and the maximum number of transactions that are observed within a day. This clearly indicates that, there is a great variety in the liquidity of these stocks, although all these stocks were included in the Nasdaq-100 index. The diversity in liquidity is also confirmed by the average duration between transactions. For the most liquid stocks in the sample this duration is below 15 seconds, whereas for the least liquid stock in the sample this lies around one minute. The maximum duration between transactions illustrates that the distribution of durations is right skewed (see Engle and Russell (1997)).

Additional to the data set in tick time we also sample the data at 2 ticks, 5 ticks and 10 ticks, where the 10 tick level is only used for the traditional realized volatility measure. We consider these data sets to analyze the effect of data aggregation on realized volatility. To compare our results to the traditional method of computing realized volatility, we also construct a price series sampled at one, two, five and ten minute intervals. Sampling in fixed time intervals is done using the previous-tick rule.

# 5 Results

In this section we present the results. Instead of reporting Realized Variance we report the square root of this, also called the realized volatility ($rv$). Further, instead of using the superscripts ($h$) and ($k(t)$) we will use the sampling frequency as the superscript. We start by looking at $rv$ at different sampling frequencies, when no correction for the microstructure noise has been done. Next, we address the parameter estimates of the structural model introduced in section 3.2, for different sampling frequencies and report the realized volatilities found by this approach. Finally, we report realized volatilities when sampling is done in both calendar time and tick time, with the correction proposed in section 3.3.

## 5.1 Uncorrected Realized Volatility

In this subsection we discuss the results for the realized volatility when no correction has been done for the microstructure noise present. We only report the results for sampling in calendar time as tick time sampling leads to similar results. The realized volatilities have been computed at different sampling frequencies.

INSERT TABLE 2

In Table 2 we report the results for average uncorrected realized volatilities at a 1, 10, 30, 60, 120, 300, 600 and 1800 second sampling frequency and represent the so called volatility signature of a stock. The Table reports the average realized volatility over the 124 trading days and reports the standard deviations of *rv* in brackets. These results confirm the general notion that at very high sampling frequencies microstructure noise contaminates the observed prices, which leads to an upward bias for *rv*. At lower frequencies *rv* decreases. Similar results have been documented by Barndorff-Nielsen and Shephard (2002) and Hansen and Lunde (2004b).

There are some interesting features observed from Table 2. Firstly, we find that for most stock the lowest value for *rv* is found at the 30 minute (1800 sec) sampling frequency. This even holds for the most liquid stocks in the sample like Dell and Intel. This entails that when sampling at e.g. a five minute frequency, a frequency often adopted, there is still a considerable amount of microstructure noise present in the measure of *rv*. It further shows that the effects of microstructure are persistent in aggregated returns and even affect lower frequency returns.

Another interesting feature of Table 2 is that the lowest standard deviations for average *rv* are found around the 30 and 60 second sampling frequency. When sampling at higher frequencies the bias in *rv* due to the microstructure noise likely causes daily *rv*'s to be more volatile. At lower frequencies the increase in the standard deviations of *rv* can be explained by the measurement error introduced by sampling at lower frequencies (see Barndorff-Nielsen and Shephard (2002)).

Hansen and Lunde (2004b) use the 30 minute *rv* as a benchmark volatility, to which they compare the realized volatilities computed with bias corrections. In the remainder of the discussion we will use this level of volatility as a benchmark.

The next section addresses the results of a structural model, designed to correct for the noise in *rv*.

## 5.2 Structural Model: Parameter Estimates

In this subsection we discuss the parameter estimates of the model developed in section 2. The parameters are calibrated using the full sample and are not re-estimated on a daily basis. Therefore the subscript $t$ for the parameters is omitted. We estimate this model at three different sampling frequencies, the 1-tick, 2-tick and 5-tick level.[3] In tables 3 - 5 we report the results of these parameters.

The results of the model when sampled at every tick are reported in Table 3. The first column reports the duration parameter ($\delta_1$) on the innovation of the efficient price. The two values of interest for this parameter are $\frac{1}{2}$ and 0. When $\delta_1 = \frac{1}{2}$ the random walk evolves in calendar time and the variance of the efficient price expands proportional to the time interval between transactions. When $\delta_1 = 0$ the random walk evolves in transaction time. This time scale has more recently been motivated by Ané and Geman (2000). The results for $\delta_1$ differ substantially over the stocks in the sample but are on average positive with a value of 0.04. For 14 stocks this value is significantly larger than 0 and only 3 stocks report a significantly negative value. The parameter estimates are in all cases significantly different from $\frac{1}{2}$ indicating that the hypothesis of the price process evolving in calendar time is clearly rejected. These results are in line with the findings of Frijns and Schotman (2004).

INSERT TABLE 3 HERE

Consequently, we are also interested in the impact of time on the correlation between the microstructure noise and the innovation in the efficient price, measure by the sum of $\delta_1$ and $\delta_2$. The second column of Table 3 reports the parameter estimates of $\delta_2$. Again

[3] Higher aggregation levels are not considered as the Kalman Filter routine uses 50 observations at the start of the day to initialize the system. At the higher aggregation level this leads to a loss of data and does not incorporate the first part of the day in the estimation of the parameters.

there is a large difference between the parameter values for the individual stocks. The average of these values is 0.03, however there are 6 stocks for which the parameter is significantly negative compared to 8 significant positive values. We typically find that $\delta_2$ takes the opposite sign of the parameter $\delta_1$ in all but three cases. In combination with $\delta_1$ this reduces the effect of time on the correlation (see Equation (18)). However, for all stocks the impact of time on the correlation is positive, which refers to higher correlations at longer durations.

The third column of Table 3 reports the $\alpha$ of the model. This parameter measures the extend to which the innovation in the random walk correlates with the microstructure noise. A first conclusion from these parameters is that all values for $\alpha$ are positive, although not all parameter values are significant. The average value for $\alpha$ is 2.18. Values of $\alpha$ larger than 2 indicate that the noise around the efficient price is more than twice the innovation in the random walk. One point to consider is that whenever $\delta_2$ is not significantly different from zero, the identification of $\alpha$ is not guaranteed and vice versa.

The next column reports the results on the innovation parameter of the random walk ($\sigma$), expressed as the square root of the variance of the innovation in the efficient price. These parameters are estimated with very high precision as indicated by their low standard errors. The most interesting result for these parameters is that the innovation in the efficient price is inversely related to the liquidity of the asset. For the most liquid stocks the innovation in the random walk per transaction is smaller. For the illiquid stocks the innovation is larger. Therefore, there seems to be a liquidity effect in the innovation in the efficient price, but no relationship with the time between transactions.

The last three columns in the Table report the microstructure noise around the efficient price. The parameter $\omega_1$ measures the noise near the open of the market (all trades between 9.30 and 11.00), $\omega_2$ measures the noise during the day (11.00 until 14.30) and $\omega_3$ measures the noise around the close (from 14.30 - 16.00). Overall we observe that the noise is highest near the opening of the market, decreases and is lowest in the middle of the trading day and increases again at the close of the day. These results are in line with Madhavan, Richardson, and Roomans (1997). Again we observe the same relationship as with the innovation in the random walk. The more liquid the stock, the lower the microstructure noise.

INSERT TABLE 4 HERE

For the aggregated data at the 2-tick level and the 5-tick level we discuss the results simultaneously. A first overall result that we observe is that the $\delta_1$ parameter decreases, where at the 5-tick level we observe only significant negative values. Where the value for $\delta_1$ was 0.04 for the 1-tick level, we now have -0.05 and -0.15 for the 2- and 5-tick level respectively. These results indicate that when for the aggregate data the duration between multiple transactions is short, the innovation in the efficient price is larger and vice versa.

The opposite result is observed for the $\delta_2$ parameter. When data is aggregated this term increases. Where the average value for this parameter was 0.03 for the 1-tick level, this increases to 0.16 and 0.51 for the 2- and the 5-tick level respectively. The impact of time on the correlation between the innovation in the efficient price and the microstructure noise, defined as the sum of $\delta_1$ and $\delta_2$, tends to be larger when durations are longer than average. Again the overall result holds in each case that $\delta_2$ take the opposite sign of $\delta_1$, but the sum remains positive.

For the $\alpha$'s we see a decreasing trend. The average of 2.18 for the 1-tick level, decreases to 1.34 and 0.58 for the 2- and the 5-tick level respectively.

INSERT TABLE 5 HERE

An interesting result comes from the idiosyncratic noise component. This noise does not change at higher aggregation levels. However, as the innovation term of the efficient price becomes larger when data is aggregated, the signal to noise ratio increases. Therefore, the total impact of microstructure noise disappears at higher aggregation levels, as expected.

## 5.3 Structural Model: Realized Volatility

In this subsection, we discuss the results of realized volatility given the structural model proposed in section 3.2. On each of the 124 trading days we calculate $rv$ based on the parameter estimates discussed in the previous section. These $rv$'s are computed using a 1, 2 and 5 tick sampling frequency. Additionally, we compute $rv$ by estimating (17) on a daily basis. This is only done for the 1 tick sampling frequency.

Panel A in Table 6 presents the average realized volatility estimates obtained from (17) using the transactions data when the Kalman Filter is run over the whole sample period. The standard deviations of these realized volatilities over the 124 trading days are reported in brackets. The results found are not in line with previous research and offer a lot of new insights. Sampling at the highest frequency results in the lowest average level of realized volatility among the frequencies analyzed. Interestingly, the average levels of realized volatilities are considerably lower compared to the 30 minute *rv*'s presented in Table 2. Moreover the highest sampling frequency also yields the lowest standard deviations for *rv*'s. When the model is estimated at a 2 and 5 tick frequency we see that *rv* increases.

INSERT TABLE 6 HERE

The results of the model presented above assume fixed parameters over the whole sample period. As a robustness test of our results we relax this restriction and allow the parameters to vary on a daily basis. Panel B in Table 6 presents the results when the model is estimated on a daily basis at a 1 tick frequency.[4] The results suggest that estimating the model over the whole sample period smooths out some of the daily variations as the standard deviation of average *rv* is somewhat higher than for the model estimated over all days. However, the average levels of the realized volatilities are within the same range. Our general result that at the highest frequency, realized volatilities are lower are not affected when estimating the model on a daily basis.

As the realized volatilities increase at lower sampling frequencies, we expect some sort of dependence between tick returns. This, however, contradicts the assumption that prices follow a random walk. In the next section we follow a different approach to correcting for the bias in *rv*, which might reveal more of the properties of *rv*.

## 5.4 Model-Free Correction

In this section we present the results for *rv* using a model-free correction recently introduced by Hansen and Lunde (2004b). We start with the correction on the data sampled in calendar time, after which we present the results when sampling is done in tick time.

One of the main findings of Hansen and Lunde (2004b) is that the best measure for *rv* is found at the highest sampling frequency. Best in this sense means that the measurement error made between the estimated and true volatility is smallest. The correction to the data, however, is done over longer lags, where Hansen and Lunde (2004b) find the correct lag length when the corrected *rv* equals the average 30 minute *rv*.

INSERT TABLE 7 HERE

[4]Numerical issues burden the estimation of the model at lower frequencies.

In Table 7, we report the average realized volatilities sampled in calendar time at every second, where the correction has been done over several lag lengths. Standard deviations of average *rv*'s are again reported in brackets. Interestingly, a one lag correction is by far not enough to correct for the impacts of the microstructure noise, which rejects the theoretically assumed MA(1) structure for calendar time returns. The next step is to find the *rv* that best corresponds to the 30 minute *rv* in Table 2. We find that the lags needed to correct for the microstructure noise to some extent depends on the liquidity of the asset. For frequently traded stocks like Cisco and Microsoft a five second correction seems to be sufficient. However, for the less liquid ones, such as Starbucks and Compuware, a correction of 5 minutes is required. These findings are all consistent with the findings of Hansen and Lunde (2004b), but there are some other interesting features.

Firstly, when we consider the standard deviations of *rv*, we find that the lowest standard deviation is often found at the value for the average *rv* that closest corresponds to the 30 minute *rv*. This may again represent the trade-off point between the bias that is induced by the microstructure noise and the measurement error that is made in *rv*. However, when corrections are done over lags longer than the one that that matches with 30 minute *rv*, we often observe that *rv* drops below the 30 minute *rv*, after which it corrects upwards again to a number close to the 30 minute level. A clear example of this is the stock of Dell, which drops below the 30 minute level of *rv* after a 10 second correction and moves back to the 30 minute *rv* reaching it again after having applied a 60 to 90 second correction. Even more interesting is the fact that the lowest value for *rv* found at a 15 second correction is close to the *rv* found with the structural model. These effect are more or less present for all the stocks in the sample but are more pronounced for the liquid stocks than the illiquid ones.[5]

INSERT FIGURE 1 HERE

The impact of the correction on *rv* is best observed when considering the autocorrelation function for the observed 1 second returns. In Figure 1 we plot the autocorrelation function up to 50 lags for four representative stocks in the sample. All stocks display the negative autocorrelation induced by the microstructure noise. For the less liquid stocks like Apple and Amgen, we observe that this negative autocorrelation persists longer than for the more

[5]Likely, the correction for the illiquid stock needs to be done over even longer lags to observe this pattern.

liquid stocks (Amazon and Dell). At longer lags the two liquid stocks reveal some positive autocorrelations, which can explain the upward movement observed in *rv*.

INSERT TABLE 8 HERE

Additional to addressing the correction in calendar time sampling we also apply the correction in tick time. In Table 8 we present the results for average *rv* when sampling at every transaction and correcting for different lag lengths. Standard deviations of average *rv* are reported in brackets.

The results for the correction in tick time reveal a very different pattern from the results obtained with calendar time sampling. For all of the stocks we find the lowest value for *rv* at a one or five tick correction on the data. Moreover, these values for *rv* closely correspond to the values for *rv* found in the structural model (Table 6). Finding *rv* that closest approximates the 30 minute realized volatility leads in most cases to a minimum correction of 20 to 30 ticks. Correcting beyond 30 ticks does not have an impact on the level of *rv*. Hence, there seems to be persistence in tick returns of up to about 30 transactions. Another remarkable feature is that the standard deviations for average *rv* are almost always lowest when *rv* is lowest.

These results indicate that there are two sources affecting returns in tick time. The first source is the microstructure noise, which induces negative autocorrelations at very short lag lengths and causes the upward bias in *rv*. The second one is a persistence in returns, which leads to positive autocorrelation up to much longer lags. This decreases the variance of *rv* when the correction is done over short lags lengths. These two effects clearly have different lag lengths at which they manifest themselves.

INSERT FIGURE 2 HERE

In Figure 2 we plot the autocorrelation function for the four stocks that we considered before. The autocorrelation functions confirm the findings in Table 8. At the first lag we observe a large negative autocorrelation induced by the microstructure noise. There are several lags beyond the first lag which are also negative, which can explain the lowest value of *rv* to be found at a five lag correction. Although there are some positive autocorrelation found at longer lags, there seems to be no clear pattern in these autocorrelations (except for Amazon).

The question that arises now is whether the positive autocorrelation that increases *rv* is a source of microstructure noise or whether this long persistency is inherent in the price process. Some persistence in returns can be explained by the presence of informed traders (see Glosten and Harris (1988)). The presence of informed traders causes order flow to be positively correlated. This results in a higher probability of a buy transaction to be followed by another buy transaction, which causes transaction returns to display autocorrelation. However, this positive autocorrelation wa only found up to five lags in transaction returns (see Hasbrouck and Ho (1987)). As we find the persistence to pertain up to the thirtieth lag, it is doubtful whether this can be explained by the presence of informed traders. Moreover, as the autocorrelations persist up to such long lags, it is unlikely to be a source of microstructure noise. However, if this persistence is no microstructure noise, then the martingale property assumed for the price process does not hold. It is further remarkable that many of these small positive, unstructured autocorrelations have such a huge impact on the level of *rv*.

# 6 Conclusion

In this paper, we analyze Realized Variance in the presence of non- i.i.d. microstructure noise. Most research relies on the assumption that the microstructure noise is i.i.d., but this is shown not to hold in practice. In this paper we follow two approaches to decompose the microstructure noise from the price process.

Firstly, using ultra-high frequency transaction data, we decompose the stock price into a component referring to the efficient price and one referring to microstructure noise. As we sample in transaction time, we observe the durations between transactions. This allows us to test for the impact of time between transactions on the innovation in the efficient price. The explicit modelling of the price process and the microstructure noise gives us an insight into their properties. We find high correlations between the efficient price and the microstructure noise when sampling at the highest frequency. At lower frequencies this correlation diminishes, although the idiosyncratic noise persists. In line with previous reserach, we find a larger impact of the microstructure noise near the open and the close and a lower impact in the middle of the trading day. We additionally test for the impact of time between trades on the correlation between the microstructure noise and the innovation in the efficient price. At the highest sampling frequency these durations have a negative impact, but this impact becomes positive at lower frequencies.

However, the realized volatilities found at the highest frequencies are substantially lower than the realized volatilities at lower frequencies. Additionally, we find the standard deviations for average $rv$ to be almost always lowest when $rv$ is lowest. In order to understand our results of the structural model, we follow a second model-free approach to correct for the microstructure noise, that was recently introduced by Hansen and Lunde (2004b). We show that when sampling is done in calendar time, we mostly observe negative autocorrelation, which lead to an upward biased measure for the realized volatility, even when correcting over longer lags. We argue that these negative autocorrelations are induced by the particular sampling technique and is not an inherent feature of the data. Our results suggest that for the evolution of the efficient price, calendar time is not the appropriate time scale.

Applying the correction in tick time shows that most of the negative autocorrelation is centered at the first or the second lag. Applying this approach with a one lag correction leads to similar results as found in our structural approach. We further find that correction for longer lags in tick time increases the realized volatility up to a level equal to the low frequency realized volatilities. However, as the lag length needed for realized volatility to be equal to the low frequency realized volatility (up to 30 lags), it remains questionable whether this is microstructure noise. We argue that this is an inherent persistency in the price process, which should not affect the measure of $rv$.

# A Identification of the univariate state space model

This appendix addresses the identification of the state space model (17). We start with the equation for log prices

$$\begin{aligned} p_{t+jk(t)} &= p^*_{t+jk(t)} + u_{t+jk(t)}, \\ p^*_{t+jk(t)} &= p^*_{t+(j-1)k(t)} + \tau^{\delta_1}_{t+jk(t)} \varepsilon_{t+jk(t)}, \\ u_{t+jk(t)} &= \alpha \tau^{\delta_2}_{t+jk(t)} \varepsilon_{t+jk(t)} + D_{1,t+jk(t)} e_{1t+jk(t)} + \\ D_{2,t+jk(t)} & e_{2,t+jk(t)} + D_{3,t+jk(t)} e_{3,t+jk(t)}. \end{aligned} \tag{23}$$

The three idiosyncratic noise terms $e_{1,t+jk(t)}, e_{2,t+jk(t)}$ and $e_{3,t+jk(t)}$ depend on the time of the day. The dummy variable $D_{1,t+jk(t)}$ equals 1 between 9.30 and 11.00 and otherwise zero, $D_{2,t+jk(t)}$ equals 1 between 11.00 and 14.30 and otherwise zero and $D_{3,t+jk(t)}$ equals 1 between 14.30 and 16.00 and otherwise zero.

On day $t$, consider we are between 9.30 and 11.00. Returns are then given as

$$\begin{aligned} r_{t+jk(t)} &= p_{t+jk(t)} - p_{t+(j-1)k(t)} \\ &= \tau^{\delta_1}_{t+jk(t)} \varepsilon_{t+jk(t)} + \alpha \tau^{\delta_2}_{t+jk(t)} \varepsilon_{t+jk(t)} + e_{1,t+jk(t)} \\ &- \alpha \tau^{\delta_2}_{t+(j-1)k(t)} \varepsilon_{t+(j-1)k(t)} - e_{1,t+(j-1)k(t)}. \end{aligned} \tag{24}$$

The variance of returns is given by

$$\begin{aligned} & E[\{r_{t+jk(t)}\}^2] = \\ & \sigma^2 (\tau^{2\delta_1}_{t+jk(t)} + 2\alpha \tau^{(\delta_1+\delta_2)}_{t+jk(t)} + \alpha^2 (\tau^{2\delta_2}_{t+jk(t)} + \tau^{2\delta_2}_{t+(j-1)k(t)})) + 2\omega_1^2, \end{aligned} \tag{25}$$

where $E[\varepsilon_{t+jk(t)}] = \sigma^2$ and $E[e_{1,t+jk(t)}] = \omega_1^2$. Consequently, first order autocorrelations are given as

$$\begin{aligned} & E[\{r_{t+jk(t)}\}\{r_{t+(j-1)k(t)}\}] = \\ & - \sigma^2 (\alpha \tau^{(\delta_1+\delta_2)}_{t+(j-1)k(t)} + \alpha^2 \tau^{2\delta_2}_{t+(j-1)k(t)}) - \omega_1^2. \end{aligned} \tag{26}$$

To show that $\alpha$ and $\omega_1^2$ are uniquely identified consider a linear translation on $\alpha$

$$\dot{\alpha} = \alpha + x. \tag{27}$$

Given this translation, the variance of returns is given by

$$
\begin{aligned}
&E[\{r_{t+jk(t)}\}^2] = \\
&\quad \sigma^2(\tau^{2\delta_1}_{t+jk(t)} + 2\dot{\alpha}\tau^{(\delta_1+\delta_2)}_{t+jk(t)} + \dot{\alpha}^2(\tau^{2\delta_2}_{t+jk(t)} + \tau^{2\delta_2}_{t+(j-1)k(t)})) + 2\dot{\omega}_1^2 \\
&\quad = \sigma^2(\tau^{2\delta_1}_{t+jk(t)} + 2\alpha\tau^{(\delta_1+\delta_2)}_{t+jk(t)} + \alpha^2(\tau^{2\delta_2}_{t+jk(t)} + \tau^{2\delta_2}_{t+(j-1)k(t)})) \\
&\quad + \sigma^2(2x\tau^{(\delta_1+\delta_2)}_{t+jk(t)} + 2\alpha x(\tau^{2\delta_2}_{t+jk(t)} + \tau^{2\delta_2}_{t+(j-1)k(t)}) + x^2(\tau^{2\delta_2}_{t+jk(t)} + \tau^{2\delta_2}_{t+(j-1)k(t)})) \\
&\quad + 2\dot{\omega}_1^2.
\end{aligned}
\tag{28}
$$

The first order autocorrelation is given by

$$
\begin{aligned}
&E[\{r_{t+jk(t)}\}\{r_{t+(j-1)k(t)}\}] = \\
&\quad -\sigma^2(\dot{\alpha}\tau^{(\delta_1+\delta_2)}_{t+(j-1)k(t)} + \dot{\alpha}^2\tau^{2\delta_2}_{t+(j-1)k(t)}) - \dot{\omega}_1^2 \\
&\quad = -\sigma^2(\alpha\tau^{(\delta_1+\delta_2)}_{t+(j-1)k(t)} + \alpha^2\tau^{2\delta_2}_{t+(j-1)k(t)}) \\
&\quad - \sigma^2(x\tau^{(\delta_1+\delta_2)}_{t+(j-1)k(t)} + 2\alpha x\tau^{2\delta_2}_{t+(j-1)k(t)} + x^2\tau^{2\delta_2}_{t+(j-1)k(t)}) - \dot{\omega}_1^2.
\end{aligned}
\tag{29}
$$

As there is no linear translation found on $\omega_1^2$ which solves for (28) and (29), the parameters $\alpha$ and $\omega_1^2$ are uniquely identified, as long as $\delta_2 \neq 0$ *and* $(\delta_1 + \delta_2) \neq 0$. A further necessity is that $Var(\tau_{t+jk(t)}) \neq 0$.

# References


Aït-Sahalia, Yacine, Per Mykland, and Lan Zhang, 2003, How often to sample a continuous-time process in the presence of market microstructure noise, Working Paper.

Andersen, Torben G., 2004, Discussion of power and bipower variation, *Journal of Financial Econometrics* 2, 37–48.

———, and Tim Bollerslev, 1998, Answering the skeptics: Yes, standard volatilit models do provide accurate forecasts, *International Economic Review* 39, 885–905.

———, Francis X. Diebold, and Paul Labys, 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 96, 42–55.

Andreou, Elena, and Eric Ghysels, 2002, Rolling-sample volatility estimators: Some new theoretical, simulation and empirical results, *Journal of Business and Economic Statistics* 20, 363–376.

Ané, Thierry, and Hélyette Geman, 2000, Order flow, transaction clock, and normality of asset returns, *Journal of Finance* 55, 2259–2284.

Bandi, Frederico M., and Jeffrey R. Russell, 2003, Microstructure noise, realized volatility and optimal sampling, Mimeo Working Paper, University of Chicago.

———, 2004, Separating microstructure noise from volatility, Mimeo Working Paper.

Barndorff-Nielsen, Ole E., and Neil Shephard, 2002, Estimating quadratic variation using realized volatility, *Journal of Applied Econometrics* 17, 457–477.

Durbin, J., and S. J. Koopman, 2001, *Time Series Analysis by State Space Methods* (Oxford University Press: Oxford).

Engle, Robert, 2000, The econometrics of ultra-high frequency data, *Econometrica* 68, 1–22.

Engle, Robert F., and Jeffrey R. Russell, 1997, Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model, *Journal of Empirical Finance* 4, 187–212.

Frijns, Bart P. M., 2004, The information content of trades and quotes, LIFE Working Paper, WP04-013.

——— , and Peter C. Schotman, 2004, Price discovery in tick time, CEPR Discussion Paper Nr. 4456.

Glosten, L. R., and L.E. Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123 – 142.

Hansen, Peter R., and Asger Lunde, 2004a, Realized variance and iid market microstructure noise, Working Paper, Brown University.

——— , 2004b, An unbiased measure of realized variance, Working Paper, Brown University.

Harvey, A., 1989, *Forecasting, structural time series models and the Kalman Filter* (Cambridge University Press: Cambridge).

Hasbrouck, Joel, 1993, Assessing the quality of a security market: A new approach to transaction-cost measurement, *Review of Financial Studies* 6, 191–212.

——— , and Thomas S. Y. Ho, 1987, Order arrival, quote behavior, and the return-generating process, *Journal of Finance* 42, 1035–1048.

Madhavan, Ananth, Matthew Richardson, and Mark Roomans, 1997, Why do security prices change? a transaction level analysis of NYSE stocks, *Review of Financial Studies* 10, 1035 – 1064.

Meddahi, Nour, 2002, A theoretical comparison between integated and realized volatility, *Journal of Applied Econometrics* 17, 479 – 508.

Oomen, Roel C. A., 2002, Modeling realized variance when returns are serially correlated, Working Paper, University of Warwick.

Revuz, Daniel, and Marc Yor, 1999, *Continuous Martingales and Brownian Motion* (Springer Verlag: Berlin, HeidelBerg, New York).

Wasserfallen, Walter, and Heinz Zimmermann, 1985, The behavior of intraday exchange rates, *Journal of Banking and Finance* 9, 55–72.

White, Halbert, 1982, Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1–25.

Zhang, Lan, Per Mykland, and Yacine Aït-Sahalia, 2003, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, Working Paper, University of Chicago.

Table 1: List of Company Ticker symbols and Company names

| Ticker Symbol | Company name | Total Trades | Mean(Obs) | Min | Max | Average(duration) | Max(duration) |
|---|---|---|---|---|---|---|---|
| AAPL | Apple Computer Inc. | 229,790 | 2933 | 1205 | 9919 | 48.42 | 549 |
| AMAT | Applied Materials Inc. | 651,363 | 6714 | 3893 | 12914 | 22.27 | 419 |
| AMGN | Amgen Inc. | 387,701 | 3815 | 1697 | 9061 | 37.43 | 579 |
| AMZN | Amazon.com, Inc. | 893,946 | 9793 | 4732 | 17145 | 16.23 | 182 |
| ATHM | At Home Corporation | 641,855 | 6760 | 2332 | 17134 | 22.60 | 604 |
| CMGI | CMGI Inc. | 787,951 | 8462 | 2588 | 17809 | 18.40 | 339 |
| COMS | 3Com Corporation | 529,780 | 5598 | 2296 | 16907 | 27.39 | 485 |
| CPWR | Compuware Corporation | 314,099 | 3044 | 1304 | 12596 | 46.20 | 543 |
| CSCO | Cisco Systems Inc. | 1,006,795 | 11233 | 6831 | 18241 | 14.41 | 225 |
| DELL | Dell Computer Corporation | 1,218,852 | 13842 | 7808 | 22121 | 11.90 | 170 |
| INTC | Intel Corp. | 1,024,335 | 11442 | 7066 | 17870 | 14.16 | 523 |
| MSFT | Microsoft Corp. | 1,099,617 | 12672 | 7637 | 19342 | 13.19 | 195 |
| NOVL | Novell Inc. | 298,136 | 2940 | 1371 | 11436 | 48.68 | 572 |
| NXTL | Nextel Communications CL-A | 331,468 | 3284 | 1257 | 11430 | 43.78 | 614 |
| ORCL | Oracle Corporation | 771,807 | 8215 | 4815 | 18463 | 18.80 | 511 |
| PSFT | Peoplesoft Inc. | 262,246 | 2596 | 1327 | 8194 | 55.35 | 591 |
| QWST | Qwest Communications Intl Inc. | 497,824 | 5096 | 1687 | 18939 | 29.13 | 592 |
| SBUX | Starbucks Corporation | 213,749 | 2051 | 639 | 18854 | 67.90 | 637 |
| SUNW | Sun Microsystems Inc. | 721,561 | 7743 | 4361 | 14459 | 20.11 | 475 |
| WCOM | MCI WorldCom Inc. | 698,576 | 7493 | 4551 | 17571 | 20.77 | 345 |

This table reports some summary statistics about the stocks used in this study. The first two columns report the ticker symbols and the names of the companies used in this study. The next column reports the total number of trades over the 124 trading days. The next two columns report the minimum and the maximum number of transactions that occur per day. The last two columns report the average duration and maximum duration between transactions in seconds.

Table 2: Realized Volatilities at different Sampling Frequencies

| | 1 sec | 10 sec | 30 sec | 60 sec | 120 sec | 300 sec | 600 sec | 1800 sec |
|---|---|---|---|---|---|---|---|---|
| AAPL | 6.15 *(1.34)* | 4.78 *(0.88)* | 3.86 *(0.65)* | 3.40 *(0.59)* | 3.11 *(0.63)* | 2.87 *(0.72)* | 2.73 *(0.77)* | 2.66 *(1.00)* |
| AMAT | 6.62 *(1.32)* | 4.47 *(0.83)* | 3.66 *(0.78)* | 3.50 *(0.84)* | 3.45 *(0.89)* | 3.36 *(0.98)* | 3.36 *(1.05)* | 3.37 *(1.19)* |
| AMGN | 5.98 *(1.44)* | 4.47 *(0.86)* | 3.47 *(0.66)* | 3.02 *(0.65)* | 2.81 *(0.70)* | 2.73 *(0.83)* | 2.60 *(0.92)* | 2.48 *(1.11)* |
| AMZN | 9.79 *(2.35)* | 6.02 *(1.32)* | 5.09 *(1.21)* | 4.98 *(1.27)* | 4.92 *(1.32)* | 4.95 *(1.43)* | 4.83 *(1.44)* | 4.60 *(1.55)* |
| ATHM | 10.33 *(3.03)* | 6.77 *(1.64)* | 5.41 *(1.46)* | 5.05 *(1.52)* | 4.93 *(1.63)* | 4.84 *(1.83)* | 4.70 *(1.90)* | 4.54 *(2.00)* |
| CMGI | 11.54 *(3.89)* | 7.31 *(2.24)* | 6.17 *(2.07)* | 5.96 *(2.16)* | 5.85 *(2.27)* | 5.83 *(2.49)* | 5.61 *(2.62)* | 5.30 *(2.92)* |
| COMS | 10.42 *(2.72)* | 7.15 *(1.18)* | 5.07 *(0.75)* | 4.11 *(0.76)* | 3.53 *(0.89)* | 3.10 *(1.18)* | 2.89 *(1.39)* | 2.76 *(1.65)* |
| CPWR | 9.75 *(3.40)* | 7.35 *(2.14)* | 5.74 *(1.48)* | 4.96 *(1.28)* | 4.44 *(1.29)* | 4.07 *(1.40)* | 3.89 *(1.58)* | 3.80 *(1.93)* |
| CSCO | 5.64 *(1.03)* | 3.26 *(0.44)* | 2.55 *(0.41)* | 2.42 *(0.47)* | 2.37 *(0.52)* | 2.33 *(0.60)* | 2.27 *(0.64)* | 2.23 *(0.81)* |
| DELL | 10.40 *(2.45)* | 5.43 *(1.16)* | 3.78 *(0.80)* | 3.33 *(0.71)* | 3.11 *(0.75)* | 2.99 *(0.93)* | 2.88 *(1.01)* | 2.78 *(1.12)* |
| INTC | 6.45 *(1.68)* | 3.63 *(0.78)* | 2.69 *(0.50)* | 2.49 *(0.47)* | 2.43 *(0.52)* | 2.38 *(0.59)* | 2.36 *(0.63)* | 2.28 *(0.84)* |
| MSFT | 5.23 *(1.17)* | 2.92 *(0.50)* | 2.24 *(0.40)* | 2.10 *(0.42)* | 2.07 *(0.48)* | 2.05 *(0.54)* | 1.99 *(0.58)* | 2.01 *(0.66)* |
| NOVL | 8.55 *(1.86)* | 6.65 *(1.06)* | 5.19 *(0.72)* | 4.34 *(0.64)* | 3.72 *(0.65)* | 3.26 *(0.75)* | 3.04 *(0.81)* | 2.77 *(0.90)* |
| NXTL | 7.11 *(2.06)* | 5.39 *(1.24)* | 4.20 *(0.93)* | 3.62 *(0.88)* | 3.21 *(0.96)* | 2.92 *(1.02)* | 2.77 *(1.10)* | 2.56 *(1.18)* |
| ORCL | 11.04 *(3.25)* | 6.69 *(1.46)* | 4.60 *(0.88)* | 3.89 *(0.81)* | 3.51 *(0.85)* | 3.22 *(1.01)* | 3.08 *(1.10)* | 3.03 *(1.37)* |
| PSFT | 12.29 *(3.06)* | 9.73 *(1.97)* | 7.59 *(1.36)* | 6.22 *(1.09)* | 5.16 *(1.09)* | 4.16 *(1.15)* | 3.69 *(1.20)* | 3.25 *(1.35)* |
| QWST | 8.37 *(2.64)* | 5.84 *(1.31)* | 4.41 *(1.00)* | 3.89 *(1.05)* | 3.65 *(1.20)* | 3.43 *(1.35)* | 3.28 *(1.38)* | 3.26 *(1.62)* |
| SBUX | 7.13 *(2.43)* | 5.77 *(1.50)* | 4.80 *(1.13)* | 4.16 *(0.96)* | 3.67 *(0.90)* | 3.21 *(0.96)* | 2.95 *(1.03)* | 2.77 *(1.14)* |
| SUNW | 6.30 *(1.71)* | 4.13 *(0.91)* | 3.23 *(0.72)* | 3.05 *(0.74)* | 2.98 *(0.80)* | 2.90 *(0.91)* | 2.84 *(1.00)* | 2.92 *(1.24)* |
| WCOM | 4.82 *(0.84)* | 3.16 *(0.49)* | 2.33 *(0.43)* | 2.11 *(0.48)* | 2.05 *(0.54)* | 1.95 *(0.58)* | 1.88 *(0.62)* | 1.81 *(0.71)* |

*Note:* This Table reports average uncorrected realized volatilities ($\overline{rv}_C^{(h)}$) sampled at different calendar time frequencies $h$. These realized volatilities are computed as

$$\overline{rv}_C^{(h)} = \frac{1}{N} \sum_{t=1}^{N} \sqrt{\sum_{i=1}^{h^{-1}} \{r_{C,t-1+ih}^{(h)}\}^2},$$

where $N = 124$ are the total number of trading days. Standard deviations of $\overline{rv}_C^{(h)}$ over the 124 days are reported in brackets.

Table 3: Parameter Estimates sampled at 1-tick

| Stock | $\delta_1$ | $\delta_2$ | $\alpha$ | $\sigma$ | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|---|---|---|---|
| AAPL | 0.10 *(0.01)* | -0.03 *(0.01)* | 2.58 *(0.04)* | 4.90 *(0.03)* | 4.91 *(0.10)* | 0.00 *(0.00)* | 2.68 *(0.16)* |
| AMAT | 0.08 *(0.01)* | -0.06 *(0.01)* | 2.34 *(0.03)* | 3.61 *(0.01)* | 3.78 *(0.06)* | 0.38 *(0.21)* | 2.16 *(0.08)* |
| AMGN | 0.05 *(0.01)* | 0.03 *(0.05)* | 2.37 *(1.46)* | 3.82 *(0.02)* | 5.72 *(1.86)* | 3.36 *(3.17)* | 4.11 *(2.59)* |
| AMZN | 0.00 *(0.01)* | 0.04 *(0.01)* | 1.83 *(0.10)* | 4.52 *(0.02)* | 6.98 *(0.16)* | 1.50 *(0.72)* | 4.70 *(0.23)* |
| ATHM | 0.08 *(0.01)* | -0.06 *(0.01)* | 1.80 *(1.09)* | 5.54 *(0.02)* | 8.88 *(1.87)* | 3.92 *(4.26)* | 6.13 *(2.71)* |
| CMGI | 0.11 *(0.01)* | -0.08 *(0.01)* | 1.79 *(0.14)* | 6.13 *(0.02)* | 8.62 *(0.29)* | 1.27 *(1.91)* | 5.08 *(0.48)* |
| COMS | -0.07 *(0.01)* | 0.28 *(0.04)* | 1.96 *(0.39)* | 3.68 *(0.02)* | 11.00 *(0.24)* | 9.67 *(0.27)* | 10.08 *(0.25)* |
| CPWR | 0.00 *(0.01)* | 0.06 *(0.03)* | 1.94 *(0.74)* | 6.72 *(0.06)* | 12.62 *(1.30)* | 8.50 *(1.94)* | 9.24 *(1.81)* |
| CSCO | 0.03 *(0.01)* | 0.02 *(0.01)* | 1.87 *(0.33)* | 2.04 *(0.01)* | 3.73 *(0.18)* | 2.85 *(0.24)* | 3.12 *(0.22)* |
| DELL | -0.17 *(0.03)* | 0.36 *(0.05)* | 1.85 *(0.45)* | 2.25 *(0.02)* | 8.17 *(0.14)* | 6.23 *(0.17)* | 6.50 *(0.16)* |
| INTC | 0.04 *(0.01)* | 0.02 *(0.01)* | 1.82 *(0.08)* | 2.01 *(0.01)* | 4.72 *(0.04)* | 3.81 *(0.04)* | 4.07 *(0.04)* |
| MSFT | 0.03 *(0.01)* | 0.02 *(0.01)* | 1.83 *(0.41)* | 1.66 *(0.01)* | 3.61 *(0.15)* | 2.77 *(0.20)* | 3.08 *(0.18)* |
| NOVL | 0.16 *(0.01)* | -0.04 *(0.06)* | 2.02 *(0.87)* | 5.88 *(0.05)* | 9.71 *(1.55)* | 7.93 *(1.90)* | 8.78 *(1.71)* |
| NXTL | -0.01 *(0.01)* | 0.12 *(0.09)* | 2.02 *(1.50)* | 4.51 *(0.03)* | 8.37 *(1.81)* | 6.28 *(2.40)* | 7.33 *(2.06)* |
| ORCL | -0.06 *(0.01)* | 0.13 *(0.02)* | 2.02 *(0.50)* | 3.16 *(0.02)* | 9.77 *(0.26)* | 8.60 *(0.29)* | 8.81 *(0.29)* |
| PSFT | 0.05 *(0.01)* | 0.00 *(0.01)* | 5.42 *(0.14)* | 7.86 *(0.08)* | 8.91 *(0.26)* | 0.00 *(0.00)* | 2.25 *(0.94)* |
| QWST | 0.04 *(0.01)* | -0.03 *(0.01)* | 2.01 *(0.04)* | 4.39 *(0.02)* | 8.42 *(0.09)* | 5.77 *(0.11)* | 6.51 *(0.11)* |
| SBUX | 0.12 *(0.01)* | -0.06 *(0.02)* | 2.01 *(0.57)* | 6.71 *(0.06)* | 9.49 *(1.38)* | 5.85 *(2.24)* | 7.50 *(1.72)* |
| SUNW | 0.04 *(0.01)* | -0.01 *(0.01)* | 2.01 *(0.12)* | 2.99 *(0.01)* | 4.54 *(0.12)* | 3.09 *(0.18)* | 3.64 *(0.15)* |
| WCOM | 0.10 *(0.01)* | -0.05 *(0.04)* | 2.03 *(1.72)* | 2.13 *(0.01)* | 3.64 *(1.07)* | 2.76 *(1.41)* | 3.07 *(1.26)* |

This table reports the parameter estimates with standard errors in brackets, of the model

$$\begin{aligned} p_{t+jk(t)} &= p^*_{t+jk(t)} + u_{t+jk(t)}, \\ p^*_{t+jk(t)} &= p^*_{t+(j-1)k(t)} + \tau^{\delta_1}_{t+jk(t)}\varepsilon_{t+jk(t)}, \\ u_{t+jk(t)} &= \alpha\tau^{\delta_2}_{t+jk(t)}\varepsilon_{t+jk(t)} + e_{1,t+jk(t)} + e_{2,t+jk(t)} + e_{3,t+jk(t)}. \end{aligned}$$

where $Var(\varepsilon_{t+jk(t)}) = \sigma^2$ and $Var(e_{1,t+jk(t)}) = \omega_1^2$, $Var(e_{2,t+jk(t)}) = \omega_2^2$, $Var(e_{3,t+jk(t)}) = \omega_3^2$. The sampling frequency is at the 1 tick level. The first column reports the estimates for the duration parameter on the random walk innovation $\delta_1$. The second column reports the parameter estimates for the time impact on the correlation between the efficient price and the microstructure noise. The next column reports the $\alpha$ in the model which measures the correlation between efficient price innovation and microstructure noise at average durations. The next column reports the volatility of the efficient price for a single transaction at average durations. The last three columns report the standard deviation of the microstructure noise, where $\omega_1$ measures the noise near the opening (9.30 - 11.00), $\omega_2$ in the middle of the trading day (11.00 - 14.30) and $\omega_3$ near the close of the market (14.30 - 16.00). The standard deviations in the last 4 columns are multiplied by 100.

Table 4: Parameter Estimates sampled at 2-tick

| Stock | $\delta_1$ | $\delta_2$ | $\alpha$ | $\sigma$ | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|---|---|---|---|
| AAPL | 0.03 *(0.01)* | 0.10 *(0.01)* | 1.32 *(0.03)* | 6.70 *(0.04)* | 4.95 *(0.14)* | 0.00 *(0.01)* | 2.91 *(0.21)* |
| AMAT | 0.01 *(0.01)* | 0.01 *(0.01)* | 1.02 *(0.01)* | 5.22 *(0.02)* | 3.74 *(0.07)* | 0.37 *(0.20)* | 2.28 *(0.11)* |
| AMGN | -0.03 *(0.01)* | 0.17 *(0.13)* | 1.07 *(0.98)* | 5.32 *(0.03)* | 5.74 *(2.39)* | 3.67 *(3.76)* | 4.46 *(3.09)* |
| AMZN | -0.04 *(0.01)* | 0.07 *(0.02)* | 0.78 *(0.01)* | 6.66 *(0.03)* | 6.75 *(0.07)* | 0.11 *(0.04)* | 4.42 *(0.08)* |
| ATHM | 0.03 *(0.01)* | -0.03 *(0.02)* | 0.77 *(0.18)* | 7.97 *(0.04)* | 8.76 *(0.62)* | 3.98 *(1.37)* | 6.11 *(0.88)* |
| CMGI | 0.07 *(0.01)* | -0.07 *(0.03)* | 0.72 *(0.60)* | 8.95 *(0.04)* | 8.50 *(2.81)* | 1.64 *(14.63)* | 5.51 *(4.34)* |
| COMS | -0.19 *(0.01)* | 0.51 *(0.03)* | 1.01 *(0.11)* | 4.93 *(0.03)* | 11.06 *(0.13)* | 9.81 *(0.14)* | 10.18 *(0.13)* |
| CPWR | -0.07 *(0.01)* | 0.16 *(0.02)* | 0.97 *(0.25)* | 9.02 *(0.07)* | 12.86 *(0.79)* | 8.88 *(1.15)* | 9.49 *(1.07)* |
| CSCO | -0.06 *(0.01)* | 0.17 *(0.11)* | 0.78 *(0.74)* | 2.96 *(0.01)* | 3.67 *(0.88)* | 2.89 *(1.12)* | 3.15 *(1.02)* |
| DELL | -0.28 *(0.02)* | 0.77 *(0.05)* | 0.81 *(0.16)* | 3.20 *(0.02)* | 8.06 *(0.13)* | 6.21 *(0.13)* | 6.59 *(0.13)* |
| INTC | -0.07 *(0.01)* | 0.32 *(0.02)* | 0.65 *(0.06)* | 2.90 *(0.01)* | 4.80 *(0.05)* | 3.97 *(0.06)* | 4.20 *(0.06)* |
| MSFT | -0.08 *(0.01)* | 0.23 *(0.02)* | 0.62 *(0.13)* | 2.39 *(0.01)* | 3.68 *(0.10)* | 2.92 *(0.13)* | 3.26 *(0.12)* |
| NOVL | 0.06 *(0.01)* | 0.05 *(0.01)* | 2.29 *(0.05)* | 7.66 *(0.05)* | 5.41 *(0.23)* | 0.00 *(0.00)* | 3.88 *(0.29)* |
| NXTL | -0.11 *(0.01)* | 0.21 *(0.04)* | 2.15 *(0.76)* | 6.02 *(0.05)* | 5.25 *(2.58)* | 1.76 *(7.67)* | 3.98 *(3.40)* |
| ORCL | -0.17 *(0.01)* | 0.21 *(0.02)* | 2.18 *(0.22)* | 4.39 *(0.02)* | 8.24 *(0.27)* | 7.18 *(0.31)* | 7.43 *(0.30)* |
| PSFT | -0.08 *(0.01)* | 0.17 *(0.01)* | 3.66 *(0.11)* | 9.73 *(0.11)* | 8.71 *(0.39)* | 0.00 *(0.00)* | 2.17 *(1.45)* |
| QWST | -0.01 *(0.01)* | 0.00 *(0.01)* | 1.68 *(0.51)* | 6.23 *(0.03)* | 6.20 *(1.65)* | 1.68 *(6.00)* | 3.68 *(2.76)* |
| SBUX | 0.05 *(0.01)* | 0.02 *(0.01)* | 1.62 *(0.04)* | 8.83 *(0.07)* | 7.33 *(0.26)* | 0.15 *(0.14)* | 4.40 *(0.34)* |
| SUNW | -0.04 *(0.01)* | 0.08 *(0.01)* | 1.32 *(0.76)* | 4.30 *(0.02)* | 3.39 *(2.07)* | 1.37 *(5.15)* | 2.33 *(3.01)* |
| WCOM | -0.01 *(0.01)* | 0.11 *(0.07)* | 1.36 *(0.88)* | 3.02 *(0.01)* | 3.06 *(1.30)* | 1.87 *(2.13)* | 2.36 *(1.69)* |

*Note:* This table reports the parameter estimates with standard errors in brackets, of the model

$$\begin{aligned} p_{t+jk(t)} &= p^*_{t+jk(t)} + u_{t+jk(t)}, \\ p^*_{t+jk(t)} &= p^*_{t+(j-1)k(t)} + \tau^{\delta_1}_{t+jk(t)}\varepsilon_{t+jk(t)}, \\ u_{t+jk(t)} &= \alpha\tau^{\delta_2}_{t+jk(t)}\varepsilon_{t+jk(t)} + e_{1,t+jk(t)} + e_{2,t+jk(t)} + e_{3,t+jk(t)}. \end{aligned}$$

where $Var(\varepsilon_{t+jk(t)}) = \sigma^2$ and $Var(e_{1,t+jk(t)}) = \omega_1^2$, $Var(e_{2,t+jk(t)}) = \omega_2^2$, $Var(e_{3,t+jk(t)}) = \omega_3^2$. The sampling frequency is at the 2 tick level. The first column reports the estimates for the duration parameter on the random walk innovation $\delta_1$. The second column reports the parameter estimates for the time impact on the correlation between the efficient price and the microstructure noise. The next column reports the $\alpha$ in the model which measures the correlation between efficient price innovation and microstructure noise at average durations. The next column reports the volatility of the efficient price for a single transaction at average durations. The last three columns report the standard deviation of the microstructure noise, where $\omega_1$ measures the noise near the opening (9.30 - 11.00), $\omega_2$ in the middle of the trading day (11.00 - 14.30) and $\omega_3$ near the close of the market (14.30 - 16.00). The standard deviations in the last 4 columns are multiplied by 100.

Table 5: Parameter Estimates sampled at 5-tick

| Stock | $\delta_1$ | $\delta_2$ | $\alpha$ | $\sigma$ | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|---|---|---|---|
| AAPL | -0.06 *(0.01)* | 0.23 *(0.02)* | 0.49 *(0.02)* | 10.51 *(0.09)* | 3.97 *(0.35)* | 0.00 *(0.00)* | 2.41 *(0.47)* |
| AMAT | -0.08 *(0.01)* | 0.22 *(0.03)* | 0.27 *(0.01)* | 8.68 *(0.04)* | 3.18 *(0.15)* | 0.01 *(0.00)* | 2.12 *(0.19)* |
| AMGN | -0.13 *(0.01)* | 0.32 *(0.04)* | 0.42 *(0.12)* | 8.57 *(0.07)* | 4.88 *(0.89)* | 2.89 *(1.52)* | 3.71 *(1.19)* |
| AMZN | -0.11 *(0.01)* | 0.32 *(0.06)* | 0.15 *(0.01)* | 11.36 *(0.06)* | 6.76 *(0.13)* | 0.00 *(0.00)* | 5.04 *(0.14)* |
| ATHM | -0.07 *(0.01)* | 1.69 *(0.14)* | 0.01 *(0.00)* | 13.16 *(0.08)* | 10.09 *(0.16)* | 6.57 *(0.12)* | 8.33 *(0.14)* |
| CMGI | -0.02 *(0.01)* | 1.35 *(0.17)* | 0.02 *(0.01)* | 15.06 *(0.08)* | 9.76 *(0.17)* | 5.12 *(0.19)* | 7.74 *(0.17)* |
| COMS | -0.32 *(0.01)* | 0.45 *(0.02)* | 2.10 *(0.05)* | 7.62 *(0.06)* | 4.58 *(0.28)* | 0.09 *(0.02)* | 2.31 *(0.47)* |
| CPWR | -0.16 *(0.01)* | 0.20 *(0.02)* | 0.76 *(0.03)* | 13.92 *(0.14)* | 8.56 *(0.48)* | 1.27 *(1.25)* | 4.32 *(0.74)* |
| CSCO | -0.19 *(0.01)* | 0.36 *(0.03)* | 0.50 *(0.02)* | 4.93 *(0.02)* | 2.22 *(0.11)* | 0.55 *(0.34)* | 1.54 *(0.15)* |
| DELL | -0.40 *(0.02)* | 0.66 *(0.03)* | 1.56 *(0.02)* | 5.30 *(0.03)* | 4.95 *(0.20)* | 0.00 *(0.00)* | 2.11 *(0.14)* |
| INTC | -0.19 *(0.01)* | 0.39 *(0.04)* | 0.73 *(0.12)* | 4.89 *(0.02)* | 2.65 *(0.54)* | 0.83 *(1.69)* | 1.74 *(0.82)* |
| MSFT | -0.21 *(0.01)* | 0.37 *(0.03)* | 0.63 *(0.05)* | 4.02 *(0.02)* | 2.16 *(0.19)* | 0.44 *(0.85)* | 1.44 *(0.27)* |
| NOVL | -0.05 *(0.01)* | 0.23 *(0.02)* | 0.99 *(0.03)* | 11.56 *(0.10)* | 4.52 *(0.54)* | 0.00 *(0.00)* | 3.50 *(0.58)* |
| NXTL | -0.23 *(0.01)* | 0.36 *(0.07)* | 0.84 *(0.43)* | 9.40 *(0.09)* | 3.90 *(4.91)* | 1.46 *(13.17)* | 3.89 *(4.95)* |
| ORCL | -0.30 *(0.01)* | 0.38 *(0.01)* | 0.85 *(0.26)* | 7.08 *(0.04)* | 7.53 *(0.87)* | 6.96 *(0.96)* | 7.12 *(0.93)* |
| PSFT | -0.22 *(0.01)* | 0.34 *(0.02)* | 1.78 *(0.07)* | 14.06 *(0.18)* | 6.10 *(0.97)* | 0.00 *(0.00)* | 1.05 *(4.80)* |
| QWST | -0.08 *(0.01)* | 1.29 *(2.54)* | 0.00 *(0.02)* | 10.14 *(0.07)* | 9.46 *(0.14)* | 7.73 *(0.14)* | 8.33 *(0.14)* |
| SBUX | 0.02 *(0.01)* | 1.28 *(0.94)* | 0.00 *(0.00)* | 13.82 *(0.16)* | 12.50 *(0.28)* | 10.41 *(0.16)* | 11.61 *(0.22)* |
| SUNW | -0.15 *(0.01)* | -0.04 *(0.69)* | -0.17 *(0.60)* | 7.19 *(0.04)* | 6.22 *(2.49)* | 5.44 *(2.85)* | 5.77 *(2.69)* |
| WCOM | -0.15 *(0.01)* | -0.19 *(0.24)* | -0.35 *(0.26)* | 4.98 *(0.03)* | 5.18 *(0.62)* | 4.73 *(0.68)* | 4.96 *(0.65)* |

*Note:* This table reports the parameter estimates with standard errors in brackets, of the model

$$
\begin{aligned}
p_{t+jk(t)} &= p^*_{t+jk(t)} + u_{t+jk(t)}, \\
p^*_{t+jk(t)} &= p^*_{t+(j-1)k(t)} + \tau^{\delta_1}_{t+jk(t)}\varepsilon_{t+jk(t)}, \\
u_{t+jk(t)} &= \alpha\tau^{\delta_2}_{t+jk(t)}\varepsilon_{t+jk(t)} + e_{1,t+jk(t)} + e_{2,t+jk(t)} + e_{3,t+jk(t)}.
\end{aligned}
$$

where $Var(\varepsilon_{t+jk(t)}) = \sigma^2$ and $Var(e_{1,t+jk(t)}) = \omega_1^2$, $Var(e_{2,t+jk(t)}) = \omega_2^2$, $Var(e_{3,t+jk(t)}) = \omega_3^2$. The sampling frequency is at the 5 tick level. The first column reports the estimates for the duration parameter on the random walk innovation $\delta_1$. The second column reports the parameter estimates for the time impact on the correlation between the efficient price and the microstructure noise. The next column reports the $\alpha$ in the model which measures the correlation between efficient price innovation and microstructure noise at average durations. The next column reports the volatility of the efficient price for a single transaction at average durations. The last three columns report the standard deviation of the microstructure noise, where $\omega_1$ measures the noise near the opening (9.30 - 11.00), $\omega_2$ in the middle of the trading day (11.00 - 14.30) and $\omega_3$ near the close of the market (14.30 - 16.00). The standard deviations in the last 4 columns are multiplied by 100.

Table 6: Realized volatility using the Kalman Filter

| | Panel A | | | Panel B |
|---|---|---|---|---|
| | KF 1 tick | KF 2 ticks | KF 5 ticks | Daily KF 1 tick |
| AAPL | 2.25 *(0.37)* | 2.25 *(0.43)* | 2.36 *(0.52)* | 2.21 *(0.44)* |
| AMAT | 2.53 *(0.26)* | 2.64 *(0.30)* | 2.90 *(0.38)* | 2.49 *(0.46)* |
| AMGN | 2.07 *(0.31)* | 2.11 *(0.36)* | 2.28 *(0.44)* | 2.02 *(0.44)* |
| AMZN | 3.82 *(0.49)* | 4.02 *(0.54)* | 4.46 *(0.66)* | 3.71 *(0.89)* |
| ATHM | 3.80 *(0.69)* | 3.93 *(0.78)* | 4.27 *(0.98)* | 3.70 *(1.08)* |
| CMGI | 4.65 *(0.79)* | 4.84 *(0.89)* | 5.29 *(1.10)* | 4.36 *(1.57)* |
| COMS | 2.44 *(0.57)* | 2.48 *(0.66)* | 2.70 *(0.80)* | 2.38 *(0.74)* |
| CPWR | 3.31 *(0.64)* | 3.30 *(0.71)* | 3.49 *(0.84)* | 3.23 *(0.92)* |
| CSCO | 1.82 *(0.15)* | 1.91 *(0.18)* | 2.11 *(0.23)* | 1.78 *(0.34)* |
| DELL | 2.33 *(0.29)* | 2.44 *(0.33)* | 2.69 *(0.40)* | 2.19 *(0.50)* |
| INTC | 1.81 *(0.15)* | 1.89 *(0.19)* | 2.11 *(0.24)* | 1.75 *(0.35)* |
| MSFT | 1.55 *(0.12)* | 1.62 *(0.14)* | 1.81 *(0.18)* | 1.52 *(0.29)* |
| NOVL | 2.65 *(0.38)* | 2.53 *(0.45)* | 2.57 *(0.55)* | 2.64 *(0.48)* |
| NXTL | 2.29 *(0.48)* | 2.31 *(0.55)* | 2.55 *(0.69)* | 2.25 *(0.65)* |
| ORCL | 2.53 *(0.36)* | 2.61 *(0.42)* | 2.88 *(0.52)* | 2.45 *(0.60)* |
| PSFT | 3.46 *(0.52)* | 3.30 *(0.61)* | 3.49 *(0.76)* | 3.37 *(0.93)* |
| QWST | 2.67 *(0.58)* | 2.73 *(0.65)* | 2.92 *(0.77)* | 2.58 *(0.78)* |
| SBUX | 2.51 *(0.60)* | 2.40 *(0.68)* | 2.42 *(0.73)* | 2.43 *(0.65)* |
| SUNW | 2.24 *(0.24)* | 2.35 *(0.29)* | 2.61 *(0.36)* | 2.19 *(0.47)* |
| WCOM | 1.55 *(0.13)* | 1.60 *(0.16)* | 1.77 *(0.22)* | 1.52 *(0.37)* |

*Note:* This table reports the Realized volatility for different aggregation levels. Realized volatilities are based on the parameter estimates of equation (17) and are computed as

$$rv^{(a)}_{T,p^*,t} = \sqrt{\sum_{j=1}^{k(t)^{-1}} \tau^{2\delta_1}_{t+jak(t)} \sigma^{(a),2}}, \; rv^{(a)}_{p^*,t} = \sqrt{\sum_{j=1}^{k(t)^{-1}} \tau^{2\delta_{1,t}}_{t+jak(t)} \sigma^{(a),2}_t},$$

for the full sample and daily calculations, respectively. We report the average values of $rv$,

$$\overline{rv}^{(a)}_{T,p^*} = \frac{1}{N} \sum_{t=1}^{N} rv^{(a)}_{T,p^*,t},$$

where $N = 124$ is the total number of trading days. Standard deviations of $rv$ over the trading days are reported in brackets. Panel A reports the results where parameters are estimated over the full sample, panel B reports the results for the daily estimates.

Table 7: Realized Volatility: Calendar time corrected

| | q = 1 | q = 5 | q = 10 | q = 15 | q = 30 | q = 45 | q = 60 | q = 90 | q = 120 | q = 300 |
|---|---|---|---|---|---|---|---|---|---|---|
| AAPL | 5.79 *(1.16)* | 4.54 *(0.79)* | 3.78 *(0.63)* | 3.37 *(0.54)* | 2.92 *(0.51)* | 2.78 *(0.58)* | 2.71 *(0.65)* | 2.69 *(0.71)* | 2.72 *(0.73)* | 2.62 *(0.80)* |
| AMAT | 5.85 *(1.04)* | 3.71 *(0.49)* | 3.06 *(0.46)* | 2.93 *(0.49)* | 3.09 *(0.65)* | 3.25 *(0.73)* | 3.30 *(0.78)* | 3.35 *(0.85)* | 3.35 *(0.85)* | 3.35 *(1.02)* |
| AMGN | 5.60 *(1.26)* | 4.19 *(0.71)* | 3.31 *(0.55)* | 2.85 *(0.51)* | 2.45 *(0.55)* | 2.45 *(0.67)* | 2.45 *(0.72)* | 2.55 *(0.78)* | 2.59 *(0.83)* | 2.56 *(0.93)* |
| AMZN | 8.26 *(1.81)* | 4.70 *(0.98)* | 4.24 *(0.95)* | 4.28 *(1.08)* | 4.65 *(1.27)* | 4.80 *(1.31)* | 4.83 *(1.34)* | 4.91 *(1.39)* | 4.98 *(1.46)* | 4.90 *(1.55)* |
| ATHM | 9.15 *(2.37)* | 5.71 *(1.28)* | 4.68 *(1.25)* | 4.39 *(1.24)* | 4.50 *(1.49)* | 4.65 *(1.57)* | 4.75 *(1.69)* | 4.82 *(1.83)* | 4.82 *(1.85)* | 4.65 *(2.02)* |
| CMGI | 9.97 *(3.13)* | 6.08 *(1.89)* | 5.28 *(1.73)* | 5.24 *(1.83)* | 5.57 *(2.11)* | 5.75 *(2.26)* | 5.74 *(2.29)* | 5.74 *(2.32)* | 5.83 *(2.41)* | 5.62 *(2.63)* |
| COMS | 9.47 *(2.05)* | 6.35 *(0.91)* | 4.54 *(0.62)* | 3.68 *(0.58)* | 2.84 *(0.95)* | 2.79 *(1.01)* | 2.75 *(1.12)* | 2.79 *(1.21)* | 2.85 *(1.29)* | 2.76 *(1.33)* |
| CPWR | 9.12 *(2.98)* | 6.98 *(1.84)* | 5.62 *(1.25)* | 4.87 *(1.03)* | 4.10 *(0.93)* | 3.91 *(1.04)* | 3.86 *(1.18)* | 3.86 *(1.25)* | 3.87 *(1.41)* | 3.77 *(1.56)* |
| CSCO | 4.74 *(0.75)* | 2.39 *(0.31)* | 1.93 *(0.39)* | 1.94 *(0.42)* | 2.16 *(0.52)* | 2.25 *(0.57)* | 2.28 *(0.56)* | 2.30 *(0.58)* | 2.31 *(0.61)* | 2.24 *(0.65)* |
| DELL | 8.37 *(1.93)* | 3.36 *(0.82)* | 2.36 *(0.68)* | 2.28 *(0.74)* | 2.57 *(0.84)* | 2.75 *(0.93)* | 2.76 *(0.82)* | 2.82 *(0.83)* | 2.88 *(0.88)* | 2.86 *(1.02)* |
| INTC | 5.39 *(1.28)* | 2.54 *(0.54)* | 1.93 *(0.39)* | 1.88 *(0.44)* | 2.08 *(0.51)* | 2.23 *(0.55)* | 2.30 *(0.57)* | 2.33 *(0.61)* | 2.36 *(0.63)* | 2.33 *(0.68)* |
| MSFT | 4.31 *(0.86)* | 1.97 *(0.30)* | 1.60 *(0.28)* | 1.61 *(0.34)* | 1.85 *(0.42)* | 1.94 *(0.45)* | 1.95 *(0.48)* | 1.98 *(0.49)* | 2.01 *(0.51)* | 2.03 *(0.61)* |
| NOVL | 8.08 *(1.60)* | 6.38 *(0.91)* | 5.20 *(0.75)* | 4.49 *(0.72)* | 3.50 *(0.68)* | 3.15 *(0.60)* | 3.03 *(0.67)* | 2.96 *(0.77)* | 2.92 *(0.84)* | 2.86 *(0.86)* |
| NXTL | 6.66 *(1.77)* | 5.07 *(1.10)* | 4.10 *(0.87)* | 3.55 *(0.81)* | 3.01 *(0.79)* | 2.81 *(0.84)* | 2.72 *(0.92)* | 2.68 *(1.00)* | 2.72 *(1.10)* | 2.65 *(1.17)* |
| ORCL | 9.61 *(2.50)* | 5.22 *(1.12)* | 3.49 *(0.66)* | 2.83 *(0.71)* | 2.69 *(0.85)* | 2.91 *(0.91)* | 3.00 *(0.95)* | 3.01 *(1.03)* | 3.06 *(1.06)* | 2.89 *(1.14)* |
| PSFT | 11.70 *(2.72)* | 9.51 *(1.81)* | 7.81 *(1.27)* | 6.77 *(1.12)* | 5.06 *(1.05)* | 4.33 *(1.03)* | 3.93 *(1.15)* | 3.73 *(1.31)* | 3.60 *(1.47)* | 3.40 *(1.44)* |
| QWST | 7.59 *(2.08)* | 5.12 *(0.99)* | 3.87 *(0.78)* | 3.39 *(0.88)* | 3.07 *(1.02)* | 3.13 *(1.18)* | 3.20 *(1.22)* | 3.27 *(1.29)* | 3.31 *(1.37)* | 3.19 *(1.46)* |
| SBUX | 6.77 *(2.01)* | 5.60 *(1.39)* | 4.85 *(1.13)* | 4.38 *(1.01)* | 3.65 *(0.92)* | 3.33 *(0.86)* | 3.12 *(0.89)* | 3.01 *(0.98)* | 2.94 *(0.94)* | 2.79 *(1.13)* |
| SUNW | 5.56 *(1.33)* | 3.37 *(0.56)* | 2.63 *(0.45)* | 2.51 *(0.49)* | 2.68 *(0.65)* | 2.76 *(0.69)* | 2.84 *(0.79)* | 2.85 *(0.80)* | 2.83 *(0.85)* | 2.80 *(1.00)* |
| WCOM | 4.32 *(0.66)* | 2.65 *(0.44)* | 1.90 *(0.42)* | 1.72 *(0.40)* | 1.73 *(0.46)* | 1.83 *(0.51)* | 1.90 *(0.57)* | 1.93 *(0.55)* | 1.96 *(0.57)* | 1.87 *(0.66)* |

*Note:* This Table reports average daily realized volatilities ($\overline{rv}_{C,q}^{(1\,sec)}$) sampled at a 1 second sampling frequency, corrected for different lags lengths ($q$),

$$\overline{rv}_{C,q}^{(1\,sec)} = \frac{1}{N}\sum_{t=1}^{N}\sqrt{\sum_{i=1}^{h^{-1}}\{r_{C,t-1+ih}^{(1\,sec)}\}^2 + 2\sum_{s=1}^{q}\frac{h^{-1}}{h^{-1}-s}\sum_{i=1}^{h^{-1}-s} r_{C,t-1+ih}^{(1\,sec)} r_{C,t-1+(i+s)h}^{(1\,sec)}},$$

where $N = 124$ are the total number of trading days. Standard deviations of $\overline{rv}_{C,q}^{(1\,sec)}$ over the 124 trading days are reported in brackets.

Table 8: Realized Volatilities: Tick time corrected

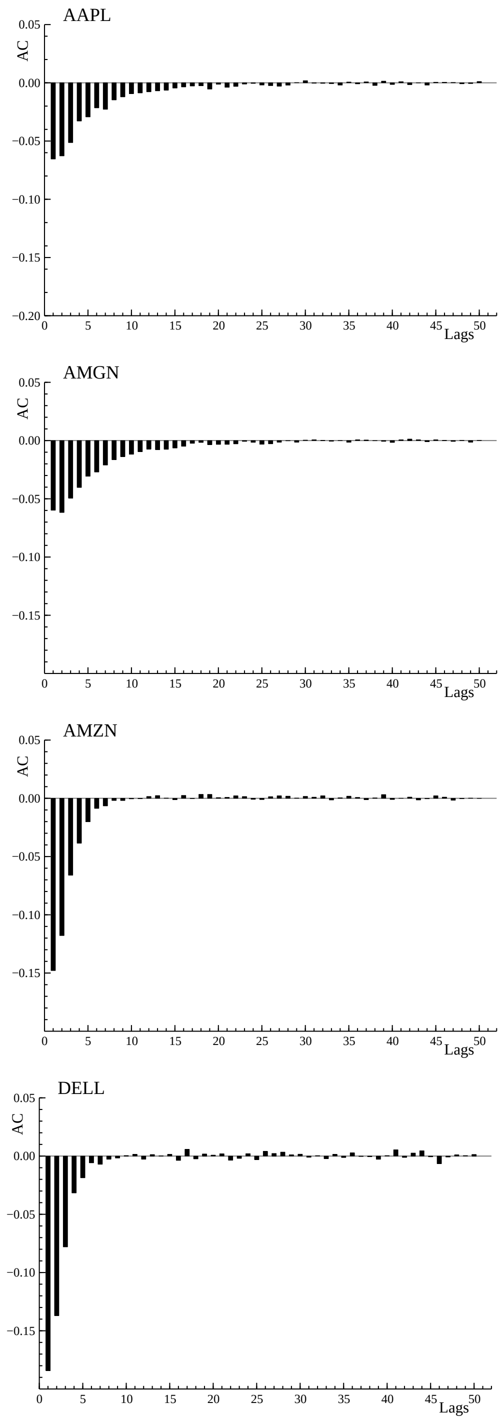
| | q = 1 | q = 5 | q = 10 | q = 20 | q = 30 | q = 40 | q = 50 | q = 60 | q = 70 | q = 80 | q = 90 | q = 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AAPL | 2.17 *(0.56)* | 2.21 *(0.54)* | 2.42 *(0.66)* | 2.58 *(0.71)* | 2.68 *(0.75)* | 2.66 *(0.79)* | 2.61 *(0.78)* | 2.65 *(0.82)* | 2.67 *(0.83)* | 2.63 *(0.83)* | 2.67 *(0.90)* | 2.67 *(0.94)* |
| AMAT | 2.35 *(0.47)* | 2.61 *(0.56)* | 2.96 *(0.67)* | 3.23 *(0.76)* | 3.36 *(0.84)* | 3.35 *(0.87)* | 3.34 *(0.89)* | 3.32 *(0.92)* | 3.30 *(0.92)* | 3.34 *(0.93)* | 3.36 *(0.96)* | 3.37 *(0.97)* |
| AMGN | 1.98 *(0.55)* | 2.00 *(0.56)* | 2.22 *(0.62)* | 2.50 *(0.77)* | 2.55 *(0.83)* | 2.56 *(0.84)* | 2.61 *(0.88)* | 2.57 *(0.94)* | 2.56 *(0.90)* | 2.61 *(0.93)* | 2.60 *(0.94)* | 2.59 *(0.97)* |
| AMZN | 3.56 *(0.99)* | 4.01 *(0.94)* | 4.48 *(1.15)* | 4.84 *(1.32)* | 4.85 *(1.35)* | 4.89 *(1.38)* | 4.95 *(1.45)* | 4.95 *(1.49)* | 4.96 *(1.48)* | 4.95 *(1.54)* | 4.94 *(1.54)* | 4.94 *(1.55)* |
| ATHM | 3.68 *(1.10)* | 3.89 *(1.22)* | 4.35 *(1.43)* | 4.73 *(1.65)* | 4.72 *(1.81)* | 4.75 *(1.82)* | 4.76 *(1.87)* | 4.76 *(1.90)* | 4.74 *(1.84)* | 4.74 *(1.90)* | 4.76 *(1.90)* | 4.74 *(1.96)* |
| CMGI | 4.37 *(1.67)* | 4.76 *(1.75)* | 5.36 *(2.07)* | 5.68 *(2.27)* | 5.74 *(2.32)* | 5.71 *(2.38)* | 5.71 *(2.42)* | 5.73 *(2.51)* | 5.72 *(2.51)* | 5.76 *(2.54)* | 5.73 *(2.57)* | 5.70 *(2.58)* |
| COMS | 2.79 *(0.81)* | 2.31 *(0.83)* | 2.44 *(1.04)* | 2.65 *(1.16)* | 2.84 *(1.29)* | 2.85 *(1.33)* | 2.76 *(1.34)* | 2.77 *(1.33)* | 2.81 *(1.31)* | 2.75 *(1.32)* | 2.74 *(1.40)* | 2.74 *(1.35)* |
| CPWR | 3.44 *(1.25)* | 3.23 *(1.10)* | 3.32 *(1.14)* | 3.59 *(1.35)* | 3.75 *(1.44)* | 3.74 *(1.46)* | 3.80 *(1.47)* | 3.78 *(1.55)* | 3.81 *(1.58)* | 3.78 *(1.66)* | 3.75 *(1.71)* | 3.74 *(1.64)* |
| CSCO | 1.71 *(0.37)* | 1.81 *(0.42)* | 2.04 *(0.51)* | 2.21 *(0.55)* | 2.26 *(0.57)* | 2.30 *(0.59)* | 2.30 *(0.63)* | 2.30 *(0.65)* | 2.27 *(0.63)* | 2.29 *(0.64)* | 2.28 *(0.63)* | 2.25 *(0.64)* |
| DELL | 2.69 *(1.55)* | 2.15 *(0.74)* | 2.32 *(0.71)* | 2.67 *(0.81)* | 2.78 *(0.84)* | 2.82 *(0.83)* | 2.80 *(0.85)* | 2.77 *(0.93)* | 2.83 *(0.95)* | 2.83 *(0.97)* | 2.82 *(1.02)* | 2.76 *(0.96)* |
| INTC | 1.72 *(0.47)* | 1.74 *(0.41)* | 1.96 *(0.51)* | 2.20 *(0.56)* | 2.33 *(0.58)* | 2.34 *(0.59)* | 2.34 *(0.64)* | 2.32 *(0.66)* | 2.31 *(0.64)* | 2.34 *(0.63)* | 2.32 *(0.64)* | 2.33 *(0.64)* |
| MSFT | 1.52 *(0.40)* | 1.53 *(0.35)* | 1.70 *(0.36)* | 1.90 *(0.47)* | 1.99 *(0.48)* | 1.98 *(0.50)* | 2.02 *(0.50)* | 1.99 *(0.53)* | 2.02 *(0.55)* | 2.02 *(0.54)* | 2.04 *(0.58)* | 2.03 *(0.58)* |
| NOVL | 2.78 *(0.71)* | 2.40 *(0.61)* | 2.58 *(0.68)* | 2.74 *(0.86)* | 2.83 *(0.86)* | 2.87 *(0.93)* | 2.94 *(0.95)* | 2.93 *(0.90)* | 2.83 *(0.90)* | 2.83 *(0.91)* | 2.81 *(1.00)* | 2.91 *(0.98)* |
| NXTL | 2.36 *(0.90)* | 2.20 *(0.72)* | 2.41 *(0.91)* | 2.63 *(1.06)* | 2.67 *(1.11)* | 2.66 *(1.17)* | 2.69 *(1.11)* | 2.70 *(1.16)* | 2.78 *(1.15)* | 2.77 *(1.22)* | 2.76 *(1.15)* | 2.74 *(1.21)* |
| ORCL | 2.75 *(0.87)* | 2.37 *(0.78)* | 2.47 *(0.93)* | 2.89 *(0.96)* | 2.92 *(1.07)* | 2.92 *(1.07)* | 3.02 *(1.09)* | 2.95 *(1.07)* | 2.98 *(1.15)* | 2.99 *(1.13)* | 2.90 *(1.17)* | 2.85 *(1.14)* |
| PSFT | 3.96 *(1.33)* | 3.21 *(1.18)* | 3.33 *(1.27)* | 3.57 *(1.49)* | 3.50 *(1.45)* | 3.46 *(1.54)* | 3.41 *(1.62)* | 3.46 *(1.50)* | 3.41 *(1.42)* | 3.30 *(1.55)* | 3.31 *(1.47)* | 3.28 *(1.47)* |
| QWST | 2.68 *(0.86)* | 2.61 *(0.87)* | 2.86 *(1.10)* | 3.17 *(1.21)* | 3.20 *(1.25)* | 3.23 *(1.30)* | 3.21 *(1.35)* | 3.15 *(1.33)* | 3.15 *(1.41)* | 3.24 *(1.36)* | 3.22 *(1.46)* | 3.32 *(1.41)* |
| SBUX | 2.62 *(0.88)* | 2.54 *(0.85)* | 2.51 *(0.97)* | 2.73 *(1.00)* | 2.86 *(1.06)* | 2.85 *(1.08)* | 2.90 *(1.13)* | 2.92 *(1.19)* | 2.83 *(1.14)* | 2.91 *(1.10)* | 2.91 *(1.17)* | 2.94 *(1.17)* |
| SUNW | 2.07 *(0.53)* | 2.29 *(0.54)* | 2.58 *(0.64)* | 2.80 *(0.72)* | 2.83 *(0.80)* | 2.82 *(0.85)* | 2.82 *(0.90)* | 2.83 *(0.89)* | 2.79 *(0.90)* | 2.80 *(0.93)* | 2.81 *(0.94)* | 2.82 *(0.96)* |
| WCOM | 1.52 *(0.45)* | 1.51 *(0.43)* | 1.68 *(0.48)* | 1.87 *(0.56)* | 1.92 *(0.54)* | 1.91 *(0.57)* | 1.89 *(0.59)* | 1.87 *(0.60)* | 1.86 *(0.60)* | 1.86 *(0.64)* | 1.84 *(0.62)* | 1.87 *(0.63)* |

*Note:* This Table reports average daily realized volatilities ($\overline{rv}_{T,q}^{(1\,tick)}$) sampled at every tick, corrected for different lags lengths ($q$),

$$\overline{rv}_{T,q}^{(1\,tick)} = \frac{1}{N}\sum_{t=1}^{N}\sqrt{\sum_{j=1}^{k(t)^{-1}}\{r_{T,t-1+jk(t)}^{(1\,tick)}\}^2 + 2\sum_{s=1}^{q}\frac{k(t)^{-1}}{k(t)^{-1}-s}\sum_{j=1}^{k(t)^{-1}-s} r_{T,t-1+jk(t)}^{(1\,tick)} r_{T,t-1+(j+s)k(t)}^{(1\,tick)}},$$

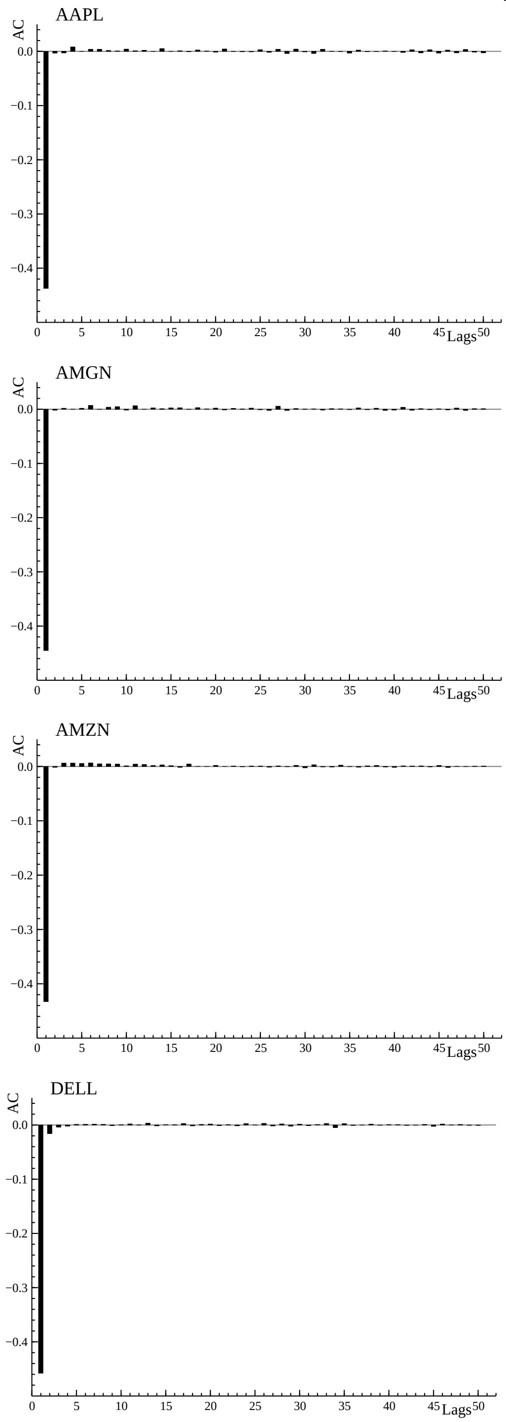
where $N = 124$ are the total number of trading days. Standard deviations of $\overline{rv}_{T,q}^{(1\,tick)}$ over the 124 trading days are reported in brackets.

Figure 1: Autocorrelation Functions for observed returns sampled at 1 second




*Note:* These graphs show the autocorrelation function for four representative stocks in the sample, sampled at 1 second intervals.

Figure 2: Autocorrelation Functions for observed returns sampled at every transaction




*Note:* These graphs show the autocorrelation function for four representative stocks in the sample, sampled at every transaction.