APPLICATIONS OF GENERALIZED FIDUCIAL INFERENCE IN HIGH FREQUENCY DATA

Dimitris Katsoridas

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2015

Approved by:

Jan Hannig

Jonathan B. Hill

Valentin Verdier

Ju Hyun Kim

Shankar Bhamidi

ABSTRACT

DIMITRIS KATSORIDAS: Applications of Generalized Fiducial Inference in High Frequency Data
(Under the direction of Jan Hannig)

Fiducial inference was introduced by R.A. Fisher Fisher (1930) as a response to the Bayesian approach to inference. The Bayesian paradigm begins by assuming a prior distribution on the parameter space and inference is conducted via the posterior distribution. Fisher, however, was concerned about the choice of the prior distribution, especially when there is insufficient information about the parameters of interest. To overcome this weakness, Fisher introduced the fiducial argument which is based on the following idea: randomness is transferred from the model space to the parameter space and a distribution on the parameter space can is defined that captures all of the information the data contains about these parameters. Fisher's idea, however, soon fell into disfavor since some of the properties Fisher claimed did not hold.

Recently, Fisher's inferential framework was revived through its connection to generalized inference. Hannig (2009) generalized Fisher's idea and introduced a framework where fiducial distributions can be defined properly. The main topic of this dissertation is to apply generalized fiducial inference methods to study intraday volatility using high frequency stock market data. In particular, we apply a generalized fiducial framework that is designed for interval data to study high frequency volatility, Hannig (2013). Our approach allows us to view the bid-ask spread as a natural interval around the latent price and use high frequency quotes for estimation. Modeling the spread in this manner allows us to take advantage of the features of the observed prices inherent to the trading process, such as rounding, and reduce the impact microstructure frictions cause to estimation. We demonstrate that our approach is very effective in estimating volatility and outperforms all alternative estimators. In chapter 2, we apply this idea, assuming that rounding errors are the only source

of microstructure frictions. In chapter 3, we extend our framework to allow for additive components. In the final chapter, we perform an empirical study to compare alternative realized volatility estimators through option pricing formulas. We find that the choice of volatility estimators does matter.

ACKNOWLEDGEMENTS

I take this opportunity to express gratitude to my advisor, Professor Jan Hannig. His passion about research, his encouragement and his generosity has been a tremendous inspiration for me. Without his guidance and support this dissertation would not have been possible.

I would like to thank Professors Jonathan Hill, Shankar Bhamidi, Ju Hyun Kim and Valentin Verdier for their support and participation on my dissertation committee.

Finally, I would like to thank my family and Marianna Feretzaki for their endless love and support throughout my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
Introduction

1.1 Thesis overview

In this thesis, there are main three contributions divided in three chapters. In chapter 2, we apply a generalized fiducial framework that is designed for interval data to study high frequency volatility. We assume that the only source of microstructure noise is rounding errors. In chapter 3, we extend our framework to allow for additive components. In the final chapter, we perform an empirical study to compare alternative realized volatility estimators through option pricing formulas. We find that the choice of volatility estimators does matter. All supplementary material is included in the appendices.

1.2 Generalized Fiducial Inference

Fiducial inference was introduced by R.A. Fisher (Fisher, 1930) as a response to the Bayesian approach to inference. The Bayesian paradigm begins by assuming a prior distribution on the parameter space and inference is conducted via the posterior distribution. Fisher, however, was concerned about the choice of the prior distribution, especially when there is insufficient information about the parameters of interest. To overcome this weakness, Fisher introduced the fiducial argument which is based on the following idea: randomness is transferred from the model space to the parameter space and a distribution on the parameter space can is defined that captures all of the information the data contains about these parameters. Subsequently, the fiducial distribution, which resembles the Bayesian posterior, can be used for inference procedures such as parameter estimation and confidence sets.

As a simple illustration of the idea we consider the following example: Let y be a realization of a random variable Y where $Y \sim N(\mu, 1)$. We know we can express the random variable Y as $Y = \mu + Z$ where Z is a standard normal random variable. Given

the observed value y, the fiducial argument solves for the unknown parameter μ, that is, $\mu = y - Z$. Even though the actual value of Z is unknown, its distribution is fully known and can be used to construct a distribution on the unknown parameter . This distribution on is known as the fiducial distribution, which in this example is simply $\mu \backsim N(y, 1)$. Hannig (2009) provides a solid introduction to the fiducial argument, together with facts about the historical development of the idea.

Soon after its inception, fiducial inference fell into disrepute among statisticians since it was shown that some of the properties Fisher claimed did not hold. In particular, not only statistical procedures based on the fiducial argument were non-exact in the frequentist sense, but also, there were non-uniqueness issues associated with the specification of these probability measures, see for example Lindley (1958) and Zabell (1992). Even though some recent attempts were made to revive fiducial procedures (Fraser (1961a,b, 1966, 1968), Dempster (1968), Dawid and Stone (1982), Barnard (1995)), it was until recently when Hannig et al. (2006) connected fiducial inference to generalized inference, introduced by Tsui and Weerahandi (1989). Tsui and Weerahandi (1989) performed hypothesis testing by introducing the concept of generalized p-values and Weerahandi (1993) constructed generalized confidence intervals by introducing the notion of a generalized pivotal quantity, based on the former idea of generalized p-values. Hannig et al. (2006) showed that most generalized pivotal inference procedures are identical to procedures obtained using fiducial inference. In fact, their recipe was introduced as a generalization of the idea of a generalized pivot.

The generalized fiducial argument expresses the data $\mathbf{X}$ through a data generating equation of the form

$$\mathbf{X} = G(\mathbf{U}, \xi) \tag{1.2.1}$$

where $G(\centerdot, \centerdot)$ is a jointly measurable structural equation based on the model under consideration, $\xi \in \Xi$ are the parameters of interest, and $\mathbf{U}$ is the random component of the structural equation; a random vector whose distribution is completely known, independent of any parameters. The data generating equation defines a set function through the inverse

image of G as follows

$$Q(\mathbf{x}, \mathbf{u}) = \{\xi \in \Xi : \mathbf{x}{=}G(\mathbf{u}, \xi)\} \tag{1.2.2}$$

where $\mathbf{x}$ is the observed data and $\mathbf{u}$ is a realization of $\mathbf{U}$.

The next step is to use the set function Q and define a distribution on the parameter space Ξ. The distribution of $\mathbf{U}$ will be used to draw samples of $Q(\mathbf{x}, \mathbf{u})$ given the data $\mathbf{x}$. However, using equation 1.2.2 to define a fiducial distribution needs caution. There are three sources of non-uniqueness that may arise in this framework and one needs to address them in order to define fiducial distributions properly. In particular, non-uniqueness can occur if Q has more than one element, if Q is empty, or due to the selection of the structural equation.

In the case where there is more than one element in Q, non uniqueness can be resolved by defining a rule, say V, for selecting an element in Q. Therefore, $V(Q(\mathbf{x}, \mathbf{u}))$ will be used to define the fiducial distribution. Discussion on how to select a rule can be found in Hannig (2009). In the case where Q is empty, non-uniqueness can be resolved by conditioning on the event $\{Q(x, \mathbf{U}) \neq \emptyset\}$. This can be achieved by removing realizations of $\mathbf{U}$ for which there is no ξ solving equation $\mathbf{x}{=}G(\mathbf{u}, \xi)$ and then re-normalizing the probabilities, i.e., use the distribution of $\mathbf{U}$ conditional on the event {there is at least one ξ solving equation $\mathbf{x}{=}G(\mathbf{u}, \xi)$}. A generalized fiducial distribution for parameter ξ is then defined as

$$V(Q(\mathbf{x}, \mathbf{U}^*)) \mid \{Q(\mathbf{x}, \mathbf{U}^*) \neq \emptyset\} \tag{1.2.3}$$

where $\mathbf{U}^*$ represents an independent copy of U.

A random element R_ξ with distribution 1.2.3 is termed generalized fiducial quantity (GFQ). Hannig (2009) showed that if the data generating equation 1.2.1 can be re-written as $X_i = g_i(\mathbf{U}, \xi)$ for $i = 1, \ldots, n$, implying that $G = (g_1, \ldots, g_n)$, then one can identify the generalized fiducial density of R_ξ. Assuming that the dimension of parameter ξ is p, where $p \leq n$, we can define the following quantities. Let $\mathbf{i} = (i_1, \ldots, i_p)$ denote a random selection of indices from $\{1, \ldots, n\}$. Let $\mathbf{X_i} = \left(X_{i_1}, \ldots, X_{i_p}\right)$ and $\mathbf{U_i} = \left(U_{i_1}, \ldots, U_{i_p}\right)$ denote the corresponding data and random elements so that $\mathbf{X_i} = G_{\mathbf{i}}(\mathbf{U_i}, \xi)$. Moreover, let $\mathbf{X_{i^c}}$, $\mathbf{U_{i^c}}$

and $G_{\mathbf{i}^c}$ denote the remaining components of $\mathbf{X}$, $\mathbf{U}$ and G respectively. Then, assuming that $G_{\mathbf{i}}$ is invertible and differentiable with respect to $\mathbf{U}_{\mathbf{i}}$, the generalized fiducial density is given by

$$f_{R_\xi}(\xi) = \frac{f_X(\mathbf{x}|\xi)\, J(\mathbf{x},\xi)}{\int_\Xi f_X(\mathbf{x}|\xi')\, J(\mathbf{x},\xi')\, d\xi'}$$

where

$$J(\mathbf{x},\xi) = \binom{n}{p}^{-1} \sum_{i=(i_1,\ldots,i_p)} \left| \frac{\det \frac{\partial}{\partial \xi} G_{\mathbf{i}}^{-1}(y_i,\xi)}{\det \frac{\partial}{\partial y_i} G_{\mathbf{i}}^{-1}(y_i,\xi)} \right|$$

It should be clear that when the dimension of the parameter space is p, we can arbitrarily select p equations to solve for ξ and condition the remaining $n-p$ equations of G on the solution of the first p.

From the discussion above, it is evident that the event $\{Q(\mathbf{x}, U^*) \neq \emptyset\}$ has zero probability only if the probability of generating the data $\{\mathbf{X} = \mathbf{x}\}$ is zero. However, the probability of observing data coming from a continuous distribution is always zero and conditioning on sets of probability zero may not be well defined (Borel paradox). This problem can be addressed by taking advantage of the fact that observed data have some degree of known uncertainty. For example, most data-sets are discretized due to the resolution of the instrument that collects them, computers store discretized data due to memory limitations and financial prices move in minimum increments (ticks). Using this "known" uncertainty, we can create a small interval around the observed value set and replace the zero probability event $\{\mathbf{X} = \mathbf{x}\}$ with the event $\{\mathbf{X} \in \mathbf{A}_{\mathbf{x}}\}$ where $\mathbf{A}_{\mathbf{x}}$ contains the observed value $\mathbf{x}$. Since $P_\xi(\mathbf{X} \in \mathbf{A}_{\mathbf{X}}) > 0$ the Borel paradox is resolved. Hannig (2013) presents the generalized fiducial inference framework for discretized data, pointing out that the nature of this data provide an attractive way to define generalized fiducial distributions, overcoming the non-uniqueness due to Borel paradox.

A simple illustration on how discretized/interval data can be used in the context of generalized fiducial inference is the following: Consider $Y = \mu + \sigma Z$ where (μ, σ) are unknown and $Z \sim N(0,1)$. Suppose now we observe $a_i \leq y_i \leq b_i$ for $i = 1, \ldots, n$ instead of

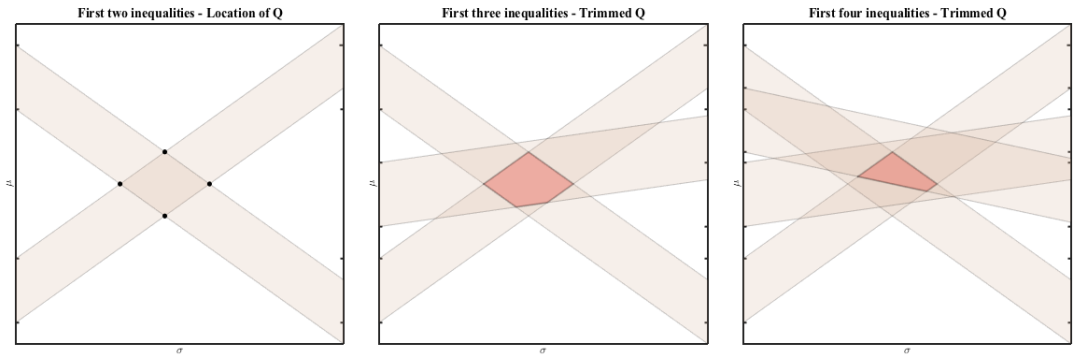


Figure 1.1: **Illustration of the sampling scheme.** The left panel illustrates how the first two inequalities can be used to generate Q, given that we generated z_1 and z_2 (with no restriction). The middle panel illustrates how the third inequality trims the polygon Q, given that we generated z_3 conditional on the first two inequalities. The right panel illustrates the update of Q using the fourth inequality.

$(y_1, \ldots, y_n)$. In order to identify the set Q, equation 1.2.2, we can use the first two inequalities[1] and solve for the unknown parameters (μ, σ). This generates four pairs of (μ, σ), given that we generated z_1 and z_2 (with no restriction). The only requirement for this step is to have $z_1 \neq z_2$, but we note that $P(Z_1 \neq Z_2) = 1$. Subsequently, we generate z_3 such that the first three inequalities are satisfied and we update the set Q (a polygon in this setup) in a way that all points (μ, σ) satisfy all inequalities. We repeat the same steps for all remaining inequalities. Finally, Picking randomly points (μ, σ) from Q amounts to sampling from the fiducial distribution (2). Repeating the procedure several times generates a fiducial sample. Figure 1.1 illustrates the sampling scheme for this simple model.

1.3 Volatility Estimation and Microstructure Noise

Estimating volatility using high frequency data (HFD) has been in the forefront of research in financial econometrics. However, the enormous availability of HFD has been a blessing and a curse to researchers since recorded prices are contaminated by market microstructure frictions. As a result, the maintained hypothesis that efficient prices are semimartingales is not consistent with observed data. In fact, observed prices resemble semimartingales

[1]For simplicity, we select the first two inequalities. As mentioned above, any other pair can be used to solve for the parameters.

recorded with error (MS noise).

In the standard microstructure setup, the efficient/unobserved log-price process, denoted by $X_t = log\,(S_t)$, is assumed to follow an Ito process:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

where W_t is a Brownian motion, μ_t is the drift of the process and σ_t is the instantaneous variance of the returns. Both μ_t and σ_t are adapted locally bounded random processes. The process is assumed to evolve in $[0, T]$ and is observed in the grid $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$. The quantity of interest is integrated volatility over the time period $[0, T]$, namely,

$$\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt$$

In the absence of market microstructure frictions, integrated volatility can be estimated consistently with the so called "realized volatility" estimator. This estimator is nothing but the quadratic variation relative to the grid G_n. That is

$$[X, X]_t^{G_n} = \sum_{t_{j+1} \leq t} \left(X_{t_{j+1}} - X_{t_j} \right)^2 \xrightarrow{p} [X, X]_t = \langle X, X \rangle_t = \int_0^t \sigma_s^2 ds$$

However, microstructure noise is present and the estimator above is heavily biased. The first remedy to MS noise was to use sparse samples. For example, Andersen et al. (2001) showed that sampling every five minutes helps mitigate the effects of microstructure noise. However, this amounts to discarding most the data available. For instance, if we have available transaction records on a liquid stock traded once every second, then the sample consists of 23,400 observations[2]. Therefore, if sampling takes place once every 5 minutes, then - whether or not this is the optimal thing to do - it amounts to retaining only 78 observations. Stated differently, one is throwing away 299 out of every 300 transactions. From a statistical perspective, this is unlikely to be the optimal solution, see Aït-Sahalia et al. (2005).

[2]A trading day has 6.5 hours = 23,400 seconds.

The first solution to MS noise was to assume that observed prices are the sum of the efficient log-price and a stochastic component capturing all microstructure frictions. Typically, the observed log-prices Y_{t_m} are assumed to be versions of X_{t_m} under the usual representation

$$Y_{t_m} = X_{t_m} + U_{t_m}$$

where U_{t_m} is introduced to capture a variety of frictions. These include frictions inherent in the trading process, such as, bid–ask bounces and discreteness of price changes, as well as, frictions attributed to informational effects, such as, differences in trade sizes, informational content of price changes, gradual response of prices to a block trade, strategic component of the order flow and inventory control effects, see for example Aït-Sahalia et al. (2005).

Microstructure frictions are responible for most of the stylized facts of the high frequency returns. For example, price discreteness leads to transaction changes of zero, 1 cent, 2 cents, etc. which may result in a very small number of log-returns. As a result, log-returns excibit high kurtosis (most tick-by-tick transactions equal their most recent transaction) and temporal dependence. Moreover, bid-ask bounces[3] bias upwards the variance of the log-returns and cause negative first order autocorrelation, see for example Engle and Russell (2004).

Estimation of volatility accounting for microstructure noise has been studied both parametrically and non-parametrically. Parametric modeling includes the framework by Aït-Sahalia et al. (2005) and Xiu (2010). Non-parametric modeling consists mainly of three different approaches. Zhang et al. (2005) and Zhang (2006) developed the Two-Scale (TSRV) and Multi-Scale (MSRV) realized volatility estimators, Barndorff-Nielsen et al. (2008) developed the Realized Kernel (RK) volatility estimators and Podolskij and Vetter (2009) use the pre-averaging method. Most of the aforementioned estimators were originally developed on the assumption that noise U_{t_m} is iid with mean zero and variance σ_u^2, independent of X. Below, we discuss briefly the most commonly used approaches in this literature.

[3]Bid-ask bounces are attributed to price discreteness. Transactions occur either on the bid or the ask. Usually we expect buy orders to be executed "close" to the ask, and sell orders "close" to the bid. The opposite occurs when traders are impatient, making the price oscillate between the the two quotes.

1.3.1 Quasi Maximum Likelihood Estimation

Following Aït-Sahalia et al. (2005), the process is assumed to evolve in $[0,T]$ and is observed/quoted in the grid $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$. The observed log-prices Y_{t_m} are assumed to be versions of X_{t_m} under the usual representation $Y_{t_m} = X_{t_m} + U_{t_m}$. The latent efficient price follows the process $X_t = \sigma W_t$ and MS noise is Gaussian, independent of the price process. Inference for this model is conducted through the log-likelihood of the log-returns

$$l\left(\sigma^2, \sigma_u^2\right) = -\frac{n}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left(det\Sigma\right) - \frac{1}{2}Y'\Sigma^{-1}Y$$

where

$$\Sigma = \begin{bmatrix} \sigma^2\Delta t_1 & -\sigma_u^2 & \cdots & 0 \\ -\sigma_u^2 & \sigma^2\Delta t_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2\Delta t_m \end{bmatrix}$$

and $\Delta t_m = t_m - t_{m-1}$.

In the case where observation times are equally spaced (calendar time sampling), that is $\Delta t_m \equiv \Delta$, and MS noise is independent of the price process, then the MLE is consistent and its asymptotic variance is given by

$$AVar\left(\hat{\sigma}^2\right) = 8\sigma^3\sigma_u\Delta^{\frac{1}{2}} + 2\sigma^4\Delta + o\left(\Delta\right)$$

In the case where Δt_m is random, independent of the process the asymptotic variance needs a further approximation, see section 8 in Aït-Sahalia et al. (2005). In our applications estimate the asymptotic variance as if observation times are equally spaced, approximating Δ by $1/n$, where n is the number of observations in $[0,T]$ and $T = 1$. Xiu (2010) showed that when volatility is stochastic, but assumed constant, the MLE is a Quasi-Maximum Likelihood Estimator (QMLE) of integrated volatility. Specifically, the MLE is a consistent, efficient and robust estimator of integrated volatility.

1.3.2 Pre-averaging Approach

The second estimator we are considering is the pre-averaging estimator introduced by Jacod et al. (2009). The pre-averaging estimator is designed to estimate integrated volatility when the underlying efficient price process is a continuous semimartingale.

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

where W_t is a standard Wiener process, μ_t is the drift of the process and σ_t is the instantaneous variance of the returns. Both μ_t and σ_t are adapted locally bounded random processes. Assuming equally spaced observation times, up to time t, we observe $n = [t/\Delta_n]$ contaminated prices. As before, $Y_{i\Delta_n} = X_{i\Delta_n} + U_{i\Delta_n}$. The error term U_t, conditional on X is centered and independent, that is $E(U_t|X) = 0$ and $U_t \perp U_s$, $t \neq s$, conditional on X. Moreover, the conditional variance of the noise process U_t, defined as $a_t = E\left(U_t^2|X\right)$, is adapted with the process $E\left(U_t^8|X\right)$ being locally bounded. The attractive feature of the pre-averaging method is that it allows for noise structure that can incorporate rounding errors explicitly. More details on the assumptions can be found in Jacod et al. (2009). The pre-averaging estimator is based on the idea of replacing the observed returns $\Delta_i^n Y = Y_{i\Delta_n} - Y_{(i-1)\Delta_n}$ by the weighted averaged returns

$$\overline{Y}_i^n = \sum_{j=1}^{k_n} g\left(\frac{j}{k_n}\right) \Delta_{i+j}^n Y$$

in an attempt to reduce the impact on noise. Here, k_n that satisfies $k_n \Delta_n^{1/2} = \theta + o\left(\Delta_n^{1/4}\right)$, where θ is selected by the modeler. Function $g : [0,1] \to \mathbb{R}$ is nonzero, continuous, piecewise continuously differentiable, such that g' is piecewise Lipschitz, with $g(0) = g(1) = 0$. Usually $g(x) = x \wedge (1-x)$. The estimator is given by

$$C_t^n = \frac{\sqrt{\Delta_n}}{\theta \psi_2} V(Y,2)_t^n - \frac{\psi_1 \Delta_n}{2\theta^2 \psi_2} RV_t^n$$

where

$$V(Y,2)_t^n = \sum_{i=0}^{[t/\Delta_n]-k_n} \left|\overline{Y}_i^n\right|^2$$

is the RV estimator based on the pre-averaged returns,

$$RV_t^n = \sum_{i=0}^{[t/\Delta_n]} \left|\Delta_i^n Y\right|^2$$

is the RV and for $i = 1, 2$, $\psi_i = \varphi_i(0)$ where $\varphi_1(s) = \int_s^1 g'(u)\, g'(u-s)\, du$ and $\varphi_2(s) = \int_s^1 g(u)\, g(u-s)\, du$. The pre-averaging estimator is a consistent and asymptotically mixed normal estimator of integrated volatility, that is

$$\Delta^{-\frac{1}{4}} \left(C_t^n - IV_t\right) \longrightarrow MN(0, \Gamma_t)$$

where converge is stable. The asymptotic variance process is given by

$$\Gamma_t = \int_0^t \gamma_s^2 ds$$

where $\gamma_s^2 = \frac{4}{\psi_2^2}\left(\Phi_{22}\theta\sigma_s^4 + 2\Phi_{12}\frac{\sigma_s^2 a_s^2}{\theta} + \Phi_{11}\frac{a_s^4}{\theta^3}\right)$, $\Phi_{ij} = \int_0^1 \phi_i(s)\,\phi_j(s)\, ds$, $i, j = 1, 2$. The consistent estimator of the conditional variance Γ_t is given by

$$\begin{aligned}
\Gamma_t^n &= \frac{4\Phi_{22}}{3\theta\psi_2^4} \sum_{i=0}^{[t/\Delta_n]-k_n} \left|\overline{Y}_i^n\right|^4 \\
&+ \frac{4\Delta_n}{\theta^3}\left(\frac{\Phi_{12}}{\psi_2^3} - \frac{\Phi_{22}\psi_1}{\psi_2^4}\right) \sum_{i=0}^{[t/\Delta_n]-2k_n+1} \left|\overline{Y}_i^n\right|^2 \sum_{j=i+k_n}^{i+2k_n-1} \left|\Delta_j^n Y\right|^2 \\
&+ \frac{\Delta_n}{\theta^3}\left(\frac{\Phi_{11}}{\psi_2^2} - 2\frac{\Phi_{12}\psi_1}{\psi_2^3} + \frac{\Phi_{22}\psi_1^2}{\psi_2^4}\right) \sum_{i=1}^{[t/\Delta_n]-2} \left|\Delta_i^n Y\right|^2 \left|\Delta_{i+2}^n Y\right|^2
\end{aligned}$$

In practice we use the adjusted version of the estimator which are given by

$$C_t^{n,adj} = \left(1 - \frac{\psi_1\Delta_n}{2\theta^2\psi_2}\right)^{-1}\left(\frac{[t/\Delta_n]\sqrt{\Delta_n}}{([t/\Delta_n] - k_n + 2)\,\theta\psi_2} V(Y,2)_t^n - \frac{\psi_1\Delta_n}{2\theta^2\psi_2} RV_t^n\right)$$

$$
\begin{aligned}
\Gamma_t^{n,adj} &= \left(1-\frac{\psi_1\Delta_n}{2\theta^2\psi_2}\right)^{-2}\left(\frac{4\Phi_{22}\left[t/\Delta_n\right]}{3\theta\psi_2^4\left(\left[t/\Delta_n\right]-k_n+2\right)}\sum_{i=0}^{[t/\Delta_n]-k_n}\left|\overline{Y}_i^n\right|^4\right.\\
&+\frac{4\Delta_n\left[t/\Delta_n\right]}{\theta^3\left(\left[t/\Delta_n\right]-k_n+2\right)}\left(\frac{\Phi_{12}}{\psi_2^3}-\frac{\Phi_{22}\psi_1}{\psi_2^4}\right)\sum_{i=0}^{[t/\Delta_n]-2k_n+1}\left|\overline{Y}_i^n\right|^2\sum_{j=i+k_n}^{i+2k_n-1}\left|\Delta_j^nY\right|^2\\
&\left.+\frac{\Delta_n\left[t/\Delta_n\right]}{\theta^3\left(\left[t/\Delta_n\right]-2\right)}\left(\frac{\Phi_{11}}{\psi_2^2}-2\frac{\Phi_{12}\psi_1}{\psi_2^3}+\frac{\Phi_{22}\psi_1^2}{\psi_2^4}\right)\sum_{i=1}^{[t/\Delta_n]-2}\left|\Delta_i^nY\right|^2\left|\Delta_{i+2}^nY\right|^2\right)
\end{aligned}
$$

The quantities ψ_i and Φ_{ij} for $i,j=1,2$ can be replaced by their finite-sample analogs which is beneficial for the finite sample bias.

1.3.3 Realized Kernels

Another class of estimators we are considering is the Realized Kernel estimators, introduced by Barndorff-Nielsen et al. (2008). The setup is similar to the pre-averaging framework, where $Y_{i\Delta_n}=X_{i\Delta_n}+U_{i\Delta_n}$ and MS noise is independent of the process with $E\left(U_t\right)=0$, $E\left(U_t^2\right)=\omega^2$ and $Var\left(U_t^2\right)=\lambda\omega^4$, for some $\lambda>0$. The flat-top realized kernel estimator is

$$
K\left(Y_{\Delta_n}\right)=\gamma_0\left(Y_{\Delta_n}\right)+\sum_{h=1}^{H}k\left(\frac{h-1}{H}\right)\left(\gamma_h\left(Y_{\Delta_n}\right)+\gamma_{-h}\left(Y_{\Delta_n}\right)\right)
$$

where $\gamma_h\left(Y_{\Delta_n}\right)=\sum_{i=1}^{n}\left(Y_{i\Delta_n}-Y_{(i-1)\Delta_n}\right)\left(Y_{(i-h)\Delta_n}-Y_{(i-h-1)\Delta_n}\right)$ is realized autocovariation process and $k\left(\cdot\right)$ is the kernel weight function, which is twice continuously differentiable on $[0,1]$. Further, if $k\left(0\right)+k\left(1\right)=0$ and $H=c_0n^{2/3}$, where c_0 can be estimated, the convergence rate of these estimators is $n^{1/6}$. If $k'\left(0\right)^2+k'\left(1\right)^2=0$ and $H=c_0n^{1/2}$ then the convergence rate is $n^{1/4}$ which is the optimal. For example, the Tukey-Hanning$_2$kernel

$$
k\left(x\right)=\sin^2\left\{\frac{\pi}{2}\left(1-x\right)^2\right\}
$$

has the optimal convergence rate. Barndorff-Nielsen et al. (2008) showed for the "faster" estimators that

$$
n^{1/4}\left(K\left(Y_{\Delta_n}\right)-\int_0^t\sigma_u^2du\right)\longrightarrow MN\left(0,4t\int_0^t\sigma_u^4du\left(c_0k_\bullet^{0,0}+2c_0^{-1}k_\bullet^{1,1}\rho\xi^2+c_0^{-3}k_\bullet^{2,2}\xi^4\right)\right)
$$

where $k_{\bullet}^{0,0} = \int_0^1 k(x)^2 dx$, $k_{\bullet}^{1,1} = \int_0^1 k'(x)^2 dx$ and $k_{\bullet}^{2,2} = \int_0^1 k''(x)^2 dx$. Also,

$$\xi^2 = \frac{\omega^2}{\left(t\int_0^t \sigma_u^4 du\right)^{1/2}} \qquad \rho = \frac{\int_0^t \sigma_u^2 du}{\left(t\int_0^t \sigma_u^4 du\right)^{1/2}}$$

The choice of H requires an estimate of c_0 in a way that it minimizes the asymptotic variance . Rewriting $H = c_0 n^{1/2} = c\xi n^{1/2}$ the asymptotic variance becomes

$$\begin{aligned} &4t\int_0^t \sigma_u^4 du \left(c_0 k_{\bullet}^{0,0} + 2c_0^{-1} k_{\bullet}^{1,1}\rho\xi^2 + c_0^{-3} k_{\bullet}^{2,2}\xi^4\right) \\ &= \omega\left(t\int_0^t \sigma_u^4 du\right)^{3/4} 4\left(c k_{\bullet}^{0,0} + 2c^{-1} k_{\bullet}^{1,1}\rho + c^{-3} k_{\bullet}^{2,2}\right) \end{aligned}$$

and c is chosen to minimize it[4]. That is,

$$c^* = \left(\rho\frac{k_{\bullet}^{1,1}}{k_{\bullet}^{0,0}}\left(1 + \sqrt{1 + \frac{3d}{\rho}}\right)\right)^{\frac{1}{2}}, \qquad d = \frac{k_{\bullet}^{0,0} k_{\bullet}^{2,2}}{\left(k_{\bullet}^{0,0}\right)^2}$$

Moreover, an estimate of ξ is required, therefore, estimates of ω^2 and the integrated quarticity are necessary. So, ξ^2 is estimated by $\hat{\xi}^2 = \hat{\omega}^2/\sqrt{\hat{IQ}}$, where $\hat{\omega} = \mathrm{RV_{all}}/2\mathrm{n}$ and $\hat{IQ} \simeq \hat{IV}^2$[5].

1.3.4 Two-Scales and Multi-Scales Realized Volatility

The Two-Scales Realized Volatility estimator, introduced by Zhang et al. (2005), uses the following setup. The efficient/unobserved log-price process. X_t is assumed to follow an Ito precess:

$$dX_t = \mu_t dt + \sigma_t dW_t,\ X_0 = x_0$$

where W_t is a Brownian motion, μ_t is the drift of the process and σ_t is the instantaneous variance of the returns. Both μ_t and σ_t are adapted (to the underlying filtration $(\mathcal{F}_t)$) locally bounded random processes. The process is assumed to evolve in $[0, T]$ and is observed in the grid $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$. We additionally assume that observation times are non-random, therefore ignoring their potential explanatory power over the process, and

[4]In the case of the Tukey-Hanning$_2$kernel, $k_{\bullet}^{0,0} = 0.219$, $k_{\bullet}^{0,0} = 1.71$, $k_{\bullet}^{0,0} = 41.7$ and $c^* = 5.74$.

[5]Here $\hat{IV} = RV_{sparse}$, where usually 20 minutes returns are used. RV_{all} uses all available data.

allow the observations to be irregularly spaced. Also we require $\max_{1\leq j\leq n}|t_j - t_{j-1}| = o_p(1)$.

Zhang et al. (2005) use the following setup. The observed log-price Y is denoted by

$$Y_{t_j} = X_{t_j} + \epsilon_{t_j}$$

where $\epsilon \overset{iid}{\sim} N\left(0, E\epsilon^2\right)$. First, they point out that $[Y,Y]_t^{G_n} = \sum_{t_{j+1}\leq t}\left(Y_{t_{j+1}} - Y_{t_j}\right)^2$ is estimating integrated volatility, but MS noise[6]. If we divide by $2n$ we will be getting a consistent estimate of the variance of the MS noise since[7]

$$[Y,Y]_t^{G_n} = \sum_{t_{j+1}\leq t}\left(Y_{t_{j+1}} - Y_{t_j}\right)^2 = 2nE\epsilon^2 + O_p\left(\sqrt{n}\right)$$

The first solution to this problem is to sample sparsely at some lower frequency and reduce the effect of MS noise. The choice of the sampling frequency is ad hoc. Let $\mathcal{H}_m = \{0 = \tau_0 < \tau_1 < ... < \tau_m = T\}$ be a sparse grid of times, not necessarily corresponding to observation times[8]. Then, they show that the quantity $[Y,Y]_t^{\mathcal{H}_m}$ has the following approximate distribution

$$[Y,Y]_T^{\mathcal{H}_m} \overset{L}{\approx} \langle X,X\rangle_T + 2mE\epsilon^2 + \left[4mE\epsilon^4 + \frac{2T}{m}\int_0^T \sigma_t^4 dt\right]^{\frac{1}{2}} Z$$

where Z is a standard normal random variable. The second term in the RHS denotes the bias of the estimator due to noise. The estimator based on the sparse grid can be further improved if we select the grid optimally. This can be done by minimizing the MSE with respect to m. he optimal sampling frequency is

$$m^\star = \left(\frac{T}{4\left(E\epsilon^2\right)^2}\int_0^T \sigma_t^4 dt\right)^{\frac{1}{3}}$$

---

[6]This was also noticed by Bandi and Russell (2006)

[7] $[Y,Y]_t^{G_n} = [X,X]_t^{G_n} + [\epsilon,\epsilon]_t^{G_n} + 2[X,\epsilon]_t^{G_n}$

[8]If at a particular sampling time an observation does not exist, we can built one using either the previous tick method or an interpolation method

It is evident that the sparse estimator based on the optimal sampling frequency still remains biased.

The idea of sparse sampling lead them introduce the Average Realized Volatility (ARV) estimator who uses the full sample. This is achieved by averaging estimators based on sparse samples of on non overlapping grids. Let $G = \{0 = t_0 < t_1 < ... < t_n = T\}$ denote the full grid. The full grid will be partitioned in K non overlapping grids $G^{(k)}$ such that

$$G = \bigcup_{k=1}^{K} G^{(k)}, \;\; where \;\; G^{(k)} \bigcap G^{(l)} = \text{ï£¡ } for \; k \neq l$$

Usually, these grids have the following form $G^{(k)} = \{t_{k-1}, t_{k-1+K}, t_{k-1+2K}, ..., t_{k-1+n_k K}\}$ for $k = 1, ..., K$. That is, we start sampling at t_{k-1} and pick every K^{th} sample point, until we exhaust the full grid. n_k is the integer making $t_{k-1+n_k K}$ the last element of the grid $G^{(k)}$ and, also, denotes the sample size of $G^{(k)}$. The new estimator based on these grids is

$$[Y,Y]_T^{(avg)} = \frac{1}{K}\sum_{k=1}^{K}[Y,Y]_T^{(k)}, \;\; where \;\; [Y,Y]_T^{(k)} = \sum_{t_j, t_{j+} \in G^{(k)}} \left(Y_{t_{j+}} - Y_{t_j}\right)^2$$

where t_{j+} denotes the following element of t_j in $G^{(k)}$. The quantity $[Y,Y]_t^{(avg)}$ has the following approximate distribution

$$[Y,Y]_T^{(avg)} \overset{L}{\approx} \langle X,X\rangle_T + 2\bar{n}E\epsilon^2 + \left[4\frac{\bar{n}}{K}E\epsilon^4 + \frac{4T}{\bar{n}}\int_0^T \sigma_t^4 dt\right]^{\frac{1}{2}} Z$$

where $\bar{n}$ denotes the average size of the K grids. As before, this estimator is biased and can be improved if we set $K^\star \approx {}^{n}/_{\bar{n}^\star}$ where

$$\bar{n}^\star = \left(\frac{T}{6\left(E\epsilon^2\right)^2}\int_0^T \sigma_t^4 dt\right)^{\frac{1}{3}}$$

Since the bias can be estimated, a bias corrected version of the ARV estimator. This estimator is called the Two Scales Realized Volatility (TSRV) and is given by

$$\widehat{\langle X,X\rangle}_T = [Y,Y]_T^{(avg)} - \frac{\bar{n}}{n}[Y,Y]_T^{(all)}$$

If the sub-grids are selected by $K = cn^{2/3}$ then

$$\widehat{\langle X,X\rangle}_T \overset{L}{\approx} \langle X,X\rangle_T + \frac{1}{n^{1/6}}\left[\frac{8}{c^2}\left(E\epsilon^2\right)^2 + c\frac{4T}{3}\int_0^T \sigma_t^4 dt\right]^{\frac{1}{2}} Z$$

where c can be optimally selected

$$c^\star = \left(\frac{T}{12\left(E\epsilon^2\right)^2}\int_0^T \sigma_t^4 dt\right)^{-\frac{1}{3}}$$

Clearly this estimator is centered. The only disadvantage of this estimator is that it converges with the small rate of $n^{-1/6}$.

Zhang (2006) extended the TSRV estimator to MSRV estimator. The MSRV is a weighted average of ARV estimators of the form $[Y,Y]_T^{(k)}$, namely

$$\widehat{\langle X,X\rangle}_T = \sum_{j=1}^{M} \alpha_j \left[Y,Y\right]_T^{(K_j)}$$

where M denotes the number of scales used. The weights have the form

$$\alpha_j = \frac{1}{M} w_M\left(\frac{j}{M}\right), \quad j = 1,...,M$$

and

$$w_M(x) = xh(x) + M^{-1}xh_1(x) + M^{-2}xh_2(x) + M^{-3}xh_3(x) + o_p\left(M^{-3}\right)$$

where the functions h and h_1 are independent of M. The conditions these functions satisfy can be found in Zhang (2006). The MSRV satisfies

$$n^{-\frac{1}{4}}\left(\widehat{\langle X,X\rangle}_T - \langle X,X\rangle_T\right) \to \nu_h Z$$

where

$$\nu_h = 4c^{-3}\left(E\epsilon^2\right)^2\int_0^1 h(x)^2 dx + c\frac{4}{3}T\eta^2\int_0^1 dx\int_0^x h(y)\,h(x)\,y^2(3x-y)\,dy$$
$$+4c^{-1}var\left(\epsilon^2\right)\int_0^1\int_0^y xh(x)\,h(y)\,dxdy + 8c^{-1}E\left(\epsilon^2\right)\int_0^1\int_0^1 h(x)\,h(y)\,min(x,y)\,dxdy\,\langle X,X\rangle$$

1.4 Sequential Monte Carlo (SMC) methods

As we mentioned above, in order to sample from the generalized fiducial distribution of the parameters, we will utilize SMC techniques. In this section we provide a very basic introduction to these algorithms, in order to stimulate the discussion below. A thorough introduction and applications of SMC methods can be found in Doucet et al. (2001).

SMC algorithms, or particle filters, are techniques for iteratively obtaining samples from an evolving target distribution (i.e. the distribution of interest) by employing importance sampling, and resampling, techniques. The principle application of these techniques is the approximate solution of the filtering, prediction and smoothing problems in Hidden Markov Models (HMMs). SMC methods are based on importance sampling (IS) techniques. IS is a technique for approximating integrals under one probability distribution (target distribution) using a collection of samples from another, instrumental distribution (proposal distribution). This can be presented using the importance sampling identity: given a distribution of interest π with support $\mathbb{R}_X$, and some instrumental distribution $\widetilde{\pi}$ with support $\mathbb{R}'_X$, such that $\mathbb{R}_X \subset \mathbb{R}'_X$, and any integrable function $h : \mathbb{R}_X \to \mathbb{R}$

$$E_\pi\left(h(X)\right) = \int h\left(x\right) \frac{\pi\left(x\right)}{\widetilde{\pi}\left(x\right)} \widetilde{\pi}\left(x\right) dx = \int h\left(x\right) \omega\left(x\right) \widetilde{\pi}\left(x\right) dx = E_{\widetilde{\pi}}\left(\omega\left(X\right) h(X)\right)$$

The reason we are considering this identity is the following: If we have a random sample$(X_1, ..., X_n)$ from π, we can estimate $E_\pi\left(h(X)\right)$ by calculating $\hat{E}_\pi\left(h(X)\right) = \frac{1}{n}\sum_{i=1}^{n} h\left(x_i\right)$. The law of large numbers in this case guarantees a good approximation of $E_\pi\left(h(X)\right)$. In the absence of $(X_1, ..., X_n)$, we can estimate the same integral by using a sample $(Y_1, ..., Y_N)$ from $\widetilde{\pi}$, and evaluate $\hat{E}_{\widetilde{\pi}}\left(\omega\left(X\right) h(X)\right) = \frac{1}{N}\sum_{i=1}^{N} \omega\left(y_i\right) h\left(y_i\right)$. Again the law of large numbers guarantees a good approximation of $E_\pi\left(h(X)\right)$. Therefore, if in the problem under consideration it is difficult to sample from the target distribution, we can use another distribution (proposal) from which we can easily sample and use the new sample together with the weights to estimate the relevant quantity.

We will now present the basic SMC algorithm. Suppose we are observing data $Y_{1:t} = (y_1, ..., y_t)$ sequentially in time. We are assuming that there is an underlying process (signal

process) $Z_{0:t} = (z_0, ..., z_t)$ that causes $Y_{1:t}$. The signal process is latent and our goal is to obtain a sample from it given $Y_{1:t}$. In other words, or goal is to sample from the density $\pi_{1:t}(Z_{0:t}|Y_{1:t})$. If it is not possible to sample from $\pi_{1:t}$, then a proposal distribution $\widetilde{\pi}_{1:t}$ will be utilized. The proposal distribution is selected in a manner so that the importance weights $\pi_{1:t}/\widetilde{\pi}_{1:t}$ can be updated recursively with the arrival of a new data point y_{t+1}. In the IS setting, the unnormalized importance weight at time t for particle $k = 1, ..., N$ would be written

$$W_{1:t}\left(Z_{0:t}^{(k)}\right) = \frac{\pi_{1:t}\left(Z_{0:t}^{(k)}|Y_{1:t}\right)}{\tilde{\pi}_{1:t}\left(Z_{0:t}^{(k)}|Y_{1:t}\right)} \tag{1.4.1}$$

In our setup we can derive the following relationships: First, the target distribution at time $t+1$, $\pi_{1:t+1}(Z_{0:t+1}|Y_{1:t+1})$ can be written in terms of a marginal and conditional distribution

$$\pi_{1:t+1}(Z_{0:t+1}|Y_{1:t+1}) = \pi_{1:t}(Z_{0:t}|Y_{1:t})\,\pi_{t+1|1:t}(Z_{t+1}|Z_{0:t}, Y_{1:t}) \tag{1.4.2}$$

If we write the proposal in the same manner, then we can derive a recursive formula for the weights

$$W_{1:t+1} = W_{1:t}\frac{\pi_{t+1|1:t}(Z_{t+1}|Z_{0:t}, Y_{1:t})}{\widetilde{\pi}_{t+1|1:t}(Z_{t+1}|Z_{0:t}, Y_{1:t})} \tag{1.4.3}$$

The generated values $Z_{0:t}^{(k)}$ for $k = 1, 2, ..., N$ are called particles and together with their associated normalized importance weights $\hat{W}_{1:t}^{(k)} = \frac{W_{1:t}^{(k)}}{\sum_{k=1}^{N} W_{1:t}^{(k)}}$, they form a particle system, namely $\left\{Z_{0:t}^{(k)}, \hat{W}_{1:t}^{(k)}\right\}_{k=1}^{N}$. Unfortunately, this method is destined to fail as t increases. The importance weights degenerate to a single particle (i.e. one particle has a normalized importance weight equal to one, the rest zero), making the particles useless for practical purposes. The reason of this degeneracy has to do with the fact that the variance of the importance weights increases with t making them inefficient. This degeneracy is usually measured by the effective sample size (ESS), a measure of the distribution of the weights of the particles. The ESS at time t is often estimated by

$$ESS_t = \left(\sum_{k=1}^{N} W_{1:t}^{(k)}\right)^2 / \sum_{k=1}^{N}\left(W_{1:t}^{(k)}\right)^2$$

If the ESS has dropped below some designated threshold (usually $ESS_t \leq N/2$), then the

particle system is resampled removing the particles with low weights and replicating the particles with higher weights. There are several ways to resample the particle system, all of which are based on the normalized importance weights.

CHAPTER 2

Generalized Fiducial Inference for High Frequency Data in the Presence of Rounging Errors

2.1 Summary

In this chapter, we adapt a generalized fiducial framework to study volatility using high frequency data. Our framework, which is designed for interval data, allows to view the bid-ask spread as a natural interval around the latent efficient price and use high frequency quotations as our dataset. Unlike the standard microstructure literature our modeling approach does not require additive components to model noise. In fact, our framework takes advantage of the features of the observed prices inherent to the trading process, such as rounding, and reduces the impact microstructure frictions cause to estimation.

Generalized fiducial methods produce distribution estimators that can be used to obtain quantities beyond point estimators, such as approximate confidence intervals. Inference is performed by splitting the trading day into blocks where volatility is assumed constant. A novel combination scheme allows to join the information from all blocks. Both our simulation study and empirical application demonstrate that the proposed volatility estimator performs remarkably well even at very high frequencies.

Moreover, we prove a Bernstein - von Mises theorem establishing that, under some regularity conditions, the generalized fiducial distribution can be approximated by a normal distribution.

2.2 Introduction

Recently, volatility estimation using high frequency data (HFD) has received considerable attention in financial econometrics. However, HFD are contaminated by market microstructure frictions and, as a result, the maintained hypothesis that the underlying efficient price

process is an Itï£¡ semimartingale is not consistent with observed data. In fact, observed prices resemble semimartingales recorded with error (MS noise). Consequently, volatility estimates ignoring microstructure can be heavily biased and, therefore, unreliable for inference procedures. Moreover, the bias is amplified as the sampling frequency increases, since market microstructure noise accumulates.

In the standard microstructure setup, the efficient/unobserved log-price process, denoted by $X_t = log\,(S_t)$, is assumed to follow an Ito process:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

where W_t is a standard Wiener process, μ_t is the drift of the process and σ_t is the instantaneous variance of the returns. Both μ_t and σ_t are adapted locally bounded random processes. The process is assumed to evolve in $[0, T]$ and is observed in the grid $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$. The quantity of interest is integrated volatility over the time period $[0, T]$, namely,

$$\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt$$

Typically, the observed log-prices Y_{t_m} are assumed to be versions of X_{t_m} under the usual representation

$$Y_{t_m} = X_{t_m} + U_{t_m}$$

where U_{t_m} is introduced to capture a variety of effects, including frictions inherent to the trading process, informational effects and other type of measurement errors, see for example Aït-Sahalia et al. (2005).

Modeling noise in this simple setup can be unsatisfactory since microstructure frictions include rounding[1]. For example, Li and Mykland (2007) analyzed the effect of rounding on the TSRV estimator and showed that it may not be a robust estimator of integrated volatility. Rounded Ito processes were initially studied by Jacod (1996) and Delattre and Jacod (1997) in a non-financial setup. Subsequently, Li and Mykland (2007) and Jacod et al. (2009) introduced a transition probability so that, conditional on X, observed log-prices Y

[1]Stocks are traded on grids and, therefore, observed prices are multiples of the tick size, usually a cent.

are distributed around X. This new approach allows to endogenize rounding and construct estimators that are robust in its presence, for example, Jacod et al. (2009). However, their estimator cannot accommodate the case where the source of error is mainly rounding. Recently, Li and Mykland (2014) proposed a bias corrected RV estimator when rounding is the only source of noise. This case is particularly interesting for less expensive stocks, since rounding is the main source of noise. They showed that the new estimator performs better than the traditional RV estimator as the sampling frequency increases, but cannot reach very high frequencies such as 1-5 seconds.

In this chapter our goal is to study volatility by taking advantage of the rounding errors. In particular, we work under the assumption that the latent efficient price process is between the (rounded) bid-ask prices. Namely, we assume that at any arrival time t_m, the process is contained in the interval $[b_{t_m}, a_{t_m}]$, that is

$$b_{t_m} \leq X_{t_m} \leq a_{t_m}$$

where both b_{t_m} and a_{t_m} are the log versions of the observed bid-ask prices. Volatility modeling in this setup is clearly not affected by rounding and the spread related microstructure frictions , making the additive component introduced in the aforementioned literature redundant. The assumption that the latent efficient price process is between the bid-ask prices has been used primarily in classical microstructure literature, see Roll (1984), Harris (1990) and Hasbrouck (1999). However, other intervals that can be justified to contain the latent price can be a possible candidates. In our empirical study below, besides the direct use of the spread, we propose a simple way to combine transactions and bid-ask prices to identify such intervals. Moreover, we work under the additional assumption that volatility is constant , that is

$$X_t = X_0 + \sigma W_t$$

This approach is similar in spirit with Aït-Sahalia et al. (2005) and Xiu (2010).

In the high frequency volatility literature the parametric approach assumes constant volatility in the entire interval $[0, T]$. We deviate from this assumption and consider constant

volatility over a short period of time[2]. In other words, we split the daily data into blocks of successive observations and generate samples for each block. Subsequently, we estimate daily volatility using two methods. The first method computes daily volatility by simply aggregating the block point estimates and the second method relies on a novel combination scheme. That is, inference is conducted through a distribution generated by combining the block distributions into one that summarizes all information from all the blocks under consideration. In a sense, the combination scheme works as an importance sampler by re-weighing all particles with weights computed through a metric that utilizes the Gaussian kernel. As we demonstrate, the combined distribution approximates remarkably well the distribution we would have generated if we had used all data together in one sample.

We test our methodology by conducting a simulation study employing a realistic simulation scheme. We generate our data by simulating the efficient price process in the original scale and, at observation times, we round the process upwards and downwards, towards the two nearest ticks. This type of contamination incorporates rounding errors explicitly and is similar in spirit with the two stage contamination scheme of Li and Mykland (2007) and Jacod et al. (2009). The proposed simulation scheme renders the choice of the starting price X_0 relevant, since the magnitude of the spread increases for less expensive stocks, due to the log-transformation, see Li and Mykland (2007, 2014). Therefore, our simulation study uses different starting prices to capture this effect. For the volatility parameter (signal), in addition to the standard values in the literature, we use low values since a weak signal introduces price sluggishness, intensifying the effect of rounding errors. Our simulation study shows that we can effectively estimate true volatility, constant or stochastic, even in cases where rounding errors dominate, and outperform the competing parametric and non-parametric estimators.

Finally, our empirical study reveals that robust volatility estimation is possible without having to assume unrealistic microstructure noise structures. We compare our estimator

[2]Mykland and Zhang (2009) studied this type of local constancy and showed that in sufficiently small neighborhoods of observations one can act as if volatility is constant. In this article we do not explore the extent where volatility is held constant. We consider blocks that contain at most 500 observations. For example, a block of 300 observations observed on average every two seconds amounts to a time frame of 10 minutes. In this small time frame we assume that volatility, even if stochastic, does not vary much.

with the standard parametric and non-parametric alternatives and show that it is very competitive.

2.3 Generalized Fiducial Inference for HF data

In our setup, we will assume that the efficient log-price follows a Geometric Brownian motion. The process is assumed to evolve in $[0, T]$ and is observed/quoted in the grid $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$. In addition to the grid G_n, we will consider the sub-grid $H_n = \{0 = \tau_0 < \tau_1 < ... < \tau_{M_n} = T\} \subseteq G_n$ where volatility is assumed constant for all $t_{i,m} \in (\tau_{i-1}, \tau_i]$. Specifically, in the interval $(\tau_{i-1}, \tau_i]$ the log-price, given $X_{\tau_{i-1}} = x_{\tau_{i-1}}$, evolves according to

$$dX_t = \sigma_{\tau_{i-1}} dW_t$$

In the high frequency literature, it is common practice to to assume that $\mu = 0$. The order of magnitude of the diffusive component ($\sqrt{dt}$) is much larger than the order of magnitude of the drift component (dt), making the drift component mathematically negligible at high frequencies. Moreover, maintaining it, may have adverse effects on the estimation procedure since it is estimated with a large standard error. Our preliminary simulation study showed that maintaining the drift component does not have any impact on the quality of the generated fiducial distributions. It adds though computational burden and, therefore, it is not included in our reported simulations.

At observation times, within the interval $(\tau_{i-1}, \tau_i]$, the process is assumed to be between the bid and ask log-prices

$$b_{t_{i,m}} \leq X_{t_{i,m}} \leq a_{t_{i,m}} \tag{2.3.1}$$

Noting that $W_t = \sqrt{t}Z$ where $Z \sim N(0, 1)$ we can re-write equation 2.3.1 as

$$b_{t_{i,m}} \leq \sigma_{\tau_{i-1}} \sqrt{t_{i,m}} Z_{t_{i,m}} \leq a_{t_{i,m}} \tag{2.3.2}$$

In terms of the fiducial argument, the structural equation is

$$G(\mathbf{U}, \xi) = G(Z, \sigma) = \sigma\sqrt{t}Z$$

The inverse image of $G(\mathbf{z}, \sigma)$ is

$$Q\left((\mathbf{b}, \mathbf{a}], \mathbf{Z}\right) = \{\sigma \in \mathbb{R}_+ : \mathbf{b} < G(\mathbf{z}, \sigma) \leq \mathbf{a}\} \tag{2.3.3}$$

The corresponding generalized fiducial distribution is

$$V\left(Q\left((\mathbf{b}, \mathbf{a}], Z\right)\right) \mid \{Q\left((\mathbf{b}, \mathbf{a}], Z\right) \neq \emptyset\} \tag{2.3.4}$$

Generating samples from the generalized fiducial distribution requires the use of Sequential Monte Carlo (SMC) methods. The SMC algorithm is based on the algorithm developed by Cisewski and Hannig (2012), where they performed inference for the parameters of normal linear mixed models and is presented in section 4. In section 5, we prove a Bernstein - von Mises theorem establishing that, under some regularity conditions, the generalized fiducial distribution 2.3.4 can be approximated by a normal distribution.

2.4 Estimation Method

2.4.1 Sequential Monte Carlo Algorithm

In this section we present the Sequential Monte Carlo algorithm to generate samples from the generalized fiducial distribution of the parameters of interest. We consider the interval $(\tau_{i-1}, \tau_i]$ where the process is assumed to be between the bid and ask log-prices, as give by equation 3.3.2. To ease notation we will drop the dependence i and we will embed $\sqrt{t}$ in Z_t such that $Z_t \sim N(0, t)$. We will be denoting $Z_{t_1:t_m} = (Z_{\Delta t_1}, ..., Z_{\Delta t_m})$, where $t_i \in \{t_1, ..., t_m\}$, $\Delta t_i = t_i - t_{i-1}$, $t_0 = 0$, and $m = 1, ..., n$ where $n \equiv n_i = \#\{j, \tau_{i-1} < t_{i,j} \leq \tau_i\}$. In our setup, $Z_{1,\Delta t_j} \sim N(0, \Delta t_j)$. The generalized fiducial distribution of the parameters will be

$$V\left(Q\left((\mathbf{b}_n,\mathbf{a}_n],Z_{t_1:t_n}^{(K)}\right)\right)\mid\left\{Q\left((\mathbf{b}_n,\mathbf{a}_n],Z_{t_1:t_n}^{(K)}\right)\neq\emptyset\right\}\tag{2.4.1}$$

where $Q_n^{(K)}=Q\left((\mathbf{b}_n,\mathbf{a}_n],Z_{t_1:t_n}^{(K)}\right)$ is the set function containing the values of the parameters that satisfy the structural equation 3.3.2 for all $m\leq n$, given the data $(\mathbf{b}_n,\mathbf{a}_n]$ and the generated $Z_{t_1:t_n}^{(K)}$ for particle K , where $K=1,2,...,N$. Generating a sample from 2.4.1 is equivalent to simulating sequentially for each m, $Z_{t_1:t_m}^{(K)}$ such that $Q_m^{(K)}$ is non-empty until we reach n. The corresponding target distribution up to time t_m, denoted by $\pi_{t_1:t_m}$, is

$$\begin{aligned}\pi_{t_1:t_m}\left(Z_{t_1:t_m}\mid(\mathbf{b},\mathbf{a}]_{t_1:t_m}\right) &\equiv \pi_{t_1:t_m}\left(Z_{t_1:t_m}\right)\\ &\propto \prod_{j=1}^{m}\frac{1}{\left(\Delta t_j\right)^{1/2}}\exp\left(-\frac{1}{2\Delta t_j}Z_{\Delta t_j}^2\right)\mathbf{I}_{\mathbf{C_m}}\left(Z_{t_1:t_m}\right)\end{aligned}\tag{2.4.2}$$

where $\mathbf{I}_{\mathbf{C_m}}\left(Z_{t_1:t_m}\right)$ is an indicator random variable of the set

$$\mathbf{C_m}=\left\{Z_{t_1:t_m}:b_{t_j}\leq\sigma\sum_{k=1}^{j}Z_{\Delta t_k}\leq a_{t_j},\ for\,all\,j=1,...,m\right\}\tag{2.4.3}$$

Restriction 2.4.3 is required in order to generate a representative sample from the fiducial distribution. It ensures that all inequalities up to time t_m are satisfied simultaneously. In practice, this can be achieved easily if at time t_m, given that we have sampled $Z_{t_1:t_{m-1}}^{(K)}$, we sample $Z_{\Delta t_m}^{(K)}$ by truncating it between the two values

$$L_m\left(Z_{t_1:t_{m-1}}^{(K)}\right)=\min\left\{\frac{b_{t_m}-\sigma\sum_{j=1}^{m-1}Z_{1,\Delta t_j}^{(K)}}{\sigma},\ \sigma\in Q_{m-1}^{(K)}\right\}$$

$$R_m\left(Z_{t_1:t_{m-1}}^{(K)}\right)=\max\left\{\frac{a_{t_m}-\sigma\sum_{j=1}^{m-1}Z_{1,\Delta t_j}^{(K)}}{\sigma},\ \sigma\in Q_{m-1}^{(K)}\right\}$$

Utilizing these type of restrictions we can write

$$\mathbf{I}_{\mathbf{C_m}}\left(Z_{t_1:t_m}\right) = \mathbf{I}_{\mathbf{C_{m-1}}}\left(Z_{t_1:t_m}\right)\mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta t_m}\right)=\prod_{j=1}^{m}\mathbf{I}_{(L_j,R_j)}\left(Z_{1,\Delta t_j}\right)$$

The proposal distribution for our SMC algorithm utilizes the Cauchy distribution

$$\begin{aligned}\tilde{\pi}_{t_1:t_m}\left(Z_{t_1:t_m} \mid (\mathbf{b},\mathbf{a}]_{t_1:t_m}\right) &\equiv \tilde{\pi}_{t_1:t_m}\left(Z_{t_1:t_m}\right) \\ &\propto \pi_{t_1}\left(Z_{t_1}\right)\prod_{j=2}^{m}\frac{1}{\left(\Delta t_j\right)^{1/2}\left(1+\frac{Z_{t_j}^2}{\Delta t_j}\right)\left(F\left(R_j\right)-F\left(L_j\right)\right)}\mathbf{I}_{(L_j,R_j)}\left(Z_{\Delta t_j}\right)\end{aligned} \tag{2.4.4}$$

where F denotes the cdf of the Cauchy distribution. It is important to point out that the proposal distribution treats Z_{t_1} as unrestricted, that is, Z_{t_1} is drawn from the target distribution. The reason is that Z_{t_1} will be used to identify $Q_1^{(K)} = Q\left((b_{t_1}, a_{t_1}], Z_{t_1}^{(K)}\right)$, for each particle K , where $K = 1, 2, ..., N$. This generates the interval

$$\left[\frac{b_{t_1}}{Z_{1,\Delta t_1}}, \frac{a_{t_1}}{Z_{1,\Delta t_1}}\right]$$

and, clearly, any σ in this interval satisfies the first set of inequalities. The rest of the inequalities will be used sequentially to trim the interval in a way that all inequalities will be satisfied.

The conditional proposal distribution for $m > 1$, which will be used to draw samples in the algorithm is

$$\widetilde{\pi}_{t_m|t_1:t_{m-1}} \propto \frac{\mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta t_m}\right)}{\left(\Delta t_m\right)^{1/2}\left(1+\frac{Z_{t_m}^2}{\Delta t_m}\right)\left(F\left(R_m\right)-F\left(L_m\right)\right)} \tag{2.4.5}$$

The final component of the algorithm is the importance weights. The weights are computed as

$$W_{t_1:t_m} = \frac{\pi_{t_1:t_m}}{\tilde{\pi}_{t_1:t_m}} = \frac{\pi_{t_m|t_1:t_{m-1}}\pi_{t_1:t_{m-1}}}{\tilde{\pi}_{t_m|t_1:t_{m-1}}\tilde{\pi}_{t_1:t_{m-1}}} = W_{t_m}W_{t_1:t_{m-1}} \tag{2.4.6}$$

where

$$W_{t_m} = \frac{\pi_{t_m|t_1:t_{m-1}}}{\tilde{\pi}_{t_m|t_1:t_{m-1}}}$$

is the incremental weight. The incremental weight is given by

$$W_{t_m} \propto \exp\left(-\frac{1}{2\Delta t_m} Z_{t_m}^2\right)\left(1+\frac{Z_{t_m}^2}{\Delta t_m}\right)\left(F\left(R_m\right)-F\left(L_m\right)\right) \tag{2.4.7}$$

2.4.2 Resampling - Alteration Step

In our setup the resampling step resembles that of a general SMC algorithm. To overcome the degeneracy of the particle system as t_m increases, we measure the effective sample size (ESS) at time t_m

$$ESS_{t_m} = \left(\sum_{k=1}^{N} W_{t_1:t_m}^{(k)}\right)^2 / \sum_{k=1}^{N}\left(W_{t_1:t_m}^{(k)}\right)^2 \tag{2.4.8}$$

and if the ESS for the particle system has dropped below a designated threshold (usually $N/2$), the particle system is resampled removing the particles with low weights and replicating the particles with higher weights. In this setup, replicating particles will not generate a representative sample of the fiducial distribution. As mentioned above, each of the particles forms an interval in the parameter space, and therefore, if the particles are simply copied, the intervals will be concentrated in a narrow area, due to particles with initially higher weight. Moreover, as the algorithm progresses, the particles will not be able to move from those regions. A solution to this issue is to alter the particles selected from resampling in a way that they will maintain their heavy weight, while still allowing for an appropriate sample of the fiducial distribution.

The alteration step is performed as follows: Suppose that at time t_m particle K is selected in the resampling step and will be copied. Up to time t_m, we have observed the following inequalities (We are suppressing the dependence in K):

$$\mathbf{b}_m \leq \sigma \mathbf{V}_m \mathbf{Z}_m \leq \mathbf{a}_m$$

We perform the following decomposition for $\mathbf{Z}_m$

$$\mathbf{Z}_m = \|\mathbf{Z}_m\| \frac{\mathbf{Z}_m}{\|\mathbf{Z}_m\|} \tag{2.4.9}$$

The SMC algorithm

Step	Action
1. Initialization	For $k = 1, 2, ..., K$ draw $Z_{t_1}^{(k)} \sim \pi_{t_1}$ (Eq. 2.4.2) and set $W_{t_1}^{(k)} = 1$
2. For $t_m > t_1$ and $t_m \leq t_n$	For $k = 1, 2, ..., K$ draw $Z_{t_m}^{(k)} \sim \widetilde{\pi}_{t_m \mid t_1:t_{m-1}}$ (Eq. 2.4.5)
3. For $t_m > t_1$ and $t_m \leq t_n$	Calculate weights $W_{t_1:t_m}^{(k)} = W_{t_m}^{(k)} W_{t_1:t_{m-1}}^{(k)}$ (Eq. 2.4.6)
4. For $t_m > t_1$ and $t_m \leq t_n$	Calculate ESS_{t_m} (Eq. 2.4.8). If $ESS_{t_m} \leq threshold$ go to step 5. Else go to step 2 and set $m = m + 1$
5. For t_m given $ESS_{t_m} \leq thd$	Resample particles and set $W_{t_1:t_m} = N^{-1}$. Go to step 6.
6. For t_m	Perform alteration as described above and set $m = m + 1$

where $\|\mathbf{Z}_m\|$ denotes L^2 norm of $\mathbf{Z}_m$. By setting $D = \|\mathbf{Z}_m\|$ and $\kappa = \frac{\mathbf{Z}_m}{\|\mathbf{Z}_m\|}$ decomposition 2.4.9 can be re-written as

$$\mathbf{Z}_m = D\kappa \tag{2.4.10}$$

Moreover, by assumption $\mathbf{Z}_m \sim N(0, \mathbf{I}_m)$, therefore, $D \sim \sqrt{\chi_m^2}$.

Decomposition 2.4.9 allows us to alter $\mathbf{Z}_m$ by sampling new values of D according to its distribution. If we denote by $\widetilde{D}$ the generated values, then we can use $\widetilde{\mathbf{Z}}_m = \widetilde{D}\kappa$ to update the set

$$Q_m = \{\sigma : \mathbf{b}_m \leq \sigma \mathbf{V}_m \mathbf{Z}_m \leq \mathbf{a}_m\}$$

We achieve that by noting that, since σ solves $\mathbf{b}_m \leq \sigma \mathbf{V}_m \mathbf{Z}_m \leq \mathbf{a}_m$, then we need to identify $\widetilde{\sigma}$ that solve $\mathbf{b}_m \leq \widetilde{\sigma} \mathbf{V}_m \widetilde{\mathbf{Z}}_m \leq \mathbf{a}_m$. Using the following equality

$$\sigma \mathbf{V} \mathbf{Z}_m = \sigma \mathbf{V} D\kappa = \widetilde{\sigma} \mathbf{V}_m \widetilde{D}\kappa = \widetilde{\sigma} \mathbf{V}_m \widetilde{\mathbf{Z}}_m$$

we can solve for $\widetilde{\sigma}$, that is

$$\widetilde{\sigma} = \sigma \frac{D}{\widetilde{D}}$$

The table below gives an outline of the steps of the algorithm.

2.5 Theoretical Results

2.5.1 Preliminaries - Likelihood of Exact Data

We assume that the efficient log-price follows a Geometric Brownian motion. The process is assumed to evolve in $[0, T]$ and is observed on the grid $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$.

Furthermore, assume that $t_i - t_{i-1} = \Delta$, fixed. Specifically, the log-price, given $X_{t_0} = x_0$, evolves according to

$$X_t = x_0 + \sigma W_t$$

Suppose we fully observe the process. Then, the corresponding likelihood is

$$L\left(\sigma^2, \mathbf{X}_n\right) = \prod_{i=1}^{n} p\left(X_{t_i} \mid X_{t_{i-1}}\right)$$

where, $p\left(X_{t_i} \mid X_{t_{i-1}}\right) = \left(2\pi\sigma^2\Delta\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2\Delta}\left(X_{t_i} - X_{t_{i-1}}\right)^2\right\}$. The log-likelihood is

$$l\left(\sigma^2, \mathbf{X}_n\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\Delta\right) - \frac{1}{2\sigma^2\Delta}\sum_{i=1}^{n}\left(X_{t_i} - X_{t_{i-1}}\right)^2$$

The score is

$$\dot{l}\left(\sigma^2, \mathbf{X}_n\right) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4\Delta}\sum_{i=1}^{n}\left(X_{t_i} - X_{t_{i-1}}\right)^2 = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}\hat{\sigma}_n^2$$

where $\hat{\sigma}_n^2 = \frac{1}{n\Delta}\sum_{i=1}^{n}\left(X_{t_i} - X_{t_{i-1}}\right)^2$ is the MLE, i.e., the solution to the score equation. Taking the expectation under the true parameter σ_0^2 we have that $E\left(\dot{l}\left(\sigma^2, \mathbf{X}_n\right)\right) = 0$. The derivation of the Fisher information relies on

$$\ddot{l}\left(\sigma^2, \mathbf{X}_n\right) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6\Delta}\sum_{i=1}^{n}\left(X_{t_i} - X_{t_{i-1}}\right)^2 = \frac{n}{2\sigma^4} - \frac{n}{\sigma^6}\hat{\sigma}_n^2$$

Denote by $\mathcal{I}_n$ the Fisher Information, derived from

$$\mathcal{I}_n = E\left(-\ddot{l}\left(\sigma^2, \mathbf{X}_n\right)\right) = \frac{n}{2\sigma_0^4} = n\mathcal{I}_0$$

where $\mathcal{I}_0 = \frac{1}{2\sigma_0^4}$ and the expectation is taken under the true parameter.

2.5.2 Likelihood of Rounded Data

The data are observed with rounding errors which occur in the original scale. Therefore, if α_n denotes the accuracy of the measurement, then, by denoting $s^{(\alpha_n)} = \alpha_n \lfloor s/\alpha_n \rfloor$, we have that at arrival times $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$

$$S_{t_i}^{(\alpha_n)} \leq S_{t_i} \leq S_{t_i}^{(\alpha_n)} + \alpha_n$$

Therefore, we observe $R_n = [b_{t_1}, a_{t_1}] \times \cdots \times [b_{t_n}, a_{t_n}]$, such that $b_{t_i} = \log S_{t_i}^{(\alpha_n)}$ and $a_{t_i} = \log\left(S_{t_i}^{(\alpha_n)} + \alpha_n\right)$. This implies

$$b_{t_i} \leq X_{t_i} \leq a_{t_i}$$

Let $\mathbf{X}^\star$ be an independent copy of $\mathbf{X}$ such that $b_{t_i} \leq X_{t_i}^\star \leq a_{t_i}$. To state this differently,

$$\mathbf{X}^\star \sim \frac{\mathbf{1}_{R_n} L\left(\mathbf{X}_n, \sigma^2\right)}{L\left(\sigma^2, R_n\right)}$$

We are interested in the probability/likelihood of the data R_n.

$$L\left(\sigma^2, R_n\right) = \int_{b_{t_1}}^{a_{t_1}} p\left(X_{t_1}^\star \mid x_{t_0}\right) dX_{t_1}^\star \cdots \int_{b_{t_n}}^{a_{t_n}} p\left(X_{t_n}^\star \mid X_{t_{n-1}}^\star\right) dX_{t_n}^\star = \int_{R_n} L\left(\sigma^2, \mathbf{X}_n^\star\right) d\mathbf{X}_n^\star \tag{2.5.1}$$

Lemma 1. *Denote by $l\left(\sigma^2, R_n\right) = \log L\left(\sigma^2, R_n\right)$ the log-likelihood 2.5.1. The score equation is given by*

$$\dot{l}\left(\sigma^2, R_n\right) = \frac{n}{2\sigma^4}\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma^2\right)$$

The solution to the score equation yields a "maximum likelihood estimator". That is

$$\sigma_{R_n}^2 = E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)$$

where $\hat{\sigma}_n^{2\star} = \frac{1}{n\Delta}\sum_{i=1}^n \left(X_{t_i}^\star - X_{t_{i-1}}^\star\right)^2$ is the MLE had we observed the the data $\mathbf{X}_n$. The expectation is taken with respect to the density $\frac{\mathbf{1}_{R_n} L\left(\mathbf{X}_n, \sigma^2\right)}{L(\sigma^2, R_n)}$. The Fisher information is

$$E\left(-\ddot{l}\left(\sigma^2, R_n\right)\right) = \mathcal{I}_n - \mathcal{I}_n^2 E\left(Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)\right)$$

where $Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) = E\left(\left(\hat{\sigma}_n^{2\star} - \sigma_{R_n}^2\right)^2 \mid \mathbf{X}_n^\star \in R_n\right)$ *and the expectation is taken under the true value* σ_0^2.

Remark 1. The second term on the RHS of the Fisher Information expresses the loss of information due to the discretization error.

Denote the local parameter by $h = \sqrt{n}\left(\sigma^2 - \sigma_0^2\right)$ and the corresponding log-likelihood by $l_{R_n,h} = l\left(\sigma_0^2 + {}^h/_{\sqrt{n}}, R_n\right)$. Expanding the log-likelihood around the local parameter we have that

$$l_{R_n,h} - l_{R_n,0} = \frac{h}{\sqrt{n}}\dot{l}\left(\sigma_0^2, R_n\right) + \frac{1}{2}\frac{h^2}{n}\ddot{l}\left(\sigma_0^2, R_n\right) + Rem_{n,h} \tag{2.5.2}$$

where $Rem_{n,h} = \frac{1}{6}\frac{h^3}{n^{3/2}}\dddot{l}_{R_n}\left(\bar{\sigma}^2, R_n\right)$ for some $\bar{\sigma}^2$ such that $\sigma_0^2 \leq \bar{\sigma}^2 \leq \sigma_0^2 + {}^h/_n$. Recall that $\mathcal{I}_n = \frac{n}{2\sigma_0^4}$, then we can easily show that

$$\dot{l}\left(\sigma_0^2, R_n\right) = \mathcal{I}_n\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2\right) \tag{2.5.3}$$

and

$$\ddot{l}_{R_n}\left(\sigma_0^2, R_n\right) = -\mathcal{I}_n\left(1 - \mathcal{I}_n Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)\right) - \frac{2}{\sigma_0^2}\mathcal{I}_n\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2\right) \tag{2.5.4}$$

Substituting equations 2.5.3 and 2.5.4 in equation 2.5.2 we have that

$$\begin{aligned} l_{R_n,h} - l_{R_n,0} &= \frac{h}{\sqrt{n}}\mathcal{I}_n\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2\right) - \frac{h^2}{2n}\mathcal{I}_n\left(1 - \mathcal{I}_n Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)\right) \\ &\quad - \frac{h^2}{n\sigma_0^2}\mathcal{I}_n\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2\right) + Rem_{n,h} \\ &= \frac{h}{\sqrt{n}}\mathcal{I}_n\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2\right) - \frac{h^2}{2n}\mathcal{I}_n\left(1 - \mathcal{I}_n Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)\right) \\ &\quad + Rem_{n,h} + o_p\left(1\right) \end{aligned} \tag{2.5.5}$$

The $o_p\left(1\right)$ term in equation 2.5.5 stems from the fact that once we establish the behavior of the score function, that is, $\frac{1}{\sqrt{n}}\dot{l}\left(\sigma_0^2, R_n\right)$, then

$$\frac{1}{n\sigma_0^2}\mathcal{I}_n\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2\right) = \frac{1}{\sqrt{n}\sigma_0^2}\frac{1}{\sqrt{n}}\dot{l}\left(\sigma_0^2, R_n\right) = o_p\left(1\right)$$

Denote

$$S_n = \frac{1}{\sqrt{n}}\mathcal{I}_n \left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2 \right)$$

and

$$F_n = \frac{1}{n}\mathcal{I}_n \left(1 - \mathcal{I}_n Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) \right)$$

Then, equation 2.5.5 becomes

$$l_{R_n,h} - l_{R_n,0} = S_n h - \frac{1}{2} F_n h^2 + Rem + o_p(1) \tag{2.5.6}$$

2.5.3 Approximation of the distribution of the score equation

To simplify the notation, we fix $T = 1$ and $t_i - t_{i-1} = \Delta = \frac{1}{n}$ or simply $t_i = \frac{i}{n}$. Before we prove the theorem, we need to understand the behavior of the quantity $\sqrt{n}\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \hat{\sigma}_n^{2\star}\right)$, since we can rewrite the score equation as

$$\frac{1}{\sqrt{n}}\mathcal{I}_n \left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2 \right) = \frac{1}{\sqrt{n}}\mathcal{I}_n \left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \hat{\sigma}_n^{2\star} \right) + \frac{1}{\sqrt{n}}\mathcal{I}_n \left(\hat{\sigma}_n^{2\star} - \sigma_0^2 \right)$$

Consider the collection of σ-fields $\mathcal{F}_{k,n} = \sigma\left\{X_1^\star, \ldots X_k^\star, R_n\right\}$ for $k \leq n$. Clearly, $\mathcal{F}_{k-1,n} \subseteq \mathcal{F}_{k,n}$. Then, denote

$$\xi_{n,k} = E\left(\hat{\sigma}_n^{2\star} \mid \mathcal{F}_{k-1,n}\right) - E\left(\hat{\sigma}_n^{2\star} \mid \mathcal{F}_{k,n}\right)$$

which is a martingale difference. Notice that

$$E\left(\hat{\sigma}_n^2 \mid \mathcal{F}_{0,n}\right) - \hat{\sigma}_n^2 = \sum_{k=1}^{n} \xi_{n,k}$$

which is a martingale. Then, there is a constant C such that

$$|\xi_{n,k}| \leq C\alpha_n^2$$

The following lemma states Azuma's inequality without proof.

Lemma 2. *(Azuma's Inequality) Let $Y_0, \ldots, Y_n$ be a martingale with bounded differences,*

namely, $|Y_i - Y_{i-1}| \leq m_i$. *Then*

$$P\left(|Y_n - Y_0| \geq t\right) \leq 2\exp\left\{-\frac{t^2}{2\sum_{i=1}^n m_i^2}\right\}$$

Using Lemma 2 we can derive an upper bound on the variance of $\sqrt{n}\sum_{k=1}^n \xi_{n,k}$

$$\begin{aligned} E\left(\sqrt{n}\sum_{k=1}^n \xi_{n,k}\right)^2 &= \int_0^\infty P\left(\left(\sqrt{n}\sum_{k=1}^n \xi_{n,k}\right)^2 \geq t\right) dt = \int_0^\infty P\left(\left|\sum_{k=1}^n \xi_{n,k}\right| \geq \frac{\sqrt{t}}{\sqrt{n}}\right) dt \\ &\leq \int_0^\infty 2\exp\left\{-\frac{t}{2n\sum_{i=1}^n \alpha_n^4 C^2}\right\} dt \\ &\leq 4n^2\alpha_n^4 C^2 \end{aligned}$$

Assuming that $\alpha_n = O_p\left(n^{-\frac{1}{2}}\right)$, then the upper bound of the variance then upper bound on the variance of $\sqrt{n}\sum_{k=1}^n \xi_{n,k}$ is $O_p(1)$. Assuming that $\alpha_n = O_p\left(n^{-\frac{1}{2}-\varepsilon}\right)$, then the bound is $o_p(1)$. Under the assumption $\alpha_n = O_p\left(n^{-\frac{1}{2}-\varepsilon}\right)$, we can easily see that

$$nVar\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) = E\left(\left(\sqrt{n}\sum_{k=1}^n \xi_{n,k}\right)^2 \mid \mathbf{X}_n^\star \in R_n\right) \to 0$$

which implies $\mathcal{I}_n Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) \to 0$. Then

$$S_n = \frac{1}{\sqrt{n}}\mathcal{I}_n\left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_0^2\right) \to N\left(0, \mathcal{I}_0\right) \tag{2.5.7}$$

and

$$F_n = \frac{1}{n}\mathcal{I}_n\left(1 - \mathcal{I}_n Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)\right) \to \mathcal{I}_0 \tag{2.5.8}$$

2.5.4 Generalized Fiducial Density

The derivation of the generalized fiducial density in the case of interval data is a rather difficult task. The reason is that the set $Q\left((\mathbf{b}_n, \mathbf{a}_n], \mathbf{Z}\right)$ is an interval, and therefore, the are more than one values of σ. However, we can derive a generalized fiducial distribution of one of its extremal points. We note that both these points are the outcome of using some of data inequalities with equality. Then, we know that the extremal point was generated by some

$x_0 + \sigma\sqrt{m/n}Z_{m/n} = a_{m/n}$[3]. Then, using this information and properties of the brownian path we can arrive at a unique solution for σ, namely

$$\sigma = \frac{a_{m/n} - x_0}{\sqrt{m/n}Z_{m/n}} \tag{2.5.9}$$

Denote $R_{-m} = \left[b_{1/n}, a_{1/n}\right] \times \cdots \times \left[b_{(m-1)/n}, a_{(m-1)/n}\right] \times \left[b_{(m+1)/n}, a_{(m+1)/n}\right] \times \cdots \times [b_1, a_1]$. That particular σ and observations R_{-m}, will be used to derive the generalized fiducial of σ.

We start by determining the joint density of (σ, R_{-m}), that is

$$f_{\sigma,R_{-m}}\left(\sigma, R_{-m}|a_{m/n}\right) = f\left(\sigma, |a_{m/n}\right)\tilde{L}\left(\sigma^2, R_{1:m-1}|x_0, a_{m/n}\right)\tilde{L}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right) \tag{2.5.10}$$

where $\tilde{L}\left(\sigma^2, R_{1:m-1}|x_0, a_{m/n}\right)$ is the likelihood of the process X_t starting at x_0, going through $R_{1:m-1} = \left[b_{1/n}, a_{1/n}\right] \times \cdots \times \left[b_{(m-1)/n}, a_{(m-1)/n}\right]$ and ending at $a_{m/n}$. In other words, X_t, for this section, behaves like a brownian bridge. $\tilde{L}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right)$ is the probability likelihood of the process X_t starting at $a_{m/n}$ and going through $R_{m+1:n} = \left[b_{(m+1)/n}, a_{(m+1)/n}\right] \times \cdots \times [b_1, a_1]$. Finally, $f\left(\sigma, |a_{m/n}\right)$ is the fiducial density of σ derived from equation 2.5.9, namely

$$f\left(\sigma, |a_{m/n}\right) = \varphi\left(g^{-1}\left(\sigma, a_{m/n}\right)\right)\left|\det\left(\frac{\partial}{\partial\sigma}g^{-1}\left(\sigma, a_{m/n}\right)\right)\right| = \frac{\left|a_{m/n} - x_0\right|}{\sigma}p\left(a_{m/n} \mid x_0\right)$$

where $g^{-1}\left(\sigma, a_{m/n}\right) = \frac{a_{m/n} - x_0}{\sigma\sqrt{m/n}}$ and φ is the density of the standard normal. Using the properties of the brownian bridge, we can show that the joint density of a process starting at $X_{t_0} = x_0$, observed at times $\{0 = t_0 < t_1 < ... < t_n < T\}$ and ending at $X_T = x_T$ is

$$\frac{p\left(x_T \mid x_{t_n}\right)}{p\left(x_T \mid x_0\right)}\prod_{i=1}^{n}p\left(x_{t_i} \mid x_{t_{i-1}}\right)$$

Therefore, denote $\tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}|x_0, a_{m/n}\right) = L\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}\right)p\left(a_{m/n} \mid x_{(m-1)/n}\right)/p\left(a_{m/n} \mid x_0\right)$. So,

$$\tilde{L}\left(\sigma^2, R_{1:m-1}|x_0, a_{m/n}\right) = \int_{R_{1:m-1}}\tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}|x_0, a_{m/n}\right)d\mathbf{X}^{\star}_{1:m-1} \tag{2.5.11}$$

[3]We could have used $b_{m/n}$.

and

$$\tilde{L}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right) = \int_{R_{m+1:n}} p\left(X^{\star}_{(m+1)/n} \mid a_{m/n}\right) L\left(\sigma^2, \mathbf{X}^{\star}_{m+2:n}\right) d\mathbf{X}^{\star}_{m+1:n} \tag{2.5.12}$$

Using 2.5.11 and 2.5.12, we can re-write 2.5.10, namely,

$$\begin{aligned}
&f_{\sigma,R_{-m}}\left(\sigma, R_{-m}|a_{m/n}\right) \\
&= \int_{R_{1:m-1}} \frac{\left|a_{m/n} - x_0\right|}{\sigma} p\left(a_{m/n} \mid x_0\right) \tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}|x_0, a_{m/n}\right) d\mathbf{X}^{\star}_{1:m-1} \times \tilde{L}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right) \\
&= \int_{R_{1:m-1}} \frac{\left|a_{m/n} - x_0\right|}{\sigma} \tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right) d\mathbf{X}^{\star}_{1:m-1} \times \tilde{L}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right) \\
&= \mathcal{J}\left(a_{m/n}, \sigma\right) \tilde{L}\left(\sigma^2, R_{1:m-1}, a_{m/n}|x_0\right) \tilde{L}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right)
\end{aligned} \tag{2.5.13}$$

where $\mathcal{J}\left(a_{m/n}, \sigma\right) = \frac{\left|a_{m/n} - x_0\right|}{\sigma}$ is the Jacobian factor and $\tilde{L}\left(\sigma^2, R_{1:m-1}, a_{m/n}|x_0\right) = \int_{R_{1:m-1}} \tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right) d\mathbf{X}^{\star}_{1:m-1}$. The generalized fiducial density of the parameter σ is given by

$$g_n\left(\sigma|a_{m/n}, R_{-m}\right) = \frac{f_{\sigma,R_{-m}}\left(\sigma, R_{-m}|a_{m/n}\right)}{\int_{\Theta} f_{\sigma,R_{-m}}\left(\sigma, R_{-m}|a_{m/n}\right) d\sigma} \tag{2.5.14}$$

where $\Theta = \mathbb{R}_+$.

2.5.5 Bernstein-von Mises theorem

Previously, we determined the expansion of the likelihood of rounded data. The same derivations apply if we replace $L\left(\sigma^2, R_n\right)$ with $\tilde{L}\left(\sigma^2, R_{1:m-1}, a_{m/n}|x_0\right) \tilde{L}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right)$. Denote by

$$l\left(\sigma^2, \tilde{R}_n\right) = l\left(\sigma^2, R_{-m}, a_{m/n}\right) = \log\left(\tilde{L}\left(\sigma^2, R_{1:m-1}, a_{m/n}|x_0\right) \tilde{L}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right)\right) \tag{2.5.15}$$

It is very easy to see that Lemma 1 applies at modified likelihood. In this case

$$\sigma^2_{R_n} = E\left(\hat{\sigma}^{2\star}_n \mid \mathbf{X}^{\star}_{-m} \in R_{-m}, X^{\star}_{m/n} = a_{m/n}\right)$$

where $\hat{\sigma}_n^{2\star}$ is simply the MLE when we replace $X^{\star}_{m/n}$ with $a_{m/n}$. The expansion 2.5.6 for the modified likelihood remains the same, namely,

$$l_{R_n,h} - l_{R_n,0} = S_n h - \frac{1}{2} F_n h^2 + Rem + o_p(1) \tag{2.5.16}$$

where

$$S_n = \frac{1}{\sqrt{n}} \mathcal{I}_n \left(E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}^{\star}_{-m} \in R_{-m}, X^{\star}_{m/n} = a_{m/n} \right) - \sigma_0^2 \right) \tag{2.5.17}$$

and

$$F_n = \frac{1}{n} \mathcal{I}_n \left(1 - \mathcal{I}_n Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}^{\star}_{-m} \in R_{-m}, X^{\star}_{m/n} = a_{m/n} \right) \right) \tag{2.5.18}$$

Under $\alpha_n = O_p\left(n^{-\frac{1}{2}-\varepsilon} \right)$, the results 2.5.7 and 2.5.8 extend to the modified versions, namely

$$S_n \to N(0, \mathcal{I}_0) \tag{2.5.19}$$

and

$$F_n \to \mathcal{I}_0 \tag{2.5.20}$$

We now collect the required conditions for the theorem.

1. The likelihood 2.5.15 satisfies the standard regularity conditions. That is, is thrice differentiable with respect to σ^2 in a neighborhood $\left(\sigma_0^2 - \delta, \sigma_0^2 + \delta\right)$.

2. In view of lemma 1, $E\left(\dot{l}\left(\sigma^2, R_{-m}, a_{m/n} \right) \right) = 0$ and

$$E\left(-\ddot{l}\left(\sigma^2, R_{-m}, a_{m/n} \right) \right) = \mathcal{I}_n - \mathcal{I}_n^2 E\left(Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}^{\star}_{-m} \in R_{-m}, X^{\star}_{m/n} = a_{m/n} \right) \right)$$

   is finite.

3. Under $\alpha_n = O_p\left(n^{-\frac{1}{2}-\varepsilon} \right)$, the results 2.5.19 and 2.5.20 hold. Moreover,

$$\sup_{\bar{\sigma}^2 \in \left(\sigma_0^2 - \delta, \sigma_0^2 + \delta\right)} \frac{1}{n} \dddot{l}\left(\bar{\sigma}^2, R_{-m}, a_{m/n} \right) = O_p(1)$$

4. For any $\delta > 0$ there is $\varepsilon > 0$ such that

$$\sup_{\sigma^2 \notin \left(\sigma_0^2-\delta, \sigma_0^2+\delta\right)} P\left\{\frac{1}{n}\left(\tilde{l}\left(\sigma^2, R_{1:m-1}, a_{m/n}|x_0\right) - \tilde{l}\left(\sigma_0^2, R_{1:m-1}, a_{m/n}|x_0\right)\right) \leq -\varepsilon\right\} \to 1$$

and similarly

$$\sup_{\sigma^2 \notin \left(\sigma_0^2-\delta, \sigma_0^2+\delta\right)} P\left\{\frac{1}{n}\left(\tilde{l}\left(\sigma^2, R_{m+1:n}|a_{m/n}\right) - \tilde{l}\left(\sigma_0^2, R_{m+1:n}|a_{m/n}\right)\right) \leq -\varepsilon\right\} \to 1$$

5. Under the local parameter $h = \sqrt{n}\left(\sigma^2 - \sigma_0^2\right)$, let $\mathcal{J}_{n,h} = \mathcal{J}\left(a_{m/n}, h\right)$. Then $\mathcal{J}_{n,h} \to \mathcal{J}_0$, where $\mathcal{J}_0$ is continuous and positive on a neighborhood of σ_0^2.

The conditions are verified below.

We now re-write the generalized fiducial density 2.5.14 in terms of the local parameter $h = \sqrt{n}\left(\sigma^2 - \sigma_0^2\right)$ and the expansion 2.5.16 that is

$$g_n\left(h|\tilde{R}_n\right) = \frac{\mathcal{J}_{n,h} \exp\left\{l_{\tilde{R}_n,h} - l_{\tilde{R}_n,0}\right\}}{\int_{H_n} \mathcal{J}_{n,h} \exp\left\{l_{\tilde{R}_n,h} - l_{\tilde{R}_n,0}\right\} dh} \tag{2.5.21}$$

where H_n is the parameter space under the local parameter h.

Theorem 1. *(Bernstein-von Mises) Let $\alpha_n = O_p\left(n^{-\frac{1}{2}-\varepsilon}\right)$. Under the conditions 1-5, the total variation between the density 2.5.21 and the density of the normal distribution $N\left(X, \mathcal{I}_0^{-1}\right)$, with $X \sim N\left(0, \mathcal{I}_0\right)$, converges to zero in probability. That is*

$$\int_{H_n} \left|g_n\left(h|\tilde{R}_n\right) - \varphi_{\mathcal{I}_0^{-1}X, \mathcal{I}_0^{-1}}(h)\right| dh \xrightarrow{P} 0$$

Proof. Since

$$\begin{aligned}\int_{H_n} \left|g_n\left(h|\tilde{R}_n\right) - \varphi_{\mathcal{I}_0^{-1}X, \mathcal{I}_0^{-1}}(h)\right| dh \leq C_n^{-1} \int_{H_n} \left|\frac{\mathcal{J}_{n,h} \exp\left\{l_{\tilde{R}_n,h}\right\}}{\exp\left\{l_{\tilde{R}_n,0}\right\}} - \mathcal{J}_0 e^{S_n h - \frac{1}{2}F_n h^2}\right| dh \\ + \int_{H_n} \left|C_n^{-1}\mathcal{J}_0 e^{S_n h - \frac{1}{2}F_n h^2} - \varphi_{\mathcal{I}_0^{-1}X, \mathcal{I}_0^{-1}}(h)\right| dh\end{aligned} \tag{2.5.22}$$

where $C_n = \int_{H_n} \mathcal{J}_{n,h} \exp\left\{l_{\tilde{R}_n,h} - l_{\tilde{R}_n,0}\right\} dh$ and $\mathcal{J}_{n,h} \to \mathcal{J}_0$, it suffices to show that the

right hand side (RHS) of 2.5.22 converges to zero in probability. This can be established by showing that

$$\int_{H_n} \left| \frac{\mathcal{J}_{n,h} \exp\left\{ l_{\tilde{R}_n,h} \right\}}{\exp\left\{ l_{\tilde{R}_n,0} \right\}} - \mathcal{J}_0 e^{S_n h - \frac{1}{2} F_n h^2} \right| dh \xrightarrow{P} 0 \tag{2.5.23}$$

and $C_n^{-1} = O_P(1)$. The latter can be easily established since the result in 2.5.23 implies $\left| C_n - \int_{H_n} \mathcal{J}_0 e^{S_n h - \frac{1}{2} F_n h^2} dh \right| \xrightarrow{P} 0$. Also, we can easily show that

$$C_1 e^{-\frac{1}{2} F_n^{-1} S_n^2} \leq \int_{H_n} e^{S_n h - \frac{1}{2} F_n h^2} dh \leq C_2 e^{-\frac{1}{2} F_n^{-1} S_n^2}$$

for some constants C_1 and C_2[4], therefore $C_n = O_P(1)$. For the second integral of the RHS of 2.5.22, let $I_1 = \int_{H_n} e^{S_n h - \frac{1}{2} F_n h^2} dh$ and $I_2 = \left| C_n^{-1} \mathcal{J}_0 - I_1^{-1} \right|$. It follows that second integral of the RHS 2.5.22 of can be expressed as $I_1 I_2$ and is $o_P(1)$, since I_1 is $O_P(1)$ and I_2 is $o_P(1)$ by 2.5.23. To establish 2.5.22, we split the parameter space H_n into 3 regions and establish convergence in each one separately. The regions are,

$$A_{1,n} = \left\{ h : |h| < c \log\left(\sqrt{n}\right) \right\} \cap H_n$$

$$A_{2,n} = \left\{ h : c \log\left(\sqrt{n}\right) < |h| < \delta \sqrt{n} \right\} \cap H_n$$

$$A_{3,n} = \left\{ h : |h| > \frac{\delta \sqrt{n}}{2} \right\} \cap H_n$$

for suitably chosen positive constants κ and λ.

Region $A_{2,n}$: For the integral in equation 2.5.23 we have that

$$\begin{aligned} &\int_{A_{2,n}} \left| \mathcal{J}_{n,h} e^{S_n h - \frac{1}{2} F_n h^2 + Rem_{n,h}} - \mathcal{J}_0 e^{S_n h - \frac{1}{2} F_n h^2} \right| dh \\ &\quad \leq \int_{A_{2,n}} \mathcal{J}_{n,h} e^{S_n h - \frac{1}{2} F_n h^2 + Rem_{n,h}} dh + \int_{A_{2,n}} \mathcal{J}_0 e^{S_n h - \frac{1}{2} F_n h^2} dh \\ &\quad \leq \sup_{h \in A_{2,n}} \mathcal{J}_{n,h} \int_{A_{2,n}} e^{S_n h - \frac{1}{2} F_n h^2 + Rem_{n,h}} dh + \int_{A_{2,n}} \mathcal{J}_0 e^{S_n h - \frac{1}{2} F_n h^2} dh \end{aligned} \tag{2.5.24}$$

[4]The constants C_1 and C_2 exist due to the fact that $\Theta - \left\{ \sigma_0^2 \right\} \subset H_n \subset \mathbb{R}$.

Notice that

$$\int_{A_{2,n}} \mathcal{J}_0 e^{-\frac{1}{2}F_n h^2} dh \leq 2\mathcal{J}_0 e^{-\frac{1}{2}F_n c\log\left(\sqrt{n}\right)} \left(\delta\sqrt{n} - c\log\left(\sqrt{n}\right)\right) \leq K\mathcal{J}_0 n^{\frac{1}{2}-\frac{1}{4}cF_n}$$

therefore, choosing c large enough, the integral goes to zero. As a result, the second integral in the RHS of 2.5.24 also goes to zero. For the first integral in the RHS of 2.5.24, we need a bound for the remainder. Since, $Rem_{n,h} = \frac{1}{6}\frac{h^3}{n^{3/2}} \dddot{l}_{R_n}\left(\bar{\sigma}^2\right)$ for some $\bar{\sigma}^2$ such that $\bar{\sigma}^2 \in \left(\sigma_0^2 - \delta, \sigma_0^2 + \delta\right)$, then, in the region $A_{2,n}$

$$|Rem_{n,h}| = \frac{1}{6}\frac{h^3}{n^{3/2}} \dddot{l}_{\tilde{R}_n}\left(\bar{\sigma}^2\right) \leq \delta\frac{1}{6}h^2\frac{1}{n} \dddot{l}_{\tilde{R}_n}\left(\bar{\sigma}^2\right)$$

Since $\sup_{\bar{\sigma}^2 \in \left(\sigma_0^2-\delta, \sigma_0^2+\delta\right)} \frac{1}{n}\left|\dddot{l}_{\tilde{R}_n}\left(\bar{\sigma}^2\right)\right| = O_p\left(1\right)$, we can choose δ small enough to ensure

$$P\left(Rem_{n,h} < \frac{1}{4}F_n h^2;\ for\ all\ h \in A_{2,n}\right) > 1 - \epsilon$$

for some $\epsilon > 0$ or equivalently

$$P\left(S_n h - \frac{1}{2}F_n h^2 + Rem_{n,h} < S_n h - \frac{1}{4}F_n h^2;\ for\ all\ h \in A_{2,n}\right) > 1 - \epsilon$$

therefore, with probability $1 - \epsilon$ we have that

$$\sup_{h \in A_{2,n}} \mathcal{J}_{n,h} \int_{A_{2,n}} e^{S_n h - \frac{1}{2}F_n h^2 + Rem_{n,h}} dh \leq \sup_{h \in A_{2,n}} \mathcal{J}_{n,h} \int_{A_{2,n}} e^{S_n h - \frac{1}{4}F_n h^2 +} dh$$

which also goes to zero using a similar argument as before.

Region $A_{1,n}$: For the integral in equation 2.5.23 we have that

$$\begin{aligned}
&\int_{A_{1,n}} \left| \mathcal{J}_{n,h} \exp\left\{ l_{\tilde{R}_n,h} - l_{\tilde{R}_n,0} \right\} - \mathcal{J}_0 e^{S_n h - \frac{1}{2} F_n h^2} \right| dh \\
&\quad \leq \int_{A_{1,n}} \left| \mathcal{J}_{n,h} e^{S_n h - \frac{1}{2} F_n h^2 + Rem_{n,h}} - \mathcal{J}_{n,h} e^{S_n h - \frac{1}{2} F_n h^2} \right| dh + \int_{A_{1,n}} \left| \mathcal{J}_{n,h} - \mathcal{J}_0 \right| e^{S_n h - \frac{1}{2} F_n h^2} dh \\
&\quad \leq \int_{A_{1,n}} \mathcal{J}_{n,h} e^{S_n h - \frac{1}{2} F_n h^2} \left| e^{Rem_{n,h}} - 1 \right| dh + \int_{A_{1,n}} \left| \mathcal{J}_{n,h} - \mathcal{J}_0 \right| e^{S_n h - \frac{1}{2} F_n h^2} dh \\
&\quad \leq \sup_{h \in A_{1,n}} \mathcal{J}_{n,h} \int_{A_{1,n}} e^{S_n h - \frac{1}{2} F_n h^2} e^{Rem_{n,h}} \left| Rem_{n,h} \right| dh + \int_{A_{1,n}} \left| \mathcal{J}_{n,h} - \mathcal{J}_0 \right| e^{S_n h - \frac{1}{2} F_n h^2} dh
\end{aligned} \tag{2.5.25}$$

For the first integral in the RHS of 2.5.25, as before, we need a bound for the remainder. In the region $A_{1,n}$

$$\left| Rem_{n,h} \right| = \frac{1}{6} \frac{h^3}{n^{3/2}} \dddot{l}_{\tilde{R}_n}\left(\bar{\sigma}^2\right) \leq \frac{1}{6} \frac{\left(c \log\left(\sqrt{n}\right)\right)^3}{n^{3/2}} \dddot{l}_{\tilde{R}_n}\left(\bar{\sigma}^2\right) = \frac{c^3}{6} \frac{\left(\log\left(\sqrt{n}\right)\right)^3}{\sqrt{n}} \frac{1}{n} \dddot{l}_{\tilde{R}_n}\left(\bar{\sigma}^2\right) = o_p\left(1\right)$$

thus, the first integral in the RHS of 2.5.25 is $o_p\left(1\right)$. The second integral in the RHS of 2.5.25 is also $o_p\left(1\right)$ since $\mathcal{J}$ is continuous at σ_0^2.

Region $A_{3,n}$: For the integral in equation 2.5.23 we have that

$$\begin{aligned}
&\int_{A_{3,n}} \left| \mathcal{J}_{n,h} e^{S_n h - \frac{1}{2} F_n h^2 + Rem_{n,h}} - \mathcal{J}_0 e^{S_n h - \frac{1}{2} F_n h^2} \right| dh \\
&\quad \leq \int_{A_{3,n}} \mathcal{J}_{n,h} e^{S_n h - \frac{1}{2} F_n h^2 + Rem_{n,h}} dh + \int_{A_{3,n}} \mathcal{J}_0 e^{S_n h - \frac{1}{2} F_n h^2} dh
\end{aligned} \tag{2.5.26}$$

Clearly, the second integral is $o_p\left(1\right)$. For the first integral we have that

$$\begin{aligned}
&\int_{A_{3,n}} \mathcal{J}_{n,h} e^{S_n h - \frac{1}{2} F_n h^2 + Rem_{n,h}} dh \\
&\quad = \int_{A_{3,n}} \mathcal{J}_{n,h} p\left(a_{m/n} \mid x_0\right) e^{l_{\tilde{R}_n,h} - l_{\tilde{R}_n,0} - \log p\left(a_{m/n} | x_0\right)} dh \\
&\quad = \int_{A_{3,n}} \mathcal{J}_{n,h} p\left(a_{m/n} \mid x_0\right) e^{\tilde{l}\left(h, R_{1:m-1}, a_{m/n} | x_0\right) - \tilde{l}\left(\sigma_0^2, R_{1:m-1}, a_{m/n} | x_0\right) - \log p\left(a_{m/n} | x_0\right)} \times \\
&\qquad\qquad e^{\tilde{l}\left(h, R_{m+1:n} | a_{m/n}\right) - \tilde{l}\left(\sigma_0^2, R_{m+1:n} | a_{m/n}\right)} dh
\end{aligned}$$

Notice that $\int_{H_n} \mathcal{J}_{n,h} p\left(a_{m/n} \mid x_0\right) dh = 1$ since $\mathcal{J}_{n,h} p\left(a_{m/n} \mid x_0\right)$ is a density. Due to conditions 5, the integral goes to zero. □

2.5.6 Proofs and Auxiliary Results

Proof of Lemma 1

Denote the likelihood as $L\left(\sigma^2, \mathbf{X}_n^\star\right) = \prod_{i=1}^n p\left(x_i^\star \mid x_{i-1}^\star\right)$. Then we can easily show that

$$\frac{\partial}{\partial \sigma^2} \log L\left(\sigma^2, \mathbf{X}_n^\star\right) = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}\hat{\sigma}_n^{2\star}$$

We are interested in the probability/likelihood

$$L\left(\sigma^2, R_n\right) = \int_{R_n} L\left(\sigma^2, \mathbf{X}_n^\star\right) d\mathbf{X}_n^\star$$

$$\begin{aligned}\frac{\partial L\left(\sigma^2, R_n\right)}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2}\int_{R_n} L\left(\sigma^2, \mathbf{X}_n^\star\right) d\mathbf{X}_n^\star = \int_{R_n} L\left(\sigma^2, \mathbf{X}_n^\star\right)\left\{-\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}\hat{\sigma}_n^{2\star}\right\} d\mathbf{X}_n^\star \\ &= -\frac{n}{2\sigma^2} L\left(\sigma^2, R_n\right) + \frac{n}{2\sigma^4} L\left(\sigma^2, R_n\right) E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)\end{aligned}$$

This implies

$$\frac{\partial \log L\left(\sigma^2, R_n\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4} E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}\sigma_{R_n}^2$$

where $\sigma_{R_n}^2 = E\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)$. In order to calculate the Fisher information we will use the following facts. First

$$\begin{aligned}\frac{\partial}{\partial \sigma^2} \log \frac{L\left(\sigma^2, \mathbf{X}_n^\star\right)}{L\left(\sigma^2, R_n\right)} &= \frac{\partial}{\partial \sigma^2} \log L\left(\sigma^2, \mathbf{X}_n^\star\right) - \frac{\partial}{\partial \sigma^2} \log L\left(\sigma^2, R_n\right) \\ &= -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}\hat{\sigma}_n^{2\star} - \left(-\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}\sigma_{R_n}^2\right) \\ &= \frac{n}{2\sigma^4}\left(\hat{\sigma}_n^{2\star} - \sigma_{R_n}^2\right)\end{aligned}$$

and second

$$\begin{aligned}
\frac{\partial}{\partial\sigma^2}\sigma^2_{R_n} &= \frac{\partial}{\partial\sigma^2}E\left(\hat{\sigma}^2_n \mid \mathbf{X}^\star_n \in R_n\right) = \frac{\partial}{\partial\sigma^2}\int_{R_n}\hat{\sigma}^2_n\frac{L\left(\sigma^2,\mathbf{X}^\star_n\right)}{L\left(\sigma^2,R_n\right)}d\mathbf{X}^\star_n \\
&= \frac{\partial}{\partial\sigma^2}\int_{R_n}\hat{\sigma}^{2\star}_n\exp\left\{\log\frac{L\left(\sigma^2,\mathbf{X}^\star_n\right)}{L\left(\sigma^2,R_n\right)}\right\}d\mathbf{X}^\star_n \\
&= \int_{R_n}\hat{\sigma}^2_n\frac{L\left(\sigma^2,\mathbf{X}^\star_n\right)}{L\left(\sigma^2,R_n\right)}\left\{\frac{\partial}{\partial\sigma^2}\log\frac{L\left(\sigma^2,\mathbf{X}^\star_n\right)}{L\left(\sigma^2,R_n\right)}\right\}d\mathbf{X}^\star_n \\
&= \frac{n}{2\sigma^4}\int_{R_n}\hat{\sigma}^{2\star}_n\left(\hat{\sigma}^{2\star}_n-\sigma^2_{R_n}\right)\frac{L\left(\sigma^2,\mathbf{X}^\star_n\right)}{L\left(\sigma^2,R_n\right)}d\mathbf{X}^\star_n \\
&= \frac{n}{2\sigma^4}\left\{E\left(\hat{\sigma}^{4\star}_n \mid \mathbf{X}^\star_n \in R_n\right)-\sigma^4_{R_n}\right\} = \frac{n}{2\sigma^4}Var\left(\hat{\sigma}^{2\star}_n \mid \mathbf{X}^\star_n \in R_n\right)
\end{aligned}$$

Therefore

$$\begin{aligned}
\ddot{l}_{R_n}\left(\sigma^2\right) &= \frac{\partial}{\partial\sigma^2}\left\{-\frac{n}{2\sigma^2}+\frac{n}{2\sigma^4}\sigma^2_{R_n}\right\} \\
&= \frac{n}{2\sigma^4}-\frac{n}{\sigma^6}\sigma^2_{R_n}+\frac{n^2}{4\sigma^8}Var\left(\hat{\sigma}^{2\star}_n \mid \mathbf{X}^\star_n \in R_n\right) \qquad (2.5.27)
\end{aligned}$$

The Fisher information is simply the expectation under the true value σ^2_0 of the negative of 2.5.27 and the fact that

$$E\left(\sigma^2_{R_n}\right) = E\left(E\left(\hat{\sigma}^{2\star}_n \mid \mathbf{X}^\star_n \in R_n\right)\right) = E\hat{\sigma}^{2\star}_n = \sigma^2_0$$

.

Verification of Conditions

Condition 3: For the third derivative we need $\frac{\partial}{\partial\sigma^2}Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) = \frac{\partial}{\partial\sigma^2}\left\{E\left(\hat{\sigma}_n^{4\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_{R_n}^4\right\}$. Therefore

$$\begin{aligned}
\frac{\partial}{\partial\sigma^2}E\left(\hat{\sigma}_n^{4\star} \mid \mathbf{X}_n^\star \in R_n\right) &= \frac{\partial}{\partial\sigma^2}\int_{R_n}\hat{\sigma}_n^{4\star}\frac{L\left(\sigma^2,\mathbf{X}_n\right)}{L\left(\sigma^2,R_n\right)}d\mathbf{X}_n^\star \\
&= \int_{R_n}\hat{\sigma}_n^{4\star}\frac{L\left(\sigma^2,\mathbf{X}_n^\star\right)}{L\left(\sigma^2,R_n\right)}\left\{\frac{\partial}{\partial\sigma^2}\log\frac{L\left(\sigma^2,\mathbf{X}_n^\star\right)}{L\left(\sigma^2,R_n\right)}\right\}d\mathbf{X}_n^\star \\
&= \frac{n}{2\sigma^4}\int_{R_n}\hat{\sigma}_n^{4\star}\left(\hat{\sigma}_n^{2\star}-\sigma_{R_n}^2\right)\frac{L\left(\sigma^2,\mathbf{X}_n^\star\right)}{L\left(\sigma^2,R_n\right)}d\mathbf{X}_n^\star \\
&= \frac{n}{2\sigma^4}\left\{E\left(\hat{\sigma}_n^{6\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_{R_n}^2 E\left(\hat{\sigma}_n^{4\star} \mid \mathbf{X}_n^\star \in R_n\right)\right\}
\end{aligned}$$

and

$$\frac{\partial}{\partial\sigma^2}\sigma_{R_n}^4 = 2\sigma_{R_n}^2\frac{\partial}{\partial\sigma^2}\sigma_{R_n}^2 = \frac{n}{\sigma^4}\sigma_{R_n}^2\left\{E\left(\hat{\sigma}_n^{4\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_{R_n}^4\right\}$$

So

$$\begin{aligned}
\frac{\partial}{\partial\sigma^2}Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) &= \frac{n}{2\sigma^4}\left\{E\left(\hat{\sigma}_n^{6\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_{R_n}^2 E\left(\hat{\sigma}_n^{4\star} \mid \mathbf{X}_n^\star \in R_n\right)\right\} \\
&\quad - \frac{n}{\sigma^4}\sigma_{R_n}^2\left\{E\left(\hat{\sigma}_n^{4\star} \mid \mathbf{X}_n^\star \in R_n\right) - \sigma_{R_n}^4\right\} \\
&= \frac{n}{2\sigma^4}E\left[\left(\hat{\sigma}_n^{2\star}-\sigma_{R_n}^2\right)^3 \mid \mathbf{X}_n^\star \in R_n\right]
\end{aligned}$$

Then,

$$\begin{aligned}
\dddot{l}_{R_n}\left(\sigma^2\right) &= \frac{\partial}{\partial\sigma^2}\left\{\frac{n}{2\sigma^4} - \frac{n}{\sigma^6}\sigma_{R_n}^2 + \frac{n^2}{4\sigma^8}Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right)\right\} \\
&= -\frac{n}{\sigma^6} + \frac{3n}{\sigma^8}\sigma_{R_n}^2 - \frac{n^2}{2\sigma^{10}}Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) \\
&\quad - \frac{n^2}{\sigma^{10}}Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) + \frac{n^2}{4\sigma^8}\frac{\partial}{\partial\sigma^2}Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) \\
&= -\frac{n}{\sigma^6} + \frac{3n}{\sigma^8}\sigma_{R_n}^2 - \frac{3n^2}{2\sigma^{10}}Var\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) + \frac{n^3}{8\sigma^{12}}E\left[\left(\hat{\sigma}_n^{2\star}-\sigma_{R_n}^2\right)^3 \mid \mathbf{X}_n^\star \in R_n\right]
\end{aligned}$$

Then, since $nVar\left(\hat{\sigma}_n^{2\star} \mid \mathbf{X}_n^\star \in R_n\right) = o_p\left(1\right)$ and $E\left[\left(\hat{\sigma}_n^{2\star}-\sigma_{R_n}^2\right)^3 \mid \mathbf{X}_n^\star \in R_n\right] = 0$

$$\frac{1}{n}\dddot{l}_{R_n}\left(\sigma^2\right) = O_p\left(1\right)$$

Condition 4: Recall that

$$\tilde{L}\left(\sigma^2, R_{1:m-1}, a_{m/n}|x_0\right) = \int_{R_{1:m-1}} \tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right) d\mathbf{X}^{\star}_{1:m-1}$$

For the log-likelihood $\tilde{l}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}\right) = \log \tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}\right)$, the following condition holds. For any $\delta > 0$, there is $\varepsilon > 0$ such that

$$\sup_{\sigma^2 \notin \left(\sigma_0^2-\delta, \sigma_0^2+\delta\right)} P\left\{\frac{1}{n}\left(\tilde{l}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right) - \tilde{l}\left(\sigma_0^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right)\right) \leq -\varepsilon\right\} \to 1$$

since this is a standard Gaussian log-likelihood. On this set,

$$\tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right) \leq e^{-n\varepsilon} \tilde{L}\left(\sigma_0^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right)$$

Then, the first result follows because

$$\begin{aligned}
\tilde{l}\left(\sigma^2, R_{1:m-1}, a_{m/n}|x_0\right) &= \log \int_{R_{1:m-1}} \tilde{L}\left(\sigma^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right) d\mathbf{X}^{\star}_{1:m-1} \\
&\leq \log \int_{R_{1:m-1}} e^{-n\varepsilon} \tilde{L}\left(\sigma_0^2, \mathbf{X}^{\star}_{1:m-1}, a_{m/n}|x_0\right) d\mathbf{X}^{\star}_{1:m-1} \\
&= -n\varepsilon + \tilde{l}\left(\sigma_0^2, R_{1:m-1}, a_{m/n}|x_0\right)
\end{aligned}$$

The second results can be derived in a similar manner.

2.6 Combinations of Fiducial Distributions

In this section we present a novel method to combine the generated fiducial samples from different blocks of data into one sample that summarizes all information. Combination schemes for fiducial distributions are not very common in the literature and most of them rely on ideas drawn from combination schemes for confidence distributions, see for example Hannig and Xie (2012) and the references therein. Additionally, all current combination schemes are for exact data and utilize the underlying likelihood functions. Applying a similar approach in our framework is a challenging task, since we are using interval data.

Our scheme aggregates the generated samples from all blocks under consideration by re-weighing all particles. The new weights are computed through a metric that utilizes the Gaussian kernel, together with point estimates of the sample mean and the Fisher information matrix. Traditionally, the Fisher information matrix, if it is not known in closed form, is approximated numerically from the log-likelihood. In our case, the corresponding log-likelihood is based on the transition probabilities $P_\sigma\left(b_{t_j} \leq \tilde{X}_{t_j} \leq a_{t_j} \mid b_{t_{j-1}} \leq \tilde{X}_{t_{j-1}} \leq a_{t_{j-1}}\right)$, where $\tilde{X}_t = X_t + U_t$. These probabilities require require a further approximation, for example, $P_\sigma\left(b_{t_j} \leq \tilde{X}_{t_j} \leq a_{t_j} \mid \tilde{X}_{t_{j-1}} = \frac{a_{t_{j-1}}+b_{t_{j-1}}}{2}\right)$, which renders the estimation numerically unstable. Our approach, utilizes the fiducial samples to estimate the Fisher information matrix. The procedure is the following.

Let (σ_i, W_i) denote the generated particle systems in block $(\tau_{i-1}, \tau_i]$, $i = 1, ...M$, where $\sigma_i = \left\{\sigma_i^{(k)}\right\}_{k=1}^N$, $W_i = \left\{w_i^{(k)}\right\}_{k=1}^N$ and N is the number of particles used in the simulation. Suppose we want to combine fiducial distributions from M segments. For each particle k in block i and every other block $j = 1, ..., M$, $i \neq j$, we calculate the following weight

$$W_{i,j}^{(k)} = \exp\left\{-\frac{1}{2} I_j \left(\sigma_i^{(k)} - \hat{\sigma}_j\right)^2 + \frac{1}{2}\log I_j\right\}$$

where $\hat{\sigma}_j$ denotes the point estimate $\hat{\sigma}_j = \sum_{k=1}^N W_j^{(k)} \sigma_j^{(k)}$ in block j and I_j denotes

$$I_j = \left(\sum_{k=1}^N W_j^{(k)} \left(\sigma_j^{(k)} - \hat{\sigma}_j\right)^2\right)^{-1}$$

which is the point estimate of Fisher information. In other words, for every particle k in block i, we calculate its "weighted" distance from the point estimate in block j through a Gaussian kernel, where $j = 1, ..., M$, $i \neq j$.

Once we generate weights $W_{i,j}^{(k)}$, the final weight for particle k in block i, is calculated by

$$\bar{W}_i^{(k)} = W_i^{(k)} \prod_{h \neq i} W_{i,h}^{(k)}$$

for all $i = 1, ..., M$.

To illustrate the effectiveness of the combination scheme we simulate a small samples

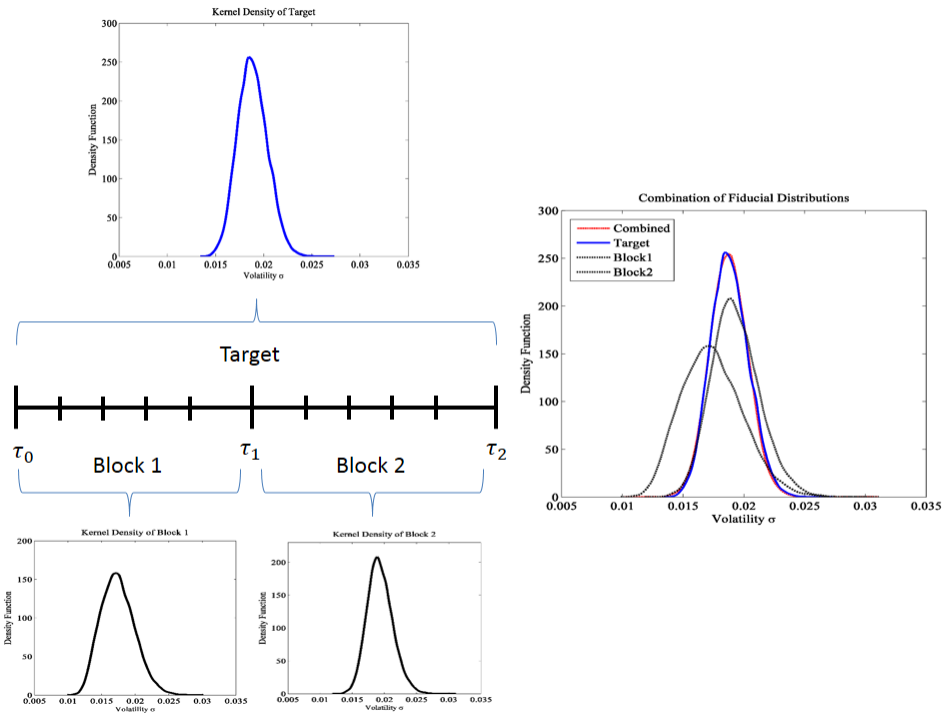


Figure 2.1: **Combination of block distributions.** A sample of 200 observations is split into two blocks of 100 observations. The fiducial distributions of the entire sample (target) and the two blocks are generated and subsequently combined. The combined distribution approximates the target remarkably well.

of 200. We split the sample in two blocks of 100 observations and generate the block distributions for each block and the entire sample. The distribution of the entire sample will serve as target distributions. Figure 2.1 illustrates the combination scheme. It is evident that the combined distributions approximate the target distributions remarkably well.

2.7 Simulation and Robustness checks

In this section we investigate our procedure by reporting a Monte Carlo simulation study under constant and stochastic volatility. We generate our data by simulating the efficient price process in the original scale and, at observation times, we round the process upwards and downwards, towards the two nearest ticks. Subsequently, we take the log transformation.

In other words, the upper and lower (ask and bid) prices at time t_m are given by

$$\begin{aligned} a_{t_m} &= \gamma \left\lceil \frac{exp(X_{t_m})}{\gamma} \right\rceil \\ b_{t_m} &= \gamma \left\lfloor \frac{exp(X_{t_m})}{\gamma} \right\rfloor \end{aligned}$$

where $\gamma = 0.01$ reflects that rounding occurs to the nearest tick.

This type of contamination is similar in spirit with the two stage contamination scheme of Li and Mykland (2007) and Jacod et al. (2009). In their framework, observed log-prices are given by $Y_{t_m} = \log\left(\gamma \left\lfloor \frac{exp(X_{t_m}+U_{t_m})}{\gamma} \right\rfloor\right)$ with the additive component being essential, since in a pure rounding model integrated variance cannot be estimated. Moreover, the proposed simulation scheme renders the choice of the starting price X_0 relevant, since the magnitude of the spread increases for less expensive stocks. As noted by Li and Mykland (2007, 2014) rounding errors, when modeled explicitly, are intensified for less expensive stocks. In our simulation study we wish to capture this effect by using different starting prices, namely $S_0 \in \{10, 30\}$.

Based on the simulation scheme, we generate daily samples which we split into blocks of observations of difference size. We consider blocks of at most 200, 300 and 500 observations[5]. Additionally, observation times are randomly generated to create non-synchronicity, a common feature high frequency data. Following Aït-Sahalia and Yu (2009), observation times follow a Poisson process with intensity λ, independent of the process X. λ ranges from 2 to 20 seconds. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations[6].

As mentioned above, we can perform inference with two different ways. Daily integrated variance can be estimated by either aggregating the block point estimates or by applying the combination scheme above. We call the first point estimator, block point estimator and the second, combined point estimator. Unlike the simple aggregation scheme, the combination scheme generates a distribution for the entire day and, therefore, confidence intervals are

[5]For example, if we observe 10,000 observation in a day and we decide to split the sample in blocks of at most 200 observations, then we generate 50 blocks.

[6]In other words, we use 40,000 particles for the SMC algorithm.

immediate.

We demonstrate the effectiveness of our methodology by comparing our estimators with the parametric and non-parametric alternatives. In particular, we employ the quasi maximum likelihood estimator (QMLE) of Aït-Sahalia et al. (2005), the pre-averaging estimator of Jacod et al. (2009) and the realized kernel estimator of Barndorff-Nielsen et al. (2008). The observations for these estimators use the same sample paths for efficient price process, with rounding occurring towards the nearest tick. We note that out of these estimators, only the pre-averaging estimator is designed to incorporate rounding errors explicitly. Discussion about the implementation of these estimators can be found in the appendix.

2.7.1 Constant Volatility

We start with constant volatility. Table 3.1 reports the performance of the combined point estimator together with the parametric and non-parametric alternatives. The block point estimator is reported in table 3.2 since there are no confidence intervals. Intensity λ is 2 seconds. Table 3.1 reports the coverage[7], the average length of the confidence intervals, the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$.

A close inspection of the results reveals that our estimator outperforms all other estimators. In terms of coverage, the most competitive estimator is the pre-averaging, since it is robust to rounding errors and the calculation of the asymptotic variance is adjusted in a way to reduce finite sample bias. However, the variance of our estimator is significantly lower and, therefore, our estimator is much more accurate. This is reflected by the smaller RMSE and the narrower confidence intervals. The QMLE when the starting price is low ($S_0 = 10$) or the signal is low ($\sigma_0 = 15\%$) is heavily upward biased. This observation is similar to the observation by Li and Mykland (2007), where the Two-Scales RV estimator was suffering by the same issue[8]. The realized kernel estimators do not perform well in terms of coverage. The asymptotic variance is not estimated accurately and, therefore, coverage

[7]The nominal is 95%.

[8]The performance of the QMLE can be improved if a two stage contamination scheme is employed. However, as in the case of the TSRV, the variance of additive component needs to be sufficiently large.

Performance of the combined fiducial distribution estimator under constant volatility.

	Annual volatility $\sigma = 15\%$ and $\lambda = 2$ seconds													
	$S_0 = 10$							$S_0 = 30$						
RV	Covg	ALCI	RMSE	Bias	Q1	Median	Q3	Covg	ALCI	RMSE	Bias	Q1	Median	Q3
Comb200	74.50%	2.02	0.95	0.39	-0.19	0.41	0.89	84.00%	0.94	0.31	-0.18	-0.37	-0.19	0.02
Comb300	62.00%	2.08	1.16	0.76	0.10	0.72	1.47	89.50%	0.94	0.29	-0.13	-0.29	-0.14	0.07
Comb500	52.00%	2.07	1.27	0.94	0.35	0.92	1.50	92.00%	0.94	0.26	-0.07	-0.24	-0.07	0.11
QMLE	0.00%	3.33	21.39	21.26	19.76	21.22	22.61	0.00%	1.19	2.28	2.25	2.01	2.27	2.51
PAV	14.00%	3.26	2.51	2.38	1.77	2.43	2.92	95.00%	2.52	0.62	0.04	-0.37	-0.01	0.51
RK TH2	1.00%	0.96	2.73	2.58	1.98	2.61	3.19	48.50%	0.72	0.54	0.03	-0.35	0.04	0.42
RK Parzen	61.00%	1.85	1.20	0.50	-0.22	0.45	1.13	55.50%	1.47	0.93	0.00	-0.65	-0.01	0.70
	Annual volatility $\sigma = 30\%$ and $\lambda = 2$ seconds													
	$S_0 = 10$							$S_0 = 30$						
RV	Covg	ALCI	RMSE	Bias	Q1	Median	Q3	Covg	ALCI	RMSE	Bias	Q1	Median	Q3
Comb200	91.00%	4.83	1.41	-0.30	-1.25	-0.35	0.81	90.50%	2.68	0.84	-0.51	-0.94	-0.52	-0.13
Comb300	91.00%	4.86	1.37	-0.12	-0.97	-0.02	0.87	94.50%	2.68	0.74	-0.33	-0.81	-0.37	0.10
Comb500	92.00%	4.85	1.32	0.00	-0.70	0.03	0.89	95.00%	2.68	0.70	-0.21	-0.64	-0.23	0.22
QMLE	0.00%	6.72	25.54	25.41	23.58	25.39	27.14	93.00%	3.31	0.91	0.36	-0.17	0.33	0.88
PAV	94.50%	10.24	2.82	0.79	-1.09	0.67	2.32	93.00%	10.10	2.79	0.22	-1.59	0.34	2.13
RK TH2	57.50%	4.14	2.56	0.99	-0.57	1.03	2.39	56.00%	3.75	2.29	0.06	-1.59	0.15	1.72
RK Parzen	65.00%	7.80	4.03	-0.01	-2.57	-0.26	2.76	64.50%	7.21	4.02	0.07	-2.48	0.04	2.66

Table 2.1: **Performance of the combined fiducial distribution estimator under constant volatility.** The upper panel reports the performance of the combined distribution estimator when annual volatility is 15% and starting prices are $S_0 = 10$ and $S_0 = 30$. The lower reports the same information when annual volatility is 30%. Intensity λ is set to be 2 seconds. For each combination of volatility and starting price we report the coverage of the combined distributions (nominal coverage is 95%), the average length of the confidence intervals, the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is the fiducial point estimator (sample mean). The first three rows report the results for different block sizes. For example, column Comb200 reports the results when the process is split in blocks of at most 200 observations and the results for the daily integrated variance are generated through the combination scheme. We simulate 200 sample paths of the process. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations. All entries except those in the columns Covg are multiples of 10^{-5}.

Performance of the block fiducial distribution estimator under constant volatility.

	Annual volatility $\sigma = 15\%$ and $\lambda = 2$ seconds													
	$S_0 = 10$							$S_0 = 30$						
RV	RMSE	Bias	Q1	Median	Q3	Block Covg	ALCI	RMSE	Bias	Q1	Median	Q3	Block Covg	ALCI
Block200	2.62	2.44	1.81	2.36	3.11	80.21%	17.21	0.32	0.18	0.00	0.19	0.36	94.32%	7.33
Block300	2.15	1.96	1.26	1.95	2.60	81.13%	13.85	0.28	0.11	-0.05	0.12	0.27	94.45%	5.94
Block500	1.75	1.54	0.97	1.56	2.05	81.37%	10.43	0.26	0.07	-0.11	0.07	0.24	93.95%	4.57
	Annual volatility $\sigma = 30\%$ and $\lambda = 2$ seconds													
	$S_0 = 10$							$S_0 = 30$						
RV	RMSE	Bias	Q1	Median	Q3	Block Covg	ALCI	RMSE	Bias	Q1	Median	Q3	Block Covg	ALCI
Block200	2.41	1.98	1.12	1.90	3.01	91.48%	38.59	0.75	0.32	-0.14	0.29	0.77	94.76%	20.71
Block300	1.90	1.35	0.51	1.42	2.26	92.03%	31.17	0.70	0.20	-0.27	0.17	0.64	95.17%	16.84
Block500	1.55	0.87	0.16	0.86	1.68	92.36%	23.85	0.68	0.11	-0.37	0.09	0.53	95.08%	13.00

Table 2.2: **Performance of the block fiducial distribution estimator under constant volatility.** The upper panel reports the performance of the block distribution estimator when annual volatility is 15% for starting prices $S_0 = 10$ and $S_0 = 30$. The lower reports the same information when annual volatility is 30%. Intensity is $\lambda = 2$. For each combination of volatility and starting price we report the coverage of the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is the fiducial point estimator (sample mean). Since we cannot report coverage and average length of the confidence intervals (ALCIs) for the daily estimates, we report the coverage of the block distributions by aggregating all distributions from all blocks (nominal coverage is 95%), as well as the corresponding ALCIs. We report the results for different block sizes. For example, row Block200 reports the results when the process is split in blocks of at most 200 observations and the results for the daily integrated variance are generated by aggregating the block estimates. We simulate 200 sample paths of the process. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations. All entries except those in the columns Covg are multiples of 10^{-5}.

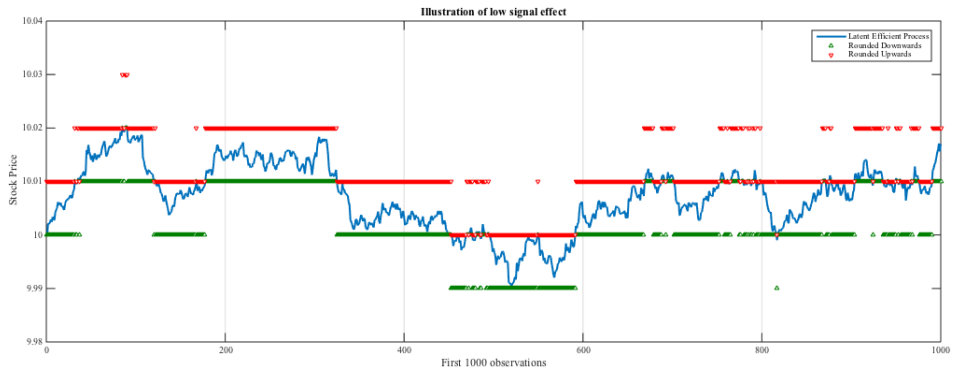


Figure 2.2: **Rounding Issue.** The figure demonstrates the case where the starting price is $S_0 = 10$ and the signal is $\sigma_0 = 15\%$. In this case the signal is so weak that is obliterated by rounding errors.

is low and RMSEs are high.

Out of the four different combinations of starting prices and parameters, the case where the starting price is low ($S_0 = 10$) and the signal is low ($\sigma_0 = 15\%$) is of particular interest. This is the case where all estimators underperform. This is no surprise, since the aforementioned combination together with arrival intensity $\lambda = 2$ seconds generates sample paths where the signal is so weak that is obliterated by rounding errors. Figure 2.2 illustrates the effect. In view of this issue, sampling sparsely is a quick fix, therefore, table 3.3 reports the performance of the combined point estimator in a similar manner as before, with the arrival intensity varying. That is, intensity λ is set at 5, 10 and 20 seconds. The block point estimator is reported in table 3.4. As before our estimator outperforms all other estimators.

In addition to the simulation study above, we include a small sample simulation study for the combination scheme. That is, for the generated block distributions we apply the scheme progressively, by combining the first two blocks, then the first three blocks, and so on, until we use all blocks. Tables B.1-B.6 in appendix B report the coverage, the average length of the confidence intervals, the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$. A quick look reveals that in most cases the procedure performs remarkably well.

Performance of the combined fiducial distribution estimator under constant volatility and different intensity levels.

	Starting price is $S_0 = 10$ and annual volatility is $\sigma = 15\%$																			
	$\lambda = 5$							$\lambda = 10$							$\lambda = 20$						
RV	Covg	ALCI	RMSE	Bias	Q1	Med	Q3	Covg	ALCI	RMSE	Bias	Q1	Med	Q3	Covg	ALCI	RMSE	Bias	Q1	Med	Q3
C200	83.00%	2.29	0.86	0.07	-0.51	0.11	0.52	92.50%	2.60	0.74	0.03	-0.47	-0.06	0.53	91.50%	3.04	0.91	0.12	-0.53	0.13	0.76
C300	88.50%	2.31	0.79	0.23	-0.32	0.24	0.71	89.50%	2.60	0.76	0.06	-0.42	-0.01	0.54	93.00%	3.04	0.99	0.22	-0.41	0.20	0.82
C500	77.00%	2.31	0.91	0.31	-0.36	0.25	0.96	85.50%	2.60	0.90	0.22	-0.41	0.17	0.67	91.50%	3.04	0.90	0.23	-0.41	0.20	0.95
QMLE	0.00%	3.54	13.03	12.85	11.42	12.69	14.30	0.00%	3.71	8.18	7.94	6.65	7.80	8.93	5.00%	4.01	4.86	4.53	3.38	4.52	5.71
PAV	83.50%	3.57	1.36	1.07	0.50	1.10	1.55	92.00%	4.08	1.24	0.67	-0.11	0.62	1.34	95.00%	4.92	1.43	0.68	-0.15	0.51	1.42
TH2	21.00%	0.95	1.34	0.99	0.39	0.99	1.55	30.00%	0.95	1.14	0.44	-0.33	0.34	1.14	31.50%	0.97	1.30	0.33	-0.49	0.33	1.10
Parzen	55.00%	1.85	1.28	0.04	-0.79	0.04	0.81	48.00%	1.84	1.45	-0.06	-1.14	-0.23	0.75	38.50%	1.88	1.71	0.20	-1.02	0.14	1.37

Table 2.3: **Performance of the combined fiducial distribution estimator under constant volatility and different intensity levels.** The table reports the performance of the combined distribution estimator when the starting price is $S_0 = 10$ and annual volatility is 15%. Intensity λ varies by taking values 5, 10 and 20. For each λ we report the coverage of the combined distributions (nominal coverage is 95%), the average length of the confidence intervals, the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is the fiducial point estimator (sample mean). The first three rows report the results for different block sizes. For example, row Comb200 reports the results when the process is split in blocks of at most 200 observations and the results for the daily integrated variance are generated through the combination scheme. We simulate 200 sample paths of the process. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations. All entries except those in the columns Covg are multiples of 10^{-5}.

Performance of the block fiducial distribution estimator under constant volatility and different intensity levels.

	Starting price is $S_0 = 10$ and annual volatility is $\sigma = 15\%$																			
	$\lambda = 5$							$\lambda = 10$							$\lambda = 20$					
RV	RMSE	Bias	Q1	Med	Q3	Covg	ALCI	RMSE	Bias	Q1	Med	Q3	Covg	ALCI	RMSE	Bias	Q1	Med	Q3	Covg	ALCI
B200	1.36	1.04	0.47	1.01	1.52	86.37%	12.04	0.98	0.60	0.03	0.57	1.21	89.96%	9.47	1.03	0.48	-0.17	0.44	1.21	91.47%	7.81
B300	1.11	0.82	0.24	0.79	1.34	88.14%	9.71	0.87	0.41	-0.13	0.36	0.90	90.87%	7.63	1.07	0.44	-0.19	0.43	1.07	91.93%	6.37
B500	1.01	0.64	0.12	0.57	1.26	87.75%	7.56	0.94	0.40	-0.22	0.34	0.88	89.50%	5.92	0.94	0.35	-0.29	0.32	1.07	92.50%	5.33

Table 2.4: **Performance of the block fiducial distribution estimator under constant volatility and different intensity levels.** The table reports the performance of the block distribution estimator when the starting price is $S_0 = 10$ and annual volatility is 15%. Intensity λ varies by taking values 5, 10 and 20. For each λ we report the coverage of the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is the fiducial point estimator (sample mean). Since we cannot report coverage and average length of the confidence intervals (ALCIs) for the daily estimates, we report the coverage of the block distributions by aggregating all distributions from all blocks (nominal coverage is 95%), as well as the corresponding ALCIs. We report the results for different block sizes. For example, row Block200 reports the results when the process is split in blocks of at most 200 observations and the results for the daily integrated variance are generated by aggregating the block estimates. We simulate 200 sample paths of the process. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations. All entries except those in the columns Covg are multiples of 10^{-5}.

2.7.2 Stochastic Volatility

Our methodology assumes that volatility is constant, at least locally. However, it is well known that volatility is in fact time varying. Therefore, we conduct a small simulation study where the true data generating mechanism exhibits stochastic volatility. Specifically, we adapt a standard stochastic volatility model where the efficient price process follows the diffusion:

$$dX_t = \sigma_t dW_{1t}$$
$$d\sigma_t^2 = \kappa\left(\upsilon - \sigma_t^2\right)dt + s\sigma_t dW_{2t}$$

where W_{1t} and W_{2t} are dependent Brownian Motions with $E\left(dW_{1t}dW_{2t}\right) = \rho dt$. We select $\upsilon = 0.04$ (which amounts to 20% volatility per year), $\kappa = 5$, $s = 0.5$ and $\rho = -0.5$.. These parameters belong in the range of values used in Aït-Sahalia and Kimmel (2007) and Aït-Sahalia and Yu (2009). It is common in this type of simulations to initialize the volatility process by drawing σ_0^2 from its stationary distribution. However, due to the small size of our simulation, we set $\sigma_0^2 = \upsilon$. Parameter $\upsilon = 0.04$ together with parameter $s = 0.5$ can lead to low values of the spot volatility and, therefore, intensify the effect of rounding. Integrated variance is approximated by $\int_0^1 \sigma_t^2 dt \simeq \frac{1}{n}\sum_{i=1}^{n}\sigma_{t_i}^2$, where $\Delta t_i = \frac{1}{23{,}400}$.

In a similar manner as above, table 3.5 reports the performance of the combined point estimator together with the parametric and non-parametric alternatives. The block point estimator is reported in table 3.6. Tables 3.7 and 3.8 report the outcome of the simulation study with time varying intensity; λ takes the values 5 and 10. The results reveal a clear advantage of the proposed estimators.

2.8 Empirical Study

In our empirical analysis we focus, initially, on the illustration of the methodology, together with data handling issues and concerns. Subsequently, we demonstrate that our approach is sufficient to estimate integrated volatility without additive components. The data were collected from the TAQ database and were cleaned according to the filters found in Barndorff-Nielsen et al. (2009). A short description of these filters can be found in appendix C.

Performance of the combined fiducial distribution estimator under stochastic volatility.

	$S_0 = 10$							$S_0 = 30$						
RV	Covg	ALCI	RMSE	Bias	Q1	Median	Q3	Covg	ALCI	RMSE	Bias	Q1	Median	Q3
Comb200	78.00%	2.83	1.15	0.20	-0.61	0.16	0.90	85.00%	1.43	0.51	-0.33	-0.59	-0.35	-0.07
Comb300	77.00%	2.92	1.24	0.44	-0.37	0.38	1.22	90.50%	1.43	0.46	-0.23	-0.48	-0.24	0.01
Comb500	75.00%	2.91	1.23	0.44	-0.24	0.31	1.10	93.50%	1.43	0.42	-0.16	-0.43	-0.15	0.06
QMLE	0.00%	4.43	24.34	24.20	22.27	24.15	25.92	12.00%	1.77	1.60	1.51	1.11	1.50	1.93
PAV	79.50%	4.94	1.98	1.54	0.82	1.50	2.38	97.00%	4.54	1.15	0.12	-0.59	0.10	0.78
RK TH2	13.00%	1.74	2.28	1.90	1.10	1.84	2.68	52.50%	1.41	1.00	0.06	-0.66	0.07	0.62
RK Parzen	61.00%	3.35	1.85	0.27	-0.97	0.25	1.62	62.00%	2.81	1.64	0.06	-1.05	-0.13	1.11

Table 2.5: **Performance of the combined fiducial distribution estimator under stochastic volatility.** The table reports the performance of the combined distribution estimator when volatility is stochastic and starting prices are $S_0 = 10$ and $S_0 = 30$. Intensity λ is set to be 2 seconds. For each price we report the coverage of the combined distributions (nominal coverage is 95%), the average length of the confidence intervals, the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is the fiducial point estimator (sample mean). The first three rows report the results for different block sizes. For example, row Comb200 reports the results when the process is split in blocks of at most 200 observations and the results for the daily integrated variance are generated through the combination scheme. We simulate 200 sample paths of the process. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations. All entries except those in the columns Covg are multiples of 10^{-5}.

Performance of the block fiducial distribution estimator under stochastic volatility.

	$S_0 = 10$							$S_0 = 30$						
RV	RMSE	Bias	Q1	Median	Q3	Block Covg	ALCI	RMSE	Bias	Q1	Median	Q3	Block Covg	ALCI
Block200	2.58	2.32	1.66	2.32	2.96	84.52%	23.75	0.44	0.21	-0.03	0.18	0.46	94.93%	11.15
Block300	2.06	1.69	0.92	1.64	2.54	85.97%	19.11	0.41	0.14	-0.09	0.12	0.39	94.71%	9.06
Block500	1.56	1.13	0.48	1.00	1.79	85.79%	14.46	0.40	0.08	-0.15	0.07	0.29	94.50%	6.98

Table 2.6: **Performance of the block fiducial distribution estimator under stochastic volatility.** The table reports the performance of the block distribution estimator when volatility is stochastic 15% and starting prices are $S_0 = 10$ and $S_0 = 30$. Intensity λ is set to be 2. For each starting price we report the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is the fiducial point estimator (sample mean). Since we cannot report coverage and average length of the confidence intervals (ALCIs) for the daily estimates, we report the coverage of the block distributions by aggregating all distributions from all blocks (nominal coverage is 95%), as well as the corresponding ALCIs. We report the results for different block sizes. For example, row Block200 reports the results when the process is split in blocks of at most 200 observations and the results for the daily integrated variance are generated by aggregating the block estimates. We simulate 200 sample paths of the process. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations. All entries except those in the columns Block Covg are multiples of 10^{-5}.

Performance of the combined fiducial distribution estimator under stochastic volatility and different intensity levels.

	$\lambda = 5$							$\lambda = 10$						
RV	Covg	ALCI	RMSE	Bias	Q1	Med	Q3	Covg	ALCI	RMSE	Bias	Q1	Med	Q3
C200	89.00%	3.32	1.01	-0.18	-0.90	-0.19	0.47	89.50%	3.88	1.20	-0.05	-0.85	0.03	0.83
C300	92.00%	3.36	0.97	0.02	-0.65	0.00	0.67	90.50%	3.87	1.18	0.00	-0.95	-0.02	0.85
C500	89.50%	3.33	1.00	0.00	-0.67	0.14	0.63	89.50%	3.89	1.15	0.06	-0.79	0.05	0.87
QMLE	0.00%	4.81	13.97	13.81	12.41	13.80	15.03	0.50%	5.20	8.34	8.06	6.70	7.85	9.35
PAV	93.50%	5.95	1.73	0.81	-0.25	0.75	1.77	95.50%	7.08	1.97	0.57	-0.69	0.61	1.87
TH2	39.50%	1.73	1.63	0.71	-0.35	0.64	1.63	32.00%	1.74	1.86	0.32	-1.00	0.49	1.49
Parzen	55.50%	3.33	2.12	0.07	-1.37	0.06	1.54	49.00%	3.34	2.55	0.01	-1.60	0.02	1.71

Table 2.7: **Performance of the combined fiducial distribution estimator under stochastic volatility and different intensity levels.** The table reports the performance of the combined distribution estimator when the starting price is $S_0 = 10$ and volatility is stochastic. Intensity λ varies by taking values 5 and 10. For each λ we report the coverage of the combined distributions (nominal coverage is 95%), the average length of the confidence intervals, the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is the fiducial point estimator (sample mean). The first three rows report the results for different block sizes. For example, row Comb200 reports the results when the process is split in blocks of at most 200 observations and the results for the daily integrated variance are generated through the combination scheme. We simulate 200 sample paths of the process. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations. All entries except those in the columns Covg are multiples of 10^{-5}.

Performance of the combined fiducial distribution estimator under stochastic volatility and different intensity levels.

	$\lambda = 5$							$\lambda = 10$						
RV	RMSE	Bias	Q1	Med	Q3	Covg	ALCI	RMSE	Bias	Q1	Med	Q3	Covg	ALCI
B200	1.38	0.92	0.10	0.93	1.58	90.42%	17.12	1.37	0.67	-0.14	0.74	1.52	91.38%	13.98
B300	1.23	0.72	-0.07	0.69	1.39	89.95%	13.88	1.25	0.46	-0.38	0.45	1.23	92.50%	11.29
B500	1.09	0.43	-0.27	0.52	1.10	90.30%	10.80	1.17	0.32	-0.48	0.29	1.13	92.70%	8.82

Table 2.8: **Performance of the combined fiducial distribution estimator under stochastic volatility and different intensity levels.** The table reports the performance of the block distribution estimator when the starting price is $S_0 = 10$ and volatility is stochastic. Intensity λ varies by taking values 5 and 10. For each λ we report the coverage of the RMSE, the bias, which is defined as the average of $\hat{\sigma}^2 - \sigma_0^2$ and the three quantiles of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is the fiducial point estimator (sample mean). Since we cannot report coverage and average length of the confidence intervals (ALCIs) for the daily estimates, we report the coverage of the block distributions by aggregating all distributions from all blocks (nominal coverage is 95%), as well as the corresponding ALCIs. We report the results for different block sizes. For example, row Block200 reports the results when the process is split in blocks of at most 200 observations and the results for the daily integrated variance are generated by aggregating the block estimates. We simulate 200 sample paths of the process. We generate the samples using the SMC algorithm discussed in the appendix and for each block we generate a sample of 40,000 observations. All entries except those in the columns Covg are multiples of 10^{-5}.

Moreover, we apply our methodology on bid and ask quotes that have a corresponding transaction.

2.8.1 Illustration

We analyze data for Alcoa Inc. (AA) collected on May 2007. The particular month contains daily data, studied by Barndorff-Nielsen et al. (2009), and is ideal for a direct comparison. We split our daily samples in blocks of at most 500 observations[9]. The upper panel of figure 2.3displays the high frequency quotes on May 4th, which according to Barndorff-Nielsen et al. (2009), was reported as an exemplary day in terms of the stability of the volatility signature plots. After applying the filters, we arrive at 14,630 quotations and 5,203 transactions. Therefore, the sample under consideration consists of 5,203 quotations that have a corresponding transaction. This amounts to a new observation roughly every 4.5 seconds. The sample in 11 blocks of observations, each one containing 473 observations.

The left lower panel of figure 2.3 displays the block estimates of σ^2 with the confidence intervals superimposed. In the same panel, we include the daily estimates of the other competing. At first glance, the block point estimates reveal a U-shaped pattern. Volatility is quite high in the beginning of the day, much lower in the middle of the day and higher towards the end. This pattern is the reason we do not apply the combination scheme on all block distributions. Volatility for this particular day exhibits sudden changes, indicating the presence of volatility jumps and a combined distribution using all blocks leads to a nearly degenerate distribution, as indicated by the small effective sample size. Instead, we combine distributions from adjacent blocks, provided that the efficient sample size of the generated distribution is high. Typically, distributions located within a band of 10% annual volatility can be combined. The right lower panel displays the combined point estimates. Figure 2.4 displays all point estimates of the integrated variance for the entire month. It is evident that most of the noise is attributed to rounding and spread related frictions.

[9]The analysis using smaller block sizes did not reveal any substantial differences and, therefore is not reported in the paper.

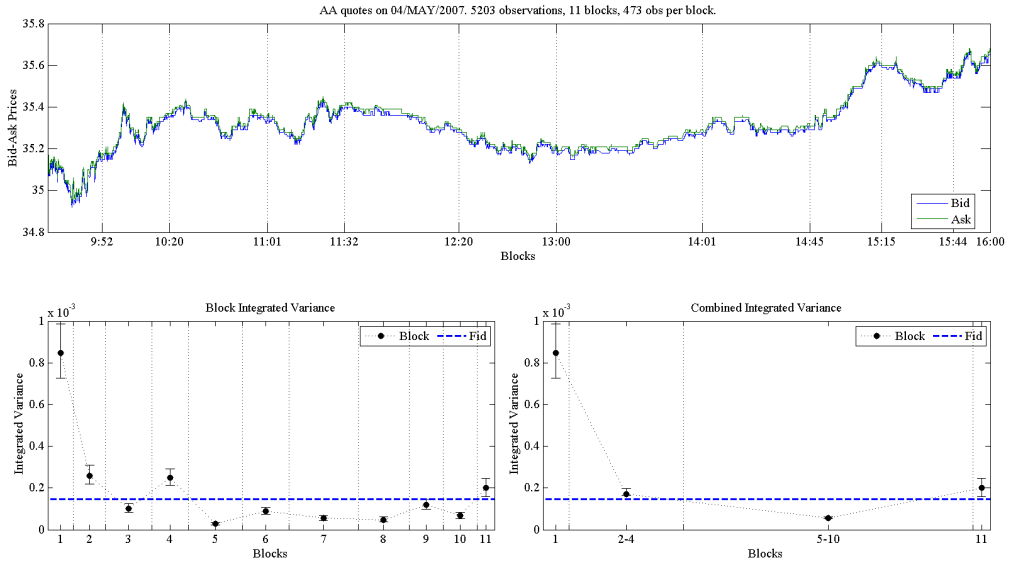


Figure 2.3: **Alcoa Inc. (AA) on May4, 2007.** The upper panel illustrates the high-frequency quotes for Alcoa Inc. (AA) on May4, 2007. The sample is split in 11 blocks of 473 observations. The lower left panel illustrates the block volatility estimates with the confidence intervals superimposed. The right panel illustrates the combined volatility estimates with the confidence intervals superimposed. The daily volatility estimates are also included.

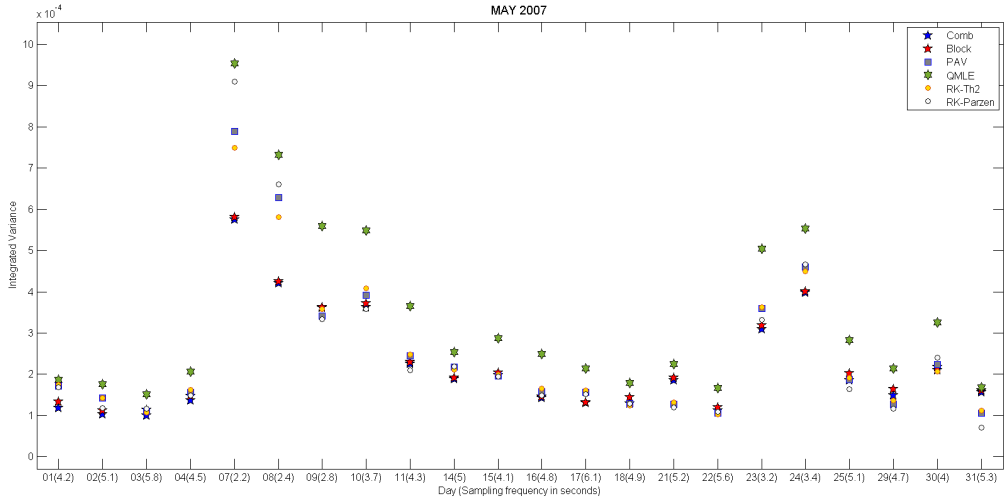


Figure 2.4: **Volatility Estimates for Alcoa Inc. (AA) on May 2007.** The figure displays all point estimates of the integrated variance for the entire month.

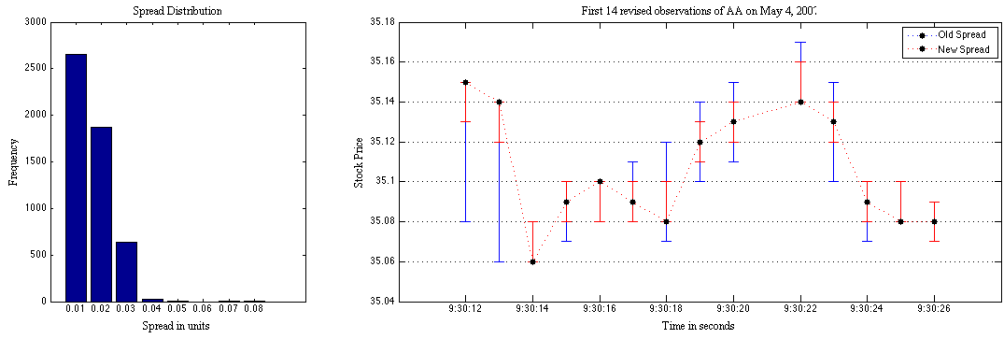


Figure 2.5: **Spread distribution and revised quotes of Alcoa Inc. (AA), on May 4, 2007.**

2.8.2 Time varying spread

In this section we discuss the implications of time varying spread in our methodology. In our modeling approach, we do not account for the relationship between the spread and volatility. If the variation of the spread was independent with volatility, then our methodology would be robust in the presence of time varying spreads. However, information based market microstructure theory suggests that the spread is positively correlated with volatility. For instance, Zhang et al. (2008) show that high volatility in the transaction price process widens the spread symmetrically about the efficient price. This suggests that an extension of the current framework that accounts for this relationship is needed, but will be considered in future work.

Our concern is that wide spreads, usually more than three of four cents, increase our uncertainty about the location of the latent efficient price. Given that our goal is to form an interval about the latent efficient price, a wide spread may have adverse effects on the estimation procedure. Moreover wide intervals add to the uncertainty of the generated distributions, especially if several of these observations are in the same block. The left panel of figure 2.5 displays the spread distribution of the day under consideration. The spread for roughly 87% of these observations is smaller or equal to two cents, roughly 12% is three cents and the rest is above three cents, with only five observations above five cents.

To address this potential issue we propose to replace the quotes where the spread is higher than some pre-specified threshold with new quotes, ensuring that they contain the

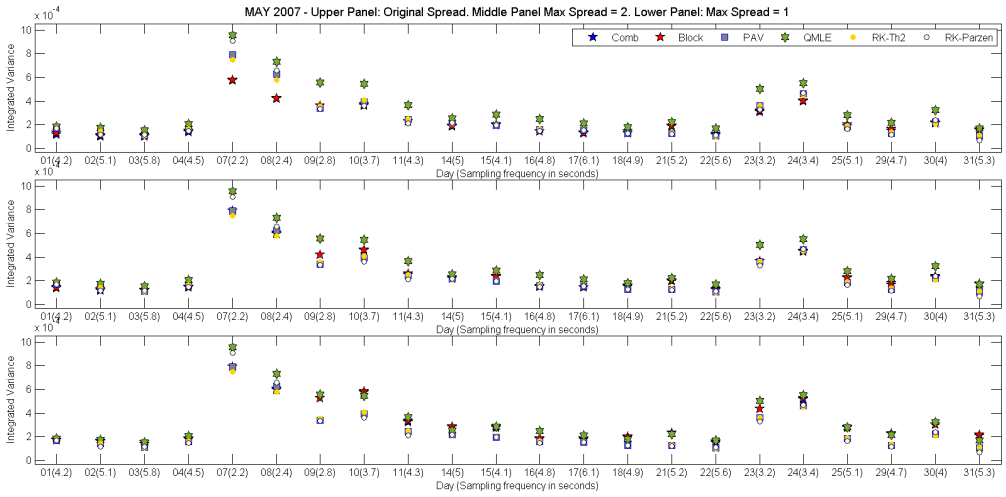


Figure 2.6: **Volatility Estimates for Alcoa Inc. (AA) on May 2007**. The upper panel displays all point estimates of the integrated variance for the entire month without altering the spread; this is the same as in figure 4. In the middle panel, the maximum spread is 2 cents and at the lower panel the maximum spread is one cent.

latent process. A simple way to achieve this is to use the corresponding transactions as a guide to reduce the spread. For instance, if we do not wish to allow observations with spread higher or equal than 3 cents, we can reduce the spread by selecting an interval about the transaction, with length equal to, say, two cents, provided that we do not exceed the quotes. For example, if the recorded spread is three cents and the corresponding transaction occurs at the ask price, the spread can be reduced by shifting the bid price upwards by one cent. The right panel of figure 2.5 displays the revision of the quotes for the first 14 observations. A closer inspection at the revised quotes reveals that reducing the spread may introduce price jumps, which may have an adverse effect in the estimation.

We apply this idea to the monthly data analyzed above by restricting the spread. The upper panel of figure 2.6 displays all point estimates of the integrated variance for the entire month without altering the spread; this is the same as in figure 2.4. In the middle panel, the maximum spread is 2 cents and at the lower panel the maximum spread is 1. As expected, the reduction of the spread has increased volatility slightly, however, the point estimates are remarkably close to the competing ones.

2.9 Conclusion

In this chapter we propose a novel framework to study high frequency financial data for volatility estimation, by taking advantage the features of the observed prices, especially rounding errors. Assuming that the efficient price-process follows a time homogenous diffusion process, we proposed a framework that allows us to specify a generalized fiducial distribution on the parameter space. The attractive feature of the framework is that it enables us to view the bid-ask spread as a natural interval around the latent efficient price and use quotes for volatility estimation, instead of transactions or mid-points. However, other intervals that are justified to contain the latent efficient price may be used. Moreover, our framework deviates from the current literature, since there are no unrealistic additive noise components, introduced to explain the deviations of the efficient price from the observed one.

The new methodology was applied assuming that volatility remains constant for a short period of time. The daily samples were split in blocks of observations and inference was performed on each block. Joint inference became feasible through our novel combination scheme which is a tool to combine the block distributions into one that summarizes the information from all the blocks under consideration. Our simulation study, which was designed to incorporate features of observed data, verified that volatility estimation is feasible at very high frequencies in the presence of rounding errors and outperforms both the parametric and non-parametric alternatives. Our empirical study showed that in very high frequencies microstructure noise is attributed to rounding effects.

CHAPTER 3

Generalized Fiducial Inference for High Frequency Data in the Presence of Presence of Rounging and Additive Errors

3.1 Summary

In this chapter, we extend the generalized fiducial inference framework introduced in the previous chapter. In particular, we study volatility using a generalized fiducial framework that is designed for interval data, but allows for an additive component. The additive component is assumed to be a Gaussian white noise, independent of the process. White noise components have been used extensively in the literature to model microstructure noise. However, since our modeling approach is robust in rounding errors, which is a major component of MS noise, we term the additive component as residual MS noise. It's functionality is to remove any remaining noise, since it allows the latent process to lie outside the observed interval (potentially the bid-ask spread). As before, we perform inference by splitting the trading day into blocks where volatility is assumed constant. The combination scheme is extended and joins the information from all blocks. Both our simulation and empirical studies demonstrate that the proposed volatility estimator performs remarkably well even at very high frequencies.

3.2 Introduction

We remain in the standard microstructure setup used in the previous chapters. Briefly, in the standard microstructure setup, the efficient/unobserved log-price process, denoted by $X_t = log\,(S_t)$, is assumed to follow an Ito process:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

where W_t is a Brownian motion, μ_t is the drift of the process and σ_t is the instantaneous variance of the returns. Both μ_t and σ_t are adapted locally bounded random processes. The process is assumed to evolve in $[0, T]$ and is observed in the grid $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$.

In this chapter we apply generalized fiducial inference to study volatility using high frequency data. The generalized fiducial recipe defines the distributions of interest through a data generating equation of the form $\mathbf{X} = G(\mathbf{U}, \xi)$, where $G(\boldsymbol{.}, \boldsymbol{.})$ is a jointly measurable structural equation based on the model under consideration, $\xi \in \Xi$ are the parameters of interest, and $\mathbf{U}$ is the random component of the structural equation. Following the recipe for interval data, our first task is to specify the data generating equation. The role of G will play the contaminated price process and, since we intend to work in a parametric framework, we employ the framework of Aït-Sahalia et al. (2005), where the log of the efficient price follows a time homogenous diffusion process and microstructure noise is additive, Gaussian, iid and independent of the process. So, at any arrival time t_m, the process is contaminated by additive noise and belongs in the interval $[b_{t_m}, a_{t_m}]$, that is

$$b_{t_m} \leq X_{t_m} + U_{t_m} \leq a_{t_m}$$

where $U_{t_m} \sim N\left(0, \sigma_u^2\right)$ and $X \perp U$. The structural equation is the function $G(\mathbf{U}, \xi) = \mu t + \sigma W_t + \sigma_u Z_t$, with $\mathbf{U} = (W_t, Z_t)$ where $Z_t \sim N(0, 1)$ and $\xi = (\mu, \sigma, \sigma_u)$.

The additive component in our framework has a different functionality from its original use. Initially, the additive component was introduced to capture a variety of microstructure frictions, including rounding and spread related effects. In our framework, the effect of frictions related to rounding is expected to attenuate, however, other sources of noise exist and the additive component may help to account for them. For this reason, we term the additive component as residual microstructure noise. Also, since we work in a parametric framework, we would like to have a direct comparison of our methodology with the standard likelihood approach and the presence of the additive component will serve this purpose. Moreover, from an empirical point of view we expect, at least in high frequencies, residual noise to be virtually zero. The reason is that a frequently traded or quoted stock implies

a liquid market, therefore, informational frictions may be insignificant relative to rounding errors. The proposed model will allow us to test this hypothesis. The case where the model has no additive component and frictions are exclusively attributed to rounding will be considered in future work.

The parametric framework under consideration assumes constant volatility in the entire interval $[0, T]$. We will deviate and study this model assuming local constancy of the volatility over blocks of successive observations. Mykland and Zhang (2009) studied this type of local constancy and showed that in sufficiently small neighborhoods of observations one can act as if volatility is constant. In this article we do not explore the extent where volatility is held constant. The size of the blocks we consider consists of at most 400 observations. For example, a block of 300 observations observed on average every two seconds amounts to a time frame of 10 minutes. In this small time frame we assume that volatility, even if stochastic, does not vary much.

In the literature research primarily focuses on daily volatility estimates. In that spirit, we present a novel method to combine the generated fiducial samples from different blocks of data, into one sample that summarizes all information from all the blocks under consideration. In a sense, the combination scheme works as an importance sampler by re-weighing all particles with weights computed through a metric that utilizes the Gaussian kernel. As we demonstrate, the combined distribution approximates remarkably well the distribution we would have generated if we had used all data in one sample.

We test our methodology by conducting a simulation study employing a realistic simulation scheme. The scheme is based on the two stage contamination scheme of Li and Mykland (2007), which incorporates rounding errors explicitly. Initially, at sampling/arrival times, additive noise contaminates the efficient price and subsequently, the contaminated process will be rounded upwards and downwards to the two nearest ticks. The proposed simulation scheme renders the choice of the starting price X_0 relevant, since the magnitude of the spread increases for less expensive stocks, due to the log-transformation, see Li and Mykland (2007, 2014). Therefore, we select different starting prices to capture this effect. We also use a variety of combinations of the parameters $\{\sigma, \sigma_u\}$ to illustrate the effectiveness of our methodology. Contrary to the standard values used in the literature, we also

use low values for volatility and additive noise. Low volatility introduces price sluggishness, intensifying the effect of rounding errors. Low additive noise renders rounding errors the only source of noise, which makes our framework ideal in the case where rounding errors is the only source of noise. Our simulation study shows that we can effectively capture true volatility, even in cases where rounding errors dominate, and outperform the competing ML estimator of Aït-Sahalia et al. (2005), Xiu (2010). We also examine our methodology in the presence of stochastic volatility with equally positive results.

Finally, we conduct a small empirical study to illustrate the use of our methodology in real data. The results indicate that after accounting for rounding, residual MS noise is virtually zero.

This chapter is organized as follows. In section 3, we apply the generalized fiducial inference framework for interval data for the model under consideration. In Section 4 we present the sequential Monte Carlo algorithm used to sample from the generalized fiducial distributions and establish its consistency. Namely, we illustrate that proposed algorithm, as the number of particles approaches infinity, targets the generalized fiducial distribution. In Section 5 we extend the the combination scheme introduces in the previous chapter in the case where there are more that one parameters. The simulation study is in section 6 and the empirical study is in section 7. Section 8 concludes.

3.3 Generalized Fiducial Inference for HF data

In our setup, we will assume that the efficient log-price follows a Geometric Brownian motion. The process is assumed to evolve in $[0, T]$ and is observed/quoted in the grid $G_n = \{0 = t_0 < t_1 < ... < t_n = T\}$. In addition to the grid G_n, we will consider the sub-grid $H_n = \{0 = \tau_0 < \tau_1 < ... < \tau_{M_n} = T\} \subseteq G_n$ where volatility is assumed constant for all $t_{i,m} \in (\tau_{i-1}, \tau_i]$. Specifically, in the interval $(\tau_{i-1}, \tau_i]$ the log-price, given $X_{\tau_{i-1}} = x_{\tau_{i-1}}$, is

$$dX_t = \mu_{\tau_{i-1}} dt + \sigma_{\tau_{i-1}} dW_t$$

where, W_t is a Brownian motion, $\mu_{\tau_{i-1}}$ is the drift of the process, $\sigma^2_{\tau_{i-1}}$ the diffusion coefficient. In the high frequency literature, it is common practice to to assume that $\mu = 0$. The

order of magnitude of the diffusive component ($\sqrt{dt}$) is much larger than the order of magnitude of the drift component (dt), making the drift component mathematically negligible at high frequencies. Moreover, maintaining it, causes estimation problems since it is estimated with a large standard error. Our preliminary simulation study showed that maintaining the drift component does not affect the quality of the generated fiducial distributions. It adds though computational burden and, therefore, it is not included in our reported simulations.

At observation times, the process is contaminated by residual microstructure noise and is between the bid and ask log-prices

$$b_{t_{i,m}} \leq X_{t_{i,m}} + U_{t_{i,m}} \leq a_{t_{i,m}}$$

Additionally, we assume that for all $t_{i,m} \in (\tau_{i-1}, \tau_i]$, $U_{t_{i,m}} \sim N\left(0, \left(\sigma_u^i\right)^2\right)$ and $X_{t_{i,m}} \perp U_{t_{i,m}}$. So, for $t_{i,m} \in (\tau_{i-1}, \tau_i]$ and every i we have that

$$b_{t_{i,m}} \leq \mu_{\tau_{i-1}} t_{i,m} + \sigma_{\tau_{i-1}} W_{t_{i,m}} + U_{t_{i,m}} \leq a_{t_{i,m}} \tag{3.3.1}$$

Noting that $W_t = \sqrt{t} Z_{1,t}$ where $Z_{1,t} \sim N(0,1)$ and $U_t = \sigma_u Z_{2,t}$ where $Z_{2,t} \sim N(0,1)$, we can re-write equation 3.3.1 as

$$b_{t_{i,m}} \leq \mu_{\tau_{i-1}} t_{i,m} + \sigma_{\tau_{i-1}} \sqrt{t_{i,m}} Z_{1,t_{i,m}} + \sigma_u^i Z_{2,t_{i,m}} \leq a_{t_{i,m}} \tag{3.3.2}$$

In terms of the fiducial argument, the structural equation $G(\mathbf{z}, \xi)$ is the linear part of equation 3.3.2, the parameters $\xi = \left(\mu_{\tau_{i-1}}, \sigma_{\tau_{i-1}}, \sigma_{u,\tau_{i-1}}\right) \in \mathbb{R} \times \mathbb{R}_+^2$ and $\mathbf{z}$ is a realization of $\mathbf{Z} = (Z_1, Z_2)$. The inverse image of $G(\mathbf{z}, \xi)$ is

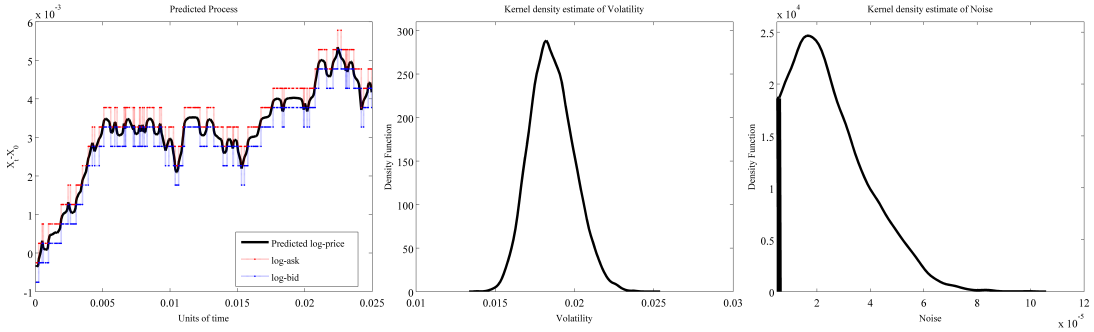
$$Q\left((\mathbf{b}, \mathbf{a}], \mathbf{Z}\right) = \left\{(\mu, \sigma, \sigma_u,) \in \mathbb{R} \times \mathbb{R}_+^2 : \mathbf{b} < G(\mathbf{z}, \xi) \leq \mathbf{a}\right\} \tag{3.3.3}$$

The corresponding generalized fiducial distribution is

$$V\left(Q\left((\mathbf{b}, \mathbf{a}], Z\right)\right) \mid \left\{Q\left((\mathbf{b}, \mathbf{a}], Z\right) \neq \emptyset\right\} \tag{3.3.4}$$

Generating samples from the generalized fiducial distribution requires the use of Sequential

Monte Carlo (SMC) methods. The SMC algorithm is based on the algorithm developed by Cisewski and Hannig (2012), where they performed inference for the parameters of normal linear mixed models. The SMC algorithm is presented analytically in appendix A. Figure 3.1 illustrates our methodology for simulated data.



Parms.	Point Est.	Median	Mode	Conf. Interval	Variance	Skewness	Kurtosis
σ_u	4.66*10^{-5}	4.88*10^{-5}	8.33*10^{-5}	[0, 1*10^{-4}]	7.81*10^{-10}	0.1313	2.295
σ	0.0185	0.0186	0.0174	[0.0152, 0.0222]	3.28*10^{-6}	0.2315	3.017

Figure 3.1: **Illustration of the fiducial methodology.** The left panel displays the simulated quotes used to generate the fiducial distributions with the predicted price superimposed on it. The middle and right panels show the distributions for σ and σ_u respectively. The block under consideration has 200 observations and the parameters used are $\sigma = 0.0189$ ($= {}^{30\%}/_{\sqrt{252}}$) and $\sigma_u = .005\%$. The table below the figure contains summary statistics for the fiducial distributions.

3.4 Estimation

3.4.1 The SMC Algorithm

In this section we present the Sequential Monte Carlo algorithm we developed to generate samples from the generalized fiducial distribution of the parameters of interest. We present the algorithm in the case where all parameters are present, namely $\xi = \left(\mu_{\tau_{i-1}}, \sigma_{\tau_{i-1}}, \sigma_{u,\tau_{i-1}}\right)$ for each i, however the nested case where $\mu_{\tau_{i-1}} = 0$ can be easily reproduced. To ease notation we will drop the dependence in i and we will embed $\sqrt{t}$ in $Z_{1,t}$ such that $Z_{1,t} \sim N(0,t)$. We will be denoting $Z_{t_1:t_m} = (Z_{t_1}, ..., Z_{t_m})$, where $Z_{t_j} = \left(Z_{1,t_j}, Z_{2,\Delta t_j}\right)$, $t_i \in \{t_1, ..., t_m\}$, $\Delta t_i = t_i - t_{i-1}$, $t_0 = 0$, and $m = 1, ..., n$ where $n \equiv n_i = \#\{j, \tau_{i-1} < t_{i,j} \leq \tau_i\}$. In our setup, $Z_{1,\Delta t_j} \sim N(0, , \Delta t_j)$ and $Z_{2,t_j} \sim N(0,1)$. The generalized fiducial distribution of the parameters will be

$$V\left(Q\left((\mathbf{b}_n,\mathbf{a}_n],Z_{t_1:t_n}^{(K)}\right)\right)\mid\left\{Q\left((\mathbf{b}_n,\mathbf{a}_n],Z_{t_1:t_n}^{(K)}\right)\neq\emptyset\right\}\tag{3.4.1}$$

where $Q_n^{(K)}=Q\left((\mathbf{b}_n,\mathbf{a}_n],Z_{t_1:t_n}^{(K)}\right)$ is the set function containing the values of the parameters that satisfy the structural equation 3.3.2, given the data $(\mathbf{b}_n,\mathbf{a}_n]$ and the generated $Z_{t_1:t_n}^{(K)}$ for particle K , where $K=1,2,...,N$. Generating a sample from 3.4.1 is equivalent to simulating sequentially for each m, $Z_{t_1:t_m}^{(K)}$ such that $Q_m^{(K)}$ is non-empty until we reach n. The corresponding target distribution up to time t_m, denoted by $\pi_{t_1:t_m}$, is

$$\begin{aligned}&\pi_{t_1:t_m}\left(Z_{t_1:t_m}\mid(\mathbf{b},\mathbf{a}]_{t_1:t_m}\right)\equiv\pi_{t_1:t_m}\left(Z_{t_1:t_m}\right)\\&\quad\propto\prod_{j=1}^{m}\frac{1}{\left(\Delta t_j\right)^{1/2}}\exp\left(-\frac{1}{2\Delta t_j}Z_{1,\Delta t_j}^2\right)\exp\left(-\frac{1}{2}Z_{2,t_j}^2\right)\mathbf{I}_{\mathbf{C_m}}\left(Z_{t_1:t_m}\right)\end{aligned}\tag{3.4.2}$$

where $\mathbf{I}_{\mathbf{C_m}}\left(Z_{t_1:t_m}\right)$ is an indicator random variable of the set

$$\mathbf{C_m}=\left\{Z_{t_1:t_m}:b_{t_j}\leq\mu t_j+\sigma\sum_{k=1}^{j}Z_{1,\Delta t_k}+\sigma_u Z_{2,t_j}\leq a_{t_j},\ for\ all\ j=1,...,m\right\}\tag{3.4.3}$$

Restriction 3.4.3 is required in order to generate a representative sample from the fiducial distribution. It ensures that all inequalities up to time t_m are satisfied simultaneously. In practice, this can be achieved easily if at time t_m, given that we have sampled $Z_{t_1:t_{m-1}}^{(K)}$ and $Z_{2,t_m}^{(K)}$, we sample $Z_{1,\Delta t_m}^{(K)}$ by truncating it between the two values

$$L_m\left(Z_{t_1:t_{m-1}}^{(K)},Z_{1,t_m}^{(K)}\right)=\min\left\{\frac{b_{t_m}-\mu t_m-\sigma\sum_{j=1}^{m-1}Z_{2,\Delta t_j}^{(K)}-\sigma_u Z_{1,t_m}^{(K)}}{\sigma},\ (\mu,\sigma,\sigma_u)\in Q_{m-1}^{(K)}\right\}$$

$$R_m\left(Z_{t_1:t_{m-1}}^{(K)},Z_{1,t_m}^{(K)}\right)=\max\left\{\frac{a_{t_m}-\mu t_m-\sigma\sum_{j=1}^{m-1}Z_{2,\Delta t_j}^{(K)}-\sigma_u Z_{1,t_m}^{(K)}}{\sigma},\ (\mu,\sigma,\sigma_u)\in Q_{m-1}^{(K)}\right\}$$

We pick to restrict Z_1 because it makes more statistical sense to truncate the component that drives the process. Utilizing these type of restrictions we can write

$$\mathbf{I}_{\mathbf{C_m}}\left(Z_{t_1:t_m}\right) \;=\; \mathbf{I}_{\mathbf{C_{m-1}}}\left(Z_{t_1:t_m}\right)\mathbf{I}_{\star}\left(Z_{2,t_m}\right)\mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta t_m}\right) = \prod_{j=1}^{m}\mathbf{I}_{\star}\left(Z_{2,t_j}\right)\mathbf{I}_{(L_j,R_j)}\left(Z_{1,\Delta t_j}\right)$$

where $\mathbf{I}_{\star}$ indicates the lack of restriction.

The proposal distribution for our SMC algorithm utilizes the Cauchy distribution

$$\tilde{\pi}_{t_1:t_m}\left(Z_{t_1:t_m}\right) \propto \pi_{t_1:t_3}\left(Z_{t_1:t_3}\right)\prod_{j=4}^{m}\frac{\exp\left(-\frac{1}{2}Z_{2,t_j}^2\right)\mathbf{I}_{\star}\left(Z_{2,t_j}\right)\cdot\mathbf{I}_{(L_j,R_j)}\left(Z_{1,\Delta t_j}\right)}{\left(\Delta t_j\right)^{1/2}\left(1+\frac{Z_{1,\Delta t_j}^2}{\Delta t_j}\right)\left(F\left(R_j\right)-F\left(L_j\right)\right)} \tag{3.4.4}$$

where F denotes the cdf of the Cauchy distribution. It is important to point out that the proposal distribution treats $Z_{t_1:t_3}$ as unrestricted, since these will be used to solve for the parameters and pin down the location of the particle. In particular, we sample $Z_{t_1:t_3}$ from the target distribution and use the first three inequalities as two double equalities (one for a_{t_j} and one for b_{t_j}, $j=1,2,3$) to form eight systems of equations by combining the equalities (using one for each $j=1,2,3$). Solving these system of equations (the unknowns are (μ,σ,σ_u)) we end up with eight points of the space $\mathbb{R}^3$ which form a polyhedron. Any point in the polyhedron satisfies all three inequalities[1]. The rest of the inequalities will be used sequentially to trim the polyhedron in a way that all inequalities will be satisfied. Below, there we present a more practical illustration of the sampling scheme.

The conditional proposal distribution for $m>3$, which will be used to draw samples in the algorithm is

$$\widetilde{\pi}_{t_m|t_1:t_{m-1}} \propto \frac{\exp\left(-\frac{1}{2}Z_{1,t_m}^2\right)\mathbf{I}_{\star}\left(Z_{2,\tau_m}\right)\mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right)}{\left(\Delta t_m\right)^{1/2}\left(1+\frac{Z_{1,\Delta t_m}^2}{\Delta t_m}\right)\left(F\left(R_m\right)-F\left(L_m\right)\right)} \tag{3.4.5}$$

That is, at time t_m, Z_{2,t_m} is drawn from a standard normal and $Z_{1,\Delta t_m}$ is drawn form a truncated Cauchy distribution.

---

[1]When we generate a negative value for either σ or σ_u, we flip its sign by simultaneously flipping the sign of the corresponding Z.

The final component of the algorithm is the importance weights. The weights are computed as

$$W_{t_1:t_m} = \frac{\pi_{t_1:t_m}}{\tilde{\pi}_{t_1:t_m}} = \frac{\pi_{t_m|t_1:t_{m-1}}\pi_{t_1:t_{m-1}}}{\tilde{\pi}_{t_m|t_1:t_{m-1}}\tilde{\pi}_{t_1:t_{m-1}}} = W_{t_m}W_{t_1:t_{m-1}} \tag{3.4.6}$$

where

$$W_{t_m} = \frac{\pi_{t_m|t_1:t_{m-1}}}{\tilde{\pi}_{t_m|t_1:t_{m-1}}}$$

is the incremental weight. The incremental weight is given by

$$W_{t_m} \propto \exp\left(-\frac{1}{2\Delta t_m}Z_{1,t_m}^2\right)\left(1+\frac{Z_{1,t_m}^2}{\Delta t_m}\right)\left(F\left(R_m\right)-F\left(L_m\right)\right) \tag{3.4.7}$$

Practical Illustration of the Sampling Scheme

In figure 3.2 we present the first steps of our sampling scheme. For simplification we set $\mu = 0$. Otherwise, the figures would have been 3-dimensional. The left panel displays how to use the data generating equation and derive the location of one particle. By setting $\mu = 0$ our assumptions require that

$$b_{t_j} \leq \sigma\sum_{k=1}^{j} Z_{1,\Delta t_k} + \sigma_u Z_{2,t_j} \leq a_{t_j},\ for\, all\, j = 1, ..., m$$

In this simple setup, the first two sets of inequalities will be used to solve for (σ_u, σ), given the generated without any restriction $Z_{t_1:t_2}$. For the first set of observations $\{b_{t_1}, a_{t_1}\}$, we use the first set of inequalities ($j = 1$) and solve for σ_u as a function of σ. This generates the two equations:

$$\sigma_u = \frac{c_{t_1}}{Z_{2,t_1}} - \frac{Z_{1,\Delta t_1}}{Z_{2,t_1}}\sigma,\quad c_{t_1} \in \{a_{t_1}, b_{t_1}\}\ \ (Eq.1\ and\ 2)$$

Similarly, for the second set of observations $\{b_{t_2}, a_{t_2}\}$, we use the second set of inequalities ($j = 2$) and solve for σ_u as a function of σ. This generates the two equations:

$$\sigma_u = \frac{c_{t_2}}{Z_{2,t_2}} - \frac{Z_{1,\Delta t_1} + Z_{1,\Delta t_2}}{Z_{2,t_2}}\sigma,\quad c_{t_2} \in \{a_{t_2}, b_{t_2}\}\ \ (Eq.3\ and\ 4)$$

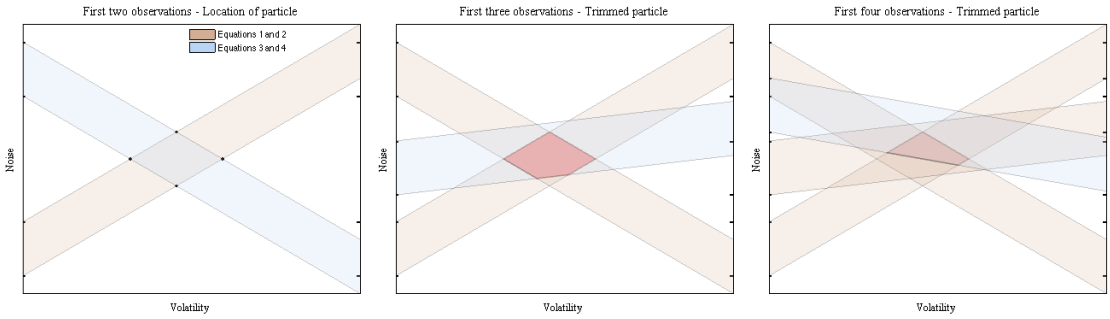


Figure 3.2: **Practical illustration of the sapling scheme.** The left panel displays the how the first four equations, Eq.1 to Eq.4, can be used to pin down the location of the particle in the space of (σ, σ_u). The middle panel displays how equations Eq.5 and Eq.6 trim the parallelogram. All points in the red shaded area satisfy the first three inequalities. The right panel in figure displays the trimming procedure using the fourth set of observations $\{b_{t_4}, a_{t_4}\}$. Similarly, all points in the red shaded area satisfy the first four inequalities.

Combining these four equations one can identify the four vertices of the parallelogram (a 2-dimensional polyhedron) and any point in the interior of the formed parallelogram satisfies the first tow inequalities.

Subsequently, Z_{t_3} is sampled from the conditional proposal distribution $\widetilde{\pi}_{t_3|t_1:t_2}$; equation 3.4.5. Given Z_{t_3} and $Z_{t_1:t_2}$, we can use the third set of inequalities, corresponding to observations $\{b_{t_3}, a_{t_3}\}$ and trim the parallelogram. The new observations generate two equations:

$$\sigma_u = \frac{c_{t_3}}{Z_{2,t_3}} - \frac{Z_{1,\Delta t_1} + Z_{1,\Delta t_2} + Z_{1,\Delta t_3}}{Z_{2,t_3}}\sigma, \quad c_{t_3} \in \{a_{t_3}, b_{t_3}\} \ \ (Eq.5 \ and \ 6)$$

The middle panel in figure 3.2 illustrates the trimming the parallelogram. The right panel in figure 3.2 displays the trimming procedure using the fourth set of observations $\{b_{t_4}, a_{t_4}\}$.

3.4.2 Resampling - Alteration Step

In our setup the resampling step resembles that of a general SMC algorithm. To overcome the degeneracy of the particle system as t_m increases, we measure the effective sample size (ESS) at time t_m

$$ESS_{t_m} = \left(\sum_{k=1}^{N} W_{t_1:t_m}^{(k)}\right)^2 / \sum_{k=1}^{N} \left(W_{t_1:t_m}^{(k)}\right)^2 \tag{3.4.8}$$

and if the ESS for the particle system has dropped below a designated threshold (usually $N/2$), the particle system is resampled removing the particles with low weights and replicating the particles with higher weights. In this setup, replicating particles will not generate a representative sample of the fiducial distribution. As mentioned above, each of the particles forms a polyhedron in the parameter space, and therefore, if the particles are simply copied, the polyhedrons will be concentrated in a narrow area, due to particles with initially higher weight. Moreover, as the algorithm progresses, the particles will not be able to move from those regions. A solution to this issue is to alter the particles selected from resampling in a way that they will maintain their high weight, while still allowing for an appropriate sample of the fiducial distribution.

The alteration step is performed as follows: Suppose that at time t_m particle K is selected in the resampling step and will be copied. Up to time t_m, we have observed the following inequalities in vector form (for simplicity dependence in K is suppressed):

$$\mathbf{b}_m \leq \mathbf{T}_m \mu + \sigma \mathbf{V}_{1,m} \mathbf{Z}_{1,m} + \sigma_u \mathbf{V}_{2,m} \mathbf{Z}_{2,m} \leq \mathbf{a}_m$$

where $\mathbf{T}_m = (t_1, t_2, \ldots, t_m)'$, $\mathbf{V}_{2,m} = \mathbf{I}_m$ and

$$\mathbf{V}_{1,m} = \begin{bmatrix} \sqrt{t_1} & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \sqrt{t_1} & \sqrt{t_2} & \cdots & \sqrt{t_m} \end{bmatrix}$$

Then we form matrix $A = [-\mathbf{T}_m \ -\mathbf{V}_{1,m}\mathbf{Z}_{1,m} \ \mathbf{V}_{2,m}]$ and decompose $\mathbf{Z}_{2,m}$ the following way

$$\mathbf{Z}_{2,m} = \Pi_{n(A)} \mathbf{Z}_{2,m} + \left\| \mathbf{Z}_{2,m} - \Pi_{n(A)} \mathbf{Z}_{2,m} \right\| \frac{\mathbf{Z}_{2,m} - \Pi_{n(A)} \mathbf{Z}_{2,m}}{\left\| \mathbf{Z}_{2,m} - \Pi_{n(A)} \mathbf{Z}_{2,m} \right\|} \tag{3.4.9}$$

where $\Pi_{n(A)} \mathbf{Z}_{2,m}$ denotes the projection of $\mathbf{Z}_{2,m}$onto the null space of matrix A, which is

$n(A) = \{\eta : A\eta = 0\}$ and $\|\cdot\|$ denotes the L^2 norm. Since A has the form $A = \left[\widetilde{A}\ \mathbf{V}_{2,m}\right]$, where $\widetilde{A} = [-\mathbf{T}_m\ -\mathbf{V}_{1,m}\mathbf{Z}_{1,m}]$, we can further decompose the basis of the null space as $\eta = (\eta_1, \eta_2)'$ and write $A\eta = \widetilde{A}\eta_1 + \mathbf{V}_{2,m}\eta_2 = 0$, where η_2 is orthonormal, i.e. $\eta_2^\intercal\eta_2 = \mathbf{I}$. From this decomposition we can write $\Pi_{n(A)}\mathbf{Z}_{2,m} = \eta_2\eta_2^\intercal\mathbf{Z}_{2,m}$. By setting $C = \eta_2^\intercal\mathbf{Z}_{2,m}$, $D = \left\|\mathbf{Z}_{2,m} - \Pi_{n(A)}\mathbf{Z}_{2,m}\right\|$ and $\kappa = \frac{\mathbf{Z}_{2,m} - \Pi_{n(A)}\mathbf{Z}_{2,m}}{\left\|\mathbf{Z}_{2,m} - \Pi_{n(A)}\mathbf{Z}_{2,m}\right\|}$ decomposition 3.4.9 can be rewritten as

$$\mathbf{Z}_{2,m} = \eta_2 C + D\kappa \tag{3.4.10}$$

Moreover, by assumption $\mathbf{Z}_{2,m} \sim N(0, \mathbf{I}_m)$, $C \sim N(0, \mathbf{I}_m)$ and $D \sim \sqrt{\chi^2_{m-d}}$ where $d = rank\,(n(A))$. Also C and D are independent by design.

Decomposition 3.4.9 allows us to alter $\mathbf{Z}_{2,m}$ by sampling new values for C and D according to their distributions. If we denote by $\widetilde{C}$ and $\widetilde{D}$ the generated values, then we have a new value for $\mathbf{Z}_{2,m}$, namely $\widetilde{\mathbf{Z}}_{2,m} = \eta_2\widetilde{C} + \widetilde{D}\kappa$. The final step in the alteration procedure is to update the set

$$Q = \left\{(\mu, \sigma, \sigma_u) : \mathbf{b}_m \leq \mathbf{T}_m\mu + \sigma\mathbf{V}_{1,m}\mathbf{Z}_{1,m} + \sigma_{\mathbf{u}}\mathbf{V}_{2,m}\mathbf{Z}_{2,m} \leq \mathbf{a}_m\right\}$$

for the new $\widetilde{\mathbf{Z}}_{2,m}$. We achieve that by noting that, since (μ, σ, σ_u) solve $\mathbf{b}_m \leq \mathbf{T}_m\mu + \sigma\mathbf{V}_{1,m}\mathbf{Z}_{1,m} + \sigma_{\mathbf{u}}\mathbf{V}_{2,m}\mathbf{Z}_{2,m} \leq \mathbf{a}_m$, then we need to identify $(\widetilde{\mu}, \widetilde{\sigma}, \widetilde{\sigma}_u)$ that solve $\mathbf{b}_m \leq \mathbf{T}_m\widetilde{\mu} + \widetilde{\sigma}\mathbf{V}_{1,m}\mathbf{Z}_{1,m} + \widetilde{\sigma}_{\mathbf{u}}\mathbf{V}_{2,m}\widetilde{\mathbf{Z}}_{2,m} \leq \mathbf{a}_m$. Furthermore, if we consider the following equality

$$\mathbf{T}_m\mu + \sigma\mathbf{V}_{1,m}\mathbf{Z}_{1,m} + \sigma_{\mathbf{u}}\mathbf{V}_{2,m}(\eta_2 C + D\kappa) = \mathbf{T}_m\widetilde{\mu} + \widetilde{\sigma}\mathbf{V}_{1,m}\mathbf{Z}_{1,m} + \widetilde{\sigma}_{\mathbf{u}}\mathbf{V}_{2,m}\left(\eta_2\widetilde{C} + \widetilde{D}\kappa\right)$$

together with the fact that $[-\mathbf{T}_m\ -\mathbf{V}_{1,m}\mathbf{Z}_{1,m}]\,\eta_1 + \mathbf{V}_{2,m}\eta_2 = 0$ and the orthogonality of C and D, then we can easily identify the updated $(\widetilde{\mu}, \widetilde{\sigma}, \widetilde{\sigma}_u)$

$$\widetilde{\sigma}_u = \sigma_u\frac{D}{\widetilde{D}}$$

$$\begin{pmatrix} \widetilde{\mu} \\ \widetilde{\sigma} \end{pmatrix} = \begin{pmatrix} \mu \\ \sigma \end{pmatrix} - \sigma_u\eta_1\left(\frac{\widetilde{C}D}{\widetilde{D}} - C\right)$$

After altering $\mathbf{Z}_{2,m}$, the procedure is repeated to alter $\mathbf{Z}_{1,m}$, which, as mentioned above, is

The SMC algorithm

Step	Action
1. Initialization	For $k = 1, 2, ..., K$ draw $Z^{(k)}_{t_1:t_3} \sim \pi_{t_1:t_3}$ (Eq. 3.4.2) and set $W^{(k)}_{t_1:t_3} = 1$
2. For $t_m > t_3$ and $t_m \leq t_n$	For $k = 1, 2, ..., K$ draw $Z^{(k)}_{t_m} \sim \widetilde{\pi}_{t_m \mid t_1:t_{m-1}}$ (Eq. 3.4.5)
3. For $t_m > t_3$ and $t_m \leq t_n$	Calculate weights $W^{(k)}_{t_1:t_m} = W^{(k)}_{t_m} W^{(k)}_{t_1:t_{m-1}}$ (Eq. 3.4.6)
4. For $t_m > t_3$ and $t_m \leq t_n$	Calculate ESS_{t_m} (Eq. 3.4.8). If $ESS_{t_m} \leq threshold$ go to step 5. Else go to step 2 and set $m = m + 1$
5. For t_m given $ESS_{t_m} \leq thd$	Resample particles and set $W_{t_1:t_m} = N^{-1}$. Go to step 6.
6. For t_m	Perform alteration as described above and set $m = m + 1$

now $N(0, \mathbf{I}_m)$. The table below gives an outline of the steps of the algorithm.

3.4.3 Convergence of the Algorithm

In this section, we are interested in establishing the behavior of the particle system as the number of particles approaches infinity. The following result is based on the results for general SMC methods by Douc and Moulines (2008) and follows closely the proof by Cisewski and Hannig (2012). For a particle system generated by the proposal distribution 3.4.4 targeting distribution 3.4.2, we want to establish the following result.

Theorem 2. *Given a particle system* $\left\{W^{(J)}_{\tau_1:\tau_n}, Z^{(J)}_{\tau_1:\tau_n}\right\}_{J=1}^{N}$ *targeting 3.4.2, then for any function* f *belonging in a proper set (defined below),*

$$\left(\sum_{J=1}^{N} W^{(J)}_{\tau_1:\tau_n}\right)^{-1} \sum_{J=1}^{N} W^{(J)}_{\tau_1:\tau_n} f\left(Z^{(J)}_{\tau_1:\tau_n}\right) \xrightarrow{P} \int f\left(Z_{\tau_1:\tau_n}\right) d\pi_{\tau_1:\tau_n}, \ as \ N \to \infty \tag{3.4.11}$$

The theorem simply states that, as the number of particles grows, the weighted sample consistently estimates the target distribution. Douc and Moulines (2008) provide a framework for weighted sample consistency. We will adapt their framework in our setup, first by stating a series of definitions and results and then by applying them to our framework. We start with the definition of a proper set:

Definition 1. *Let* $C \subseteq X$ *be a subset of a general state space* X*. The set* C *is said to be proper if it satisfies the following conditions: (i) For any* $f, g \in C$ *and real numbers* a, b*,* C *is a linear space, i.e.,* $af + bg \in C$*. (ii) If some measurable function* f *satisfies* $|f| \leq |g|$ *where* $g \in C$*, then* $f \in C$ *and (iii)* C *contains all constant functions.*

Using the notion of a proper set, weighted sample consistency is defines as follows:

Definition 2. *Denote by π a probability measure on $(\Theta, \mathbb{B}(\Theta))$. A weighted sample $\{W_i, Z_i\}_{i=1}^N$ on Θ is said to be consistent for the probability measure π and the proper set C if for any $f \in C$, as $N \to \infty$, $\Omega_N^{-1} \sum_{i=1}^N W_i f(Z_i) \xrightarrow{P} \int_\Theta f(z)\,\pi(dz)$ and $\Omega_N^{-1} \max_{1 \le i \le N} W_i \xrightarrow{P} 0$, where $\Omega_N = \sum_{i=1}^N W_i$.*

Next, we state, without proof, theorem A.11 by Douc and Moulines (2008), which is essential for establishing consistency.

Theorem 3. *Let $\{U_j\}_{j=1}^N$ denote a triangular array of random variables defined on $(\Omega, \mathcal{F}, P)$ and $\{\mathcal{F}_j\}_{j=1}^N$ denote a triangular array of sigma-fields, such that $\mathcal{F}_{j-1} \subseteq \mathcal{F}_j$. Each random variable U_j is measurable with respect to $\mathcal{F}_j$. Moreover we assume that $E[|U_j|\,|\mathcal{F}_{j-1}] < \infty$ for every $j = 1, ..., N$, and*

$$\sup_N P\left(\sum_{j=1}^N E[|U_j|\,|\mathcal{F}_{j-1}] \ge \lambda\right) \to 0 \quad as\ \lambda \to \infty \tag{3.4.12}$$

$$\sum_{j=1}^N E[|U_j|\,\mathbf{1}\{|U_j| \ge \epsilon\}\,|\mathcal{F}_{j-1}] \xrightarrow{P} 0 \quad for\ any\ \epsilon > 0 \tag{3.4.13}$$

Then

$$\max_{1 \le i \le N} \left| \sum_{j=1}^i U_j - \sum_{j=1}^i E[U_j|\mathcal{F}_{j-1}] \right| \xrightarrow{P} 0$$

Now we have the necessary tools to establish consistency. We will tackle the proof by focusing on the transition from τ_{m-1} to τ_m for the sampling step and then we will address the resampling and alteration steps at time τ_m. We begin by defining a collection of sigma fields which will be used throughout the proof. Let $\mathcal{F}_0 \triangleq \sigma\left(\left\{Z_{\tau_1:\tau_{m-1}}^{(J)}\right\}_{J=1}^N, (b,a]_{\tau_1:\tau_m}\right)$ and $\mathcal{F}_J \triangleq \mathcal{F}_0 \vee \sigma\left(\left\{Z_{\tau_1:\tau_m}^{(K)}\right\}_{K=1}^J, (b,a]_{\tau_1:\tau_m}\right)$. The first sigma field contains the information of the particles up to time τ_{m-1} and the data up to time τ_m. The J^{th} indexed sigma fields represent a triangular array of sigma fields, each augmented with the information of the particles up to time τ_m.

Since we will be focusing on the transition from τ_{m-1} to τ_m, we need to define the transition kernel from from τ_{m-1} to τ_m. The transition kernel from $\left(\Theta_{\tau_1:\tau_{m-1}}, \mathbb{B}\left(\Theta_{\tau_1:\tau_{m-1}}\right)\right) \times (\Theta_{\tau_m}, \mathbb{B}(\Theta_{\tau_m}))$ into $[0,1]$ is the map $F : \Theta_{\tau_1:\tau_{m-1}} \times \mathbb{B}(\Theta_{\tau_m}) \to [0,1]$. For any function

$f \in \mathbb{B}\left(\Theta_{\tau_1:\tau_{m-1}} \times \Theta_{\tau_m}\right)$ such that

$$\int_{\Theta_{\tau_m}} \left| f\left(Z_{\tau_1:\tau_{m-1}}, Z_{\tau_m}\right)\right| F\left(Z_{\tau_1:\tau_{m-1}}, dZ_{\tau_m}\right) < \infty$$

we define

$$F\left(Z_{\tau_1:\tau_{m-1}}, f\right) = \int_{\Theta_{\tau_m}} f\left(Z_{\tau_1:\tau_{m-1}}, Z_{\tau_m}\right) F\left(Z_{\tau_1:\tau_{m-1}}, dZ_{\tau_m}\right) \tag{3.4.14}$$

where, in our setup,

$$F\left(Z_{\tau_1:\tau_{m-1}}, A\right) = \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right)\left(\Phi\left(R_m\right) - \Phi\left(L_m\right)\right) \pi_{\tau_m|\tau_1:\tau_{m-1}}\left(A\right)$$

for any $A \in \Theta_{\tau_m}$. The following definition defines the notion os a proper set. Next, we define the following set

$$C_{\tau_m} = \left\{ f \in L^1\left(\Theta_{\tau_1:\tau_m}, \mathbb{B}\left(\Theta_{\tau_1:\tau_m}\right), \pi_{\tau_1:\tau_m}\right), F\left(\cdot, |f|\right) \in C_{\tau_{m-1}} \right\} \tag{3.4.15}$$

Lemma 3. *The set C_{τ_m}, defined above, is proper.*

Proof. Let a, b be real numbers and $f, g \in C_{\tau_m}$. The latter implies that $\int |f|\, d\pi_{\tau_1:\tau_m} < \infty$ and $\int |g|\, d\pi_{\tau_1:\tau_m} < \infty$. It follows that $\int |af + bg|\, d\pi_{\tau_1:\tau_m} < \infty$, since $|af + bg| \leq |a|\,|f| + |b|\,|g|$, therefore $af + bg \in C_{\tau_m}$. Next, we consider some measurable function f which satisfies $|f| \leq |g|$ where $g \in C_t$. Then $f \in C_{\tau_m}$ since $\int |f|\, d\pi_{\tau_1:\tau_m} \leq \int |g|\, d\pi_{\tau_1:\tau_m} < \infty$. Finally, C_{τ_m} contains all constant functions since $\pi_{\tau_1:\tau_m}\left(\Theta_{\tau_1:\tau_m}\right) = 1$. □

Consistency in our framework is defined as follows:

Definition 3. *The weighted sample $\left\{W^{(J)}_{\tau_1:\tau_m}, Z^{(J)}_{\tau_1:\tau_m}\right\}_{J=1}^{N}$ is consistent for the probability measure $\pi_{\tau_1:\tau_m}$ and the proper set C_{τ_m} if for any $f \in C_{\tau_m}$, as $N \to \infty$, $\Omega_{\tau_m}^{-1} \sum_{J=1}^{N} W^{(J)}_{\tau_1:\tau_m} f\left(Z^{(J)}_{\tau_1:\tau_m}\right) \xrightarrow{P} \int f\left(z_{\tau_1:\tau_m}\right) \pi_{\tau_1:\tau_m}\left(dz_{\tau_1:\tau_m}\right) \triangleq \pi_{\tau_1:\tau_m}\left(f\right)$ and $\Omega_{\tau_m}^{-1} \underset{1\leq J\leq N}{max} W^{(J)}_{\tau_1:\tau_m} \xrightarrow{P} 0$, where $\Omega_{\tau_m} = \sum_{J=1}^{N} W^{(J)}_{\tau_1:\tau_m}$.*

The following theorem establishes the consistency of the sampling step:

Theorem 4. *(Sampling Step) Let C_{τ_m} be as in 3.4.15 and let $\left\{W^{(J)}_{\tau_1:\tau_{m-1}}, Z^{(J)}_{\tau_1:\tau_{m-1}}\right\}_{J=1}^{N}$ be a weighted sample consistent for $\left(\pi_{\tau_1:\tau_{m-1}}, C_{\tau_{m-1}}\right)$. Then the weighted sample*

$\left\{W^{(J)}_{\tau_1:\tau_m}, Z^{(J)}_{\tau_1:\tau_m}\right\}_{J=1}^{N}$ *is consistent for* $(\pi_{\tau_1:\tau_m}, C_{\tau_m})$.

Proof. Let $f \in C_{\tau_m}$. We will consider the triangular array of random variables

$$\tilde{U}_J = E\left[W^{(J)}_{\tau_1:\tau_m} f\left(Z^{(J)}_{\tau_1:\tau_m}\right)|\mathcal{F}_{J-1}\right]$$

For these random variables the following identity holds

$$\begin{aligned}
E\left[W^{(J)}_{\tau_1:\tau_m} f\left(Z^{(J)}_{\tau_1:\tau_m}\right)|\mathcal{F}_{J-1}\right] &= W^{(J)}_{\tau_1:\tau_{m-1}} E\left[W^{(J)}_{\tau_m} f\left(Z^{(J)}_{\tau_1:\tau_m}\right)|\mathcal{F}_{J-1}\right] \\
&= W^{(J)}_{\tau_1:\tau_{m-1}} \int f\left(Z^{(J)}_{\tau_1:\tau_m}\right) \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) W^{(J)}_{\tau_m} \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right) \\
&= W^{(J)}_{\tau_1:\tau_{m-1}} \int f\left(Z^{(J)}_{\tau_1:\tau_m}\right) \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \left(\Phi\left(R_m\right) - \Phi\left(L_m\right)\right) \pi_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right) \\
&= W^{(J)}_{\tau_1:\tau_{m-1}} F\left(Z_{\tau_1:\tau_{m-1}}, f\right)
\end{aligned} \tag{3.4.16}$$

We derived this equation using 3.4.14, noting that in our setup Z_{τ_m} is drawn from $\tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}$ and the relation of the conditional proposal and target probability measures is given by $W_{\tau_m}\tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(Z_{\tau_m}\right) = \left(\Phi\left(R_m\right) - \Phi\left(L_m\right)\right)\pi_{\tau_m|\tau_1:\tau_{m-1}}\left(Z_{\tau_m}\right)$. For the triangular array $U_J = \Omega^{-1}_{\tau_{m-1}} E\left[W^{(J)}_{\tau_1:\tau_m} f\left(Z^{(J)}_{\tau_1:\tau_m}\right)|\mathcal{F}_{J-1}\right]$ we can show the following

$$\begin{aligned}
\Omega^{-1}_{\tau_{m-1}} \sum_{J=1}^{N} E\left[W^{(J)}_{\tau_1:\tau_m} f\left(Z^{(J)}_{\tau_1:\tau_m}\right)|\mathcal{F}_{J-1}\right] &= \Omega^{-1}_{\tau_{m-1}} \sum_{J=1}^{N} W^{(J)}_{\tau_1:\tau_{m-1}} F\left(Z^{(J)}_{\tau_1:\tau_{m-1}}, f\right) \\
&\xrightarrow{P} \int F\left(Z_{\tau_1:\tau_{m-1}}, f\right) \pi\left(dz_{\tau_1:\tau_{m-1}}\right) \\
&= \pi_{\tau_1:\tau_{m-1}}\left(F\left(Z_{\tau_1:\tau_{m-1}}, f\right)\right)
\end{aligned} \tag{3.4.17}$$

The first equality comes from equation 3.4.16 and convergence is established since the weighted sample $\left\{W^{(J)}_{\tau_1:\tau_{m-1}}, Z^{(J)}_{\tau_1:\tau_{m-1}}\right\}_{J=1}^{N}$ is consistent for $\left(\pi_{\tau_1:\tau_{m-1}}, C_{\tau_{m-1}}\right)$ and $F\left(Z_{\tau_1:\tau_{m-1}}, f\right) \in C_{\tau_{m-1}}$. Next we will apply theorem 3 for the triangular array $U_J =$

$\Omega_{\tau_{m-1}}^{-1} W_{\tau_1:\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)$ starting with equation 3.4.12. Note that

$$\begin{aligned}\sum_{J=1}^{N} E\left[\left|\Omega_{\tau_{m-1}}^{-1} W_{\tau_1:\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right| |\mathcal{F}_{J-1}\right] &= \Omega_{\tau_{m-1}}^{-1} \sum_{J=1}^{N} W_{\tau_1:\tau_{m-1}}^{(J)} E\left[W_{\tau_m}^{(J)} \left| f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right| |\mathcal{F}_{J-1}\right] \\ &= \Omega_{\tau_{m-1}}^{-1} \sum_{J=1}^{N} W_{\tau_1:\tau_{m-1}}^{(J)} F\left(Z_{\tau_1:\tau_{m-1}}^{(J)}, |f|\right) \\ &\to \pi_{\tau_1:\tau_{m-1}}\left(F\left(Z_{\tau_1:\tau_{m-1}}, |f|\right)\right)\end{aligned} \tag{3.4.18}$$

Therefore condition 3.4.12 (tightness)

$$\underset{N}{sup} P\left(\sum_{j=1}^{N} E\left[\left|\Omega_{\tau_{m-1}}^{-1} W_{\tau_1:\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right| |\mathcal{F}_{j-1}\right] \geq \lambda\right) \to 0 \;\; as \; \lambda \to \infty$$

follows from the fact that $F\left(\cdot, |f|\right) \in C_{\tau_{m-1}}$. We continue to establish equation 3.4.13, by introducing first another transition kernel based on the conditional proposal distribution. Namely we define

$$G\left(Z_{\tau_1:\tau_{m-1}}, A\right) = \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(A\right)$$

for for any $A \in \Theta_{\tau_m}$. In a similar manner we define, assuming integrability $G\left(Z_{\tau_1:\tau_{m-1}}, |f|\right) < \infty$

$$\begin{aligned}G\left(Z_{\tau_1:\tau_{m-1}}, f\right) &= \int_{\Theta_{\tau_m}} f\left(Z_{\tau_1:\tau_{m-1}}, Z_{\tau_m}\right) F\left(Z_{\tau_1:\tau_{m-1}}, dZ_{\tau_m}\right) \\ &= \int_{\Theta_{\tau_m}} f\left(Z_{\tau_1:\tau_{m-1}}, Z_{\tau_m}\right) \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right)\end{aligned}$$

It is easy to that

$$G\left(Z_{\tau_1:\tau_{m-1}}, W_{\tau_m}|f|\right) = F\left(Z_{\tau_1:\tau_{m-1}}, |f|\right)$$

Therefore for $\epsilon > 0$ it holds

$$G\left(Z_{\tau_1:\tau_{m-1}}, W_{\tau_m}|f| \mathbf{1}\left\{W_{\tau_m}|f| \geq \epsilon\right\}\right) \leq G\left(Z_{\tau_1:\tau_{m-1}}, W_{\tau_m}|f|\right) = F\left(Z_{\tau_1:\tau_{m-1}}, |f|\right)$$

therefore $G\left(\cdot, W_{\tau_m}|f| \mathbf{1}\left\{W_{\tau_m}|f| \geq \epsilon\right\}\right) \in C_{\tau_{m-1}}$. Using again the triangular array $U_J =$

$\Omega_{\tau_{m-1}}^{-1} W_{\tau_1:\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)$, for $\epsilon > 0$ we have the following

$$
\begin{aligned}
&\sum_{J=1}^{N} E\left[\left|\Omega_{\tau_{m-1}}^{-1} W_{\tau_1:\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right| \mathbf{1}\left\{\left|\Omega_{\tau_{m-1}}^{-1} W_{\tau_1:\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right| \geq \epsilon\right\} |\mathcal{F}_{j-1}\right] \\
&\quad \times \mathbf{1}\left\{\Omega_{\tau_{m-1}}^{-1} \underset{1\leq I\leq N}{max} W_{\tau_1:\tau_m}^{(I)} \leq \epsilon/K\right\} \\
&= \Omega_{\tau_{m-1}}^{-1} \sum_{J=1}^{N} W_{\tau_1:\tau_{m-1}}^{(J)} \int \left|W_{\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right| \mathbf{1}\left\{\left|\Omega_{\tau_{m-1}}^{-1} W_{\tau_1:\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right| \geq \epsilon\right\} \\
&\quad \times \mathbf{1}\left\{\Omega_{\tau_{m-1}}^{-1} \underset{1\leq I\leq N}{max} W_{\tau_1:\tau_m}^{(I)} \leq \epsilon/K\right\} \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right) \\
&\leq \Omega_{\tau_{m-1}}^{-1} \sum_{J=1}^{N} W_{\tau_1:\tau_{m-1}}^{(J)} \int \left|W_{\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right| \\
&\quad \times \mathbf{1}\left\{\left|\Omega_{\tau_{m-1}}^{-1} W_{\tau_1:\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right| \geq \Omega_{\tau_{m-1}}^{-1} \underset{1\leq I\leq N}{max} W_{\tau_1:\tau_m}^{(I)} K\right\} \\
&\quad \times \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right) \\
&= \Omega_{\tau_{m-1}}^{-1} \sum_{J=1}^{N} W_{\tau_1:\tau_{m-1}}^{(J)} \int \left|W_{\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right| \mathbf{1}\left\{\frac{W_{\tau_1:\tau_m}^{(J)}}{\underset{1\leq I\leq N}{max} W_{\tau_1:\tau_m}^{(I)}} \left|f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right| \geq K\right\} \\
&\quad \times \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right) \\
&\leq \Omega_{\tau_{m-1}}^{-1} \sum_{J=1}^{N} W_{\tau_1:\tau_{m-1}}^{(J)} \int \left|W_{\tau_m}^{(J)} f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right| \mathbf{1}\left\{W_{\tau_m}^{(J)} \left|f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right| \geq K\right\} \\
&\quad \times \mathbf{I}_{\star}\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right) \\
&= \Omega_{\tau_{m-1}}^{-1} \sum_{J=1}^{N} W_{\tau_1:\tau_{m-1}}^{(J)} G\left(Z_{\tau_1:\tau_{m-1}}, W_{\tau_m} \left|f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right| \mathbf{1}\left\{W_{\tau_m} \left|f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right| \geq K\right\}\right) \\
&\xrightarrow{P} \int G\left(Z_{\tau_1:\tau_{m-1}}, W_{\tau_m} \left|f\left(Z_{\tau_1:\tau_{m-1}}\right)\right| \mathbf{1}\left\{W_{\tau_m} \left|f\left(Z_{\tau_1:\tau_{m-1}}\right)\right| \geq K\right\}\right) \pi_{\tau_1:\tau_{m-1}}\left(dZ_{\tau_1:\tau_{m-1}}\right) \\
&\xrightarrow{P} 0
\end{aligned}
$$

as $K \to \infty$. The final result is derived by a direct application of the dominated convergence theorem since $\mathbf{1}\left\{W_{\tau_m} \left|f\left(Z_{\tau_1:\tau_{m-1}}\right)\right| \geq K\right\} \xrightarrow{P} 0$ as $K \to \infty$ and the dominating function

being $F\left(Z_{\tau_1:\tau_{m-1}},|f|\right)$. Also, noting that $\mathbf{1}\left\{\Omega_{\tau_{m-1}}^{-1}\underset{1\leq I\leq N}{max}W_{\tau_1:\tau_m}^{(I)}\leq\epsilon/K\right\}\xrightarrow{P}1$ since by assumption $\Omega_{\tau_{m-1}}^{-1}\underset{1\leq I\leq N}{max}W_{\tau_1:\tau_m}^{(I)}\xrightarrow{P}0$ it is easy to see that

$$\begin{aligned}
&\sum_{J=1}^{N}E\left[\left|\Omega_{\tau_{m-1}}^{-1}W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right|\mathbf{1}\left\{\left|\Omega_{\tau_{m-1}}^{-1}W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right|\geq\epsilon\right\}|\mathcal{F}_{j-1}\right]\\
&\quad\times\mathbf{1}\left\{\Omega_{\tau_{m-1}}^{-1}\underset{1\leq I\leq N}{max}W_{\tau_1:\tau_m}^{(I)}\leq\epsilon/K\right\}\\
&\quad\xrightarrow{P}\sum_{J=1}^{N}E\left[\left|\Omega_{\tau_{m-1}}^{-1}W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\right|\mathbf{1}\left\{\left|\Omega_{\tau_{m-1}}^{-1}W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_{m-1}}^{(J)}\right)\right|\geq\epsilon\right\}|\mathcal{F}_{j-1}\right]
\end{aligned}$$

These two results together imply condition 3.4.13 of the theorem above. The theorem then states that

$$\underset{1\leq I\leq N}{max}\left|\Omega_{\tau_{m-1}}^{-1}\sum_{J=1}^{I}W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_m}^{(J)}\right)-\Omega_{\tau_{m-1}}^{-1}\sum_{J=1}^{I}E\left[W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_m}^{(J)}\right)|\mathcal{F}_{j-1}\right]\right|\xrightarrow{P}0\quad(3.4.19)$$

Equation 3.4.19 together with equation 3.4.17 imply that

$$\Omega_{\tau_{m-1}}^{-1}\sum_{J=1}^{I}W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_m}^{(J)}\right)\xrightarrow{P}\pi_{\tau_1:\tau_{m-1}}\left(F\left(Z_{\tau_1:\tau_{m-1}},f\right)\right)$$

Applying this for $f=1$ we get

$$\Omega_{\tau_{m-1}}^{-1}\Omega_{\tau_m}=\Omega_{\tau_{m-1}}^{-1}\sum_{J=1}^{I}W_{\tau_1:\tau_m}^{(J)}1\xrightarrow{P}\pi_{\tau_1:\tau_{m-1}}\left(F\left(Z_{\tau_1:\tau_{m-1}},1\right)\right)$$

which will be used to complete the proof. That is

$$\begin{aligned}
\Omega_{\tau_m}^{-1}\sum_{J=1}^{I}W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_m}^{(J)}\right)&=\frac{\Omega_{\tau_{m-1}}^{-1}\sum_{J=1}^{I}W_{\tau_1:\tau_m}^{(J)}f\left(Z_{\tau_1:\tau_m}^{(J)}\right)}{\Omega_{\tau_{m-1}}^{-1}\Omega_{\tau_m}}\xrightarrow{P}\frac{\pi_{\tau_1:\tau_{m-1}}\left(F\left(Z_{\tau_1:\tau_{m-1}},f\right)\right)}{\pi_{\tau_1:\tau_{m-1}}\left(F\left(Z_{\tau_1:\tau_{m-1}},1\right)\right)}\\
&=\frac{\int\cdots\int f\left(Z_{\tau_1:\tau_m}\right)\mathbf{I}_\star\left(Z_{2,\tau_m}\right)\mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right)\pi_{\tau_1:\tau_m}\left(dz_{\tau_1:\tau_m}\right)}{\int\cdots\int\mathbf{I}_\star\left(Z_{2,\tau_m}\right)\mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right)\pi_{\tau_1:\tau_m}\left(dz_{\tau_1:\tau_m}\right)}\\
&=\pi_{\tau_1:\tau_m}f
\end{aligned}$$

The result follows since $\pi_{\tau_1:\tau_m}=\left(\Phi\left(R_m\right)-\Phi\left(L_m\right)\right)\pi_{\tau_m|\tau_1:\tau_{m-1}}\pi_{\tau_1:\tau_{m-1}}$. □

Then next theorem, Theorem 3 by Douc and Moulines (2008), ensures that the consistency of the algorithm is still valid after the resampling step.

Theorem 5. *(Resampling Step, Theorem 3 by Douc and Moulines (2008)) If the weighted sample* $\left\{W_{\tau_1:\tau_n}^{(J)}, Z_{\tau_1:\tau_n}^{(J)}\right\}_{J=1}^{N}$ *is consistent for* $(\pi_{\tau_1:\tau_n}, C_{\tau_n})$*, then the uniformly weighted sample* $\left\{\mathbf{1}_n^{(J)}, \tilde{Z}_{\tau_1:\tau_n}^{(J)}\right\}_{J=1}^{N}$ *obtained using multinomial or deterministic-plus-residual multinomial resampling is also consistent for* $(\pi_{\tau_1:\tau_n}, C_{\tau_n})$*.*

The following theorem ensures that the consistency of the algorithm is still valid after the alteration step.

Theorem 6. *(Alteration Step) If the uniformly weighted sample* $\left\{\mathbf{1}_n^{(J)}, Z_{\tau_1:\tau_n}^{(J)}\right\}_{J=1}^{N}$ *is consistent for* $(\pi_{\tau_1:\tau_n}, C_{\tau_n})$*, then the altered uniformly weighted sample* $\left\{\mathbf{1}_n^{(J)}, \tilde{Z}_{\tau_1:\tau_n}^{(J)}\right\}_{J=1}^{N}$ *is consistent for* $\left(\pi_{\tau_1:\tau_n}, \tilde{C}_{\tau_n}\right)$*.*

Proof. Consider $f \in C_{\tau_m}$ and the decomposition $\widetilde{Z}_{\tau_1:\tau_m} = \eta_2 \widetilde{C} + \widetilde{D}\kappa_{\tau_1:\tau_m}$, where $\widetilde{C}$ and $\widetilde{D}$ are the generated values used to alter the particle, $\kappa_{\tau_1:\tau_m} = \frac{Z_{\tau_1:\tau_m} - \Pi_{n(A)} Z_{\tau_1:\tau_m}}{\left\| Z_{\tau_1:\tau_m} - \Pi_{n(A)} Z_{\tau_1:\tau_m} \right\|}$ and η_2 is defined in the alteration step above. The decomposition refers to both $\mathbf{Z}_{1,m}$ and $\mathbf{Z}_{2,m}$. Also let, $\widetilde{\mathcal{F}}_0 \triangleq \sigma\left(\left\{\widetilde{Z}_{\tau_1:\tau_{m-1}}^{(J)}\right\}_{J=1}^{N}, (b,a]_{\tau_1:\tau_m}\right)$ and $\widetilde{\mathcal{F}}_J \triangleq \widetilde{\mathcal{F}}_0 \vee \sigma\left(\left\{\widetilde{Z}_{\tau_1:\tau_m}^{(K)}\right\}_{K=1}^{J}, (b,a]_{\tau_1:\tau_m}\right)$ be the collection of sigma fields corresponding to the altered sample. Then,

$$E\left(f\left(\widetilde{Z}_{\tau_1:\tau_m}^{(J)}\right)|\mathcal{F}_{j-1}\right) = \int f\left(\eta_2\widetilde{C} + \widetilde{D}\kappa_{\tau_1:\tau_m}\right) d\pi_{\widetilde{C},\widetilde{D}}$$

where $\pi_{\widetilde{C},\widetilde{D}}$ is the joint probability distribution of $\widetilde{C}$ and $\widetilde{D}$. Denote by $h_f\left(Z_{\tau_1:\tau_m}^{(J)}\right)$ this conditional expectation, which is a fiction of $Z_{\tau_1:\tau_m}$, by the definition of $\kappa_{\tau_1:\tau_m}$. We need to show that the altered particles still target the same target distribution. So it remains to show

$$\begin{aligned} &\frac{1}{N}\sum_{J=1}^{N} E\left[f\left(\widetilde{Z}_{\tau_1:\tau_m}^{(J)}\right)|\mathcal{F}_{J-1}\right] = \frac{1}{N}\sum_{J=1}^{N} h_f\left(Z_{\tau_1:\tau_m}^{(J)}\right) \\ &\quad\longrightarrow \int h_f\left(Z_{\tau_1:\tau_m}\right) d\pi_{\tau_1:\tau_m}\left(Z_{\tau_1:\tau_m}\right) = \int f\left(Z_{\tau_1:\tau_m}\right) d\pi_{\tau_1:\tau_m}\left(Z_{\tau_1:\tau_m}\right) \end{aligned} \tag{3.4.20}$$

since $E\left(f\left(\widetilde{Z}_{\tau_1:\tau_m}^{(J)}\right)|\mathcal{F}_{j-1}\right) = f\left(\widetilde{Z}_{\tau_1:\tau_m}^{(J)}\right)$. Ti establish the limit in 3.4.20, we need to sow

that $h_f \in C_{\tau_m}$. This requires f to be selected in a way such that, first

$$\int \left| h_f \left(\widetilde{Z}_{\tau_1:\tau_m} \right) \right| d\pi_{\tau_1:\tau_m} = \int \left| \int f \left(\eta_2 \widetilde{C} + \widetilde{D} \kappa_{\tau_1:\tau_m} \right) d\pi_{\widetilde{C},\widetilde{D}} \right| d\pi_{\tau_1:\tau_m} < \infty \tag{3.4.21}$$

and second

$$\begin{aligned} F\left(Z_{\tau_1:\tau_{m-1}}, h_f\right) = \int_{\Theta_{\tau_m}} & h_f\left(Z_{\tau_1:\tau_{m-1}}, Z_{\tau_m}\right) \mathbf{I}_\star\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \times \\ & \left(\Phi\left(R_m\right) - \Phi\left(L_m\right)\right) \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right) \\ = \int_{\Theta_{\tau_m}} & \int f\left(\eta_2 \widetilde{C} + \widetilde{D} \kappa_{\tau_1:\tau_m}\right) dP_{\widetilde{C},\widetilde{D}} \mathbf{I}_\star\left(Z_{2,\tau_m}\right) \mathbf{I}_{(L_m,R_m)}\left(Z_{1,\Delta\tau_m}\right) \times \\ & \left(\Phi\left(R_m\right) - \Phi\left(L_m\right)\right) \tilde{\pi}_{\tau_m|\tau_1:\tau_{m-1}}\left(dZ_{\tau_m}\right) < \infty \end{aligned} \tag{3.4.22}$$

Denote by $\tilde{C}_{\tau_n}$ the set where for $f \in C_{\tau_n}$ such that 3.4.21 and 3.4.22 hold. Then clearly, $\tilde{C}_{\tau_n}$ is a subset C_{τ_n} and is non-empty, since all bounded f satisfy 3.4.21 and 3.4.22. The final step in this proof is to show the last equality in 3.4.20. We have the following

$$\begin{aligned} \int h_f\left(Z_{\tau_1:\tau_m}\right) d\pi_{\tau_1:\tau_m} &= \int h_f^\star\left(\kappa_{\tau_1:\tau_m}\right) d\pi_{\kappa_{\tau_1:\tau_m}} = \int \left[\int f\left(\eta_2 \widetilde{C} + \widetilde{D} \kappa_{\tau_1:\tau_m}\right) d\pi_{\widetilde{C},\widetilde{D}} \right] d\pi_\kappa \\ &= \int \left[\int f\left(\eta_2 C + D \kappa_{\tau_1:\tau_m}\right) d\pi_{C,D} \right] d\pi_\kappa \\ &= \int f\left(\eta_2 C + D \kappa_{\tau_1:\tau_m}\right) d\pi_{C,D} \times d\pi_\kappa \\ &= \int f\left(Z_{\tau_1:\tau_m}\right) d\pi_{\tau_1:\tau_m} \end{aligned}$$

where simply $h_f^\star\left(\kappa_{\tau_1:\tau_m}\right) = h_f\left(Z_{\tau_1:\tau_m}\right)$. In this step, we used the fact that $\kappa_{\tau_1:\tau_m}\left(Z_{\tau_1:\tau_m}\right) = \kappa_{\tau_1:\tau_m}\left(\widetilde{Z}_{\tau_1:\tau_m}\right)$ and Fubini's theorem. That is, $\widetilde{C}$ and $\widetilde{D}$ are independent of κ by construction. □

3.5 Combinations of Fiducial Distributions

In this section we extend the combination scheme introduced in the previous chapter. Our scheme aggregates the generated samples from all blocks under consideration by re-weighing all particles. The new weights are computed through a metric that utilizes the Gaussian kernel, together with point estimates of the sample mean and the Fisher information matrix.

The procedure is the following.

Let (ξ_i, W_i) denote the generated particle systems in block $(\tau_{i-1}, \tau_i]$, $i = 1, ...M$, where $\xi_i = \left\{\xi_i^{(k)}\right\}_{k=1}^N = \left\{\mu_{\tau_{i-1}}^{(k)}, \sigma_{\tau_{i-1}}^{(k)}, \sigma_{u,\tau_{i-1}}^{(k)}\right\}_{k=1}^N$, $W_i = \left\{w_i^{(k)}\right\}_{k=1}^N$ and N is the number of particles used in the simulation. Suppose we want to combine fiducial distributions from M segments. For each particle k in block i and every other block $j = 1, ..., M$, $j \neq i$, we calculate the following weight

$$W_{i,j}^{(k)} = \exp\left\{-\frac{1}{2}\left(\xi_i^{(k)} - \hat{\xi}_j\right)' I_j^{-1}\left(\xi_i^{(k)} - \hat{\xi}_j\right) + \frac{1}{2}\log\left|\det I_j^{-1}\right|\right\}$$

where $\hat{\xi}_j$ denotes the point estimate $\hat{\xi}_j = \sum_{k=1}^N W_j^{(k)} \xi_j^{(k)}$ in block j and I_j denotes

$$I_j = \left[\sum_{k=1}^N W_j^{(k)}\left(\xi\,(l)_j^{(k)} - \hat{\xi}_j\right)\left(\xi(m)_j^{(k)} - \hat{\xi}_j\right)'\right]_{l,m=1,2,3}$$

which the point estimate of the 3×3 Fisher information matrix. In other words, for every particle k in block i, we calculate its "weighted" distance from the point estimate in block j through a Gaussian kernel, where $j = 1, ..., M$, $i \neq j$.

Once we generate weights $W_{i,j}^{(k)}$, the final weight for particle k in block i, is calculated by

$$\bar{W}_i^{(k)} = W_i^{(k)} \prod_{h \neq i} W_{i,h}^{(k)}$$

for all $i = 1, ..., M$.

3.6 Simulation and Robustness checks

In this section we report a Monte Carlo simulation study that investigates the performance of our procedure under constant volatility. Stochastic volatility is studied in section 4.2. To generate quotes data we will employ a version of the Li and Mykland (2007) two stage contamination scheme. Initially, at sampling times, additive noise contaminates the efficient price and, subsequently, the contaminated process is rounded upwards and downwards,

towards the two nearest ticks. The log-ask and bid prices at time t_m are given by

$$\begin{aligned} log(a_{t_m}) &= log\left(\gamma\left\lceil\frac{exp(X_{t_m}+U_{t_m})}{\gamma}\right\rceil\right) \\ log(b_{t_m}) &= log\left(\gamma\left\lfloor\frac{exp(X_{t_m}+U_{t_m})}{\gamma}\right\rfloor\right) \end{aligned}$$

where $\gamma = 0.01$ to reflect that rounding occurs to the nearest tick.

The proposed simulation scheme renders the choice of the starting price X_0 relevant, since the magnitude of the spread increases for less expensive stocks, due to the log-transformation. As noted by Li and Mykland (2007, 2014) rounding errors, when modeled explicitly, are intensified for less expensive stocks. In our simulation study we wish to capture this effect by using different starting prices, that is, we set $S_0 \in \{10, 20, 30\}$.

To illustrate the effectiveness of our methodology, the parameters for the simulation study include both typical values used in the literature and values relatively smaller[2]. Specifically, the diffusion coefficient is set $\sigma \in \{15\%, 30\%\}$ in annual terms and additive noise is set $\sigma_u \in \{.005\%, .01\%, .02\%\}$. Unlike the typical value $\sigma = 30\%$, the value $\sigma = 15\%$ - low signal - causes price sluggishness, intensifying the effect of rounding errors. This a common feature of observed prices (both transactions and quotations), especially in the middle of the trading day.

The values chosen for the noise component include values smaller than the typical values used in the literature. As we argued above, additive noise in our setup is considered as residual noise and, therefore, is expected to be smaller. Moreover, the value $\sigma_u = .005\%$ has small contribution, making rounding the primary source of error. This is of particular interest, since for stocks traded frequently the MS variance is smaller. For example, Aït-Sahalia and Yu (2009) applied the parametric framework in a large number of stocks and showed that MS variance is smaller for frequently traded stocks. Our empirical study verifies this stylized fact. Larger values, such as $\sigma_u = .02\%$, are also considered, however these are more relevant for moderate and low frequencies, since, in high frequencies, the value

[2]As mentioned above, for computational benefits, the drift of the process is not included in the simulation, even though it does not affect the simulation outcomes

	$\sigma_u = .005\%$, $\sigma = 15\%$		$\sigma_u = .01\%$, $\sigma = 15\%$		$\sigma_u = .02\%$, $\sigma = 15\%$	
Stock Price	σ_u	σ	σ_u	σ	σ_u	σ
$S_0 = 10$	95.67%	82.43%	88.49%	83.46%	91.25%	83.93%
$S_0 = 20$	96.13%	92.64%	92.47%	90.98%	91.16%	81.71%
$S_0 = 30$	95.25%	94.48%	92.30%	90.68%	90.73%	77.18%
	$\sigma_u = .005\%$, $\sigma = 30\%$		$\sigma_u = .01\%$,$\sigma = 30\%$		$\sigma_u = .02\%$, $\sigma = 30\%$	
Stock Price	σ_u	σ	σ_u	σ	σ_u	σ
$S_0 = 10$	98.43%	92.96%	96.32%	93.16%	92.66%	91.75%
$S_0 = 20$	98.63%	94.69%	94.19%	94.53%	92.26%	90.78%
$S_0 = 30$	98.46%	95.54%	93.02%	93.69%	92.45%	88.76%

Table 3.1: Empirical coverage of the parameters. The nominal coverage is 95%. We simulate 200 sample paths of the process, each having 10,000 observations. Each sample is split in blocks of 200 observations, generating 50 blocks per sample path and, therefore, 10,000 blocks overall. The table reports coverage based on the 10,000 generated distributions.

$\sigma_u = .02\%$ generates unreasonable sample paths due to our simulation scheme[3].

Another feature of HFD we incorporate in the simulation is the non-synchronicity of the arrival of the quotations. Following Aït-Sahalia and Yu (2009), arrival times follow a Poisson process with intensity $\lambda = 2$, independent of the process X. For different combinations of the parameters we simulate 200 sample paths of the process. We use samples of 10,000 observations, which are split in blocks of 200 observations. So, each sample path is split into 50 blocks of observations. The SMC algorithm was implemented using 40,000 particles.

Based on the generated distributions, we perform two separate studies. The first study examines the block distributions generated from all sample paths. Conditional independence of the processes allows us to aggregate the results of each block of every generated sample path to investigate block coverage. Subsequently, we investigate the performance of the daily block point estimator for all generated sample paths. That is, we calculate the a point estimate of the variance for each day by aggregating the block point estimates. The second study investigates the performance of the combined point estimator.

Table 3.1 reports the coverage for all combinations of prices and parameters, which in most cases is very close to the nominal (95%). Table 3.2 reports the performance of the daily block point estimator summarized by the RMSE of $\hat{\sigma}^2_{fid}$ and the quantiles and mean of $\hat{\sigma}^2_{fid} - \sigma^2_0$. Coverage is very close to the nominal in the cases where the signal is stronger than

[3]In our simulation scheme we keep the spread constant. In practice, the spread fluctuates, reducing the effect of residual additive noise.

Volatility		$\sigma = 15\%$					$\sigma = 30\%$				
	Noise	RMSE	Q1	Median	Q3	Mean	RMSE	Q1	Median	Q3	Mean
$S_0 = 10$	$\sigma_u = .005\%$	2.04	1.23	1.75	2.52	1.84	1.52	-0.87	0.15	1.02	0.17
	$\sigma_u = .01\%$	2.38	1.76	2.27	2.75	2.25	2.20	0.70	1.68	2.57	1.62
	$\sigma_u = .02\%$	2.72	2.06	2.64	3.10	2.63	3.95	2.39	3.66	4.70	3.58
$S_0 = 20$	$\sigma_u = .005\%$	0.57	0.18	0.44	0.67	0.44	1.02	-1.02	-0.45	0.30	-0.42
	$\sigma_u = .01\%$	1.08	0.67	1.05	1.28	0.99	2.27	1.37	1.96	2.65	2.05
	$\sigma_u = .02\%$	1.90	1.36	1.87	2.20	1.82	3.29	1.83	3.06	3.88	2.90
$S_0 = 30$	$\sigma_u = .005\%$	0.50	0.23	0.42	0.60	0.41	0.84	-0.64	-0.12	0.43	-0.12
	$\sigma_u = .01\%$	0.91	0.53	0.83	1.08	0.82	2.28	1.29	2.00	2.70	2.03
	$\sigma_u = .02\%$	1.95	1.48	1.87	2.24	1.87	5.40	2.52	3.26	4.13	5.46

Table 3.2: In this table we report the performance of the daily block point estimator. The sample size is 10,000 (= 50 blocks $\times$ 200 obs) observations. Intensity of the arrival times is $\lambda = 2$. Column 3 reports the RMSE of $\hat{\sigma}^2$. Columns 4-7 report the quantiles and the mean of $\hat{\sigma}^2 - \sigma_0^2$. All entries are multiples of 10^{-5}.

additive (residual) noise. In the presence of rounding errors and at the current sampling frequency, high additive noise makes signal discovery more difficult, especially when the stock price is low. Also, when additive noise is high, table 3.2 reveals that daily block point estimator has lower RMSE as the stock price increases, even though block coverage appears lower. This can be explained from the magnitude of the spread, since the block distributions are more dispersed when both the spread (lower stock price) and additive noise are high.

The aforementioned issue can be resolved with two approaches. Either one can reduce the sampling frequency or increase the number of observations per block. In the case where both signal and stock price level are low, prices are sluggish, which may introduce some bias in the generated distributions. We demonstrate the first approach by repeating the above simulation for moderate frequencies, focusing on the most challenging case where $S_0 = 10$ and $\sigma = 15\%$. Tables 3.3 and 3.4 report the results for all combinations of additive noise. It is evident that both coverage and RMSE improved substantially, as arrival times are less frequent. This outcome is no surprise, since the impact of additive noise and rounding errors is smaller. For different starting prices and parameters the results improve in a similar manner. For the second approach, we increase the sample size per block from 200 to 300 and 400 observations. The results in table 3.5 indicate an improvement in coverage and RMSE.

The second study investigates the performance of the combination scheme. We apply the

Ms Noise	$\sigma_u = .005\%$		$\sigma_u = .01\%$		$\sigma_u = .02\%$	
Intensity	σ_u	σ	σ_u	σ	σ_u	σ
$\lambda = 2$	95.67%	82.43%	88.49%	83.46%	91.25%	83.93%
$\lambda = 5$	96.53%	90.13%	91.83%	91.05%	92.05%	89.28%
$\lambda = 10$	96.35%	93.75%	94.45%	93.65%	91.50%	91.10%

Table 3.3: Empirical coverage of the parameters as the intensity of the arrival times becomes larger. The starting price is $S_0 = 10$, volatility is $\sigma = 15\%$ and the nominal coverage is 95%. We simulate 200 sample paths of the process. When intensity is $\lambda = 2$, we have 10,000 observations of the process, split in 50 blocks. When intensity is $\lambda = 5$, we have 4,000 observations of the process, split in 20 blocks. When intensity is $\lambda = 10$, we have 2,000 observations of the process, split in 10 blocks. The reported coverage is based on the 10,000 generated distributions when $\lambda = 2$, 4,000 generated distributions, when $\lambda = 5$ and 2,000 generated distributions, when $\lambda = 10$.

scheme progressively by combining a few blocks until we use all 50 blocks. Tables B.1-B.6 in appendix B report the coverage, the average length of the confidence intervals, the root mean square error of the point estimator of σ^2 and the average effective sample size[4] for all combinations of the parameters and starting prices, as the sample size increases. It is evident that in most cases the procedure performs remarkably well. In an analogous manner as above, coverage is lower in the cases where the starting price is lower and additive noise is high. However, when the block size increases there is substantial improvement in coverage. Table B.7 illustrates the improvement in coverage when the block size increases to 300 and 400. The next section analyzes the 50-block combined estimator together with the MLE and shows that lowering the sampling frequency may lead to substantial improvement in coverage.

3.6.1 Comparison with the MLE

We investigate the performance of the MLE in the presence of rounding errors, by employing the two stage contamination simulation scheme of Li and Mykland (2007). In order to have a direct comparison with our framework, we simulate the same sample paths for X_t. The difference with the simulation scheme above, is that now the process is rounded to the nearest tick. We focus on the case where the sample consists of 10,000 observations and

[4]The effective sample size indicates the quality of the generated distributions. It gives an estimate of the number of particles (parameters (σ_u, σ)) that have high (importance) weight in the sample.

Ms Noise		$\sigma_u = .005\%$					$\sigma_u = .01\%$					$\sigma_u = .02\%$			
Intensity	RMSE	Q1	Median	Q3	Mean	RMSE	Q1	Median	Q3	Mean	RMSE	Q1	Median	Q3	Mean
$\lambda = 2$	2.04	1.23	1.75	2.52	1.84	2.38	1.76	2.27	2.75	2.25	2.72	2.06	2.64	3.10	2.63
$\lambda = 5$	0.92	0.03	0.55	0.96	0.49	1.18	0.43	0.89	1.38	0.92	1.59	0.93	1.41	1.90	1.41
$\lambda = 10$	0.80	-0.12	0.31	0.83	0.30	0.87	-0.007	0.44	0.93	0.46	1.17	0.15	0.71	1.32	0.70

Table 3.4: In this table we report the performance of the fiducial point estimator of the block distributions as the intensity of the arrival times becomes larger. The starting price is $S_0 = 10$ and volatility is $\sigma = 15\%$. The sample size is 10,000 (= 50 blocks $\times$ 200 obs) observations when $\lambda = 2$, 4,000 (= 20 blocks $\times$ 200 obs) observations when $\lambda = 5$ and 2,000 (= 10 blocks $\times$ 200 obs) observations when $\lambda = 10$. Column 3 reports the RMSE of $\hat{\sigma}^2$. Columns 4-7 report the quantiles and the mean of $\hat{\sigma}^2 - \sigma_0^2$. All entries are multiples of 10^{-5}.

Block Size	Covg. σ_u	Covg. σ	RMSE	Q1	Median	Q3	Mean
200	98.43%	92.96%	1.52	-0.87	0.15	1.02	0.17
300	98.04%	93.53%	1.51	-0.68	0.13	1.24	0.27
400	97.94%	93.68%	1.45	-0.56	0.23	1.26	0.30

Table 3.5: In this table we report the coverage of the individual blocks and performance the daily block point estimator when the block size increases from 200 to 300 and 400. The starting price is $S_0 = 10$ and the true parameters are $\sigma = 30\%$, $\sigma_u = .005\%$. The sample size is 10,200 (= 34 blocks $\times$ 300 obs) observations when the block size is 300 and 10,000 (= 25 blocks $\times$ 400 obs) when the block size is 400. Intensity of the arrival times is $\lambda = 2$. Columns 2-3 report the coverage of the two parameters. Column 4 reports RMSE of $\hat{\sigma}^2$. Columns 5-8 report the quantiles and the mean of $\hat{\sigma}^2 - \sigma_0^2$. All entries are multiples of 10^{-5}.

compare the ML estimator with the point estimator of the combined distributions. Tables 3.6 to 3.9 report coverage, the average length of the confidence intervals of both point estimators for the three starting prices. The estimation errors are summarized through the root mean square error (RMSE) and their 1st quartile, median, 3rd quartile and the mean of $\hat{\sigma}^2 - \sigma_0^2$, where $\hat{\sigma}^2$ is either the MLE or the point estimator of the combined distribution. Also, the average efficient sample size is included to indicate the quality of the generated distributions. It is evident that the combination of rounding and small additive noise deteriorates the performance of the MLE substantially. At this relatively high frequency, the MLE overestimates the true value, and only when fundamental volatility or the initial stock price is large coverage and RMSEs improve. The effect of small σ_u on the estimation of σ is similar with the finding of Li and Mykland (2007) where the TSRV estimator requires a sufficiently large variance of the additive component to perform well.

3.6.2 Stochastic Volatility

To illustrate the robustness of our methodology, we conduct a small simulation study where the true data generating mechanism exhibits stochastic volatility. Specifically, we adapt a standard stochastic volatility model where the efficient price process follows the diffusion:

$$dX_t = \sigma_t dW_{1t}$$
$$d\sigma_t^2 = \kappa\left(\upsilon - \sigma_t^2\right)dt + s\sigma_t dW_{2t}$$

where W_{1t} and W_{2t} are independent Brownian Motions. We select $\upsilon = 0.04$ (which amounts to 20% volatility per year), $\kappa = 5$ and for the volatility of the volatility parameter $s = 1$.

$\sigma_0 = 15\%$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	73.00%	1.99	0.96	-0.30	0.36	0.86	0.35	1.26
	MLE	0.00%	2.80	13.8	12.7	13.7	14.7	13.7	NA
$\sigma_u = .01\%$	Fid	54.50%	1.96	1.13	0.28	0.81	1.32	0.79	.83
	MLE	0.00%	2.47	7.17	6.31	7.10	7.90	7.08	NA
$\sigma_u = .02\%$	Fid	65.50%	1.94	1.03	0.16	0.64	1.25	0.68	1.98
	MLE	27.00%	2.10	1.59	0.99	1.41	1.90	1.46	NA
$\sigma_0 = 30\%$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	75.00%	4.95	2.13	-2.45	-1.43	-0.47	-1.44	5.43
	MLE	0.00%	5.98	20.0	18.1	20.1	21.7	19.8	NA
$\sigma_u = .01\%$	Fid	90.00%	5.29	1.51	-1.09	-0.06	1.01	-0.07	4.36
	MLE	0.00%	5.76	11.7	10.1	11.4	13.1	11.5	NA
$\sigma_u = .02\%$	Fid	88.50%	5.61	1.82	-1.40	-0.27	0.88	-0.25	3.62
	MLE	53.50%	5.62	3.14	1.46	2.67	3.71	2.64	NA

Table 3.6: In this table we report the performance of the fiducial point estimator of the combined distributions against the MLE. Starting price is $S_0 = 10$, and the sample size is 10,000 (= 50 blocks $\times$ 200 obs) observations. Intensity of the arrival times is $\lambda = 2$. Columns 2-5 report the coverage, the average length of the confidence intervals and the RMSE of $\hat{\sigma}^2$. Columns 6-9 report the quantiles and the mean of $\hat{\sigma}^2 - \sigma_0^2$. Column 10 reports the average efficient sample size of the combined distributions. Coverage of the MLE relies on the asymptotic variance. Entries in columns 4-9 are multiples of 10^{-5}. The ESS is a multiple of 10^3.

$\sigma_0 = 15\%$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	90.00%	1.32	0.40	-0.28	-0.01	0.26	-0.01	4.01
	MLE	0.00%	1.44	3.00	2.64	2.96	3.27	2.95	NA
$\sigma_u = .01\%$	Fid	88.00%	1.40	0.51	-0.39	0.01	0.37	-0.03	3.09
	MLE	62.00%	1.39	0.70	0.26	0.63	0.84	0.57	NA
$\sigma_u = .02\%$	Fid	64.50%	1.39	0.93	-0.45	0.09	0.56	0.00	1.22
	MLE	92.50%	1.60	0.45	-0.28	-0.01	0.30	0.01	NA
$\sigma_0 = 30\%$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	83.00%	3.70	1.28	-1.50	-0.85	-0.17	-0.86	7.05
	MLE	44.50%	3.61	2.45	1.08	1.98	2.98	2.07	NA
$\sigma_u = .01\%$	Fid	91.00%	4.30	1.25	-0.37	0.41	1.26	0.48	5.34
	MLE	87.00%	3.84	1.27	-0.19	0.58	1.42	0.61	NA
$\sigma_u = .02\%$	Fid	83.50%	4.69	1.59	-1.23	-0.27	0.80	-0.32	3.31
	MLE	90.00%	4.51	1.33	-1.01	-0.23	0.74	-0.17	NA

Table 3.7: In this table we report the performance of the fiducial point estimator of the combined distributions against the MLE. Starting price is $S_0 = 20$, and the sample size is 10,000 (= 50 blocks $\times$ 200 obs) observations. Intensity of the arrival times is $\lambda = 2$. Columns 2-5 report the coverage, the average length of the confidence intervals and the RMSE of $\hat{\sigma}^2$. Columns 6-9 report the quantiles and the mean of $\hat{\sigma}^2 - \sigma_0^2$. Column 10 reports the average efficient sample size of the combined distributions. Coverage of the MLE relies on the asymptotic variance.Entries in columns 4-9 are multiples of 10^{-5}. The ESS is a multiple of 10^3.

$\sigma_0 = 15\%$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	93.50%	1.15	0.32	-0.11	0.09	0.31	0.08	4.74
	MLE	42.50%	1.09	0.71	0.37	0.61	0.88	0.62	NA
$\sigma_u = .01\%$	Fid	86.50%	1.26	0.45	-0.35	-0.07	0.20	-0.07	3.70
	MLE	91.00%	1.20	0.34	-0.19	0.03	0.28	0.04	NA
$\sigma_u = .02\%$	Fid	54.00%	1.23	0.92	-0.58	0.05	0.58	-0.04	4.70
	MLE	91.50%	1.51	0.43	-0.33	-0.01	0.24	-0.04	NA
$\sigma_0 = 30\%$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	93.00%	3.35	0.93	-0.87	-0.28	0.29	-0.28	7.35
	MLE	89.00%	3.00	0.98	-0.57	0.10	0.81	0.11	NA
$\sigma_u = .01\%$	Fid	94.00%	4.04	1.12	-0.74	0.03	0.72	0.03	5.18
	MLE	88.00%	3.84	1.30	-0.33	0.52	1.28	0.52	NA
$\sigma_u = .02\%$	Fid	80.50%	4.44	1.70	-0.84	0.21	1.31	0.22	2.66
	MLE	93.00%	4.53	1.33	-0.75	0.06	1.11	0.17	NA

Table 3.8: In this table we report the performance of the fiducial point estimator of the combined distributions against the MLE. Starting price is $S_0 = 30$, and the sample size is 10,000 (= 50 blocks $\times$ 200 obs) observations. Intensity of the arrival times is $\lambda = 2$. Columns 2-5 report the coverage, the average length of the confidence intervals and the RMSE of $\hat{\sigma}^2$. Columns 6-9 report the quantiles and the mean of $\hat{\sigma}^2 - \sigma_0^2$. Column 10 reports the average efficient sample size of the combined distributions. Coverage of the MLE relies on the asymptotic variance. Entries in columns 4-9 are multiples of 10^{-5}. The ESS is a multiple of 10^3.

$\lambda = 5$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	85.50%	2.38	0.84	-0.74	-0.05	0.39	-0.13	5.60
	MLE	0.00%	3.01	8.94	7.70	8.76	9.83	8.80	NA
$\sigma_u = .01\%$	Fid	84.00%	2.41	0.82	-0.25	0.24	0.74	0.21	3.84
	MLE	0.00%	2.77	4.73	3.70	4.55	5.31	4.59	NA
$\sigma_u = .02\%$	Fid	87.00%	2.51	0.84	-0.25	0.28	0.86	0.30	4.70
	MLE	69.50%	2.53	1.17	0.35	0.88	1.43	0.91	NA
$\lambda = 10$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	94.50%	2.78	0.76	-0.46	0.02	0.47	-0.01	11.04
	MLE	0.00%	3.23	6.11	4.55	5.77	7.05	5.87	NA
$\sigma_u = .01\%$	Fid	92.50%	2.86	0.79	-0.43	0.12	0.57	0.08	7.78
	MLE	9.00%	3.04	3.22	2.17	2.92	3.78	2.98	NA
$\sigma_u = .02\%$	Fid	87.00%	3.02	0.96	-0.61	-0.02	0.67	-0.01	7.34
	MLE	83.00%	2.95	1.07	-0.03	0.56	1.12	0.58	NA

Table 3.9: In this table we report the performance of the fiducial point estimator of the combined distributions against the MLE. Starting price is $S_0 = 10$ and $\sigma_0 = 15\%$. Intensity of the arrival times is $\lambda = 5$ and $\lambda = 10$. Sample size is 4,000 (= 20 blocks $\times$ 200 obs) and 2,000 (= 10 blocks $\times$ 200 obs) observations respectively. Columns 2-5 report the coverage, the average length of the confidence intervals and the RMSE of $\hat{\sigma}^2$. Columns 6-9 report the quantiles and the mean of $\hat{\sigma}^2 - \sigma_0^2$. Column 10 reports the average efficient sample size of the combined distributions. Coverage of the MLE relies on the asymptotic variance. Entries in columns 4-9 are multiples of 10^{-5}. The ESS is a multiple of 10^3.

$S_0 = 10$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	88.00%	2.83	0.97	-0.80	-0.15	0.52	-0.15	1.83
	MLE	0.00%	3.77	17.20	15.50	17.10	18.50	17.10	NA
$\sigma_u = .01\%$	Fid	80.00%	2.92	1.17	-0.31	0.41	1.17	0.43	1.84
	MLE	0.00%	3.51	9.43	8.02	9.21	10.30	9.27	NA
$\sigma_u = .02\%$	Fid	79.50%	2.96	1.37	-0.33	0.36	1.10	0.36	2.68
	MLE	36.00%	3.14	2.18	1.19	1.87	2.64	1.92	NA
$S_0 = 30$	$\hat{\sigma}^2$	Coverage	ALCI	RMSE	Q1	Median	Q3	Mean	ESS
$\sigma_u = .005\%$	Fid	89.50%	1.77	0.50	-0.42	-0.07	0.25	-0.08	4.13
	MLE	70.00%	1.65	0.77	0.13	0.50	0.91	0.49	NA
$\sigma_u = .01\%$	Fid	81.50%	2.04	0.72	-0.72	-0.31	0.09	-0.31	3.42
	MLE	86.50%	1.85	0.65	-0.39	0.01	0.47	0.02	NA
$\sigma_u = .02\%$	Fid	59.50%	2.05	1.20	-1.12	-0.19	0.50	-0.27	1.01
	MLE	86.50%	2.35	0.79	-0.47	-0.07	0.43	-0.08	NA

Table 3.10: In this table we report the performance of the fiducial point estimator of the combined distributions against the MLE. Starting price is $S_0 = 10$, and the sample size is 10,000 (= 50 blocks $\times$ 200 obs) observations. Intensity of the arrival times is $\lambda = 2$. Columns 2-5 report the coverage, the average length of the confidence intervals and the RMSE of $\hat{\sigma}^2$. Columns 6-9 report the quantiles and the mean of $\hat{\sigma}^2 - \frac{1}{n}\sum_{i=1}^{n}\sigma_{t_i}^2$. Column 10 reports the average efficient sample size of the combined distributions. Coverage of the MLE relies on the asymptotic variance. Entries in columns 4-9 are multiples of 10^{-5}. The ESS is a multiple of 10^3.

These parameters belong in the range of values used in Aït-Sahalia and Kimmel (2007) and Aït-Sahalia and Yu (2009). It is common in this type of simulations to initialize the volatility process by drawing σ_0^2 from its stationary distribution. Due to the small size of our simulation, we set $\sigma_0^2 = \upsilon$. Parameter $\upsilon = 0.04$ together with parameter $s = 1$ can lead to low values of the spot volatility and, therefore, intensify the effect of rounding. Similarly to the case of constant volatility, for different combinations of starting prices and additive noise we simulate 200 sample paths of the process. Arrival times have intensity $\lambda = 2$, therefore, we use samples of 10,000 observations, which roughly constitute a trading day. Each sample path is split in 50 blocks of 200 observations and the combination scheme is applied. Integrated variance is approximated by $\int_0^1 \sigma_t^2 dt \simeq \frac{1}{n}\sum_{i=1}^{n}\sigma_{t_i}^2$, where $\Delta t_i = \frac{1}{23{,}400}$. Table 3.10 reports the performance of fiducial point estimator and the MLE in a similar manner as above. Coverage of the approximate integrated variance is very high and the fiducial estimator outperforms the MLE in most cases.

Figure 3.3: **Alcoa Inc. (AA) on May4, 2007.**High-frequency quotes for Alcoa Inc. (AA) on May4, 2007, split in 18 blocks of 298 observations.

3.7 Empirical Study

Our empirical analysis focuses primarily on the illustration of the methodology, together with data handling issues and concerns. We analyze data for Alcoa Inc. (AA) collected on May 4. The particular day was studied by Barndorff-Nielsen et al. (2009) and was reported as an exemplary day in terms of the stability of the volatility signature plots. The data were collected from the TAQ database and were cleaned according to the filters found in Barndorff-Nielsen et al. (2009). Moreover, we apply our methodology on bid and ask quotes that have a corresponding transaction.

3.7.1 Analysis of Alcoa Inc. on May 4, 2007

After applying the filters, we arrive at 14,630 quotations and 5,203 transactions. Therefore, the sample under consideration consists of 5,203 quotations that have a corresponding transaction. This amounts to a new observation roughly every 4.5 seconds. Next, we split the sample in 18 blocks of observations, each one containing 298 observations. Figure 3.3 displays the high frequency quotes with the block division superimposed on it.

Figure 3.4 displays all the block density estimates together with the combined density estimates for parameter σ. The left panel displays the first six blocks and right panel the last 12 blocks. Evidently, the locations of the block distributions indicate a U-shaped pattern. Volatility is quite high in the beginning of the day, much lower in the middle of the day and higher towards the end. This pattern is the reason we do not apply the

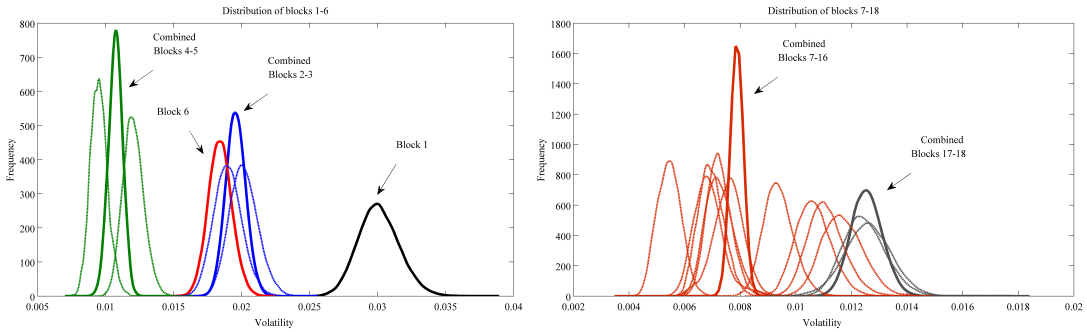


Figure 3.4: Fiducial densities for Alcoa Inc. (AA) on May4, 2007. Combined and block fiducial distributions for Alcoa Inc. (AA) on May4, 2007. Left panel displays the first six blocks. Blocks 2-3 and 4-5 can be combined. Blocks 1 and 6 cannot be combined with adjacent blocks. Left panel displays the last 12 blocks. Blocks 7-16 and 17-18 can be combined.

Blocks/Time	Parms.	Point Est.	Median	Mx. Weight	Conf. Interval	Variance	Skew.	Kurt.
Block 1	σ_u	1.91*10^{-5}	1.71*10^{-5}	5.37*10^{-5}	[0, 5.60*10^{-5}]	2.69*10^{-10}	0.7337	3.135
9:30:12-9:42:16	σ	0.0301	0.0300	0.0281	[0.0273, 0.0332]	2.26*10^{-6}	0.2390	3.102
Blocks 2-3	σ_u	1.14*10^{-5}	1.09*10^{-5}	1.23*10^{-5}	[0, 3.02*10^{-5}]	7.92*10^{-11}	0.4735	2.622
9:42:16-10:13:01	σ	0.0196	0.0196	0.0203	[0.0182, 0.0210]	5.32*10^{-7}	0.0938	3.034
Blocks 4-5	σ_u	1.26*10^{-5}	1.15*10^{-5}	3.07*10^{-5}	[0, 3.13*10^{-5}]	8.01*10^{-11}	0.5282	2.739
10:13:01-11:03:06	σ	0.0107	0.0108	0.0116	[0.0098, 0.0117]	2.44*10^{-7}	-0.0921	2.994
Block 6	σ_u	2.68*10^{-5}	2.72*10^{-5}	4.11*10^{-5}	[8.05*10^{-6}, 4.34*10^{-5}]	7.79*10^{-11}	-0.2359	3.236
11:03:06-11:22:44	σ	0.0184	0.0184	0.0180	[0.0168, 0.0202]	7.52*10^{-7}	0.1161	3.030
Blocks 7-16	σ_u	1.66*10^{-5}	1.65*10^{-5}	1.63*10^{-5}	[1.04*10^{-5}, 2.27*10^{-5}]	1.02*10^{-11}	0.0061	2.919
11:22:44-15:39:34	σ	0.0079	0.0079	0.0080	[0.0075, 0.0082]	3.00*10^{-8}	0.0048	2.903
Blocks 17-18	σ_u	1.12*10^{-5}	1.09*10^{-5}	2.06*10^{-5}	[0, 2.70*10^{-5}]	5.71*10^{-11}	0.4100	2.877
15:39:34-16:00:00	σ	0.0125	0.0125	0.0124	[0.0115, 0.0137]	3.16*10^{-7}	0.1323	3.008

Table 3.11: **Summary statistics for the combined distributions for Alcoa Inc. (AA) on May 4, 2007**.

combination scheme on all block distributions. Volatility for this particular day exhibits sudden changes, indicating the presence of volatility jumps. A combined distribution using all block distributions leads to a nearly degenerate distribution, as indicated by the small effective sample size. Instead, we combine distributions from adjacent blocks, provided that the efficient sample size of the generated distribution is high. Typically, distributions located within a band of 10% annual volatility can be combined. Figure 3.4 displays the combined distributions. Information for the combined distributions can be found in table 3.11.

Figure 3.5 displays the block, combined and ML point estimates for both σ^2 and σ_u^2.

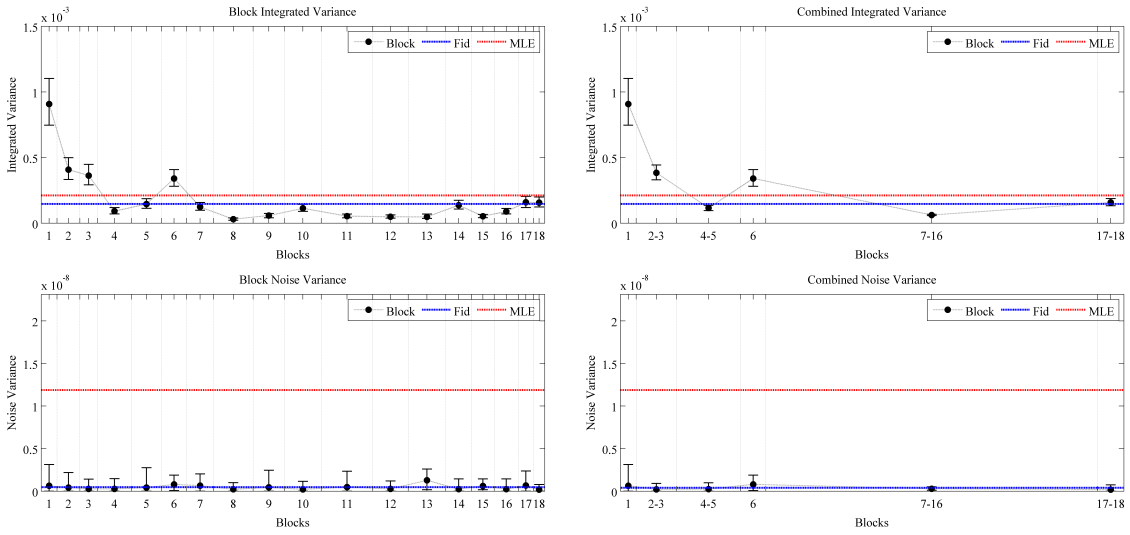


Figure 3.5: **Volatility and noise estimates for Alcoa Inc. (AA) on May 4, 2007.** The left panels display the fiducial block point estimates for σ^2 (upper) and σ_u^2 (lower), together with the ML estimates. The fiducial point estimates of σ^2 and σ_u^2 for the entire day are $1.46 * 10^{-4}$ and $4.56 * 10^{-10}$ respectively. The ML estimates are $2.11 * 10^{-4}$ and $1.19 * 10^{-8}$ respectively. The right panels display the same information for the point estimates of the combined distributions. The fiducial point estimates are $1.40 * 10^{-4}$ and $3.92 * 10^{-10}$ respectively.

The left upper panel displays the block estimates of σ^2 with the confidence intervals superimposed. In the same panel, we display the daily variance estimate, as a result of the sum of the block estimates, together with ML estimate[5]. In a similar manner, the left lower panel displays the estimates of σ_u^2. The right panels display the same information, but for the combined distribution point estimates. Although the daily estimates for σ^2 are relatively close, the estimates for σ_u^2 are quite different. In fact, the fiducial noise estimate is virtually zero, indicating that most of the noise for this particular day is attributed to rounding and spread related frictions.

We conclude this section by generating volatility signature plots. We form the subsamples by keeping every i^{th} observation in the sample, where $i = 1, 2, ..., 5$. In addition to the parametric estimators, we also include the Parzen realized kernel with optimal bandwidth. The left panel in figure 3.6 displays the signature plots for integrated variance. The signature plot for the fiducial estimator is very smooth and much closer to the RK estimates.

[5]The ML estimate was based on the corresponding transactions.

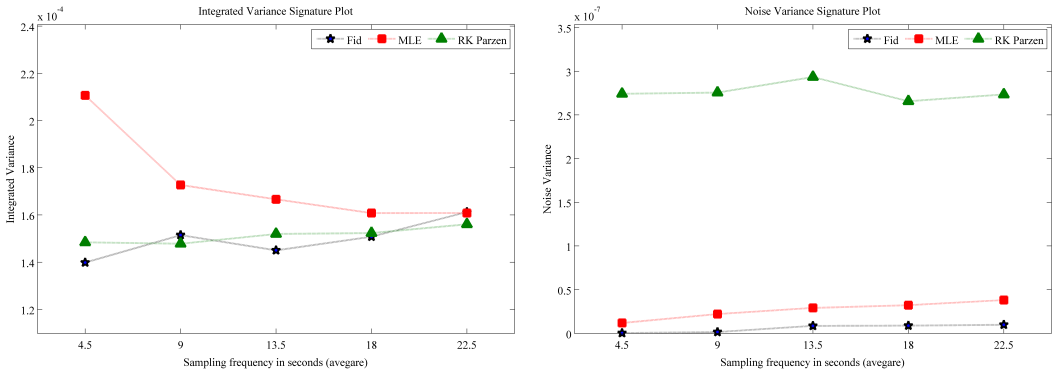


Figure 3.6: **Volatility and noise signature plots for Alcoa Inc. (AA) on May 4, 2007.** The left panel displays point estimates for σ^2 at different sampling frequencies for the fiducial (combined), ML and RK-Parzen. At the highest frequency the estimates are $1.40 * 10^{-4}$ (Fid), $2.11 * 10^{-4}$ (ML) and $1.48 * 10^{-4}$ (RK). The right panel displays the same information for σ_u^2 . At the highest frequency the estimates are $3.92 * 10^{-10}$ (Fid), $1.19 * 10^{-8}$ (ML) and $2.74 * 10^{-7}$ (RK).

The right panel displays signature plots for additive noise. Noise is virtually zero for the fiducial estimator as the sample size decreases.

3.7.2 Time varying spread

In this section we discuss the implications of time varying spread in our methodology. In the previous chapter, we explained how the time varying spread may have adverse effects on the estimation. As before, we modify the quotes where the spread is higher than some pre-specified threshold. That is, we modify the data by restricting the spread to be no larger than three cents. Figure 3.7 displays the block, combined and ML point estimates for both σ^2 and σ_u^2, in a similar manner as in figure 3.5. Table 3.12 reports the results for the first block before the reduction of the spread and after. The small difference in the results can be explained by a closer inspection of the revised spread. Before the reduction of the spread, large movements in transactions (potential jumps) were irrelevant to the estimation. However, if we use transactions as a guide to revise the spread, large changes in transactions will be carried over.

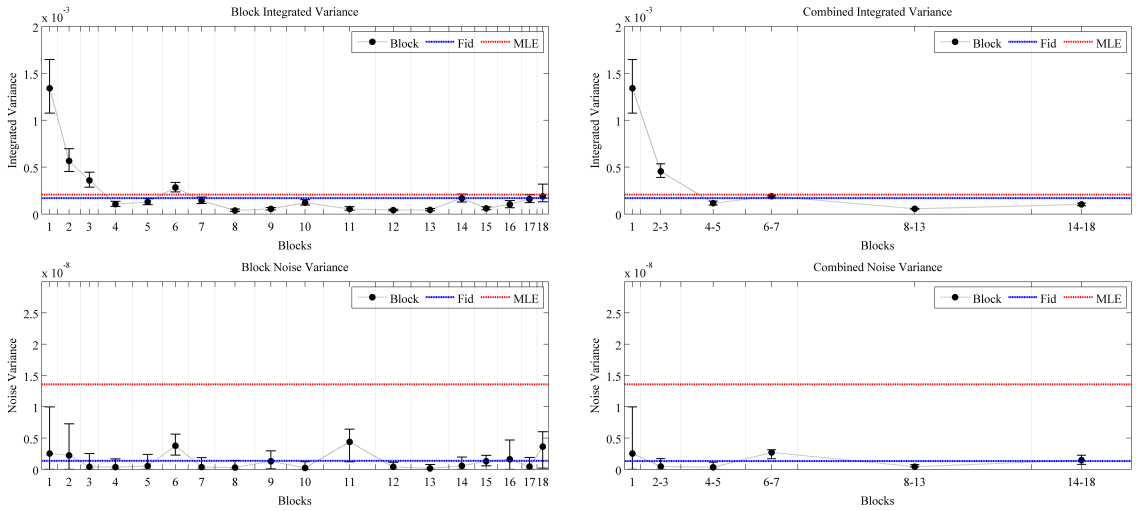


Figure 3.7:

Figure 8: Volatility and noise estimates for Alcoa Inc. (AA) on May 4, 2007 with revised quotes. The left panels display the fiducial block point estimates for σ^2 (upper) and σ_u^2 (lower), together with the ML estimates. The fiducial point estimates of σ^2 and σ_u^2 for the entire day are $1.69 * 10^{-4}$ and $1.39 * 10^{-9}$ respectively. The ML estimates are $2.11 * 10^{-4}$ and $1.19 * 10^{-8}$ respectively. The right panels display the same information for the point estimates of the combined distributions. The fiducial point estimates are $1.58 * 10^{-4}$ and $1.31 * 10^{-9}$ respectively.

Block 1	Parms	Point Est.	Median	Mx. Weight	Conf. Interval	Variance	Skew.	Kurt.
Original	σ_u	1.91*10^{-5}	1.71*10^{-5}	5.37*10^{-5}	[0, 5.60*10^{-5}]	2.69*10^{-10}	0.7337	3.135
	σ	0.0301	0.0300	0.0281	[0.0273, 0.0332]	2.26*10^{-6}	0.2390	3.102
Revised	σ_u	4.08*10^{-5}	3.85*10^{-5}	1.04E-04	[0, 9.98*10^{-5}]	8.48*10^{-10}	0.3726	2.320
	σ	0.0366	0.0365	0.0305	[0.0328, 0.0406]	3.87*10^{-6}	0.1180	3.057

Table 3.12: Summary statistics for the first block distributions for Alcoa Inc. (AA) on May 4, 2007, before and after the quote revision.

3.8 Conclusion

In this article we propose a novel framework to study high frequency financial data for volatility estimation. Assuming that the efficient price-process follows a time homogenous diffusion process, we proposed a framework that allows us to specify a generalized fiducial distribution on the parameter space. The attractive feature of the framework is that it enables us to view the bid-ask spread as a natural interval around the unobservable efficient price and use quotes for volatility estimation, instead of transactions or mid-points. Moreover, our framework is flexible enough that allows us to maintain the additive noise component, introduced in the current literature to explain the deviations of the efficient price from the observed one.

The new methodology was applied assuming that volatility remains constant for a short period of time. The samples under consideration were split in blocks of observations and inference was performed on each block. Additionally, the proposed combination scheme introduced a tool to combine the block distributions into one that summarizes the information from all the blocks under consideration. Our simulation study, which was designed to incorporate features of observed data, verified that volatility estimation is feasible at very high frequencies in the presence of rounding errors and outperforms the standard parametric approach. Our empirical study reports two findings. First, residual noise is virtually zero indicating that in very high frequencies microstructure noise is attributed to rounding effects. Second, daily volatility exhibits a rough U-shaped pattern indicating the presence of volatility jumps.

CHAPTER 4

Option Pricing with Alternative Realized Volatility Estimators

4.1 Summary

In this chapter we utilize the generalized affine realized volatility (GARV) option pricing model of Christoffersen et al. (2014), which extends the Heston and Nandi (2000) (HN) option pricing framework by incorporating realized volatility, to compare the performance of alternative realized volatility estimators. Specifically, we use 7 alternative estimators of realized volatility. Our benchmark estimator is the standard realized volatility estimator based on 5-minute returns (RV-5min). The other estimators are the Two-Scale (TSRV) and Multi-Scale (MSRV) realized volatility estimators developed by Zhang et al. (2005) and Zhang (2006) respectively, the parametric estimator (QMLE introduced by Aït-Sahalia et al. (2005) and Xiu (2010), two Realized Kernel (RK) volatility estimators developed by Barndorff-Nielsen et al. (2008), particularly the estimators using the Tukey-Hanning$_2$ kernel and the Parzen kernel, and finally, the Pre-Averaging (PAV) introduced by Podolskij and Vetter (2009). The GARV model has a closed-form option pricing formula based on Fourier transforms. Using the four realized volatility estimators, we estimate the GARV model using call option data on the S&P 500 index. The estimation method we employ is minimization of the implied volatility root mean square error (IVRMSE) criterion. We find that the QMLE, the RK estimators and the PAV outperform the standard RV-5min by reducing the option pricing errors more than the RV-5min, with the RK-Parzen performing the best.

4.2 Introduction

In many finance applications it is of paramount importance to measure volatility. In particular, when the ultimate goal is the pricing of derivatives securities, like options, volatility is one of the important drivers of the price. While volatility initially was treated as being

constant through time, for example, the classical papers by Black and Scholes (1973) and Merton (1976), it is by now well understood that the volatility of most financial return series varies through time. An extensive empirical literature has documented the empirical biases of the Black-Scholes option valuation model, which arrive in the form of differences between observed market prices and model predictions. Specifically, observed market prices for out-of-the-money put prices and in-the-money call prices are higher than Black-Scholes prices. This stylized fact is known as the volatility "smirk" or the volatility "smile". Implied volatilities for at-the-money options also contain a term structure effect that cannot be explained by the Black-Scholes model.

The most popular approach to modeling the smirk is the use of GARCH and stochastic volatility models, which extend the classical constant volatility framework and allow for time varying volatility. Additionally, these models allow for negative correlation between the level of the stock return and its volatility. This negative correlation captures the stylized fact that decreases in the stock price are associated with larger increases in volatility than similar stock price increases, for example see Black (1976) and Christie (1982). This stylized fact, known as the leverage effect, is important for equity index option valuation, because it increases the probability of a large loss and consequently the value of out-of-the-money put options. The leverage effect induces negative skewness in stock returns, which in turn yields a volatility smirk.

The GARCH framework, initially proposed by Engle (1982) and extended by Bollerslev (1986), has been applied extensively in the empirical finance literature. In this framework volatility is treated as a time varying process depending on lagged values of volatility itself and on lagged squared innovations to the returns. Hence in this setting volatility can be estimated entirely from the return data by using the return innovations as a proxy for volatility. Heston and Nandi (2000) derived the first closed form option pricing formula in a discrete time setup when volatility follows a GARCH process. Their model, although is easy to implement, suffers from the forecasting disability of the GARCH models themselves. GARCH models, although they have been very successful in describing conditional volatility, they do not perform as well in forecasting it. This apparent shortcoming in forecasting volatility, as documented by several studies, is mainly because the history of the returns is

not enough to explain the conditional volatility.

In the continuous time SV framework, volatility is in fact treated as being "truly" stochastic. Models of this type have been extensively used in the theoretical derivative pricing literature as it is often possible to derive option pricing models more elegantly in a continuous time world. Classical examples include among others the work of Heston and Nandi (2000), Hull and White (1987), Johnson and Shanno (1987), Scott (1987) and Wiggins (1987). However, when it comes to empirically applying these models, a consistent estimation of the volatility process is needed. Given that stochastic volatility is an unobserved state variable, complicated filtering techniques are required which complicates the application for empirical option pricing.

These single-factor GARCH and stochastic volatility models, as several recent studies documented, have the structure to capture some of the empirical biases of the early models, but not all. By now, it is well known that richer volatility dynamics are needed to capture the empirical discrepancies. For instance Bates (2000), Duffie et al. (2000), Christoffersen et al. (2008), Christoffersen et al. (2009a) and Christoffersen et al. (2009c) point out that more than one volatility component is needed.

A recent trend in the financial econometrics literature is the use of high frequency intraday data for constructing non-parametric measures to estimate volatility. As demonstrated in the literature, these types of measures can be more accurate estimators of ex post volatility than the traditional sample variances based on daily or coarser frequency data. The construction of these volatility measures relies on the properties of the price processes which are assumed to evolve as special semimartingales, that is, they have a unique decomposition into a local martingale part and a predictable process of finite variation. Given that the local martingale part captures the variation of the process in the form of a stochastic integral of the instantaneous volatility (called integrated volatility), based on the theoretical properties of the semimartingales, we are in the position to construct efficient estimators of the stochastic quantity using high frequency intraday data. The estimators of the integrated volatility are called realized volatility estimators/measures.

In the option pricing literature integrated volatility is a very important quantity. For example, the Hull-White stochastic volatility model demonstrates that the call option price

is the conditional expectation (under the risk neutral measure) of the Black-Scholes formula evaluated at the average future integrated volatility. It is immediate then that integrated volatility is a key component to pricing options. Therefore, incorporating realized volatility in an option pricing framework might prove to be valuable in our effort to price options efficiently.

There are very few recent papers that incorporate realized volatility in option pricing. Feunou and Meddahi (2009) extended the class of affine models with non-Markovian dynamics and proposed the pricing of options with realized volatility as one possible application of this extension. In Stentoft (2008) an Inverse Gaussian model of a 30–minute realized volatility is used to price options on some individual stocks. However, their work does not provide a formal change of measure for the RV process, but it only considers the case when the risk neutral and physical dynamics of RV are the same (i.e. when the volatility risk is not priced). Corsi et al. (2013) proposed the Heterogeneous Auto-Regressive Gamma (HARG) process as a discrete-time stochastic volatility option pricing model that exploits the historical information contained in the high frequency data. One shortcoming of these proposed models is the lack of a closed form pricing formula. Both Stentoft (2008) and Corsi et al. (2013) price options using Monte-Carlo simulations.

Christoffersen et al. (2014) attempt to combine the need for richer volatility dynamics with the forecasting improvements of realized volatility and investigate how all these can be translated into economic value added. They do that by using incorporating realized volatility into the Heston and Nandi (2000) framework. Specifically, they develop a new class of affine discrete-time models that allow for closed-form option valuation formulas using conditional moment generating functions. What is different from the HN model is that the new models contain not only a GARCH-type component of past squared returns, but also a realized volatility GARCH-type component. For their application they use only realized volatility constructed by 5 and 60 minutes returns only.

In what follows we will utilize the GARV model they developed and estimate it using alternative volatility estimators. Specifically, we use 7 alternative estimators of realized volatility. Our benchmark estimator is the standard realized volatility estimator based on 5-minute returns (RV-5min). The other estimators are the Two-Scale (TSRV) and Multi-Scale

(MSRV) realized volatility estimators developed by Zhang et al. (2005) and Zhang (2006) respectively, the parametric estimator (QMLE introduced by Aït-Sahalia et al. (2005) and Xiu (2010), two Realized Kernel (RK) volatility estimators developed by Barndorff-Nielsen et al. (2008), particularly the estimators using the Tukey-Hanning$_2$ kernel and the Parzen kernel, and finally, the Pre-Averaging (PAV) introduced by Podolskij and Vetter (2009). The GARV model has a closed-form option pricing formula based on Fourier transforms. Using the four realized volatility estimators, we estimate the GARV model using call option data on the S&P 500 index. The estimation method we employ is minimization of the implied volatility root mean square error (IVRMSE) criterion. We find that the QMLE, the RK estimators and the PAV outperform the standard RV-5min by reducing the option pricing errors more than the RV-5min, with the RK-Parzen performing the best.

The paper proceeds as follows. In part 3, we introduce the GRV model. In part 4, we present the option pricing framework. In part 5 we present the estimation results, together with the data and the estimation methodology.

4.3 The GARV Model

The log-returns process in the GARV model is specified as follows

$$r_{t+1} = X_{t+1} - X_t = r_f + \lambda \bar{h}_t - \frac{1}{2}\bar{h}_t + \sqrt{\bar{h}_t}\varepsilon_{1,t+1}$$

where $\bar{h}_t$ denotes the conditional variance for day $t+1$ which known at the end of day t and is defined as a convex combination of the following two factor GARCH-type processes:

$$\begin{aligned}
\bar{h}_t &= n h_t + (1-n)\, RV_t \\
h_{t+1} &= \omega_1 + \beta_1 h_t + \alpha_1 \left(\varepsilon_{1,t} - \gamma_1 \sqrt{\bar{h}_t}\right)^2 \\
RV_{t+1} &= \omega_2 + \beta_2 RV_t + \alpha_2 \left(\varepsilon_{2,t} - \gamma_2 \sqrt{\bar{h}_t}\right)^2
\end{aligned}$$

where the joint distribution of the innovations $(\varepsilon_{1,t}, \varepsilon_{2,t})$ is assumed to be bivariate standard normal with correlation ρ. In this setup, realized volatility RV_t is assumed to follow the same affine dynamics as in the Heston and Nandi model, which is ideal for option pricing. Additionally, this specification nests the HN model as a special case whenever $n = 1$, and

nests a purely realized-volatility based model (RV) whenever $n = 0$. The latter can be considered as a discrete-time stochastic volatility model since the innovation of the process is different from that of the returns process. This assumption allows for two separate leverage effects via γ_1 and $\rho\gamma_1$, which can be seen through the conditional covariance of the spot return and the augmented variance process

$$Cov_t\left(r_{t+1}, \bar{h}_{t+1}\right) = -2\left(n\alpha_1\gamma_1 + (1-n)\,\alpha_2\rho\gamma_2\right)\bar{h}_t$$

For option pricing purposes we will need to identify the dynamics of the model under a risk neutral probability measure, known as the equivalent martingale measure (EMM). This is achieved by utilizing a change of measure argument. The specification of the pricing kernel (Radon-Nikodym derivative) together with the no-arbitrage argument that leads to the construction of the EMM is given analytically in the appendix. Here we will only show the correspondence between the historical and the risk neutral measures. Under the risk neutral measure

$$\begin{aligned} r_{t+1} &= r_f - \tfrac{1}{2}\bar{h}_t + \sqrt{\bar{h}_t}\varepsilon^{\star}_{1,t+1} \\ h_{t+1} &= \omega_1 + \beta_1 h_t + \alpha_1\left(\varepsilon^{\star}_{1,t} - \gamma^{\star}_1\sqrt{\bar{h}_t}\right)^2 \\ RV_{t+1} &= \omega_2 + \beta_2 RV_t + \alpha_2\left(\varepsilon^{\star}_{2,t} - \gamma^{\star}_2\sqrt{\bar{h}_t}\right)^2 \end{aligned}$$

where the map between the two measures amounts to $\gamma^{\star}_1 = \gamma_1 + \lambda$ and $\gamma^{\star}_2 = \gamma_2 + \chi$. Here, λ denotes the market risk premium parameter and χ denotes the (realized) volatility risk premium parameter. That is, the risk-neutral parameters are endogenously determined by the risk premium parameters , λ and χ.

4.4 Option Pricing

4.4.1 Risk neutralization

Christoffersen et al. (2009b) studied a large class of specifications of the underlying asset return in discrete time for the purpose of the valuation of European-style contingent claims with the use of the risk neutral valuation relationship (RNVR). Their approach did not involve the characterization of the preferences underlying the RNVR. Instead, they specified a

class of Radon-Nikodym derivatives and, using no-arbitrage arguments, constructed a class of equivalent martingale measures (EMMs) suitable for pricing contingent claims for which the resulting risk-neutral return dynamics are from the same family of distributions as the physical return dynamics. The class of processes that can be studied under their approach is relatively large, provided that the conditional moment-generating function (MGF) exists. It contains conditionally non-normal heteroskedastic processes, such as GARCH processes, as well as general discrete time stochastic volatility models. As mentioned above, the GRV model can be considered as special discrete time stochastic volatility model with the second source of randomness coming from realized volatility. The candidate Radon-Nikodym derivative for the above specification is given by:

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = \exp\left(u\sum_{i=1}^{t}\left(\nu_{1,i-1}\varepsilon_{1,i} + \nu_{2,i-1}\varepsilon_{2,i} + \Psi_t\left(\nu_{1,i-1},\nu_{2,i-1}\right)\right)\right)$$

where $\nu_{1,i}$ and $\nu_{2,i}$ are predetermined sequences and $\Psi_t(u_1,u_2)$ is defined as the natural logarithm of the conditional MGF

$$\Psi_t(u_1,u_2) = \log E_{t-1}^{P}\left[\exp\left(u_1\varepsilon_{1,t} + u_2\varepsilon_{2,t}\right)\right]$$

We can easily show that $\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t}$ is a valid Radon-Nikodym derivative by noting first that by construction is non-negative and that $E_0^P\left(\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t}\right) = 1$ .The latter results from the use of the law of iterated expectations. Now we can easily show that the probability measure Q defined by the Radon-Nikodym derivative $\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t}$ is an EMM if and only if

$$\Psi_t\left(\nu_{1,t-1} - \sqrt{h_{t-1}},\nu_{2,t-1}\right) - \Psi_t\left(\nu_{1,t-1},\nu_{2,t-1}\right) + \lambda\bar{h}_t - \frac{1}{2}\bar{h}_t = 0$$

This result is derived by using the RNVR. That is, under Q, the discounted process of the asset is a martingale

$$E_{t-1}^{Q}\left[\frac{S_t}{B_t}\right] = \frac{S_{t-1}}{B_{t-1}}$$

To get the above result, all we need to do is change the measure to P using the Radon-Nikodym derivative . That is,

$$E_{t-1}^{Q}\left[\frac{\frac{S_t}{S_{t-1}}}{\frac{B_t}{B_{t-1}}}\right]=E_{t-1}^{Q}\left[\frac{\frac{dQ}{dP}\Big|_{\mathcal{F}_t}}{\frac{dQ}{dP}\Big|_{\mathcal{F}_{t-1}}}\frac{\frac{S_t}{S_{t-1}}}{\frac{B_t}{B_{t-1}}}\right]=1$$

since $E_{t-1}^{Q}\left[\frac{dQ}{dP}\Big|_{\mathcal{F}_t}\right]=\frac{dQ}{dP}\Big|_{\mathcal{F}_{t-1}}$. The last result also identifies the pricing kernel which is

$$Z_t=\frac{\frac{dQ}{dP}\Big|_{\mathcal{F}_t}}{E_{t-1}^{P}\left[\frac{dQ}{dP}\Big|_{\mathcal{F}_t}\right]}=\frac{\frac{dQ}{dP}\Big|_{\mathcal{F}_t}}{\frac{dQ}{dP}\Big|_{\mathcal{F}_{t-1}}}=\exp\left(\nu_{1,t-1}\varepsilon_{1,t}+\nu_{2,t-1}\varepsilon_{2,t}+\Psi_t\left(\nu_{1,t-1},\nu_{2,t-1}\right)\right)$$

Since we assume that the innovations $(\varepsilon_{1,t},\varepsilon_{2,t})$ are distributed bivariate standard normal with correlation ρ, we have that

$$\Psi_t\left(\nu_{1,t-1},\nu_{2,t-1}\right)=\log E_{t-1}^{P}\left[\exp\left(\nu_{1,t-1}\varepsilon_{1,t}+\nu_{2,t-1}\varepsilon_{2,t}\right)\right]=-\frac{\nu_{1,t-1}^2}{2}-\frac{\nu_{2,t-1}^2}{2}-\nu_{1,t-1}\nu_{2,t-1}\rho$$

The pricing kernel becomes

$$Z_t=\exp\left(\nu_{1,t-1}\varepsilon_{1,t}+\nu_{2,t-1}\varepsilon_{2,t}-\frac{\nu_{1,t-1}^2}{2}-\frac{\nu_{2,t-1}^2}{2}-\nu_{1,t-1}\nu_{2,t-1}\rho\right)$$

and the necessary and sufficient condition for the probability measure to be an EMM becomes

$$\nu_{1,t-1}+\nu_{2,t-1}\rho+\lambda\sqrt{\bar{h}_t}=0$$

It is interesting to note that we can reach the same conclusion by using the pricing kernel and imposing that under the risk neutral measure Q, $E_{t-1}^{Q}\left[\exp\left(r_t\right)\right]=\exp\left(r_f\right)$. This condition, although is very valuable for the characterization of the risk neutral distribution of the returns process, points out that the second error term is responsible for a new source of non-uniqueness since it is an equation with two unknowns $\nu_{1,t}$ and $\nu_{2,t}$. In order to identify the risk neutral distribution we need to compute the risk neutral conditional MGF of the

two error terms

$$\begin{aligned} E_{t-1}^{Q}\left[\exp\left(u_1\varepsilon_{1,t}+u_2\varepsilon_{2,t}\right)\right] &= E_{t-1}^{P}\left[Z_t\exp\left(u_1\varepsilon_{1,t}+u_2\varepsilon_{2,t}\right)\right] \\ &= \exp\left(u_1\left(\nu_{1,t-1}+\nu_{2,t-1}\rho\right)+u_2\left(\nu_{2,t-1}+\nu_{1,t-1}\rho\right)-\frac{u_1^2}{2}-\frac{u_2^2}{2}-u_1u_2\rho\right) \end{aligned}$$

Using this result we can identify that under the risk neutral measure Q, the two "new" error terms $\varepsilon_{1,t}^{\star}=\varepsilon_{1,t}-\left(\nu_{1,t-1}+\nu_{2,t-1}\rho\right)$ and $\varepsilon_{2,t}^{\star}=\varepsilon_{2,t}-\left(\nu_{2,t-1}+\nu_{1,t-1}\rho\right)$ are bivariate normal with correlation ρ. We can use this information to rewrite the model as follows

$$\begin{aligned} r_{t+1} &= r_f+\lambda\bar{h}_t-\frac{1}{2}\bar{h}_t+\sqrt{\bar{h}_t}\varepsilon_{1,t+1} \\ &= r_f+\lambda\bar{h}_t-\frac{1}{2}\bar{h}_t+\sqrt{\bar{h}_t}\left(\varepsilon_{1,t+1}^{\star}+\nu_{1,t}+\nu_{2,t}\rho\right) \\ &= r_f-\frac{1}{2}\bar{h}_t+\sqrt{\bar{h}_t}\varepsilon_{1,t+1}^{\star} \end{aligned}$$

after using $\nu_{1,t-1}+\nu_{2,t-1}\rho+\lambda\sqrt{\bar{h}_t}=0$. So, in the risk neutral version of the returns process, λ vanishes. The dynamics of the GARCH component of the volatility become

$$\begin{aligned} h_{t+1} &= \omega_1+\beta_1h_t+\alpha_1\left(\varepsilon_{1,t}-\gamma_1\sqrt{\bar{h}_t}\right)^2 \\ &= \omega_1+\beta_1h_t+\alpha_1\left(\varepsilon_{1,t+1}^{\star}+\nu_{1,t}+\nu_{2,t}\rho-\gamma_1\sqrt{\bar{h}_t}\right)^2 \\ &= \omega_1+\beta_1h_t+\alpha_1\left(\varepsilon_{1,t+1}^{\star}-\gamma_1^{\star}\sqrt{\bar{h}_t}\right)^2 \end{aligned}$$

where $\gamma_1^{\star}=\gamma_1+\lambda$. The dynamics of the RV-component of the volatility become

$$\begin{aligned} RV_{t+1} &= \omega_2+\beta_2RV_t+\alpha_2\left(\varepsilon_{2,t}-\gamma_2\sqrt{\bar{h}_t}\right)^2 \\ &= \omega_2+\beta_2RV_t+\alpha_2\left(\varepsilon_{2,t+1}^{\star}+\nu_{2,t}+\nu_{1,t}\rho-\gamma_2\sqrt{\bar{h}_t}\right)^2 \\ &= \omega_2+\beta_2RV_t+\alpha_2\left(\varepsilon_{2,t+1}^{\star}-\gamma_2^{\star}\sqrt{\bar{h}_t}\right)^2 \end{aligned}$$

where $\gamma_2^{\star}=\gamma_2-\frac{\nu_{2,t}+\nu_{1,t}\rho}{\sqrt{\bar{h}_t}}$.

In order to keep the model affine under the risk neutral probability measure we assume that $\nu_{2,t}+\nu_{1,t}\rho=-\chi\sqrt{\bar{h}_t}$ which implies $\gamma_2^{\star}=\gamma_2+\chi$. By this assumption we resolve the non-uniqueness problem stated above. This assumption, which is very common in the

option pricing literature, is made to preserve the affine structure in the RV component of the volatility. Parameter χ is nothing but the price of (realized) volatility risk. This in the spirit of Heston (1993).

4.4.2 Option Pricing

Using the risk neutral model above we can proceed to evaluate European call options with payoff $(S_T - K)^+$at time t and maturity T. The price of a European call option is given by the following formula

$$C_t = E_t \left(S_T - K\right)^+ = \exp\left(-r_f T\right) S_t P_{1,t} - \exp\left(-r_f T\right) K P_{2,t}$$

where

$$P_{1,t} = \frac{\exp\left(r_f T\right)}{2} + \int_0^{+\infty} Re\left[\frac{\exp\left(\Psi^Q_{t,T}\left(1+iu\right) - iu\log\left({}^{K}/_{S_t}\right)\right)}{\pi i u}\right] du$$

and

$$P_{1,t} = \frac{1}{2} + \int_0^{+\infty} Re\left[\frac{\exp\left(\Psi^Q_{t,T}\left(iu\right) - iu\log\left({}^{K}/_{S_t}\right)\right)}{\pi i u}\right] du$$

Quantity $\Psi^Q_{t,t+T}$ denotes the multi-period risk-neutral conditional characteristic function which is given by its physical counterpart. Namely,

$$\Psi^P_{t,T}\left(u\right) = E^P_t\left[\exp\left(u\sum_{j=1}^{T} r_{t+j}\right)\right]$$

In order to use these functions for option pricing all we need to do is to use their risk neutral version, which is, we have to replace the parameters in these functional forms with the risk neutral ones. Appendix C has the functional forms analytically.

4.5 Data, Methodology and Results

4.5.1 Data

In this section we perform in sample estimation of the parameters of the option pricing formula. For the estimation we will use European S&P 500 index call options together

with high frequency data on the index. For the empirical application the option data, gathered from OptionMetrics, are observed from February 9, 2000 through December 26, 2007. We choose the 7 most liquid, out-of-the-money, closing on Wednesday call options with maturity less than 150 days. The sample consists of 2,821 options. Table 4.1 below provides information about the dataset.

4.5.2 Methodology

In the option pricing literature where option prices are computed in closed form, researchers have proposed the construction of loss functions as a way to price options. In this chapter, we employ the IVRMSE criterion

$$IVRMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(IV_j^{Data} - IV\left(\theta\right)_j^{Model}\right)}$$

where IV^{Data} is the Black-Scholes implied volatility using the actual data, $IV\left(\theta\right)^{Model}$ is the Black-Scholes implied volatility using the fitted option price and θ denotes the parameters of the option pricing formula. This loss function, overcomes the weighting problems, making the estimation more tractable, since the implied volatility metric provides an intuitive weighting of options across strikes and maturities.

The procedure to estimate the option pricing model is the following. The first step is to initialize the algorithm by picking starting values. Usually, the MLE estimates are good candidates. The second step in to calculate the option prices using the formula

$$C_t = E_t\left(S_T - K\right)^+ = \exp\left(-r_f T\right) S_t P_{1,t} - \exp\left(-r_f T\right) K P_{2,t}$$

The formula requires the calculation of $P_{1,t}$ and $P_{2,t}$ which requires the calculation of the conditional characteristic function

$$\Psi_{t,T}^P\left(u\right) = E_t^P\left[\exp\left(u\sum_{j=1}^{T} r_{t+j}\right)\right] = \exp\left(C_1\left(u,T\right)h_t + C_2\left(u,T\right)RV_t + D\left(u,T\right)\right)$$

The calculation requires the knowledge of h_t (since RV_t is observed). This is calculated

By Moneyness	F/X<0.95	0.95F/X<0.96	0.96<F/X<0.97	0.97<F/X<0.98	0.98<F/X<0.99	.99<F/X<1	All Options
Number of contracts	843	260	345	392	427	554	2821
Average implied volatility	20.80%	17.26%	16.80%	16.33%	16.94%	17.57%	18.14%
By Maturity	DTM<15	15<DTM<30	30<DTM<45	45<DTM<60	60<DTM<75	DTM>74	All Options
Number of contracts	367	1197	502	313	200	242	2821
Average implied volatility	17.84%	17.79%	18.36%	19.01%	19.06%	18.03%	18.14%

Table 4.1: **SNP500 Index Option Data 2000-2007.** We use option data that are observed from February 9, 2000 through December 26, 2007. We choose the 7 most liquid, out-of-the-money, closing on Wednesday call options with maturity less than 150 days. The sample consists of 2,821 options.

easily, since it is a function of the parameters, the history of the log-returns and the history of the realized volatility process. To generate the process, we use variance targeting. That is, we set $\bar{h}_0$ equal to the model implied long run mean of the variance. That is, we set t $\bar{h}_0$ equal to the average implied variance of the "at of the money" options. The estimation of the Fourier transforms are done by the Gauss quadrature method. After estimating the option values, we estimate their Black-Scholes implied volatility. Subsequently the loss function is calculated and the parameters are updated. The procedure is repeated until we reach a minimum.

Additionally, variance targeting pins down values for the parameters ω_1 and ω_2. That is,

$$\omega_i = \left(1 - \beta_i - \alpha_i \gamma_i^2\right) \bar{h}_0 - \alpha_i$$

where $i = 1, 2$. Under this setup, $\lambda = 0$. We also estimate a stationary GARCH model, therefore the following conditions are imposed for the estimation

$$\begin{aligned} 0 < \beta_i < 1 \\ |\gamma_1| < \sqrt{\frac{1-\beta_1}{\alpha_1 n}} \\ |\gamma_2| < \sqrt{\left(\frac{1-\beta_2}{1-\beta_1}\right)\left(\frac{1-\beta_1-\alpha_1\gamma_1^2 n}{\alpha_2\,(1-n)}\right)} \end{aligned}$$

Since these conditions imply a constrained minimization problem, we employ parameter transformations to transform the problem into an unconstrained one and facilitate the estimation procedure. Specifically, for the parameters $0 < \beta_i < 1$, we can can use the transformation $t_\beta(x) = (1 + e^{-x})^{-1}$ for which we know that $t : \mathbb{R} \to (0, 1)$. For α_1 we can use the fact that $\omega_1 = \left(1 - \beta_1 - \alpha_1 \gamma_1^2\right) \bar{h}_0 - \alpha_1 > 0$, therefore $0 < \alpha_1 < (1 - \beta_1) \bar{h}_0$. This implies the transformation $t_\alpha(x) = (1 - \beta_1) \bar{h}_0 (1 + e^{-x})^{-1}$. For γ_1 we use the following transformation

$$t_\gamma(x) = \left(\frac{\sqrt{1-\alpha_1}}{\beta_1} - \frac{1}{\bar{h}_0}\right)\left(-1 - 2\left(1 + e^{-x}\right)^{-1}\right)$$

The constrains for $(\alpha_2, \beta_2, \gamma_2)$ are calculated in a similar manner.

Another important remark is that our estimation procedure does not provide a way to

compute standard errors of our estimates. The way we overcome this issue is to assume a distribution function for the observed option prices. By doing so, we can then evaluate the distribution at the optimum (derived by minimizing the RMSE), and subsequently, compute an estimate of the standard errors numerically. We can achieve that by inverting the Fisher information, computed by the outer product of the gradient of the assumed distribution at the optimum. The distribution we assume is (log-likelihood)

$$\log L = -\frac{N}{2}\log(2\pi) - \sum_{i=1}^{N}\log(Vg_i) - \frac{1}{2}\sum_{i=1}^{N}\frac{\left(C_i^{Data} - C\left(\hat{\theta}\right)_i^{Model}\right)^2}{Vg_i}$$

where C_i^{Data} are the observed option prices, $C\left(\hat{\theta}\right)_i^{Model}$ are the model implied option prices when $\hat{\theta}$ is the IVRMSE estimate and Vg is the vega of the observed option prices.

4.5.3 Results

The estimation results are reported in table 4.2. The stationarity requirements as well as the equations for variance and equity premium targeting which are used to pin down estimates of and as functions of the other parameters are given in the appendix. Based on the IVRMSE metric conclude that the model that uses the RK-Parzen realized volatility estimator outperforms all other models, in the sense that it reduces the option pricing errors the most. The model utilizing the RK-Parzen estimator has IVRMSE equal to 2.94%, whereas the benchmark model (RV-5min) 3.13%. For the rest of the estimators, the PAV is the second best with IVRMSE equal to 2.99% and the QMLE and the RK-TH2 follow with IVRMSEs equal to 3.01% and 3.02% respectively. The MSRV is marginally better than the RV5-min and TSRV underperforms.

The first important conclusion we draw from this application is that the estimators that are robust in the presence of dependence noise perform better, with the only exception the TSRV. We attribute that to the fact that the TSRV has the lower convergence rate $\frac{1}{6}$. Given that all these estimators are robust in the presence of microstructure noise we conclude that microstructure noise has to be taken into account into this type of applications. Moreover, since financial data do exhibit jumps, the jump robust volatility estimators perform better.

NLS Estimates for the GARV model.

Parms	RV5min	TSRV	MSRV	QMLE	RK-Parzen	RK-Th2	PAV
α_1	2.59E-07	3.16E-07	2.41E-07	2.87E-07	1.87E-07	1.90E-07	1.98E-07
	(5.41E-09)	(3.26E-09)	(4.62E-09)	(8.12E-09)	(6.94E-09)	(7.89E-09)	(2.99E-09)
β_1	0.95	0.91	0.98	0.98	0.97	0.97	0.97
	(3.22E-03)	(4.72E-03)	(3.91E-03)	(4.55E-03)	(8.22E-03)	(6.15E-03)	(6.77E-03)
γ_1	415.5	509.6	267.6	264.9	392.7	350.7	364.1
	(75.1)	(95.4)	(53.4)	(64.8)	(36.9)	(44.4)	(81.3)
ω_1	4.80E-12	8.81E-12	1.85E-12	2.16E-12	3.10E-12	2.50E-12	2.82E-12
α_2	4.48E-06	4.07E-07	3.98E-07	3.88E-07	1.11E-06	5.15E-07	7.64E-07
	(9.56E-09)	(7.24E-09)	2.97E-09)	(4.72E-09)	(8.51E-09)	(6.73E-09)	(4.29E-09)
β_2	4.86E-09	4.86E-09	4.86E-09	4.86E-09	4.86E-09	4.86E-09	4.86E-09
	(4.78E-10)	(4.77E-10)	(4.80E-10)	(4.79E-10)	(4.80E-10)	(4.81E-10)	(4.79E-10)
γ_2	460.2	1564.1	1582.2	1601.9	944.4	1389.4	1139.3
	(62.1)	(95.0)	(79.2)	(102.1)	(91.7)	(100.7)	(101.2)
ω_2	1.03E-10	1.08E-10	1.08E-10	1.08E-10	1.07E-10	1.08E-10	1.07E-10
ρ	-0.996	-0.995	-0.996	-0.995	-0.995	-0.995	-0.995
	(3.14E-03)	(2.64E-03)	(2.22E-03)	(4.13E-03)	(1.33E-03)	(2.41E-03)	(4.77E-03)
n	0.67	0.94	0.38	0.39	0.42	0.34	0.39
	(9.86E-03)	(1.72E-02)	(4.41E-03)	(3.66E-03)	(5.25E-03)	(3.57E-03)	(7.13E-03)
Vol Persistence							
Returns	0.997	0.996	0.997	0.997	0.998	0.998	0.998
RV	0.950	0.995	0.996	0.996	0.988	0.994	0.991
IVRMSE%	3.13%	3.27%	3.11%	3.01%	2.94%	3.02%	2.99%

Table 4.2: **NLS Estimates for the GARV model.** We estimate the GARV model for the six different realized volatility estimators using daily close-to-close returns for the S&P 500 index, for the period February 9, 2000 to December 26, 2007. Standard errors are indicated in parentheses. To estimate parameters we impose variance targeting to 7.3% per year and equity premium targeting to 5% per year, when risk-free rate is assumed to be 3%.

However, we believe that a model that incorporates a jump component together with realized volatility that is robust in both microstructure noise and jumps might reduce the option pricing errors even further.

APPENDIX A

Simulation Results for Chapter 2

Table A.1. Combined distributions. Annual volatility is 15% and observations per block is 200.

	$S_0 = 10$								$S_0 = 30$							
Blocks	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3
2	51.00%	14.58	6.77	9.45	1.87	2.04	6.22	11.36	91.50%	5.15	0.15	1.43	3.07	-0.83	0.01	1.00
3	63.00%	11.18	4.25	6.39	1.75	0.91	3.63	7.11	91.00%	4.18	0.09	1.22	3.56	-0.67	0.08	0.86
4	66.50%	9.09	3.00	4.92	1.75	0.40	2.69	5.32	91.50%	3.60	0.04	1.02	3.95	-0.66	-0.04	0.69
5	65.00%	7.85	2.50	4.14	1.89	0.09	2.43	4.65	90.00%	3.20	-0.05	0.92	4.10	-0.67	-0.05	0.49
8	66.50%	5.89	1.74	3.22	2.01	-0.36	1.46	3.32	91.00%	2.52	-0.09	0.71	4.60	-0.51	-0.16	0.35
10	69.00%	5.18	1.37	2.68	1.96	-0.24	1.10	2.91	89.00%	2.25	-0.10	0.65	4.73	-0.50	-0.13	0.28
15	75.50%	4.14	0.93	2.02	2.07	-0.44	0.77	1.97	91.50%	1.83	-0.13	0.52	5.15	-0.43	-0.14	0.15
20	72.50%	3.54	0.75	1.70	1.79	-0.30	0.60	1.70	92.50%	1.58	-0.13	0.44	5.44	-0.39	-0.14	0.16
25	71.50%	3.13	0.63	1.44	1.75	-0.21	0.57	1.56	91.00%	1.42	-0.14	0.41	5.65	-0.42	-0.13	0.11
30	75.00%	2.83	0.53	1.26	1.56	-0.26	0.50	1.18	91.00%	1.29	-0.14	0.37	6.02	-0.39	-0.14	0.06
35	76.50%	2.61	0.50	1.17	1.41	-0.19	0.44	1.10	94.00%	1.19	-0.16	0.35	6.33	-0.35	-0.18	0.04
40	73.00%	2.44	0.51	1.16	1.29	-0.11	0.49	1.16	92.00%	1.12	-0.17	0.33	6.59	-0.35	-0.17	0.00
45	73.00%	2.30	0.44	1.08	1.31	-0.21	0.43	1.01	91.00%	1.05	-0.17	0.33	6.72	-0.36	-0.18	0.03
50	69.50%	2.18	0.43	1.03	1.28	-0.18	0.41	1.05	89.00%	1.00	-0.17	0.32	7.02	-0.35	-0.17	-0.01

Tables A.1: Simulation results of the combination scheme. The first column indicates how many blocks are being combined. The results rely on 200 processes split in blocks of 200 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^4. The table displays the coverage (Covg), the average length of the confidence intervals (ALCI), the bias, the root mean square error of the point estimator of σ^2, the average effective sample size (ESS), and the quantiles of the distribution of the point estimator of $\sigma^2 - \sigma_0^2$, when $\sigma_0 = 15\%$.

Table A.2. Combined distributions. Annual volatility is 15% and observations per block is 300.

	$S_0 = 10$								$S_0 = 30$							
Blocks	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3
2	51.00%	11.18	5.74	7.68	1.98	1.75	5.00	9.16	96.50%	4.17	0.10	1.06	3.13	-0.59	0.13	0.73
3	58.50%	8.68	3.76	5.37	1.87	1.23	3.22	5.92	94.00%	3.35	-0.03	0.83	3.46	-0.52	-0.02	0.54
4	58.50%	7.17	2.89	4.33	2.02	0.63	2.42	4.57	96.50%	2.90	-0.05	0.75	3.80	-0.52	-0.05	0.49
5	64.00%	6.22	2.22	3.47	2.14	0.46	1.96	3.81	92.00%	2.59	-0.05	0.72	4.15	-0.50	-0.06	0.44
8	68.50%	4.74	1.51	2.45	2.30	0.17	1.17	2.76	95.50%	2.05	-0.04	0.54	4.60	-0.47	-0.08	0.33
10	73.00%	4.16	1.18	2.14	2.19	-0.14	0.94	2.11	93.50%	1.83	-0.06	0.48	4.93	-0.42	-0.07	0.23
15	67.50%	3.33	0.95	1.63	2.10	0.12	0.89	1.81	93.50%	1.50	-0.07	0.39	5.39	-0.32	-0.06	0.17
20	67.50%	2.87	0.83	1.40	1.98	0.10	0.85	1.59	93.50%	1.30	-0.08	0.35	5.98	-0.34	-0.07	0.19
25	69.00%	2.56	0.73	1.20	1.92	0.12	0.70	1.35	95.00%	1.16	-0.08	0.31	6.36	-0.29	-0.09	0.14
30	68.50%	2.32	0.71	1.14	1.72	0.09	0.64	1.29	93.50%	1.06	-0.08	0.29	6.66	-0.30	-0.07	0.09
34	65.50%	2.17	0.66	1.08	1.85	0.00	0.69	1.23	92.50%	0.99	-0.09	0.28	6.93	-0.27	-0.08	0.09

Tables A.2: Simulation results of the combination scheme. The first column indicates how many blocks are being combined. The results rely on 200 processes split in blocks of 300 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^4. The table displays the coverage (Covg), the average length of the confidence intervals (ALCI), the bias, the root mean square error of the point estimator of σ^2, the average effective sample size (ESS), and the quantiles of the distribution of the point estimator of $\sigma^2 - \sigma_0^2$, when $\sigma_0 = 15\%$.

Table A.3. Combined distributions. Annual volatility is 15% and observations per block is 500.

	$S_0 = 10$								$S_0 = 30$							
Blocks	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3
2	50.00%	7.84	3.88	5.60	1.95	0.78	3.45	6.64	95.00%	3.21	0.11	0.80	3.29	-0.38	0.11	0.66
3	56.50%	6.21	2.56	3.94	1.96	0.30	2.27	4.48	95.50%	2.62	0.11	0.66	3.61	-0.37	0.16	0.53
4	59.50%	5.24	1.95	3.09	1.91	0.10	1.74	3.44	94.50%	2.27	0.09	0.58	3.89	-0.28	0.11	0.42
5	60.50%	4.62	1.64	2.68	2.07	0.13	1.54	2.99	95.50%	2.02	0.04	0.50	4.12	-0.27	0.05	0.34
8	64.50%	3.57	1.19	1.94	2.33	0.05	1.00	2.38	95.00%	1.59	-0.01	0.43	4.62	-0.30	-0.04	0.28
10	64.50%	3.17	1.02	1.68	2.48	0.09	0.84	1.99	93.50%	1.42	-0.03	0.38	4.94	-0.27	-0.05	0.24
15	65.50%	2.55	0.79	1.36	2.65	-0.03	0.66	1.55	93.50%	1.16	-0.04	0.33	5.64	-0.29	-0.04	0.18
20	64.00%	2.20	0.70	1.23	2.78	-0.03	0.64	1.37	91.50%	1.00	-0.04	0.30	6.15	-0.27	-0.03	0.17

Tables A.3: Simulation results of the combination scheme. The first column indicates how many blocks are being combined. The results rely on 200 processes split in blocks of 500 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^4. The table displays the coverage (Covg), the average length of the confidence intervals (ALCI), the bias, the root mean square error of the point estimator of σ^2, the average effective sample size (ESS), and the quantiles of the distribution of the point estimator of $\sigma^2 - \sigma_0^2$, when $\sigma_0 = 15\%$.

Table A.4. Combined distributions. Annual volatility is 30% and observations per block is 200.

	$S_0 = 10$								$S_0 = 30$							
Blocks	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3
2	86.50%	27.99	4.37	9.24	2.97	-1.15	4.59	9.55	94.00%	14.48	0.15	3.68	3.60	-2.18	0.03	2.26
3	89.50%	22.07	2.43	6.64	2.93	-1.97	2.35	6.41	92.00%	11.79	0.01	3.23	3.97	-1.88	0.00	1.81
4	89.00%	18.97	1.75	5.80	3.18	-2.13	1.62	5.53	93.00%	10.20	-0.09	2.70	4.29	-1.98	-0.14	1.61
5	89.50%	16.90	1.45	4.96	3.32	-1.31	0.99	4.91	94.00%	9.10	-0.21	2.25	4.51	-1.67	-0.25	1.36
8	92.00%	13.19	0.68	3.57	3.59	-1.57	1.03	2.94	96.00%	7.19	-0.22	1.83	5.26	-1.41	-0.28	1.02
10	92.00%	11.76	0.59	3.31	3.88	-1.44	0.69	2.76	94.50%	6.42	-0.28	1.64	5.73	-1.38	-0.42	0.86
15	87.50%	9.54	0.33	2.97	4.28	-1.66	0.28	2.40	93.50%	5.22	-0.43	1.39	6.45	-1.38	-0.46	0.63
20	87.50%	8.24	0.19	2.66	4.50	-1.55	0.26	2.01	90.00%	4.51	-0.57	1.30	7.01	-1.37	-0.48	0.23
25	87.50%	7.35	0.12	2.37	4.71	-1.38	-0.02	1.77	90.00%	4.03	-0.61	1.19	7.52	-1.32	-0.56	0.15
30	89.00%	6.68	-0.02	2.06	4.88	-1.42	0.04	1.44	90.50%	3.68	-0.55	1.08	8.16	-1.23	-0.58	0.12
35	90.00%	6.19	0.02	1.88	5.07	-1.33	-0.02	1.33	90.50%	3.41	-0.54	0.99	8.73	-1.17	-0.56	0.08
40	89.50%	5.79	0.01	1.72	5.24	-1.10	0.07	1.10	90.00%	3.19	-0.56	0.96	9.10	-1.10	-0.58	-0.04
45	90.50%	5.45	-0.06	1.65	5.31	-1.09	-0.06	0.96	87.50%	3.00	-0.57	0.94	9.53	-1.07	-0.57	-0.11
50	89.50%	5.16	-0.13	1.57	5.37	-0.98	-0.15	0.83	90.00%	2.85	-0.56	0.92	9.85	-1.09	-0.61	-0.08

Tables A.4: Simulation results of the combination scheme. The first column indicates how many blocks are being combined. The results rely on 200 processes split in blocks of 200 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^4. The table displays the coverage (Covg), the average length of the confidence intervals (ALCI), the bias, the root mean square error of the point estimator of σ^2, the average effective sample size (ESS), and the quantiles of the distribution of the point estimator of $\sigma^2 - \sigma_0^2$, when $\sigma_0 = 30\%$.

Table A.5. Combined distributions. Annual volatility is 30% and observations per block is 300.

	$S_0 = 10$								$S_0 = 30$							
Blocks	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3
2	89.00%	22.30	3.21	6.90	3.09	-1.26	2.96	7.40	95.00%	11.74	-0.06	3.00	3.45	-1.92	-0.11	1.99
3	91.00%	17.78	1.95	5.23	3.34	-1.12	1.67	5.52	93.50%	9.56	-0.20	2.59	3.97	-1.98	-0.46	1.68
4	93.50%	15.27	1.33	4.42	3.59	-1.63	1.57	4.20	94.00%	8.29	-0.13	2.15	4.25	-1.56	-0.17	1.28
5	93.00%	13.57	0.99	3.99	3.69	-1.98	1.02	3.74	96.00%	7.42	-0.12	1.98	4.49	-1.57	-0.12	1.12
8	89.00%	10.64	0.51	3.18	3.87	-1.66	0.39	2.79	94.00%	5.85	-0.18	1.48	5.29	-1.17	-0.27	0.84
10	92.00%	9.50	0.26	2.75	3.99	-1.50	0.34	1.99	93.50%	5.23	-0.19	1.36	5.74	-1.08	-0.24	0.73
15	88.50%	7.75	0.25	2.46	4.45	-1.62	0.40	1.88	93.00%	4.27	-0.26	1.15	6.67	-0.98	-0.24	0.52
20	88.50%	6.72	0.26	2.15	4.87	-1.42	0.26	1.71	92.50%	3.70	-0.23	1.00	7.40	-0.86	-0.26	0.41
25	89.50%	6.02	0.16	1.92	5.05	-1.41	0.13	1.49	93.50%	3.31	-0.25	0.91	8.21	-0.83	-0.25	0.32
30	90.50%	5.48	0.06	1.68	5.28	-1.21	-0.02	1.21	91.50%	3.02	-0.25	0.82	8.63	-0.76	-0.25	0.22
34	93.00%	5.14	0.05	1.51	5.50	-1.04	0.01	1.18	93.00%	2.83	-0.25	0.78	9.02	-0.70	-0.28	0.23

Tables A.5: Simulation results of the combination scheme. The first column indicates how many blocks are being combined. The results rely on 200 processes split in blocks of 300 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^4. The table displays the coverage (Covg), the average length of the confidence intervals (ALCI), the bias, the root mean square error of the point estimator of σ^2, the average effective sample size (ESS), and the quantiles of the distribution of the point estimator of $\sigma^2 - \sigma_0^2$, when $\sigma_0 = 30\%$.

Table A.6. Combined distributions. Annual volatility is 30% and observations per block is 500.

	$S_0 = 10$								$S_0 = 30$							
Blocks	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3	Covg	ALCI	Bias	RMSE	ESS	Q1	Median	Q3
2	93.50%	16.79	1.65	4.77	3.06	-1.91	1.35	4.84	96.00%	9.12	0.13	2.39	3.66	-1.56	-0.09	1.79
3	95.00%	13.68	1.15	3.75	3.38	-1.44	1.00	3.48	96.00%	7.44	0.15	1.91	3.96	-1.30	0.05	1.62
4	93.50%	11.80	0.89	3.30	3.66	-1.29	0.74	2.99	97.00%	6.44	0.09	1.56	4.32	-0.88	0.10	1.09
5	93.00%	10.51	0.72	3.05	3.90	-1.23	0.25	3.00	98.00%	5.75	-0.01	1.33	4.51	-0.94	-0.08	0.90
8	93.50%	8.25	0.48	2.37	4.28	-1.37	0.31	2.23	96.50%	4.53	-0.10	1.06	5.25	-0.90	-0.09	0.61
10	95.00%	7.38	0.39	2.05	4.48	-1.00	0.41	1.85	94.00%	4.05	-0.11	1.03	5.61	-0.81	-0.16	0.59
15	92.50%	6.00	0.17	1.72	4.97	-1.14	0.18	1.47	93.00%	3.30	-0.16	0.89	6.38	-0.77	-0.22	0.43
20	93.00%	5.18	0.06	1.52	5.16	-1.07	0.04	1.08	94.00%	2.86	-0.18	0.74	7.01	-0.63	-0.19	0.26

Tables A.6: Simulation results of the combination scheme. The first column indicates how many blocks are being combined. The results rely on 200 processes split in blocks of 500 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^4. The table displays the coverage (Covg), the average length of the confidence intervals (ALCI), the bias, the root mean square error of the point estimator of σ^2, the average effective sample size (ESS), and the quantiles of the distribution of the point estimator of $\sigma^2 - \sigma_0^2$, when $\sigma_0 = 30\%$.

APPENDIX B

Simulation Results for Chapter 3

Table B.1. Combined distributions. Annual volatility is 15% and observations per block is 200.

Noise	$\sigma_u = .005\%$				$\sigma_u = .01\%$				$\sigma_u = .02\%$			
Blocks	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS
2	60.0%	13.20	7.11	12.41	72.5%	13.20	5.51	13.53	85.0%	11.70	4.29	14.58
3	64.0%	9.95	5.20	9.54	76.5%	9.63	3.99	8.81	79.5%	9.08	3.53	11.24
4	66.0%	8.27	4.17	7.83	79.5%	8.05	3.40	7.23	77.5%	7.75	3.11	9.73
5	67.0%	7.19	3.86	7.45	76.0%	7.05	3.02	6.62	77.0%	6.85	2.87	8.74
8	70.0%	5.41	2.55	5.56	72.5%	5.37	2.40	4.63	78.5%	5.20	2.21	6.43
10	73.0%	4.80	2.15	4.73	75.5%	4.67	2.06	3.76	77.0%	4.59	1.99	5.23
15	70.0%	3.81	1.80	3.61	71.0%	3.74	1.74	2.75	77.5%	3.72	1.62	4.07
20	71.0%	3.27	1.55	2.70	68.0%	3.19	1.63	2.02	75.5%	3.15	1.41	3.56
25	70.5%	2.87	1.37	2.39	67.5%	2.84	1.47	1.70	73.5%	2.80	1.30	3.08
30	73.5%	2.59	1.30	2.13	64.0%	2.55	1.38	1.37	66.0%	2.55	1.29	2.76
35	69.0%	2.39	1.21	1.84	60.5%	2.37	1.32	1.10	68.5%	2.34	1.21	2.52
40	70.5%	2.26	1.11	1.63	60.0%	2.20	1.21	1.00	66.5%	2.19	1.13	2.27
45	71.5%	2.11	1.02	1.40	59.0%	2.07	1.15	.95	63.5%	2.04	1.06	2.13
50	73.0%	1.99	0.96	1.26	54.5%	1.96	1.13	.83	65.5%	1.94	1.03	1.98

Tables B.1: Simulation results of the combination scheme. The table displays the coverage, the average length of the confidence intervals, the root mean square error of the point estimator of σ^2 and the average effective sample size, when $S_0 = 10$ and $\sigma_0 = 15\%$. The results rely on 200 processes split in blocks of 200 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^3.

Table B.2. Combined distributions. Annual volatility is 30% and observations per block is 200.

Noise	σ_u = .005%				σ_u = .01%				σ_u = .02%			
Blocks	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS
2	93.0%	27.20	7.65	22.65	94.0%	28.80	7.99	1.99*10^4	90.5%	30.10	9.72	15.29
3	93.0%	21.70	5.91	18.55	92.5%	22.80	6.36	1.63*10^4	88.0%	24.20	7.77	12.45
4	93.5%	18.50	5.20	16.31	96.0%	19.60	5.33	1.38*10^4	88.5%	20.80	6.67	11.62
5	91.0%	16.30	4.90	14.55	95.0%	17.30	4.89	1.29*10^4	87.5%	18.40	5.89	10.54
8	89.0%	12.90	3.72	11.83	93.5%	13.50	3.63	9.81	88.5%	14.60	4.56	8.49
10	87.5%	11.50	3.54	10.72	92.5%	12.10	3.26	9.22	88.5%	12.90	4.10	7.40
15	86.5%	9.22	2.99	9.28	92.5%	9.78	2.75	7.44	88.0%	10.40	3.29	5.98
20	82.5%	7.92	2.78	8.17	92.5%	8.46	2.28	6.71	89.5%	8.96	2.77	5.04
25	81.0%	7.05	2.70	7.45	94.0%	7.54	2.06	6.13	89.0%	8.00	2.58	4.76
30	81.0%	6.44	2.49	6.59	94.0%	6.82	1.97	5.33	87.5%	7.27	2.50	4.50
35	80.5%	5.95	2.30	6.12	89.0%	6.33	1.86	5.11	90.0%	6.71	2.22	4.34
40	79.5%	5.55	2.19	5.82	89.0%	5.93	1.76	4.84	88.0%	6.27	2.09	4.00
45	72.5%	5.22	2.21	5.54	90.0%	5.57	1.66	4.53	86.5%	5.90	1.96	3.84
50	75.0%	4.95	2.13	5.43	90.0%	5.29	1.51	4.36	88.5%	5.61	1.82	3.62

Tables B.2: Simulation results of the combination scheme. The table displays the coverage, the average length of the confidence intervals, the root mean square error of the point estimator of σ^2 and the average effective sample size, when $S_0 = 10$ and $\sigma_0 = 30\%$.. The results rely on 200 processes split in blocks of 200 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^3.

Table B.3. Combined distributions. Annual volatility is 15% and observations per block is 200.

Noise		σ_u = .005%				σ_u = .01%				σ_u = .02%		
Blocks	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS
2	88.0%	7.16	2.11	17.18	88.5%	7.50	2.37	15.92	75.5%	8.22	3.38	12.47
3	92.5%	5.73	1.68	14.64	86.5%	6.02	1.98	12.86	71.5%	6.56	2.80	8.99
4	90.0%	4.91	1.46	13.39	87.5%	5.16	1.72	11.44	73.0%	5.61	2.39	7.66
5	91.0%	4.36	1.30	11.34	89.5%	4.60	1.48	9.82	73.0%	4.94	2.16	6.21
8	87.5%	3.36	1.05	8.95	90.5%	3.62	1.15	7.90	74.5%	3.82	1.79	3.88
10	88.5%	3.01	0.94	7.98	89.5%	3.23	1.04	6.95	71.5%	3.39	1.67	3.33
15	93.0%	2.45	0.71	7.00	90.5%	2.63	0.86	5.34	70.0%	2.70	1.51	2.54
20	91.5%	2.10	0.59	6.14	90.0%	2.27	0.73	4.68	65.5%	2.27	1.35	2.21
25	91.5%	1.87	0.53	5.35	85.0%	2.03	0.69	4.19	64.5%	2.03	1.27	1.85
30	92.0%	1.71	0.49	5.16	85.5%	1.83	0.65	3.99	69.5%	1.87	1.17	1.76
35	92.0%	1.58	0.46	4.80	84.5%	1.70	0.60	3.84	64.0%	1.70	1.09	1.51
40	92.0%	1.48	0.44	4.36	86.0%	1.59	0.55	3.63	69.5%	1.59	1.01	1.46
45	88.5%	1.39	0.42	4.23	85.0%	1.48	0.55	3.37	67.0%	1.49	0.96	1.35
50	90.0%	1.32	0.40	4.01	88.0%	1.40	0.51	3.09	64.5%	1.39	0.93	1.22

Tables B.3: Simulation results of the combination scheme. The table displays the coverage, the average length of the confidence intervals, the root mean square error of the point estimator of σ^2 and the average effective sample size, when $S_0 = 20$ and $\sigma_0 = 15\%$. The results rely on 200 processes split in blocks of 200 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^3.

Table B.4. Combined distributions. Annual volatility is 30% and observations per block is 200.

Noise	$\sigma_u = .005\%$				$\sigma_u = .01\%$				$\sigma_u = .02\%$			
Blocks	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS
2	96.0%	19.30	4.84	24.02	89.5%	22.40	6.56	19.45	93.0%	25.30	6.86	14.22
3	94.0%	15.50	3.92	20.41	93.5%	18.40	5.16	16.36	93.5%	20.30	5.87	11.52
4	95.5%	13.30	3.39	18.74	93.5%	15.70	4.45	14.83	92.5%	17.40	5.05	10.53
5	93.5%	11.90	3.17	16.59	92.0%	14.00	4.03	13.55	89.5%	15.40	4.66	9.42
8	93.5%	9.35	2.44	14.25	91.5%	11.00	3.21	11.34	90.5%	12.10	3.42	7.19
10	91.5%	8.34	2.23	13.34	92.0%	9.81	2.89	9.97	90.0%	10.80	3.11	6.33
15	91.0%	6.80	1.88	10.94	89.5%	7.96	2.37	8.59	87.0%	8.74	2.69	5.46
20	93.5%	5.88	1.65	9.79	89.5%	6.89	2.10	7.71	86.0%	7.56	2.32	5.05
25	88.5%	5.24	1.58	9.09	90.0%	6.15	1.80	6.89	88.0%	6.72	2.06	4.64
30	88.5%	4.79	1.51	8.74	93.5%	5.63	1.59	6.52	88.0%	6.11	1.95	4.23
35	89.5%	4.43	1.43	8.20	93.0%	5.20	1.50	6.15	89.0%	5.66	1.87	4.13
40	86.5%	4.14	1.36	7.68	91.5%	4.82	1.40	5.82	90.5%	5.28	1.77	3.84
45	87.5%	3.90	1.30	7.17	89.5%	4.54	1.33	5.53	85.5%	4.96	1.66	3.53
50	83.0%	3.70	1.28	7.05	91.0%	4.30	1.25	5.34	83.5%	4.69	1.59	3.31

Tables B.4: Simulation results of the combination scheme. The table displays the coverage, the average length of the confidence intervals, the root mean square error of the point estimator of σ^2 and the average effective sample size, when $S_0 = 20$ and $\sigma_0 = 30\%$. The results rely on 200 processes split in blocks of 200 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^3.

Table B.5. Combined distributions. Annual volatility is 15% and observations per block is 200.

Noise	σ_u = .005%				σ_u = .01%				σ_u = .02%			
Blocks	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS
2	92.5%	5.98	1.60	20.16	88.5%	6.82	2.02	1.5.37	81.5%	7.71	2.87	11.37
3	94.5%	4.83	1.26	16.11	90.5%	5.51	1.64	12.39	74.5%	6.10	2.67	7.99
4	93.5%	4.17	1.10	14.61	91.5%	4.69	1.35	9.93	69.5%	5.03	2.38	6.00
5	93.0%	3.69	0.96	13.02	90.5%	4.13	1.21	9.04	72.5%	4.39	2.15	5.04
8	94.0%	2.92	0.81	11.08	92.5%	3.23	0.95	6.70	71.0%	3.31	1.89	3.04
10	94.5%	2.60	0.72	9.73	88.0%	2.88	0.87	6.10	66.5%	2.96	1.80	2.66
15	95.0%	2.11	0.57	8.17	89.0%	2.33	0.73	4.96	63.5%	2.37	1.60	2.03
20	95.5%	1.82	0.47	7.01	90.0%	2.02	0.63	4.65	57.0%	2.00	1.40	1.61
25	95.5%	1.63	0.42	6.63	87.5%	1.79	0.63	4.58	56.0%	1.78	1.28	1.30
30	95.0%	1.49	0.38	6.15	88.0%	1.64	0.56	4.14	56.5%	1.62	1.19	1.10
35	94.0%	1.38	0.39	5.65	90.5%	1.51	0.50	3.95	53.0%	1.49	1.14	.96
40	90.5%	1.29	0.36	5.05	90.0%	1.41	0.49	3.80	54.0%	1.38	1.07	.83
45	93.5%	1.22	0.33	4.81	87.0%	1.33	0.48	3.73	50.5%	1.30	1.00	.74
50	93.5%	1.15	0.32	4.74	86.5%	1.26	0.45	3.70	54.0%	1.23	0.92	.76

Tables B.5: Simulation results of the combination scheme. The table displays the coverage, the average length of the confidence intervals, the root mean square error of the point estimator of σ^2 and the average effective sample size, when $S_0 = 30$ and $\sigma_0 = 15\%$. The results rely on 200 processes split in blocks of 200 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^3.

Table B.6. Combined distributions. Annual volatility is 30% and observations per block is 200.

Noise	$\sigma_u = .005\%$				$\sigma_u = .01\%$				$\sigma_u = .02\%$			
Blocks	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS
2	96.0%	17.30	4.26	23.74	92.0%	21.20	6.05	18.85	90.5%	24.30	7.72	14.89
3	95.5%	14.10	3.67	21.38	93.5%	17.00	4.69	16.17	90.0%	19.50	6.17	11.69
4	92.5%	12.10	3.19	19.48	91.0%	14.60	4.13	14.17	88.5%	16.60	5.35	9.96
5	95.5%	10.80	2.60	17.39	91.5%	12.90	3.66	12.24	91.0%	14.70	4.64	8.81
8	98.0%	8.48	1.91	14.57	87.5%	10.20	3.05	10.13	88.5%	11.60	3.87	6.50
10	98.0%	7.60	1.70	13.91	90.5%	9.09	2.71	9.28	89.5%	10.20	3.37	5.87
15	94.0%	6.16	1.50	12.31	94.5%	7.41	2.03	8.16	83.0%	8.30	2.90	4.98
20	96.5%	5.33	1.28	11.22	93.5%	6.43	1.78	7.17	82.0%	7.13	2.56	4.30
25	94.5%	4.76	1.24	10.46	91.5%	5.74	1.67	6.69	82.5%	6.35	2.33	3.97
30	95.0%	4.34	1.15	9.60	92.0%	5.23	1.59	6.26	78.5%	5.77	2.21	3.70
35	92.0%	4.01	1.09	8.83	91.0%	4.84	1.44	5.94	82.5%	5.31	2.03	3.50
40	93.5%	3.75	0.97	8.44	91.0%	4.53	1.28	5.61	82.5%	4.95	1.84	3.18
45	92.5%	3.53	0.95	7.95	90.0%	4.27	1.19	5.31	82.5%	4.68	1.82	2.85
50	93.0%	3.35	0.93	7.35	94.0%	4.04	1.12	5.18	80.5%	4.44	1.70	2.66

Tables B.6: Simulation results of the combination scheme. The table displays the coverage, the average length of the confidence intervals, the root mean square error of the point estimator of σ^2 and the average effective sample size, when $S_0 = 30$ and $\sigma_0 = 30\%$. The results rely on 200 processes split in blocks of 200 observations. All columns but the ESS are multiples of 10^{-5}. The ESS is a multiple of 10^3.

Table B.7. Combined distributions when $S_0 = 10$, $\sigma = 30\%$ and $\sigma_u = .005\%$,

Obs	200				300				400			
Blocks	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS	Covg.	ALCI	RMSE	ESS
2	93.0%	27.20	7.65	22.62	91.0%	22.00	6.63	21.90	94.5%	18.60	5.03	19.91
3	93.0%	21.70	5.91	18.51	89.5%	17.70	5.50	18.35	94.0%	15.10	3.96	17.35
4	93.5%	18.50	5.20	16.30	90.0%	15.20	4.47	16.53	96.5%	13.00	3.44	15.50
5	91.0%	16.30	4.90	14.59	89.0%	13.50	4.39	14.86	93.5%	11.60	3.12	13.83
8	89.0%	12.90	3.72	11.80	87.0%	10.60	3.43	11.88	93.0%	9.12	2.50	10.97
10	87.5%	11.50	3.54	10.77	89.0%	9.41	3.06	10.71	90.5%	8.13	2.28	9.80
15	86.5%	9.22	2.99	9.28	90.5%	7.62	2.36	9.46	90.5%	6.58	1.93	8.53
20	82.5%	7.92	2.78	8.17	91.0%	6.59	2.00	8.61	91.5%	5.70	1.65	7.88
25	81.0%	7.05	2.70	7.45	90.0%	5.88	1.95	8.26	90.0%	5.08	1.56	7.08
30	81.0%	6.44	2.49	6.59	87.0%	5.36	1.78	7.93	NA	NA	NA	NA
35	80.5%	5.95	2.30	6.12	85.5%	5.02	1.73	7.51	NA	NA	NA	NA
40	79.5%	5.55	2.19	5.82	NA	NA	NA	NA	NA	NA	NA	NA
45	72.5%	5.22	2.21	5.54	NA	NA	NA	NA	NA	NA	NA	NA
50	75.0%	4.95	2.13	5.43	NA	NA	NA	NA	NA	NA	NA	NA

Tables B.7: Simulation results of the combination scheme. The table displays the coverage, the average length of the confidence intervals, the root mean square error of the point estimator of σ^2 and the average effective sample size, when $S_0 = 10$, $\sigma = 30\%$ and $\sigma_u = .005\%$. The results rely on 200 processes split in blocks of 200, 300 and 400 observations.

APPENDIX C

Data Cleaning Filters

The following gives an outline of the data cleaning filters found in Barndorff-Nielsen et al. (2009). First, entries with a time stamp outside the 9:30 am–4 pm window are deleted. Entries with a bid, ask or transaction price equal to zero are deleted. We retain entries originating from a single exchange (NYSE in our application). Quotes data only: When multiple quotes have the same time stamp, we replace all these with a single entry with the median bid and median ask price. We delete entries for which the spread is negative. We delete entries for which the spread is more that 50 times the median spread on that day. We delete entries for which the mid-quote deviated by more than 5 mean absolute deviations from a rolling centered median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after). Transactions data only: Delete entries with corrected trades. (Trades with a Correction Indicator, $CORR \neq 0$). Delete entries with abnormal Sale Condition. (Trades where COND has a letter code, except for 'E' and 'F'). If multiple transactions have the same time stamp, use the median price. Delete entries with prices that are above the 'ask' plus the bid–ask spread. Similar for entries with prices below the 'bid' minus the bid–ask spread.

APPENDIX D

Moment generating functions for the GARV model

The one-period conditional moment generating function is

$$E_t^P\left[\exp\left(u_1 r_{t+1}+u_2 h_{t+1}+u_3 RV_{t+1}\right)\right]=\exp\left(A_1\left(u_1,u_2,u_3\right)+A_2\left(u_1,u_2,u_3\right)+B\left(u_1,u_2,u_3\right)\right)$$

where

$$\begin{aligned}A_1\left(u_1,u_2,u_3\right)&=u_1\left(\lambda-\tfrac{1}{2}\right)n+u_2\beta_1+na\left(u_1,u_2,u_3\right)\\A_2\left(u_1,u_2,u_3\right)&=u_1\left(\lambda-\tfrac{1}{2}\right)(1-n)+u_3\beta_2+(1-n)\,a\left(u_1,u_2,u_3\right)\\B\left(u_1,u_2,u_3\right)&=u_1 r_f+u_2\omega_1+u_3\omega_2+b\left(u_1,u_2,u_3\right)-\tfrac{1}{2}\log\left(1-2u_3\alpha_2\left(1-\rho^2\right)\right)\end{aligned}$$

So the multi-period physical conditional characteristic function, using the above results, has the following form

$$\Psi_{t,T}^P\left(u\right)=E_t^P\left[\exp\left(u\sum_{j=1}^{T}r_{t+j}\right)\right]=\exp\left(C_1\left(u,T\right)h_t+C_2\left(u,T\right)RV_t+D\left(u,T\right)\right)$$

where

$$\begin{aligned}C_1\left(u,T+1\right)&=A_1\left(u,C_1\left(u,T\right),C_2\left(u,T\right)\right)\\C_2\left(u,T+1\right)&=A_2\left(u,C_1\left(u,T\right),C_2\left(u,T\right)\right)\\D\left(u,T+1\right)&=B\left(u,C_1\left(u,T\right),C_2\left(u,T\right)\right)+D\left(u,T\right)\end{aligned}$$

and initial conditions$C_1\left(u,1\right)=A_1\left(u,0,0\right)$, $C_2\left(u,1\right)=A_2\left(u,0,0\right)$ and $D\left(u,1\right)=B\left(u,0,0\right)$.

BIBLIOGRAPHY

Yacine Aït-Sahalia and Robert Kimmel. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83(2):413–452, 2007.

Yacine Aït-Sahalia and Jialin Yu. High frequency market microstructure noise estimates and liquidity measures. *The Annals of Applied Statistics*, 3(1):422–457, 2009.

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

Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453): 42–55, 2001.

Federico M Bandi and Jeffrey R Russell. Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3):655–692, 2006.

George A Barnard. Pivotal models and the fiducial argument. *International Statistical Review/Revue Internationale de Statistique*, pages 309–323, 1995.

Ole E Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008.

Ole E Barndorff-Nielsen, P Reinhard Hansen, Asger Lunde, and Neil Shephard. Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):C1–C32, 2009.

David S Bates. Post-'87 crash fears in the s&p 500 futures option market. *Journal of Econometrics*, 94(1):181–238, 2000.

Fischer Black. The pricing of commodity contracts. *Journal of financial economics*, 3(1): 167–179, 1976.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

Andrew A Christie. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, 10(4):407–432, 1982.

P Christoffersen, C Dorion, K Jacobs, and Y Wang. Volatility components: Affine restrictions and non-normal innovations, forthcoming in the journal of business and economic statistics. 2009a.

Peter Christoffersen, Kris Jacobs, Chayawat Ornthanalai, and Yintian Wang. Option valuation with long-run and short-run volatility components. *Journal of Financial Economics*, 90(3):272–297, 2008.

Peter Christoffersen, Redouane Elkamhi, Bruno Feunou, and Kris Jacobs. Option valuation with conditional heteroskedasticity and nonnormality. *Review of Financial studies*, page hhp078, 2009b.

Peter Christoffersen, Steven Heston, and Kris Jacobs. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 55(12):1914–1932, 2009c.

Peter Christoffersen, Bruno Feunou, Kris Jacobs, and Nour Meddahi. The economic value of realized volatility: using high-frequency returns for option valuation. *Journal of Financial and Quantitative Analysis*, 49(03):663–697, 2014.

Jessi Cisewski and Jan Hannig. Generalized fiducial inference for normal linear mixed models. *The Annals of Statistics*, 40(4):2102–2127, 2012.

Fulvio Corsi, Nicola Fusari, and Davide La Vecchia. Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics*, 107(2):284–304, 2013.

A Philip Dawid and Mervyn Stone. The functional-model basis of fiducial inference. *The Annals of Statistics*, pages 1054–1067, 1982.

Sylvain Delattre and Jean Jacod. A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli*, 3(1):1–28, 1997.

Arthur P Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 205–247, 1968.

Randal Douc and Eric Moulines. Limit theorems for weighted samples with applications to sequential monte carlo methods. *Ann. Statist.*, 36(5):2344–2376, 10 2008. doi: 10.1214/07-AOS514. URL `http://dx.doi.org/10.1214/07-AOS514`.

Arnaud Doucet, Nando De Freitas, and Neil Gordon. *An introduction to sequential Monte Carlo methods*. Springer, 2001.

Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

Robert F Engle and Jeffrey R Russell. Analysis of high frequency financial data. *Handbook of Financial Econometrics, Y Ait-Sahalia (ed)*, 2004.

Bruno Feunou and Nour Meddahi. Generalized affine models. *Available at SSRN 1367033*, 2009.

R. A. Fisher. Inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26:528–535, 10 1930. ISSN 1469-8064. doi: 10.1017/S0305004100016297. URL `http://journals.cambridge.org/article_S0305004100016297`.

Donald Alexander Stuart Fraser. Structural probability and a generalization. *Biometrika*, 53(1-2):1–9, 1966.

Donald AS Fraser. On fiducial inference. *The Annals of Mathematical Statistics*, pages 661–676, 1961a.

Donald AS Fraser. The fiducial method and invariance. *Biometrika*, pages 261–280, 1961b.

Donald AS Fraser. *The structure of inference*, volume 23. Wiley New York, 1968.

Jan Hannig. On generalized fiducial inference. *Statistica Sinica*, 19(2):491, 2009.

Jan Hannig. Generalized fiducial inference via discretization. *Statist. Sinica*, 23(2):489–514, 2013.

Jan Hannig and Min-ge Xie. A note on dempster-shafer recombination of confidence distributions. *Electronic Journal of Statistics*, 6:1943–1966, 2012.

Jan Hannig, Hari Iyer, and Paul Patterson. Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, 101(473):254–269, 2006.

Lawrence Harris. Statistical properties of the roll serial covariance bid/ask spread estimator. *The Journal of Finance*, 45(2):579–590, 1990.

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

Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

Steven L Heston and Saikat Nandi. A closed-form garch option valuation model. *Review of Financial Studies*, 13(3):585–625, 2000.

John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *The journal of finance*, 42(2):281–300, 1987.

Jean Jacod. 236(236), 1996.

Jean Jacod, Yingying Li, Per A Mykland, Mark Podolskij, and Mathias Vetter. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications*, 119(7):2249–2276, 2009.

Herb Johnson and David Shanno. Option pricing when the variance is changing. *Journal of Financial and Quantitative Analysis*, 22(02):143–151, 1987.

Yingying Li and Per A Mykland. Are volatility estimators robust with respect to modeling assumptions? *Bernoulli*, 13:601–622, 2007.

Yingying Li and Per A Mykland. Rounding errors and volatility estimation. *Journal of Financial Econometrics*, page nbu005, 2014.

Dennis V Lindley. Fiducial distributions and bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 102–107, 1958.

Robert C Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1):125–144, 1976.

Per A Mykland and Lan Zhang. Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5):1403–1445, 2009.

Mark Podolskij and Mathias Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658, 2009.

Richard Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4):1127–1139, 1984.

Louis O Scott. Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative analysis*, 22(04):419–438, 1987.

Lars Stentoft. American option pricing using garch models and the normal inverse gaussian distribution. *Journal of Financial Econometrics*, 6(4):540–582, 2008.

Kam-Wah Tsui and Samaradasa Weerahandi. Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 84(406):602–607, 1989.

Samaradasa Weerahandi. Generalized confidence intervals. *Journal of the American Statistical Association*, 88:899–905, 1993.

James B Wiggins. Option values under stochastic volatility: Theory and empirical estimates. *Journal of financial economics*, 19(2):351–372, 1987.

Dacheng Xiu. Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, 159(1):235–250, 2010.

SL Zabell. Ra fisher and fiducial argument. *Statistical Science*, pages 369–387, 1992.

Lan Zhang. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6):1019–1043, 2006.

Lan Zhang, Per A Mykland, and Yacine Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.

Michael Yuanjie Zhang, Jeffrey R Russell, and Ruey S Tsay. Determinants of bid and ask quotes and implications for the cost of trading. *Journal of Empirical Finance*, 15(4): 656–678, 2008.