

Quadratic Variation by Markov Chains

Peter Reinhard Hansen*
Stanford University
Department of Economics
Stanford, CA 94305, USA

Guillaume Horel
Merrill Lynch
New York

February 22, 2009

Abstract

We introduce a novel estimator of the quadratic variation that is based on the theory of Markov chains. The estimator is motivated by some general results concerning filtering contaminated semimartingales. Specifically, we show that filtering can *in principle* remove the effects of market microstructure noise in a general framework where little is assumed about the noise. For the practical implementation, we adopt the discrete Markov chain model that is well suited for the analysis of financial high-frequency prices. The Markov chain framework facilitates simple expressions and elegant analytical results. The proposed estimator is consistent with a Gaussian limit distribution and we study its properties in simulations and an empirical application.

Keywords: Markov chain; Filtering Contaminated Semimartingales; Quadratic Variation; Integrated Variance; Realized Variance; High Frequency Data.

JEL Classification: C10; C22; C80.

*Corresponding author, email: peter.hansen@stanford.edu. Peter Hansen is also affiliated with CREATES a center at the University of Aarhus, that is funded by the Danish National Research Foundation. The Ox language of Doornik (2001) and R were used to perform the calculations reported here. We thank seminar participants at CREATES, Oxford-Man Institute, University of Pennsylvania, and the 2008 SOFIE conference for valuable comments. We especially thank Asger Lunde for helping extract and clean the high-frequency data that are used in this paper.

1 Introduction

The advent of high-frequency financial data brought the scope for highly accurate measures of volatility over short periods of time, such as an hour or a day. The main obstacle in obtaining precise estimators has been the fact that high-frequency returns do not conform with conventional no-arbitrage models. The apparent contradiction can be explained by market microstructure noise, which give rise to the notion that the observed price is a noisy measure of the efficient price.

The leading example of high-frequency based measure of volatility is the realized variance, see e.g. Andersen, Bollerslev, Diebold & Labys (2001), Meddahi (2002) and Barndorff-Nielsen & Shephard (2002). The realized variance is simply the sum of squared returns which is sought to estimate the quadratic variation. A drawback of the realized variance is that it strongly relies on a semi-martingale assumption. While this assumption is an intrinsic feature of standard no-arbitrage models, it is also known to be at odds with the empirical properties of high-frequency prices. So the extent to which the realized variance can utilize high-frequency returns is limited. This has motivated a number of robust estimators, such as the two-scale estimator by Zhang, Mykland & Aït-Sahalia (2005) and the realized kernels by Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*a*). Common for these estimators is an assumption that market microstructure noise is iid, either at the tick-by-tick level or when sampling only includes every k -th price observation, for some k .

In this paper, we introduce a novel estimator that is build on the theory of Markov chains. The MC estimator has three distinct features. First, it utilizes the discreteness of high-frequency data. Second, the estimator permits a high degree of serial dependence in the noise as well as dependence between the efficient price and the noise. The latter is important because it allows us to compute the estimator with all available tick-by-tick data. There is no need to “skip” observations in order to meet specific assumptions about the noise. Third, simplicity is another characteristic of the Markov chain estimator. It only takes simple counting and basic matrix operations to compute the estimator and its confidence intervals.

To illustrate our estimator consider the sample of high frequency prices, X_{T_0}, \dots, X_{T_n} , where price increments, $\Delta X_{T_i} \in \{x_1, \dots, x_S\}$, are distributed as a homogeneous Markov

chain of order one. The $S \times S$ transition matrix, P , is here given by

$$P_{r,s} = \Pr(\Delta X_{T_{i+1}} = x_s | \Delta X_{T_i} = x_r), \quad r, s = 1, \dots, S,$$

and its stationary distribution, $\pi = (\pi_1, \dots, \pi_S)'$, is characterized by $\pi'P = \pi'$. We define $\Lambda_\pi = \text{diag}(\pi_1, \dots, \pi_S)$ and the *fundamental matrix* $Z = (I - P + \Pi)^{-1}$ where $\Pi = \iota\pi'$ with $\iota = (1, \dots, 1)' \in \mathbb{R}^S$. The Markov estimator is given by

$$\text{MC} = \frac{x' \Lambda_{\hat{\pi}} (2\hat{Z} - I)x}{n^{-2} \sum_{i=1}^n X_{T_i}^2},$$

where $\hat{\pi}$ and \hat{Z} are estimates of π and Z , respectively, and $x = (x_1, \dots, x_S)'$. Many properties of Markov chains can be linked to the fundamental matrix, Z , that also plays a central role in our analysis. A useful feature of our Markov framework is that standard errors for the MC estimator are readily available.

The main contribution of the present paper are as follows: First, we show that filtering can resolve the problems caused by market microstructure noise under weak assumptions that only require the noise process to be ergodic with finite first moment. A consequence is that the theory in Barndorff-Nielsen & Shephard (2002) applies to tick-by-tick returns of the filtered price process. This may sound too good to be true, and – from a practical viewpoint – it is. The reason is that the ideal filter requires more knowledge about the data generating process than is available in practice. This is where the Markov chain framework is effectual. Second, we derive expressions for the returns of filtered prices within a homogeneous Markov chain framework. This leads to a consistent estimator of the quadratic variation and we derive a feasible limit theory for the estimator. In fact, we show that confidence intervals are easy to compute, either from analytical expressions or by the use of bootstrap methods. Third, our estimator is derived within a homogeneous Markov framework. We show that a homogeneous model is justified when prices are generated by a continuous time Markov process, provided that certain conditions are met. More importantly, we show that the Markov chain estimator is robust to situations where the underlying process is inhomogeneous. Fourth, we apply the Markov chain framework to high-frequency data of an exchange traded fund that tracks the S&P 500 index.

The discreteness of financial date is a product of the so-called *tick size*, which defines the coarseness of the grid that prices are confined to. For example, the tick-size is currently 1 cent for most of the stocks that are listed on the New York Stock Exchange. The implication

is that all transaction and quoted prices are in whole cents. The Markov estimator can also be applied to time series that do not live on a grid, by forcing the process onto a grid. While this will introduce rounding error, it will not affect the long-run variance of the process. Delattre & Jacod (1997) studied the effect of rounding on realized variances for a standard Brownian motion, and Li & Mykland (2006) extended this analysis to log-normal diffusions.

In contrast to our framework, the existing literature has largely treated the discreteness of prices as a bad form of noise. An important exception is Large (2006) who was one of the first to take advantage of the discreteness of return data. He proposed an “alternation” estimator that is applicable when prices can only change by a fixed amount, e.g. 1 cent, and the noise has a particular form. We will show that the framework of Large (2006) is a special case of our Markov chain framework – specifically a Markov chain of order one with two states. The Markov framework used in this paper permits a larger number of price changes and is less restrictive in terms of assumptions made about the noise. Furthermore, the use of higher-order Markov chains will be shown to be essential for robustness to inhomogeneity.

The present paper adds to the a growing literature on volatility estimation using high-frequency data, dating back to Zhou (1996, 1998). Well known estimators include the realized variance, see Andersen, Bollerslev, Diebold & Labys (2001) and Barndorff-Nielsen & Shephard (2002); the two-scale and multi-scale estimators, see Zhang et al. (2005) and Zhang (2006); the realized kernels, see Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*a*, 2008*b*). The finite sample properties of these estimators are analyzed in Bandi & Russell (2006, 2008), and the close relation between multi-scale estimators and realized kernels is established in Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*d*). Other estimators include those based on moving average filtered returns, see Andersen, Bollerslev, Diebold & Ebens (2001), Maheu & McCurdy (2002), and Hansen, Large & Lunde (2008); the range-based estimator, see Christensen & Podolskij (2007); the pre-averaging estimator, see Jacod, Li, Mykland, Podolskij & Vetter (2008); the quantile-based estimator Christensen, Oomen & Podolskij (2008); and the duration-based estimator, see Andersen, Dobrev & Schaumburg (2008).

The stochastic properties of market microstructure noise are very important in this context. Estimators that are robust to iid noise can be adversely affected by dependent noise. Hansen & Lunde (2006) analyzed the empirical features of market microstructure noise and showed that serial dependence and endogenous noise are pronounced in high-

frequency stock prices. Endogenous noise refers to dependence between the noise and the efficient price. A major advantage of the Markov chain estimator is that dependent and endogenous noise is permitted in the framework. So estimation and inference can be done under a realistic set of assumptions about the noise.

The outline of this paper is as follows. We derive the generic results for filtering contaminated semimartingales in Section 2. We apply the Markov chain framework to filter high-frequency returns and derive our Markov chain based estimator in Section 3. The asymptotic properties of the estimator are established in Section 4, where we provide the details needed for conducting inference with the delta-method or the bootstrap. In Section 5, we show that the Markov chain framework is easily adapted for estimation of the volatility of log-prices, which is typically the object of interest. In Section 6 we consider the case where the underlying process is a continuous time Markov chain. This analysis gives an argument in favor of dropping the zero-increments from the analysis. We show that the Markov chain based estimator is robust to inhomogeneity in Section 7. In Section 8 we discuss various empirical issues related to jumps and computational aspects. Section 9 presents our empirical analysis and Section 10 various extensions, Section 11 concludes.

2 Filtering Contaminated Semimartingales

In this Section we analyze theoretical aspects of filtering observed prices in a general framework. The Section established the theoretical foundation for the Markov chain estimator that we introduce in the next Section. Readers who are primarily interested in aspects of the Markov chain estimator and its implementation can skip this Section. We show that the ideal (but infeasible) filter preserves the key features of the latent efficient price. This is true under very mild assumptions on the efficient price and the noise.

Suppose that Y_t is a semimartingale, so that

$$Y_t = M_t + FV_t,$$

where M_t is a local martingale and FV_t is a process that has finite variation almost surely.

We denote the observed process by X_t , and denote the difference between X_t and the latent process, Y_t , by U_t , so that

$$X_t = Y_t + U_t.$$

Market imperfections, rounding errors, and data errors are some of the sources for the measurement error, U_t , which we will refer to as market microstructure noise, or simply noise.

Let \mathcal{G}_t be some filtration so that (Y_t, U_t) is adapted to \mathcal{G}_t , and consider the filtered process

$$E(X_{t+h}|\mathcal{G}_t) = E(M_{t+h}|\mathcal{G}_t) + E(FV_{t+h}|\mathcal{G}_t) + E(U_{t+h}|\mathcal{G}_t).$$

An objective of this Section is to filter out U_t , so that the properties of Y_t can be inferred from those of the filtered X_t .

We make the following assumption:

Assumption 1 *The filtration \mathcal{G}_t is continuous¹ and for $Y_t = M_t + FV_t$ we assume:*

- (i) $\{M_t, \mathcal{G}_t\}$ is a martingale with finite quadratic variation;
- (ii) FV_t is continuous with locally integrable variation; and
- (iii) $\tilde{U}_t = \lim_{h \rightarrow \infty} E(U_{t+h}|\mathcal{G}_t)$ is a continuous finite variation process almost surely.

Assumption 1 is quite mild. For instance, (ii) holds if Y_t is a *quasimartingale*. Furthermore, (iii) holds if $E(U_{t+h}|\mathcal{G}_t) \xrightarrow{L^1} \mu$ as $h \rightarrow \infty$, for some $\mu \in \mathbb{R}$, which in turn is implied by the uniform mixing condition, formulated in the following Lemma.

Lemma 1 *If U_t is stationary with $E|U_t| < \infty$ and ϕ -mixing with respect to \mathcal{G}_t , that is*

$$\phi(m) = \sup\{|P(A|B) - P(B)| : A \in \sigma(U_{t+s}, s \geq m), B \in \mathcal{G}_t\} \rightarrow 0, \quad \text{as } m \rightarrow \infty.$$

Then $E(U_{t+h}|\mathcal{G}_t) \xrightarrow{L^1} E(U_t)$ as $h \rightarrow \infty$.

Note that we assume that U_t is ϕ -mixing with respect to \mathcal{G}_t , which is larger than the natural filtration for U_t , $\sigma(\{U_s\}, s \leq t)$.

When Y_t is a Brownian semimartingale, written $Y_t \in \mathcal{BSM}$, we have

$$M_t = \int_0^t \sigma_u dB_u.$$

More generally we have $M_t = \int_0^t \sigma_u dB_u + \sum_{s \leq t} J_s$ where J_s is a pure jump component.

¹A filtration is continuous if $\mathcal{G}_{t-} = \mathcal{G}_t = \mathcal{G}_{t+}$, where $\mathcal{G}_{t-} = \sigma(\cup \mathcal{G}_s, s < t)$ and $\mathcal{G}_{t+} = \sigma(\cap_{s > t} \mathcal{G}_s)$. This assumption is only used to show that $\lim_{h \rightarrow \infty} E(FV_{t+h}|\mathcal{G}_t)$ is continuous.

Theorem 1 *Given Assumption 1, then*

$$\lim_{h \rightarrow \infty} \mathbb{E}(X_{t+h} | \mathcal{G}_t) = M_t + \text{FV}_t^*,$$

where FV_t^* is a continuous finite variation process.

The proof is based on the decomposition, $\mathbb{E}(X_{t+h} | \mathcal{G}_t) = \mathbb{E}(M_{t+h} | \mathcal{G}_t) + \mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_t) + \mathbb{E}(U_{t+h} | \mathcal{G}_t)$, where the only obstacle is to show that $\tilde{\text{FV}}_t = \mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_t)$ is a finite variation process.

2.1 Filtered Estimator of Quadratic Variation

Let

$$0 = T_0 < T_1 < \dots < T_n = T,$$

be the times where X_t is observed, and we will establish asymptotic results using an in-fill asymptotic scheme, where $\sup_{1 \leq i \leq n} |T_i - T_{i-1}| \rightarrow \infty$, as $n \rightarrow \infty$. This asymptotic design is standard in this literature.

First we note that in the absence of noise, $U_t = 0$, there is no need for filtering because $X_t = Y_t$. So if we take $h = 0$, and have $\mathbb{E}(X_{T_i} | \mathcal{G}_{T_i}) = Y_{T_i}$. Next we consider the situation where noise is present.

2.1.1 A Special Case

Consider the special case where $\mathbb{E}(U_{T_{i+1}} | \mathcal{G}_{T_i}) = 0$. This assumption is far more restrictive than Assumption 1 (iii), yet weaker than the assumption that $\{U_{T_i}\}$ is iid, which is often used in this literature. Our framework does not rely on this restrictive form of noise, but the simplified framework offers valuable intuition about the general situation.

With $\mathbb{E}(U_{T_{i+1}} | \mathcal{G}_{T_i}) = 0$ it follows by Theorem 1 that the filtered price process, $\mathbb{E}(X_{T_{i+1}} | \mathcal{G}_{T_i})$, only differs from Y_{T_i} by a finite variation process. So the one-step filtered realized variance

$$\text{RV}_F^{(1)} = \sum_{i=1}^n \left\{ \mathbb{E}(X_{T_{i+1}} | \mathcal{G}_{T_i}) - \mathbb{E}(X_{T_i} | \mathcal{G}_{T_{i-1}}) \right\}^2,$$

is consistent for the quadratic variation, and we recover the same asymptotic framework as that in Barndorff-Nielsen & Shephard (2002) (henceforth BNS). We see that filtering entirely removes the unfortunate features of noise.

Note that $\text{RV}_F^{(1)}$ depends on the conditional expectation, $\text{E}(X_{T_{i+1}}|\mathcal{G}_{T_i})$, which is unknown in most practical situations. So the filtered realized variance is, in this sense, not an estimator.

Naturally we have $\text{E}(X_{T_{i+1}}|\mathcal{G}_{T_i}) = X_{T_i} + \text{E}(\Delta X_{T_{i+1}}|\mathcal{G}_{T_i})$, where $\Delta X_{T_{i+1}} = X_{T_{i+1}} - X_{T_i}$. So returns of the filtered prices can be expressed as

$$\text{E}(X_{T_{i+1}}|\mathcal{G}_{T_i}) - \text{E}(X_{T_i}|\mathcal{G}_{T_{i-1}}) = \Delta X_{T_i} + \text{E}(\Delta X_{T_{i+1}}|\mathcal{G}_{T_i}) - \text{E}(\Delta X_{T_i}|\mathcal{G}_{T_{i-1}}).$$

The implication is that the filtered realized variance, can be rewritten in the more instructive form:

$$\text{RV}_F^{(1)} = \sum_{i=1}^n \left\{ \Delta X_{T_i} + \text{E}(\Delta X_{T_{i+1}}|\mathcal{G}_{T_i}) - \text{E}(\Delta X_{T_i}|\mathcal{G}_{T_{i-1}}) \right\}^2. \quad (1)$$

This expression reveals how the filtering operates on the observed intraday returns. We see that the estimator corrects each increment by adding and subtracting anticipated changes in ΔX_{T_i} and $\Delta X_{T_{i+1}}$.

Comment. An important insights from our analysis is that one should use returns of the filtered price, $\Delta \text{E}(X_{T_{i+1}}|\mathcal{G}_{T_i})$ rather than filtered returns, $\text{E}(\Delta X_{T_{i+1}}|\mathcal{G}_{T_i})$, when computing the realized variance. The reason is that the sum of squares of $\text{E}(\Delta X_{T_{i+1}}|\mathcal{G}_{T_i}) = \Delta Y_{T_i} - U_{T_i}$, will not estimate the quadratic variation of Y_t . This is also evident from the fact that $\sum_{i=1}^n \{\text{E}(\Delta X_{T_{i+1}}|\mathcal{G}_{T_i})\}^2$ will only simplify to the expression in (1) if $\Delta X_{T_i} - \text{E}(\Delta X_{T_i}|\mathcal{G}_{T_{i-1}}) = 0$, which requires $U_{T_i} = \text{E}(U_{T_i}|\mathcal{G}_{T_{i-1}})$.

This results generalized a result in Hansen et al. (2008). For the special case where returns are filtered by a moving average model, Hansen et al. (2008) have shown that the sum of squared filtered returns do not estimate the object of interest. Consider the case where the noise is iid and the volatility is constant, so that intraday returns follow a moving average process of order one. The sum-of-squared residuals, $\sum_{i=1}^n \hat{\varepsilon}_i^2$, obtained from estimating $\Delta X_{T_i} = \varepsilon_i - \theta \varepsilon_{i-1}$, is not consistent for the volatility. The proper estimator in this framework is $(1 - \hat{\theta})^2 \sum_{i=1}^n \hat{\varepsilon}_i^2$, see Hansen et al. (2008).

2.1.2 The General Case

When we filter $X_{T_{i+h}}$ by \mathcal{G}_{T_i} we obtain the h -steps filtered realized variance,

$$\text{RV}_F^{(h)} = \sum_{i=1}^n \left\{ \text{E}(X_{T_{i+h}}|\mathcal{G}_{T_i}) - \text{E}(X_{T_{i+h-1}}|\mathcal{G}_{T_{i-1}}) \right\}^2,$$

and analogous to (1) we rewrite this expression as

$$\text{RV}_F^{(h)} = \sum_{j=1}^n \left\{ \Delta X_{T_j} + \sum_{l=1}^h \text{E}(\Delta X_{T_{j+l}} | \mathcal{G}_{T_j}) - \sum_{l=1}^h \text{E}(\Delta X_{T_{j+l-1}} | \mathcal{G}_{T_{j-1}}) \right\}^2. \quad (2)$$

This defines the class of h -steps filtered RV, and it is natural to ask which h should we use? The increments depend on h only through an expectation, so we need not observe $X_{T_{i+h}}$ in order to compute $\text{E}(X_{T_{i+h}} | \mathcal{G}_{T_i})$. So there is no obstacle in using a large h . In fact, we can take $h = \infty$ which offers robustness to the most general form of noise. In our Markov chain implementation we will use $h = \infty$.

2.2 Feasible Filtration

Before the quantity $\text{RV}_F = \lim_{h \rightarrow \infty} \text{RV}_F^{(h)}$ can be put into practical use, we need to specify $\text{E}(\cdot | \mathcal{G}_{T_i})$. In practice we must use a filtration based on observables, such as

$$\mathcal{F}_t = \sigma(X_s, s \leq t),$$

and the Markov chain framework is well suited for computing conditional expectations such as $\text{E}(X_{T_{i+h}} | \mathcal{F}_{T_i})$. However, substituting \mathcal{F} for \mathcal{G} is not innocuous, as shown by Li & Mykland (2007). Specifically we have

$$\text{E}(X_{T_{i+h}} | \mathcal{F}_{T_i}) = \text{E} \left\{ \text{E}(X_{T_{i+h}} | \mathcal{G}_{T_i}) | \mathcal{F}_{T_i} \right\} = \text{E} \left\{ Y_{T_i} + \text{E}(U_{T_{i+h}} | \mathcal{G}_{T_i}) | \mathcal{F}_{T_i} \right\},$$

which converges to $\text{E}(Y_{T_i} | \mathcal{F}_{T_i})$ as $h \rightarrow \infty$, and the quadratic variation of Y_t need not equal that of $\text{E}(Y_t | \mathcal{F}_t)$, see Li & Mykland (2007).

The potentially harmful difference between Y_t and $\text{E}(Y_t | \mathcal{F}_t)$, can be expressed as

$$\text{E}(Y_{T_i} | \mathcal{F}_{T_i}) = \text{E}(X_{T_i} - U_{T_i} | \mathcal{F}_{T_i}) = Y_{T_i} + U_{T_i} - \text{E}(U_{T_i} | \mathcal{F}_{T_i}).$$

So the relevant question is whether $U_{T_i} - \text{E}(U_{T_i} | \mathcal{F}_{T_i})$ contributes to the quadratic variation.

In our analysis we make substantially weaker assumption about U_t than do Li & Mykland (2007). So our discussion in this Section shows that the results by Li & Mykland (2007) hold quite generally. Though the results in Li & Mykland (2007) concern smoothen returns, $\text{E}(X_{T_{i+h}} | \mathcal{F}_{T_n})$, contrary to the filtered returns discussed here. We believe that the focus on $U_{T_i} - \text{E}(U_{T_i} | \mathcal{F}_{T_i})$ is a useful way to think about this issue.

3 Markov Chain Estimator

In this Section we show how the observed price process can be filtered in a Markov chain framework, using the natural filtration for $\{X_t\}$, $\mathcal{F}_t = \sigma(X_s, s \leq t)$. The realized variance of the Markov chain filtered prices defines our novel estimator of the quadratic variation, and we establish the asymptotic properties of the estimator. The estimator takes advantage of the fact that price increments are confined to a grid. We will initially assume that the observed price increments follow a homogeneous Markov chain, and later show that our estimator is robust to inhomogeneity. The Markov chain framework used in this Section is related to that in Russell & Engle (2006), see also Campbell, Lo & Mackinlay (1997, pp.107–147), but our objective and analysis are entirely different.

We seek the filtered price, $E(X_{T_{i+h}}|\mathcal{F}_{T_i})$, and preferably the filtered price as $h \rightarrow \infty$. Fortunately, (2) shows that the returns of the filtered price can be constructed from filtered increments, $E(\Delta X_{T_{i+h}}|\mathcal{F}_{T_i})$. The Markov chain framework conveniently enables us to compute $E(\Delta X_{T_{i+h}}|\mathcal{F}_{T_i})$, for any h , in as simple way.

Quoted and traded prices are typically confined to be on a grid. For instance on the New York Stock Exchange the typical *tick size* is currently 1 cent, which implies that prices move in multiples of 1 cents.² Naturally, we can always force the price data to live on a grid at the expense of rounding error. We make the following assumptions.

Assumption 2 *The increments $\{\Delta X_{T_{i+1}}\}_{i=1}^n$ are ergodic and distributed as a homogeneous Markov chain of order $k < \infty$, with $S < \infty$ states.*

The homogeneity is a restrictive assumption and likely to be at odds with high-frequency returns over longer intervals of time, such as a day. In Section 6 we analyze the case where the underlying process is a continuous time Markov chain, and formulate assumptions that will generate returns that are distributed as a homogeneous Markov chain in discrete time. More importantly, in Section 7 we show that the Markov chain estimator is surprisingly robust to inhomogeneity. Robustness is achieved by increasing the order of the Markov chain that is being estimated.

²Some stocks are now quoted and traded at prices that are multiples of half a cent. In the foreign exchange market the terminology is different, as exchange rates are quoted in *pips*. The Euro-Dollar is currently trading with a pip size of 1/100 of a cent.

3.1 Notation

The transition matrix for price increments is denoted by P . For a Markov chain of order k with S basic states, P will be an $S^k \times S^k$ matrix. We use π to denote the stationary distribution associated with P , i.e. $\pi'P = \pi'$. The *fundamental matrix* is defined by

$$Z = (I - P + \Pi)^{-1},$$

where $\Pi = \iota\pi'$ is a square matrix and $\iota = (1, \dots, 1)'$, (so all rows of Π is π'). For a vector $b \in \mathbb{R}^m$, we let Λ_b denote the diagonal matrix

$$\Lambda_b = \text{diag}(b_1, \dots, b_m),$$

and for vectors, a and b , of proper dimensions we define the inner product

$$\langle a, b \rangle_\pi = a' \Lambda_\pi b.$$

We use $(b)_r$ to denote the r -th element of the vector b , and write $\delta_{i,j} = 1_{\{i=j\}}$ where $1_{\{\cdot\}}$ is the indicator function.

3.2 Markov Chain Filtering

Let $\{x_1, \dots, x_S\}$ be the support for ΔX_{T_i} , and suppose that ΔX_{T_i} is distributed as a Markov chain of order k . We consider the k -tuple, $\Delta \mathcal{X}_{T_i} = (\Delta X_{T_{i-k+1}}, \dots, \Delta X_{T_i})$, and index the possible values for $\Delta \mathcal{X}_{T_i}$ by \mathbf{x}_s , $s = 1, \dots, S^k$, where $\mathbf{x}_s \in \{x_1, \dots, x_S\}^k \subset \mathbb{R}^k$. The transition matrix, P , is given by

$$P_{r,s} = \Pr(\Delta \mathcal{X}_{T_{i+1}} = \mathbf{x}_s | \Delta \mathcal{X}_{T_i} = \mathbf{x}_r).$$

Regardless of the order of the Markov chain, there are at most S possible transitions from any given state \mathbf{x}_r . So P will have many zeros when $k > 1$, and specifically we have

$$P_{r,s} = \begin{cases} \Pr(\Delta X_{T_{i+1}} = x_{s_1} | \Delta \mathcal{X}_{T_{i-1}} = \mathbf{x}_r) & \text{if } (x_{s_1}, \dots, x_{s_{k-1}}) = (x_{r_2}, \dots, x_{r_k}), \\ 0 & \text{if } (x_{s_1}, \dots, x_{s_{k-1}}) \neq (x_{r_2}, \dots, x_{r_k}). \end{cases}$$

We use the vector $f \in \mathbb{R}^{S^k}$ to keep track of the value of ΔX_{T_i} , as we let f_s be the last element of \mathbf{x}_s , $s = 1, \dots, S^k$. For a particular realization of $\Delta \mathcal{X}_{T_i}$, the conditional expectation of $\Delta X_{T_{i+1}}$ can be expressed as

$$\mathbb{E}(\Delta X_{T_{i+1}} | \Delta \mathcal{X}_{T_i} = \mathbf{x}_r) = \sum_{s=1}^{S^k} P_{r,s} f_s = (Pf)_r.$$

More generally we have $E(\Delta X_{T_i+h} | \Delta \mathcal{X}_{T_i} = \mathbf{x}_r) = (P^h f)_r$. So the return of the h -steps filtered price is given by

$$y^{(h)}(\Delta \mathcal{X}_{T_{i-1}}, \Delta \mathcal{X}_{T_i}) = \Delta X_{T_i} + \sum_{l=1}^h E(\Delta X_{T_{i+l}} | \Delta \mathcal{X}_{T_i}) - \sum_{l=1}^h E(\Delta X_{T_{i+l-1}} | \Delta \mathcal{X}_{T_{i-1}}),$$

where the conditional expectation is $E(\Delta X_{T_{i+h}} | \Delta \mathcal{X}_{T_i}) = \sum_{r=1}^S (P^h f)_r 1_{\{\Delta \mathcal{X}_{T_i} = \mathbf{x}_r\}}$. The contribution to RV_F , when $(\Delta \mathcal{X}_{T_{i-1}}, \Delta \mathcal{X}_{T_i}) = (\mathbf{x}_r, \mathbf{x}_s)$, is simply given by $y^{(h)}(\mathbf{x}_r, \mathbf{x}_s)^2$. So we obtain the following expression for Markov filtered realized variance:

$$\text{RV}_F^{(h)} = \sum_{r,s} n_{r,s} \{y^{(h)}(\mathbf{x}_r, \mathbf{x}_s)\}^2, \quad \text{where} \quad n_{r,s} = \sum_{i=1}^n 1_{\{\Delta \mathcal{X}_{T_{i-1}} = \mathbf{x}_r, \Delta \mathcal{X}_{T_i} = \mathbf{x}_s\}}.$$

We have the following expressions for the filtered returns.

Lemma 2 *Let e_r denote the r -th unit vector. Then*

$$y^{(h)}(\mathbf{x}_r, \mathbf{x}_s) = e_r'(I - Z^{(h)})f + e_s'Z^{(h)}f,$$

with $Z^{(h)} = I + \sum_{l=1}^h (P^l - \Pi)$ and

$$y(\mathbf{x}_r, \mathbf{x}_s) = \lim_{h \rightarrow \infty} y^{(h)}(\mathbf{x}_r, \mathbf{x}_s) = e_r'(I - Z)f + e_s'Zf,$$

where $Z = \lim_{h \rightarrow \infty} Z^{(h)} = (I - P + \Pi)^{-1}$.

Lemma 2 shows that it is straightforward to use $h = \infty$. This is convenient because $h = \infty$ offers the greatest degree of robustness to dependent noise. We will primarily focus on this case and the corresponding filtered realized variance is given by

$$\text{RV}_F = \sum_{r,s} n_{r,s} y_{(r,s)}^2, \quad \text{with } y_{(r,s)} = y(\mathbf{x}_r, \mathbf{x}_s).$$

Conditional expectations, such as those that are used in the expression for $y^{(h)}(\mathbf{x}_r, \mathbf{x}_s)$, cannot be evaluated without knowledge of P . Consequently, RV_F is not an empirical quantity in the realistic situation where P is unknown. The empirical transition matrix \hat{P} is given by $\hat{P}_{r,s} = n_{r,s}/n_{r,\cdot}$, where $n_{r,\cdot} = \sum_s n_{r,s}$, and by substituting \hat{P} for P we obtain the feasible estimator

$$\text{RV}_{\hat{F}} = \sum_{r,s} n_{r,s} \hat{y}_{(r,s)}^2, \quad \text{with } \hat{y}_{(r,s)} = e_r'(I - \hat{Z})f + e_s'\hat{Z}f, \quad (3)$$

and $\hat{Z} = (I - \hat{P} + \hat{\Pi})^{-1}$. We derive the asymptotic properties of RV_F and $\text{RV}_{\hat{F}}$ in the next Section and show that the latter is essentially equal to $\text{MC}^\# = n \langle f, (2\hat{Z} - I)f \rangle_{\hat{\pi}}$.

4 Asymptotic Analysis

In the previous Section we established a feasible filtering of high frequency prices with Markov chain methods. This lead to the quantities, RV_F , $RV_{\hat{F}}$, and (as we shall see) $MC^\# = n\langle f, (2\hat{Z} - I)f \rangle_{\hat{\pi}}$. Now we now seek the asymptotic properties of these quantities.

We shall derive limit results using the following asymptotic scheme, which is similar to the one in Delattre & Jacod (1997) and Li & Mykland (2006).

Assumption 3 (asymptotic design)

$$\text{As } n \rightarrow \infty \text{ we have: } f = \frac{1}{\sqrt{n}}\xi, \text{ where } \xi \in \mathbb{R}^S \text{ is held constant.} \quad (4)$$

It may seem odd to fiddle with the state-space as $n \rightarrow \infty$. Yet, the assumption is quite natural in the present context with infill asymptotics, and analogous to local-to-unity and local-to-zero asymptotics that are commonly seen in the literature. The asymptotic design in (4) can be motivated by the following scenario. Suppose that price changes of 10 cents in the stock price are common if we sample 100 times per day. Whereas if we sample 3,000 times per day, the typical size of a price change will be around 2 cents. To reconcile the two asymptotic theories, we need to assume that the state-values shrink towards zero with the sampling frequency. Furthermore, (4) makes our asymptotic analysis compatible with the standard BNS framework. In the BNS framework, squared intraday returns are $\Delta Y_{T_i}^2 = O_p(n^{-1})$ and the corresponding situation in the Markov framework in the absence of noise (absence of autocorrelation ΔY_{T_i}) is that where $P = \iota\pi'$. So in this Markov framework we have

$$E(\Delta Y_{T_i}^2) = \sum_{s=1}^{S^k} \pi_s x_s^2 = x' \Lambda_\pi x = \langle x, x \rangle_\pi,$$

and in order for the $RV = \sum_{i=1}^n \Delta Y_{T_i}^2$ not to diverge to infinity as $n \rightarrow \infty$, we need $x' \Lambda_\pi x \propto n^{-1}$. This is achieved by (4). Alternatively one could fiddle with P , so that the stationary distribution becomes concentrated about zero as $n \rightarrow \infty$. However, this approach seems more complicated because P_n would be specific to the state vector x , and will require one to introduce additional states as $n \rightarrow \infty$ – states that were not observed in the sample.

4.1 Asymptotic Properties of Filtered Realized Variance

First, we derive the asymptotic properties of the infeasible filtered realized variance. This will form the basis for the analysis of the feasible estimator.

Theorem 2 (Consistency of RV_F) For the Markov filtered realized variance we have

$$\text{RV}_F \xrightarrow{a.s.} \langle \xi, (2Z - I)\xi \rangle_\pi, \quad \text{as } n \rightarrow \infty.$$

where $Z = (I - P + \Pi)^{-1}$.

We now turn to the limit distribution of RV_F . A classical result within the theory of Markov chains is the following.

Proposition 1 Let $\{V_i\}$ be an ergodic Markov chain with states $\{v_1, \dots, v_S\}$, and let P_V and π_V denote the transition matrix and stationary distribution, respectively. For any real function, $\mathbf{g}(\cdot)$, we have

$$\sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{g}(V_i) - \langle g, \pi_V \rangle \right\} \xrightarrow{d} \text{N} \left\{ 0, \langle g, (2Z_V - \Pi_V - I)g \rangle_{\pi_V} \right\},$$

where $g = (\mathbf{g}(v_1), \dots, \mathbf{g}(v_S))'$, $Z_V = (I - P_V + \Pi_V)^{-1}$, and $\Pi_V = \iota \pi_V'$.

The central limit theorem for ergodic Markov chains is valid quite generally, see e.g. Duflo (1997, theorem 8.3.21) for the case with positive recurrent Markov chains on measurable spaces. In its most general formulation, the variance of the limiting normal distribution is given implicitly from a solution to a Poisson equation. For finite spaces, such as the present one, most of the formulas can be elegantly written in closed form in terms of the *fundamental matrix*, Z , of Kemeny & Snell (1976). In this paper we follow the presentation of Brémaud (1999, chapter 6).

An interesting observation that can be made from Proposition 1 is that our estimator, RV_F is related to the long-run variance. This is evident from

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \sqrt{n} \Delta X_{T_i} - \langle \xi, \pi \rangle \right) \xrightarrow{d} \text{N} \left\{ 0, \langle \xi, (2Z - \Pi - I)\xi \rangle_\pi \right\},$$

where $\langle \xi, (2Z - \Pi - I)\xi \rangle_\pi = \langle \xi, (2Z - I)\xi \rangle_\pi - \langle \xi, \Pi\xi \rangle_\pi = \langle \xi, (2Z - I)\xi \rangle_\pi - (\xi' \pi)^2$. So in the absence of a drift, $\xi' \pi = 0$, the long-run variance coincides with the probability limit of RV_F , see Theorem 2. This is a common feature of estimators in this literature.

We will now apply Proposition 1 to obtain the asymptotic distribution of the filtered realized variance, $\text{RV}_F = \sum_{i=1}^n \{y(\Delta \mathcal{X}_{T_{i-1}}, \Delta \mathcal{X}_{T_i})\}^2$. First we note that $\{y(\Delta \mathcal{X}_{T_{i-1}}, \Delta \mathcal{X}_{T_i})\}^2$ only depends on the $k + 1$ -tuple, $(\Delta X_{T-k}, \dots, \Delta X_{T_i})'$. So we have

$$\text{RV}_F = \sum_{i=1}^n \{y(\Delta \mathcal{X}_{T_{i-1}}, \Delta \mathcal{X}_{T_i})\}^2 = \sum_{i=1}^n \mathbf{g}(\Delta X_{T-k}, \dots, \Delta X_{T_i}),$$

for some real function \mathbf{g} . In our asymptotic framework, $\{n^{1/2}\Delta X_{T_i}\}$ is (for any fixed n) an ergodic Markov chain with state-values given by the vector ξ . So it follows that $\mathcal{Z}_{T_i} = (n^{1/2}\Delta X_{T-k}, \dots, n^{1/2}\Delta X_{T_i})'$ is also an ergodic Markov chain, and we denote the corresponding $S^{k+1} \times S^{k+1}$ transition matrix $P_{\mathcal{Z}}$, and let $\pi_{\mathcal{Z}}$ denote the corresponding stationary distribution. Next we note that

$$\mathbf{g}(\Delta X_{T-k}, \dots, \Delta X_{T_i}) = \frac{1}{n} \mathbf{g}(n^{1/2}\Delta X_{T-k}, \dots, n^{1/2}\Delta X_{T_i}) = \frac{1}{n} \mathbf{g}(\mathcal{Z}_{T_i}).$$

So we can apply Proposition 1 to $\text{RV}_F = n^{-1} \sum_{i=1}^n \mathbf{g}(\mathcal{Z}_{T_i})$, which yields the expression for the asymptotic variance,

$$\Sigma_{\text{RV}_F} = \langle g, (2Z_{\mathcal{Z}} - \Pi_{\mathcal{Z}} - I)g \rangle_{\pi_{\mathcal{Z}}},$$

where $Z_{\mathcal{Z}}$ is the fundamental matrix associates with $P_{\mathcal{Z}}$ and $\Pi_{\mathcal{Z}} = \iota \pi_{\mathcal{Z}}$. The vector $g \in \mathbb{R}^{S^{k+1}}$ is here defined by $g_s = \mathbf{g}(\mathbf{z}_s)$, for $s = 1, \dots, S^{k+1}$ where \mathbf{z}_s is the state value of \mathcal{Z}_{T_i} that corresponds to the s -th row of $P_{\mathcal{Z}}$. So g contains all the possible values of the squared filtered returns, $\{y(n^{1/2}\Delta \mathcal{X}_{T_{i-1}}, n^{1/2}\Delta \mathcal{X}_{T_i})\}^2 = n \{y(\Delta \mathcal{X}_{T_{i-1}}, \Delta \mathcal{X}_{T_i})\}^2$. In Theorem 3 we derive a more transparent expression for Σ_{RV_F} .

Because the underlying process for \mathcal{Z}_{T_i} is a Markov chain of order k , its transition matrix, $P_{\mathcal{Z}}$, has a particular structure, where each element of P appears S times. Naturally we have $\mathcal{Z}_{T_i} = (\Delta X_{T-k}, \Delta \mathcal{X}_{T_i})' = (\Delta \mathcal{X}_{T-1}, \Delta X_{T_i})'$, so for $\mathbf{z} = (\mathbf{x}_r, x_s)$ and $\tilde{\mathbf{z}} = (x_j, \mathbf{x}_r)$, we have

$$\Pr(\mathcal{Z}_{T_{i+1}} = \mathbf{z} | \mathcal{Z}_{T_i} = \tilde{\mathbf{z}}) = \Pr(\Delta X_{T_{i+1}} = x_s | \Delta \mathcal{X}_{T_i} = \mathbf{x}_r).$$

This is true for any $x_j \in \{x_1, \dots, x_S\}$ which simply reflects that the underlying process is a Markov chain of order k , so that the $(k+1)$ -th lagged value, ΔX_{T_i-k} , is redundant for the conditional probability.

Theorem 3 (Limit distribution for RV_F) *For the infeasible realized variance, RV_F , we have*

$$\sqrt{n} \{ \text{RV}_F - \langle \xi, (2Z - I) \xi \rangle_{\pi} \} \xrightarrow{d} \text{N} \{ 0, \Sigma_{\text{RV}_F} \},$$

where $\Sigma_{\text{RV}_F} = \sum_{r,s,u,v} \Sigma_{n_r,s,n_u,v} \{e'_r(I-Z)\xi + e'_s Z \xi\}^2 \{e'_u(I-Z)\xi + e'_v Z \xi\}^2$, with $\Sigma_{n_r,s,n_u,v} = P_{r,s} P_{u,v} (\pi_r Z_{s,u} + \pi_u Z_{v,r} - 3\pi_r \pi_u) + \pi_r P_{r,s} \delta_{r,u} \delta_{s,v}$.

Comment. Theorem 3 is stated for a homogeneous Markov chain with fixed parameters. However, in relation to Section 2 the exact specification of the Markov chain is define ex-post to the realization of $\{Y_t, U_t\}$, so both the probability limit, $\langle \xi, (2Z - I) \xi \rangle_{\pi}$, and the

asymptotic variance Σ_{RV_F} , may be thought of as random quantities, as is common in this literature.

4.2 Asymptotic Properties of Markov Chain Estimator

So far we have established results for RV_F that cannot be computed in practice because it depends on the unknown transition matrix, P . Substituting an estimate \hat{P} for P leads to the feasible estimator,

$$RV_{\hat{F}} = \sum_{r,s} n_{r,s} \hat{y}_{(r,s)}^2, \quad \text{where } \hat{y}_{(r,s)} = e'_r(I - \hat{Z})f + e'_s \hat{Z}f.$$

We will derive the asymptotic properties of this estimator and another estimator defined by,

$$MC^\# = \xi' \Lambda_{\hat{\pi}} (2\hat{Z} - I)\xi.$$

We use the #-superscript to indicate that $MC^\#$ is a volatility estimator for prices (in levels) that live on a grid. The notation MC will be reserved for our estimator of the volatility for log-prices. Next, we derive some intermediate results that serve two purposes. First to establish that $RV_{\hat{F}}$ and $MC^\#$ are asymptotically equivalent; and second, to derive the limit distribution of $MC^\#$ (and hence that of $RV_{\hat{F}}$).

4.2.1 Asymptotic Distribution by Delta Method

Since $MC^\# = \langle \xi, (2Z - I)\xi \rangle_\pi$ is a differentiable function of P alone, we can obtain a closed form expression for the asymptotic variance by the delta method. We will need the derivative of $\langle \xi, (2Z - I)\xi \rangle_\pi$ with respect to P , and the asymptotic distribution for \hat{P} . The latter is well known and stated in the following Proposition.

Proposition 2 *Given Assumptions 2 and 3, the asymptotic distribution \hat{P} is given by*

$$\sqrt{n}(\hat{P} - P) \xrightarrow{d} N(0, \Sigma_P),$$

where $(\Sigma_P)_{(r,s)(u,v)} = 1_{\{r=u\}} \frac{1}{\pi_r} \{(\Lambda_{P_{r,\cdot}} - P'_{r,\cdot} P_{r,\cdot})\}_{s,v}$.

The analogy with the covariance of the maximum likelihood estimator of a multinomial distribution is obvious, with $\pi_s n$ playing the role of the number of trials.

Next, we derive the differential that is needed for the delta method, and then the limit distribution of $MC^\#$.

Lemma 3 Let $\mu = \langle \xi, \pi \rangle$. The differential of $\langle \xi, (2Z - I)\xi \rangle_\pi$ is given by

$$\left(\frac{\partial \langle \xi, (2Z - I)\xi \rangle_\pi}{\partial P} \right)' = Z \{ \Lambda_\xi (I + P - \Pi) - 2\mu I \} Z \xi \pi' + 2Z \xi \xi' \Lambda_\pi Z.$$

The differential in Lemma 3 is not a partial derivative in the usual sense, due to the constraints on the transition matrix P . In particular, we have $\sum_s dP_{r,s} = 0$ for all r .

Theorem 4 Let $\text{MC}^\# = \langle \xi, (2\hat{Z} - I)\xi \rangle_{\hat{\pi}}$ then

$$\sqrt{n} \left\{ \text{MC}^\# - \langle \xi, (2Z - I)\xi \rangle_\pi \right\} \xrightarrow{d} \text{N}(0, \Sigma_{\text{MC}}),$$

where

$$\Sigma_{\text{MC}} = \left\{ \frac{\partial \langle \xi, (2Z - I)\xi \rangle_\pi}{\partial P} \right\}' \Sigma_P \frac{\partial \langle \xi, (2Z - I)\xi \rangle_\pi}{\partial P}.$$

with Σ_P defined as in Proposition 2.

Using Lemma 3 and the expression for Σ_P given in Proposition 2, we can now write the variance Σ_{MC} in the following way.

Corollary 1 The asymptotic variance of $\text{MC}^\#$ can be expressed as,

$$\Sigma_{\text{MC}} = \sum_r \pi_r u_r' V_r u_r, \quad (5)$$

with $V_r = \Lambda_{P_{r,\cdot}} - P_{r,\cdot}' P_{r,\cdot}$, and $u_r = Z \{ \Lambda_\xi (I + P - \Pi) - 2\mu I \} Z \xi + 2Z \xi \xi' \Lambda_\pi Z \Lambda_\pi^{-1} e_r$.

The following result shows that the two estimators, $\text{RV}_{\hat{F}}$ and $\text{MC}^\#$, are asymptotically equivalent to first order. This is convenient because it shows that our asymptotic results for $\text{MC}^\#$ apply equally to $\text{RV}_{\hat{F}}$.

Theorem 5 $\text{RV}_{\hat{F}} - \text{MC}^\# = O_p(n^{-1})$. Furthermore, if the first observed state coincides with the last observed state then $\text{RV}_{\hat{F}} = \text{MC}^\#$.

In Theorem 3 we obtained an expression for the asymptotic variance of the infeasible quantity RV_F . The difference between these two asymptotic variances, stems from the estimation error of \hat{P} . By replacing the true transition matrix with the empirical one we introduce an extrinsic variance that adds extra variation to the estimator.

4.2.2 Limit distribution of $\log(\text{MC})$

Several studies have shown that the Gaussian limit distribution can be a poor approximation to the finite sample distribution of the realized variance, see e.g. Barndorff-Nielsen & Shephard (2005) and Goncalves & Meddahi (2008). To improve the finite sample properties Barndorff-Nielsen & Shephard (2002) advocated the use of the log transform, and Goncalves & Meddahi (2006) highlight the benefits of the log transform and related transformations. For the Markov chain estimator, the asymptotic distribution of the log-transformed estimator is simple to derive.

Corollary 2 *The limit distribution of $\log(\text{MC}^\#)$ is given by,*

$$n^{1/2} \left\{ \log \langle \xi, (2\hat{Z} - I)\xi \rangle_{\hat{\pi}} - \log \langle \xi, (2Z - I)\xi \rangle_{\pi} \right\} \xrightarrow{d} \text{N} \left\{ 0, \frac{\Sigma_{\text{MC}}}{\langle \xi, (2Z - I)\xi \rangle_{\pi}^2} \right\}.$$

In our framework, the log transformation need not improve the finite sample coverage probability to the same extent as is the case for the realized variance. The reason is that the Markov based estimator can utilize all observations, unlike the realized variance that typically is limited to less than 100 intraday returns, in order to avoid the problems arising from market microstructure noise.

4.2.3 Asymptotic Distribution by Bootstrap Methods

Bootstrap method for conducting inference about the realized variance are analyzed and discussed in Goncalves & Meddahi (2008). In this Section, we discuss how bootstrap methods can be used to conduct inference about $\text{MC}^\#$. Fortunately, bootstrap results for our Markov chain framework are readily available. There are two standard ways to bootstrap the transition matrix of a discrete Markov chain given the empirical transition matrix \hat{P}_n , see Shao & Tu (1995) for an introduction. The conventional scheme is to generate synthetic samples of length n using $P^* = \hat{P}_n$, and then re-estimate the transition matrix \hat{P}_n^* . It can be shown that $\hat{P}_n^* - P^*$ will have the same asymptotic distribution as $\hat{P}_n - P$. We can then plug it into our expression for $\text{MC}^\#$. A more efficient bootstrap method is the conditional bootstrap of Basawa & Green (1990). Instead of sampling a whole run, it is recommended to bootstrap each row of the transition matrix separately, using a multinomial with parameters $(n_{r,\cdot}; \hat{P}_{r,1}, \dots, \hat{P}_{r,S})$.

We compare confidence intervals based on the bootstrap with those based on the delta method in Section 9.2. In our empirical application we find the three to produce almost identical confidence intervals.

5 Quadratic Variation of Log Prices

Up to now, we have assumed that the observed price X_{T_i} lives on a discrete grid. This is true in practice, but the object of interest is typically the quadratic variation of log-prices, and $\log X_t$ does not live on a grid. The implication is that $\text{MC}^\#$ is estimating the wrong quantity. Let $y_{T_i} = \log Y_{T_i}$. We are estimating

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \sum_{i=1}^n (Y_{T_i} - Y_{T_{i-1}})^2 &= \text{plim}_{n \rightarrow \infty} \sum_{i=1}^n \left\{ \exp(\Delta y_{T_i} + y_{T_{i-1}}) - \exp(y_{T_{i-1}}) \right\}^2 \\ &= \text{plim}_{n \rightarrow \infty} \sum_{i=1}^n \left\{ [\exp(\Delta y_{T_i}) - 1] \exp(y_{T_{i-1}}) \right\}^2 \\ &= \text{plim}_{n \rightarrow \infty} \sum_{i=1}^n \left\{ \Delta y_{T_i} Y_{T_{i-1}} \right\}^2, \end{aligned}$$

which equals $\int_0^T \sigma_u^2 Y_u^2 du$ when the observations are equidistant, i.e. $T_i - T_{i-1} = T/n$.

Fortunately, the Markov chain framework can be adapted to estimate the appropriate quantity. We consider two ways to address this issue: An exact correction method and a simpler approximation. The drawback of the “exact” estimator is that its expression is path dependent, so that our asymptotic results of the previous Section do not apply. The approximate method simply amount to a scaling of $\text{MC}^\#$, and our asymptotic distribution theory is directly applicable to this estimator.

5.1 Exact Method

At the expense of having a closed form expression for the estimator, we can get an exact formula by computing the filtered log-increments as follows. The filtered realized variance we seek is given by,

$$\text{plim}_{h \rightarrow \infty} \sum_{i=1}^n \left\{ \log \mathbb{E}(X_{T_{i+h}} | \mathcal{F}_{T_i}) - \log \mathbb{E}(X_{T_{i+h-1}} | \mathcal{F}_{T_{i-1}}) \right\}^2.$$

First we observed that

$$\mathbb{E}(X_{T_{i+h}} | X_{T_i} = y_0, X_{T_{i-1}} = y_{-1}, \dots) = y_0 + \sum_{l=1}^h \mathbb{E}(\Delta X_{T_{i+l}} | \Delta X_{T_i} = \Delta y_0, \Delta X_{T_{i-1}} = \Delta y_{-1}, \dots),$$

where $\Delta y_i = y_i - y_{i-1}$. In our Markov framework the conditional expectation only depends on the k most recent price changes, $\Delta \mathcal{X}_{T_i} = (\Delta X_{T_{i-k+1}}, \dots, \Delta X_{T_i})$. Using that $X_{T_{i+h}} = X_{T_i} + \sum_{l=1}^h \Delta X_{T_{i+l}}$ we have

$$\log \mathbb{E}(X_{T_{i+h}} | \mathcal{F}_{T_i}) = \log \left(X_{T_i} + \sum_s \sum_{l=1}^h (P^l f)_s 1_{\{\Delta \mathcal{X}_{T_i} = \mathbf{x}_s\}} \right),$$

which leads to the following result.

Lemma 4 *Let $\mu_h = h \cdot \langle \pi, f \rangle$. Then we have*

$$\lim_{h \rightarrow \infty} \log \left\{ \mathbb{E}(X_{T_{i+h}} - \mu_h | \mathcal{F}_{T_i}) \right\} = \log \left\{ X_{T_i} + \sum_{s=1}^{S^k} \{(Z - I)f\}_s 1_{\{\Delta \mathcal{X}_{T_i} = \mathbf{x}_s\}} \right\}.$$

When computing log-increments of the filtered price series, we need to subtract the drift, μ_h .³ The reason is that

$$\lim_{h \rightarrow \infty} \frac{\mathbb{E}(X_{T_{i+h}} | \Delta \mathcal{X}_{T_i} = \mathbf{x}_s)}{\mathbb{E}(X_{T_{i+h-1}} | \Delta \mathcal{X}_{T_{i-1}} = \mathbf{x}_r)} = \lim_{h \rightarrow \infty} \frac{\sum_{l=1}^h (P^l x)_s}{\sum_{l=1}^h (P^l x)_r} = 1, \quad \text{when } \langle \pi, f \rangle \neq 0.$$

Previously, where we compute increments in filtered prices (in levels), it was not necessary to subtract the drift, because the constant, μ_h , cancels out, since $\mathbb{E}(X_{T_{i+h}} | \mathcal{F}_{T_i}) - \mathbb{E}(X_{T_{i+h-1}} | \mathcal{F}_{T_{i-1}}) = \mathbb{E}(X_{T_{i+h}} - \mu_h | \mathcal{F}_{T_i}) - \mathbb{E}(X_{T_{i+h-1}} - \mu_h | \mathcal{F}_{T_{i-1}})$. By substituting our empirical estimate \hat{Z} for Z , we obtain the exact expression

$$\text{MC}^* = \sum_{i=1}^n \left[\log \frac{X_{T_i} + \sum_{s=1}^S \{(\hat{Z} - I)f\}_s 1_{\{\Delta \mathcal{X}_{T_i} = \mathbf{x}_s\}}}{X_{T_{i-1}} + \sum_{r=1}^S \{(\hat{Z} - I)f\}_r 1_{\{\Delta \mathcal{X}_{T_{i-1}} = \mathbf{x}_r\}}} \right]^2. \quad (6)$$

We note that the expression for the log increments is path dependent. So our asymptotic result in Section 4 are not easily adapted to this problem.

5.2 Approximate Method

Consider the case where $\Delta \mathcal{X}_{T_{i-1}} = \mathbf{x}_r$ and $\Delta \mathcal{X}_{T_i} = \mathbf{x}_s$ (so that $\Delta X_{T_i} = f_s$). Then

$$\log \left(\frac{X_{T_i} + \{(\hat{Z} - I)f\}_s}{X_{T_{i-1}} + \{(\hat{Z} - I)f\}_r} \right) = \log \left(1 + \frac{\{\hat{Z}f\}_s - \{(\hat{Z} - I)f\}_r}{X_{T_{i-1}} + \{(\hat{Z} - I)f\}_r} \right) \approx \frac{\{Zf\}_s - \{(\hat{Z} - I)f\}_r}{X_{T_{i-1}} + \{(\hat{Z} - I)f\}_r},$$

where we have used a Taylor expansion of $\log(1 + x)$. The numerator is simply $\hat{y}_{(r,s)}$, see (3), and the denominator is roughly equal to $X_{T_{i-1}}$, because $(Z - I)f = O(n^{-1/2})$. Since

³This is a minor issue in practice because empirically we find that $\pi'f \approx 0$.

$\sum_{i=1}^n \hat{y}_{(r,s)}^2 1_{\{\Delta \mathcal{X}_{T_{i-1}} = \mathbf{x}_r, \Delta \mathcal{X}_{T_i} = \mathbf{x}_s\}} = \text{RV}_{\hat{F}}$ and $\text{RV}_{\hat{F}} - n \langle f, (2\hat{Z} - I)f \rangle_{\hat{\pi}} = O_p(n^{-1})$, we should expect MC^* to be roughly equal

$$\text{MC} = \frac{n^2 \langle f, (2\hat{Z} - I)f \rangle_{\hat{\pi}}}{\sum_{i=1}^n X_{T_i}^2}.$$

In fact this approximation is very good as long as X_{T_i} does not fluctuate dramatically over the estimation period. The advantage of the estimator, MC, is that it allows for faster computations and it preserves the elegant asymptotic theory that we derived for $\text{MC}^\#$. In the data we have analyzed, we found that MC and MC^* are empirically indistinguishable, and the difference between the two quantities is trivial compared to the confidence intervals we obtain for MC.

6 Continuous Time Markov Chain and Zero Returns

In this Section we consider the case where the underlying process is a continuous time Markov chain (CTMC). We formulate conditions under which the increments of the discretely sampled process are stationary and homogeneous. Interestingly, in this framework we find an argument for discarding all zero-increments when estimating the discrete time Markov chain. Interesting aspects of zero-increments, also known as flat pricing, are analyzed in Phillips & Yu (2008).

Three types of time scales for the discretely sampled process are often used in this literature. *Calendar time* sampling refers to the case where prices are sampled equidistant in time, e.g. every 60 seconds. *Event time* (or tick time) is the case where we sample every K -th price observation. *Volatility time* (or business time) is a theoretical time scale with the characteristic that returns are homogeneous when sampled equidistant in volatility time. We will assume that there exist a suitable time change so that volatility time is well defined.

Sampling returns that are equidistant in volatility time is the ideal sampling scheme, because volatility time minimizes the asymptotic variance of RV, see Hansen & Lunde (2006, section 3). This provides an argument for sampling in event time rather than calendar time, because the former is thought to better approximate volatility time. This is consistent with studies that have shown that realized variance behaves much better in event time, see e.g. Griffin & Oomen (2006). The CTMC framework used in this Section will give an elegant explanation to this phenomenon. We will, under certain conditions, show that event time

and volatility time coincide as $n \rightarrow \infty$. One of the requirements is that zero increments be discarded.

Suppose that the observed process, X_t , is a CTMC with a countable state space. In calendar time, this chain is inhomogeneous with an infinitesimal generator, $A(t)$, that may be random. So $\Pr(X_{t+h} = j | X_t = i) = h[A(t)]_{ij} + o(h)$, and $\Pr(X_{t+h} = i | X_t = i) = 1 + h[A(t)]_{ii} + o(h)$. In matrix form we can formulate this with the transition matrix from X_s to X_t , which has the form

$$P(s, t) = \mathbb{E} \left(e^{\int_s^t A(u) du} | \mathcal{F}_s \right).$$

In the most general formulation, the time varying generator is infinitely dimensional, which makes it an arduous task to estimate the integrated variance without imposing some structure. A major simplification is achieved by assuming the existence of a volatility time scale, because it imposes a one dimensional time structure on the generator, i.e. $A(t) = \lambda_t A$. When $A(t)$ only depends on time through λ_t , the finite state CTMC is *uniformisable*, which enable us to represent the chain as two independent processes: a counting process N_t with intensity λ_t and jump times τ_i ; and the embedded discrete time Markov chain, $\hat{X}_i = X_{\tau_i}$, with transition matrix P . In this situation, P , $A(t)$, and λ_t are linked by the relationship $A(t) = \lambda_t(P - I)$, and we have $X_t \stackrel{d}{=} \hat{X}_{N_t}$.

In our sample, we do not know if the times where we observe transactions/quotes, $T'_0, T'_1, \dots, T'_{N'}$, coincide with actually jump times, τ_1, \dots, τ_N , of the CTMC. However, by discarding zero increments we know that we have a subset of all the values taken by the underlying CTMC. For the resulting discrete time Markov chain to be homogeneous in the limit we further need that sampling becomes *exhaustive* as $N \rightarrow \infty$, such that we do not miss a price change in the limit. This result is formulated the following theorem.

Theorem 6 *Let X_t be a CTMC on a countable state space, with generator $A(t) = \lambda_t A$ that is adapted to \mathcal{F}_t . Let the observation times T'_i , $0 \leq i \leq N'$, be jump times from a Poisson process with continuous intensity μ_t , and suppose that $\lambda_t = \lambda_t(\alpha)$ and $\mu_t = \mu_t(\alpha)$ are functions of a parameter α , (α can be tied to the grid size for instance). If $\frac{\lambda_t(\alpha)}{\mu_t(\alpha)} \rightarrow 0$, uniformly in t , as $\alpha \rightarrow \infty$, then with (T_0, T_1, \dots, T_n) given by,*

$$\{T_i : T_i = T'_j \neq T'_{j-1} \text{ for some } 0 \leq j \leq N'\},$$

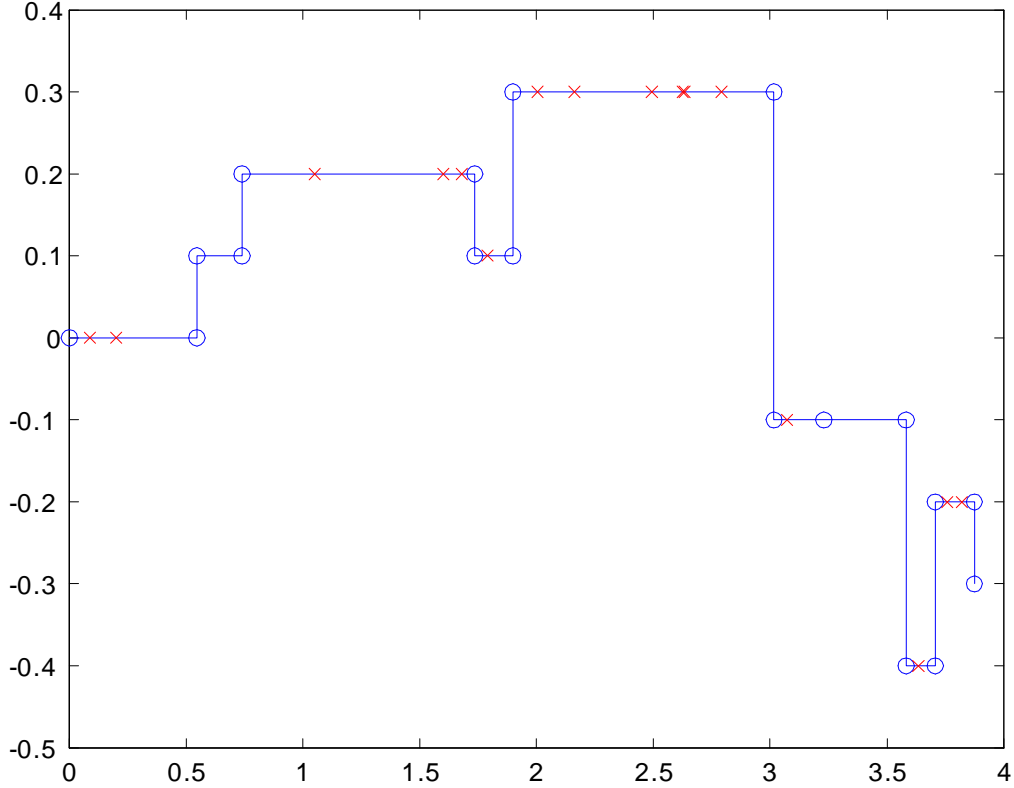


Figure 1: A continuous time Markov chain with three states. Jump times are identified by circles and observation times by crosses. These are generated by independent Poisson processes with intensities λ and μ , respectively, where $\lambda/\mu = 1/2$.

$(X_{T_0}, X_{T_1}, \dots, X_{T_n}) \stackrel{d}{=} (x_0, x_1, \dots, x_n)$ where x_i is a homogeneous Markov chain with transition distribution given by $P_{i,j} = \frac{A_{i,j}}{A_{i,i}}$ for $i \neq j$ and $P_{i,i} = 0$.

Comment. The asymptotic result is driven by $\alpha \rightarrow \infty$ that causes the number of jumps in the CTMC, $N \rightarrow \infty$, while the number of observation times, N' grows at a uniformly faster rate. The implication is that the sequence of observed states (excluding zero increments) will be identical to that of the latent CTMC, as $\alpha \rightarrow \infty$.

The assumption that $\frac{\lambda_t(\alpha)}{\mu_t(\alpha)} \rightarrow 0$ uniformly in t looks very stringent. In practice this is not an issue, because the observed sample will be indistinguishable from the sample obtained

by sampling exactly at jump times that led to a price change, as long as $\frac{\lambda_t}{\mu_t} \ll 1$. This is illustrated in Figure 1 where we have simulated a CTMC with three states. The circles are the jump times from the Markov chain (with intensity λ_t), whereas the red crosses are the observation times. Notice that the CTMC can “jump” without changing its value. The observation intensity μ_t is such that $\frac{\lambda_t}{\mu_t} = \frac{1}{2}$. The sampling only misses one change of state, so even moderate frequencies can lead to quasi time-homogeneous sampling.

Homogeneous intraday returns arises naturally in the CTMC framework laid out above. However this is within an specific framework and the homogeneity is established as an asymptotic feature of returns. So it is natural to ask what happens when the reality is far from this idealized setting. We address this in the next Section where we analyze the properties of $MC^\#$ when the underlying process is inhomogeneous.

7 Robustness to Inhomogeneous

The analysis of quantities estimated with high frequency data, are typically derived from assuming local constancy of volatility. A general treatment of this approach and its validity is analyzed in Mykland & Zhang (2008), for the case without market microstructure noise. In this Section we analyze the situation with an inhomogeneous Markov chain using this approach. Unlike the general analysis in Mykland & Zhang (2008), our analysis will be specific to that with Markov chains, but an advantage of our framework is that it allows for the presence of market microstructure noise. Local constancy of volatility and related quantities translate into a locally homogeneous Markov chain in our framework. So we consider the situation where the Markov chain is locally homogeneous, but globally inhomogeneous. Naturally, if the Markov chain is piecewise homogeneous, then we can simply compute the $MC^\#$ over the homogeneous subintervals, and add these estimates up. By increasing the number of subintervals as $n \rightarrow \infty$, this type of local estimation scheme can accommodate more general forms of inhomogeneity. We shall see that a local estimation approach is not needed here, because the full sample Markov chain estimator turns out to be surprisingly robust to inhomogeneity. The robustness is achieved by artificially increasing the order of the Markov chain that is being estimated with the full sample. After theoretical argument for the robustness, we consider a simple example where the data generating process is a Markov chain of order one with a transition matrix that is subject to structural

changes. A simulation study shows that estimating a Markov chain of order two yields an accurate estimate despite the Markov chain of order two being grossly misspecified. Due to misspecification, the framework is likely to produce poor estimates of many population quantities, but fortunately it will accurately estimate the quantity we seek. Somewhat surprisingly, we find that $\text{MC}_2^\#$ estimated over the full sample can be more accurate than an “oracle” estimator that adds up the $\text{MC}_1^\#$ estimates that are computed over homogeneous subsamples.

7.1 Theoretical Insight about Robustness

If the true model is a homogeneous Markov chain of order k , but we estimate the Markov chain with too high an order, $k' > k$, then the estimator $\text{MC}_{k'}^\#$ will still be consistent. The drawback of using too large an order is that the asymptotic variance of the estimator increases with k . A Markov chain of order k is characterized by the reduced form transition matrix,

$$Q_{i_1, \dots, i_{k+1}} = \Pr \{ \Delta X_{T_{i+1}} = i_{k+1} | \Delta \mathcal{X}_{T_i} = (i_1, \dots, i_k)' \}.$$

Suppose that the true order of the chain is of lower order, $k - 1$ say, then it is simple to verify that $Q_{i_1, \dots, i_{k+1}} = \tilde{Q}_{i_2, \dots, i_{k+1}}$, where \tilde{Q} is the reduced form transition matrix of the Markov chain of order $k - 1$. For instance, if the true order is one with transition matrix P , then

$$\lim_{n \rightarrow \infty} \hat{Q}_{i_1, \dots, i_{k+1}} = \frac{N_n(i_1, \dots, i_{k+1})}{N_n(i_1, \dots, i_k)} \xrightarrow{p} P_{i_{k-1}, i_k}.$$

Consider now the case where the transition matrix is inhomogeneous, so that $P(t) = P^{(1)}$ for $t \leq wn$ and $P(t) = P^{(2)}$ for $wn < t \leq n$, for some $w \in (0, 1)$. Let $N_n(i_1, \dots, i_k)$ be the number of time state (i_1, \dots, i_k) is observed in the full sample, and let $N_{n_j}^{(j)}(i_1, \dots, i_k)$ be the corresponding counts for the j -th subsample, $j = 1, 2$, where $n_1 = \lfloor wn \rfloor$, and $n_2 = n - n_1$. Then we have

$$\hat{Q}_{i_1, \dots, i_{k+1}} = \frac{N_n(i_1, \dots, i_{k+1})}{N_n(i_1, \dots, i_k)} = \frac{N_{n_1}^{(1)}(i_1, \dots, i_{k+1}) + N_{n_2}^{(2)}(i_1, \dots, i_{k+1})}{N_{n_1}^{(1)}(i_1, \dots, i_k) + N_{n_2}^{(2)}(i_1, \dots, i_k)},$$

and dividing by n and using the fact that $\frac{n_1}{n} \rightarrow w$, we find

$$\text{plim}_{n \rightarrow \infty} \hat{Q}_{i_1, \dots, i_{k+1}} = \frac{w P_{i_1, i_2}^{(1)} \cdots P_{i_k, i_{k+1}}^{(1)}}{w P_{i_1, i_2}^{(1)} \cdots P_{i_{k-1}, i_k}^{(1)} + (1-w) P_{i_1, i_2}^{(2)} \cdots P_{i_{k-1}, i_k}^{(2)}}$$

$$+ \frac{(1-w)P_{i_1, i_2}^{(2)} \cdots P_{i_k, i_{k+1}}^{(2)}}{wP_{i_1, i_2}^{(1)} \cdots P_{i_{k-1}, i_k}^{(1)} + (1-w)P_{i_1, i_2}^{(2)} \cdots P_{i_{k-1}, i_k}^{(2)}} = Q_{i_1, \dots, i_{k+1}}. \quad (7)$$

We will now show that $Q_{i_1, \dots, i_{k+1}}$ leads to the appropriate estimator if we increase k .

Suppose that we have n observations drawn from a transition matrix P which is either $P^{(1)}$ with probability w or $P^{(2)}$ with probability $1-w$. Practically, we toss a weighted coin with probability w of showing heads. If it is heads, we draw n observations using $P^{(1)}$, if it is tails we draw n observations from $P^{(2)}$. Given k observations, i_1, \dots, i_k , the posterior probability, $\mathcal{L}(P = P^{(1)} | i_1, \dots, i_k)$, is by Bayes' formula equal to

$$\begin{aligned} & \frac{\mathcal{L}(i_1, \dots, i_k | P = P^{(1)}) \Pr(P = P^{(1)})}{\Pr(i_1, \dots, i_k | P = P^{(1)}) \Pr(P = P^{(1)}) + \Pr(i_1, \dots, i_k | P = P^{(2)}) \Pr(P = P^{(2)})} \\ = & \frac{wP_{i_1, i_2}^{(1)} \times \cdots \times P_{i_{k-1}, i_k}^{(1)}}{wP_{i_1, i_2}^{(1)} \times \cdots \times P_{i_{k-1}, i_k}^{(1)} + (1-w)P_{i_1, i_2}^{(2)} \times \cdots \times P_{i_{k-1}, i_k}^{(2)}}. \end{aligned}$$

This shows that $Q_{i_1, \dots, i_{k+1}}$ given in (7) can be reinterpreted as the posterior mixture

$$Q_{i_1, \dots, i_{k+1}} = \mathcal{L}(P = P^{(1)} | i_1, \dots, i_{k+1})P_{i_k, i_{k+1}}^{(1)} + \mathcal{L}(P = P^{(2)} | i_1, \dots, i_{k+1})P_{i_k, i_{k+1}}^{(2)}.$$

So we have the remarkable result that everything happens as if the transition matrix was decided a priori with probability w . In the limit as k goes to infinity, the posterior probability has point mass at the true value, so that

$$\hat{Q}_{i_1, \dots, i_{k+1}} \xrightarrow{P} \begin{cases} P_{i_k, i_{k+1}}^{(1)} & \text{if } P = P^{(1)}, \\ P_{i_k, i_{k+1}}^{(2)} & \text{if } P = P^{(2)}. \end{cases}$$

The implication is that the ergodic variance using $Q_{i_1, \dots, i_{k+1}}$ is (as $k \rightarrow \infty$) identical to

$$\Pr(P = P^{(1)})\text{QV}^{(1)} + \Pr(P = P^{(2)})\text{QV}^{(2)} = w\text{QV}^{(1)} + (1-w)\text{QV}^{(2)},$$

where $\text{QV}^{(j)} = \langle \xi, (2Z^{(j)} - I)\xi \rangle_{\pi^{(j)}}$, $j = 1, 2$, which is exactly what we want. Some intuition is offered by the transition matrix for the higher order chain. By reordering the states according to the odds ratios $\pi_{i_1, \dots, i_{k+1}}^{(1)} / \pi_{i_1, \dots, i_{k+1}}^{(2)}$, the transition matrix approaches, as $k \rightarrow \infty$, a block diagonal structure.

The argument we have given in this Section can be generalized to piecewise constant non-homogeneous transition matrices. [G]

7.2 Simulation Results

To analyze the robustness of $\text{MC}^\#$ we consider a simple simulation design, where the transition matrix changes over time. Consider the class of transition matrices,

$$P(\lambda) = \frac{1}{2} \begin{pmatrix} 1 + \lambda & 1 - \lambda \\ 1 - \lambda & 1 + \lambda \end{pmatrix}, \quad \text{with } \lambda \in (-1, 1),$$

so that $\pi = (\frac{1}{2}, \frac{1}{2})'$ for all λ . When $\xi = (\kappa, -\kappa)'$ it can be shown that $V(\lambda) = \langle \xi, (2Z(\lambda) - I)\xi \rangle_\pi = \kappa^2 \frac{1+\lambda}{1-\lambda}$, where $Z(\lambda)$ is the fundamental matrix for $P(\lambda)$.

	Oracle	MC ₁ [#]	MC ₂ [#]	MC ₃ [#]	MC ₄ [#]	MC ₅ [#]	MC ₆ [#]	MC ₇ [#]
plim MC _k [#]	1.0000	0.7778	1.0370	1.0256	0.9910	0.9926	1.0055	1.0026
<i>Simulation (n = 1,000)</i>								
Average	0.9986	0.7777	1.0378	1.0290	0.9968	1.0024	1.0168	1.0168
Std.dev.	0.0793	0.0460	0.0988	0.1315	0.1505	0.1695	0.1917	0.2110
RMSE	0.0793	0.2270	0.1057	0.1346	0.1505	0.1695	0.1924	0.2117
<i>Analytical</i>								
Std.dev.	0.0789	0.0496	0.1028	0.1386	0.1594	0.1796	0.2027	0.2217
<i>Simulation (n = 23,400)</i>								
Average	1.0000	0.7778	1.0371	1.0257	0.9912	0.9930	1.0060	1.0032
Std.dev.	0.0163	0.0095	0.0203	0.0269	0.0305	0.0340	0.0381	0.0415
RMSE	0.0163	0.2224	0.0423	0.0372	0.0317	0.0347	0.0386	0.0416
<i>Analytical</i>								
Std.dev.	0.0163	0.0102	0.0213	0.0286	0.0329	0.0371	0.0419	0.0458

Table 1: Simulate and analytical quantities of, $\text{MC}_k^\#$, estimated with the full (inhomogeneous) sample, when returns are driven by different Markov chains in the first and second half of the sample. With $n = 1,000$ observations the most accurate estimator is $\text{MC}_2^\#$ (excluding the infeasible “oracle” estimator). With $n = 23,400$ observation $\text{MC}_4^\#$ is the most accurate. The results are based on 50,000 simulations.

Suppose that $\xi = (1, -1)'$ and we have a single break in the middle of the sample, where the transition matrix in the first and second half of the sample are

$$P(-\frac{1}{2}) = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}, \quad \text{and} \quad P(\frac{1}{4}) = \begin{pmatrix} \frac{5}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{5}{8} \end{pmatrix},$$

respectively. The object of interest is here given by,

$$QV = \frac{1}{2}V(-\frac{1}{2}) + \frac{1}{2}V(\frac{1}{4}) = \frac{1}{2}(\frac{1}{3} + \frac{5}{3}) = 1.$$

When estimating a Markov chain of order one, the expected frequencies of the possible transitions are given by

$$\frac{1}{2} \begin{pmatrix} \frac{1}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{1}{8} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \frac{5}{16} & \frac{3}{16} \\ \frac{3}{16} & \frac{5}{16} \end{pmatrix} = \begin{pmatrix} \frac{7}{32} & \frac{9}{32} \\ \frac{9}{32} & \frac{7}{32} \end{pmatrix},$$

which shows that the estimated transition matrix is such that

$$\hat{P}_{(k=1)} \xrightarrow{p} \begin{pmatrix} \frac{7}{32}/(\frac{7}{32} + \frac{9}{32}) & \frac{9}{32}/(\frac{7}{32} + \frac{9}{32}) \\ \frac{9}{32}/(\frac{9}{32} + \frac{7}{32}) & \frac{7}{32}/(\frac{9}{32} + \frac{7}{32}) \end{pmatrix} = \begin{pmatrix} \frac{7}{16} & \frac{9}{16} \\ \frac{9}{16} & \frac{7}{16} \end{pmatrix},$$

from which it follows that $MC_1^\# \xrightarrow{p} \frac{7}{9} = 0.7778$. So the estimator based on a homogeneous Markov chain of order one is inconsistent for QV.

Suppose instead we estimate a Markov chain of order two. We order the four states as follows, $(1, 1)$, $(1, -1)$, $(-1, 1)$, and $(-1, -1)$, so that $\xi = (1, -1, 1, -1)'$. Now, the Markov chains for the two subsamples imply different stationary distributions, $(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8})'$ and $(\frac{5}{16}, \frac{3}{16}, \frac{3}{16}, \frac{5}{16})'$, respectively, and the expected ratios of the various transition in the full sample are given by

$$\frac{1}{2} \begin{pmatrix} \frac{1}{32} & \frac{3}{32} & 0 & 0 \\ 0 & 0 & \frac{9}{32} & \frac{3}{32} \\ \frac{3}{32} & \frac{9}{32} & 0 & 0 \\ 0 & 0 & \frac{3}{32} & \frac{1}{32} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \frac{25}{128} & \frac{15}{128} & 0 & 0 \\ 0 & 0 & \frac{9}{128} & \frac{15}{128} \\ \frac{15}{128} & \frac{9}{128} & 0 & 0 \\ 0 & 0 & \frac{15}{128} & \frac{25}{128} \end{pmatrix}.$$

So that the probability limit for the estimated transition matrix is

$$\hat{P}_{(k=2)} \xrightarrow{p} \begin{pmatrix} \frac{29}{56} & \frac{27}{56} & 0 & 0 \\ 0 & 0 & \frac{5}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{5}{8} & 0 & 0 \\ 0 & 0 & \frac{27}{56} & \frac{29}{56} \end{pmatrix},$$

that implies the “stationary” distribution, $\pi = (\frac{7}{32}, \frac{9}{32}, \frac{9}{32}, \frac{7}{32})'$, and $MC_2^\# \xrightarrow{p} \frac{28}{27} = 1.037$, which is far closer to one than the limit for $MC_1^\#$. By further increasing k , the plim of $MC_k^\#$ approaches QV = 1. In Table 1 we have listed the theoretical limit of $MC_k^\#$, for $k = 1, \dots, 7$ along with the analytical standard deviation, where the latter is given as the square-root of Σ_{MC} , see (5). Table 1 also reports an “Oracle” estimator. This estimator is defined by Oracle = $MC_1^{\#(1)} + MC_1^{\#(2)}$, where $MC_1^{\#(j)}$ is the Markov chain estimator computed with the j -th subsample, $j = 1, 2$. This estimator has the advantage of knowing that a structural

change occurs in the middle of the sample, and computes $\text{MC}_1^{\#(j)}$ with homogeneous returns from the two subsamples.

Importantly, do we see that the analytical and simulation-based standard deviations are largely in agreements. So computing the asymptotic variance for $\text{MC}_k^\#$, as if the Markov chain was homogeneous continues to be valid even if the chain is, in fact, homogeneous, provided that the estimator is robustified by increasing k . Not only is the estimator robust to inhomogeneity, so is our expressions for its standard deviation.

Next we study a less homogeneous process, where the transition matrix alternates between,

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}, \quad \begin{pmatrix} \frac{4}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{4}{5} \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} \frac{1}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{1}{5} \end{pmatrix},$$

which correspond to $\lambda = 0, \frac{1}{2}, \frac{3}{5}$, and $-\frac{3}{5}$, respectively. In the first simulation design we divide the sample into four subsamples using the following ratios, $\frac{1}{4}, \frac{1}{4}, \frac{7}{30}$, and $\frac{8}{30}$, respectively. With $\xi = (1/\sqrt{2}, -1/\sqrt{2})'$ the object of interest is normalized to one, because

$$\frac{1}{2} \left(\frac{1}{4} + \frac{1}{4} \frac{1+0.5}{1-0.5} + \frac{7}{30} \frac{1+0.6}{1-0.6} + \frac{8}{30} \frac{1-0.6}{1+0.6} \right) = 1.$$

In a second design we have 16 regimes, i.e. 15 structural changes. These are obtained by shorten each of the four subintervals by a factor of four and repeating the sequence of subintervals four times. Since each of the four transition matrices are “active” in the same fraction of the sample, the object of interest is also one in this design.

The simulation results are reported in Table 2 for $n = 400, 1,000$, and $10,000$. The $\text{MC}^\#$ estimator applied to the full sample is denoted by $\text{MC}_k^\#$, where we consider $k = 1, \dots, 4$. We also compute the oracle estimator for the case with four regimes. With four regimes, we again see that $\text{MC}_1^\#$ fails at estimating QV, but interestingly, we find that the oracle estimator is inferior to $\text{MC}_2^\#$. Thus, in this design, the misspecified Markov chain of order two yields a more accurate estimate than the oracle estimator, that is based on a correctly specified models. This shows that even if we know that the Markov chain is inhomogeneous, there may not be an advantage from estimating the volatility by adding up estimators computed over homogeneous subsamples. It would be interesting to study this issue in more depth in future research.

The properties of $\text{MC}_k^\#$ for the case with 16 regimes approaches those seen with 4 regimes. This is consistent with our theoretical results, because each of the four transition matrices are “active” in the same fraction of the sample.

		4 Regimes					16 Regimes			
		Oracle	MC ₁ [#]	MC ₂ [#]	MC ₃ [#]	MC ₄ [#]	MC ₁ [#]	MC ₂ [#]	MC ₃ [#]	MC ₄ [#]
$n = 400$	Average	0.9740	0.6128	0.9880	1.0228	1.0099	0.6031	0.9465	0.9685	0.9554
	Std.dev.	0.1593	0.0544	0.1505	0.2123	0.2536	0.0529	0.1406	0.1933	0.2268
	RMSE	0.1612	0.3894	0.1509	0.2137	0.2539	0.3955	0.1487	0.1951	0.2303
$n = 1,000$	Average	0.9886	0.6151	0.9916	1.0223	1.0028	0.6099	0.9710	0.9963	0.9796
	Std.dev.	0.0969	0.0343	0.0945	0.1328	0.1564	0.0339	0.0912	0.1260	0.1491
	RMSE	0.0975	0.3858	0.0948	0.1348	0.1564	0.3856	0.0941	0.1261	0.1498
$n = 10,000$	Average	0.9989	0.6171	0.9945	1.0222	0.9987	0.6167	0.9927	1.0200	0.9968
	Std.dev.	0.0302	0.0109	0.0298	0.0415	0.0486	0.0109	0.0298	0.0416	0.0486
	RMSE	0.0303	0.3830	0.0303	0.0470	0.0486	0.3832	0.0306	0.0462	0.0486

Table 2: Estimation results for inhomogeneous Markov chains. The accuracy of the estimator, MC₂, that is based on a highly misspecified model, is on par with the Oracle estimator. The results are based on 50,000 simulations.

8 Empirical and Computational Issues

8.1 Jumps and Infrequently States

Jumps in the form of large price changes are not uncommon in high-frequency returns. Jumps of this kind have implications for the analysis, essentially because one cannot rely on asymptotics with a single observation. In the absence of noise it is well known that the finite activity jump contributions to the quadratic variation is consistently estimated by the difference between the realized variance and the bipower variation (or a related multipower variation), see Barndorff-Nielsen & Shephard (2004).

In the present context we suggest a simple ad-hoc method for classifying intraday returns as jumps. Specifically, we can define jumps to be price changes whose absolute value exceeds a certain threshold. (In our empirical application we use a 10 cents threshold). The idea is then to estimate the Markov chain with the remaining intraday returns. The

contribution to the quadratic variation from the increments classified as jumps, J_t , can then be added separately. One possibility is to model the jumps separately if the number of jumps is relatively large. For instance, one could model positive jumps and negative jumps as two independent compound Poisson processes J_t^+ and J_t^- , with rates λ_t^+ and λ_t^- , and exponentially distributed jump sizes. A simpler approach is to compute the realized variance of the jumps separately (i.e. $\sum_t J_t^2$). However, it is our experience that many of the price changes that are classified as jumps are caused by outliers, where a large increment in one direction is followed by a large increment in the opposite direction. So an alternative measure is $(\sum_t J_t)^2$ that offers robustness to outliers. A potential drawback of this measure is that it will also offset real jumps that happened to have opposite signs. Naturally this is only an issue if the number of real jumps is two or more.

Infrequent medium large returns can be aggregated in to a single state. For instance in our empirical application we bundle increments between 5 and 10 cents into a single state with their average as the common state value (increments between -10 and -5 cents are combined similarly). This aggregations can be justified by viewing the medium large returns as a single state where the actual increment is drawn from a multinomial distribution. The aggregations help reduce the number of states which is particular beneficial when estimating higher order chains. In our empirical application we found that the aggregation had a negligible impact on the MC estimator.

8.2 Dimension Reduction

The number of distinct states in a Markov chain of order k is S^k . So the dimension of the transition matrix, P , which is $S^k \times S^k$, increases dramatically with k . However, in practice, most states will not be observed, so the dimension of the matrix that must be dealt with can be quite manageable even if k is large. In such cases we define \tilde{P} to be the submatrix of P , that results from deleting the s -th row and s -th column if state s were not observed in the sample. In our empirical analysis we analyze high-frequency data for an exchange traded fund for each of the trading days in June 2008. On June 1st, 2007 we have $n = 4,065$ non-zero intraday returns with price changes ranging from -5 cents to $+4$ cents. With $S = 9$ basic states, the potential number of distinct states, S^k , is given in Table 3 for $k = 1, \dots, 7$. The number of observed states, S_k , is given right below and we see that for $k = 4$ we only observe 391 out of the 6,561 possible states, which simplifies the computational burden

substantially.

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
S^k	9	81	729	6,561	59,049	531,441	4,782,969
S_k	9	44	151	391	801	1,370	2,022

Table 3: The number of potential states, S^k , and that number of observed states, S_k , with the $n = 4,065$ high frequency returns we have in our June 1st, 2007 sample.

8.3 The Empirical Stationary Distribution and Ergodicity

Computing the empirical stationary distribution, $\hat{\pi}$, for a high-dimensional transition matrix can be computationally burdensome, in particular in simulation studies where this must be done repeatedly. A simple modification that makes the computation of π a simple matter is to ensure that the first observed state and last observed state are identical. This can be done by artificially extending the sample with k (or less) observations. For instance, with $k = 2$ and the sample is $x_1, x_2, \dots, x_{n-1}, x_n$, we extend the sample with two observation, and use the sample $x_1, x_2, \dots, x_{n-1}, x_n, x_1, x_2$ to estimate P , etc. Let $n_{r,s}$ be the number of transitions from state r to state s . There will be exactly n observed transitions, where the first transition is the one from state (x_1, x_2) to state (x_2, x_3) and the last observe transition is that from (x_n, x_1) to (x_1, x_2) . Our estimator of P is given by $\hat{P}_{r,s} = n_{r,s}/n_{r,\cdot}$, where $n_{r,\cdot} = \sum_s n_{r,s}$. Because the first and last observed states are identical, we have the same number of transitions in and out of every state, i.e. $n_{r,\cdot} = n_{\cdot,s} = \sum_r n_{r,s}$. In is now simple to verify that the stationary distribution associated with \hat{P} is given by $\hat{\pi} = (\frac{n_{1,\cdot}}{n}, \dots, \frac{n_{sk,\cdot}}{n})'$, because

$$(\hat{\pi}'\hat{P})_s = \sum_r \hat{\pi}_r \hat{P}_{r,s} = \sum_r \frac{n_{s,\cdot}}{n} \frac{n_{r,s}}{n_{r,\cdot}} = \frac{n_{s,\cdot}}{n} \frac{n_{\cdot,s}}{n_{r,\cdot}} = \hat{\pi}_s,$$

since $n_{r,\cdot} = n_{\cdot,s}$ by construction.

Another advantage from this sample extension is that no state will be absorbing under \hat{P} , which is reasonable to rule out in this context. This guarantees that the empirical transition matrix is within the class of ergodic transition matrices almost surely.

8.4 Order Selection

Selecting the order of the Markov chain, k , is not a simple matter in practice. While selection methods, such as AIC and BIC, can be justified as $n \rightarrow \infty$ they are not useful in the present context where the number of free parameters is large relative to the sample size. Even with thousands of high-frequency returns to estimate the Markov chain, the number of free parameters becomes large, even with $k = 2$ or $k = 3$. The reason is that the number of parameters grows exponentially with k , and the number of free parameters is a very poor indicator of the “effective” degrees of freedom in this context. We are currently working on theoretically sound methods for choosing k for the present problem. In our empirical application we recommend $k = 3$ or $k = 4$ because Markov chains of higher orders produce similar estimates which suggests that a higher order is not needed in order to capturing the dependence structure.

9 Empirical Analysis

In this Section we apply the MC estimator to high-frequency data for the SPDR Trust, Series 1 (AMEX:SPY) which is an exchange traded fund that tracks the S&P 500 index. We use transaction prices and bid and ask quotes for the 21 trading days in June 2007. The data were extracted from the TAQ data base, using the cleaning procedures advocated by Barndorff-Nielsen, Hansen, Lunde & Shephard (2008c). We will focus on several aspects of our estimator, specifically:

- How the MC estimator compares with our exact estimator, MC*.
- Sensitivity to the choice of k (the order of the Markov chain).
- Compare the three methods for constructing confidence intervals for MC.

Some summary statistics are computed in Table 4. We see that the majority of the price changes are small, between -2 and +2 cents. In terms of trade and quote data, the main difference is the larger number of zero intraday returns for the quote data. In fact the larger sample sizes for quote data are largely due to a larger number of zero returns. Thus by omitting zero returns the three price series will result in similar sample sizes.

Date		10 ⁻	5 ⁻	-4	-3	-2	-1	0	1	2	3	4	5 ⁺	10 ⁺	min	max	RV _{5 min}
20070601	T	0	1	5	26	297	1684	3795	1756	263	26	7	0	0	-5	4	0.2066
	B	0	2	1	25	217	1477	16091	1593	176	17	3	0	0	-6	4	0.2267
	A	0	0	3	14	172	1564	16129	1506	184	27	2	1	0	-4	6	0.2267
20070604	T	0	1	0	21	203	1353	3287	1445	194	20	1	0	0	-5	4	0.1375
	B	0	0	2	14	135	1116	16089	1232	123	5	2	1	0	-4	5	0.1636
	A	0	0	1	11	103	1219	16017	1223	128	16	1	0	0	-4	4	0.1643
20070605	T	0	3	6	46	333	2045	4433	1978	337	52	9	2	0	-6	7	0.2769
	B	0	2	2	26	258	1893	16086	1953	224	23	1	1	0	-6	5	0.2902
	A	0	1	2	22	201	1962	16235	1724	278	41	3	0	0	-5	4	0.2900
20070606	T	0	0	2	41	318	2045	4645	1980	320	33	3	0	0	-4	4	0.2834
	B	0	0	0	26	250	1900	16381	1971	188	15	1	0	0	-3	4	0.2845
	A	0	0	1	17	225	1994	16357	1885	229	23	1	0	0	-4	4	0.2834
20070607	T	0	4	16	118	642	2854	5773	2747	627	88	19	3	0	-6	6	0.7413
	B	0	2	3	68	478	2983	15236	2822	468	45	10	2	0	-5	5	0.6297
	A	0	1	10	61	458	3003	15263	2804	440	66	10	1	0	-7	5	0.6276
20070608	T	0	8	14	80	467	2400	5016	2414	512	91	20	8	0	-7	6	0.4776
	B	0	2	2	69	398	2354	15308	2510	406	57	9	1	0	-5	6	0.5157
	A	0	1	5	42	352	2498	15324	2371	435	81	6	1	0	-5	5	0.5180
20070611	T	0	1	2	25	210	1895	4286	1768	265	42	4	0	0	-5	4	0.2735
	B	0	0	0	8	176	1674	16242	1602	217	17	0	0	0	-3	3	0.2881
	A	0	0	0	7	161	1712	16244	1563	218	29	2	0	0	-3	4	0.2891
20070612	T	0	0	9	57	478	2624	5833	2534	464	63	4	4	0	-4	6	0.5528
	B	0	1	6	41	396	2350	15869	2434	317	36	4	1	0	-7	5	0.5764
	A	0	1	3	40	352	2436	15960	2214	388	55	6	0	0	-6	4	0.5772
20070613	T	0	5	9	80	417	2243	4651	2306	438	95	12	3	0	-6	6	0.4532
	B	0	3	9	51	388	2142	15643	2360	376	46	5	1	0	-6	5	0.4359
	A	0	0	6	44	323	2364	15530	2276	409	63	9	0	0	-4	4	0.4373
20070614	T	0	0	7	21	245	2047	4927	1956	309	28	7	1	0	-4	5	0.2177
	B	0	0	1	28	219	1616	16687	1772	183	21	1	0	0	-4	4	0.2378
	A	0	0	1	17	181	1726	16681	1676	211	31	4	0	0	-4	4	0.2395
20070615	T	0	0	2	13	231	1889	4891	1914	203	17	1	0	0	-4	4	0.1550
	B	0	0	1	21	167	1474	16670	1548	140	9	0	0	0	-4	3	0.1574
	A	0	0	0	16	141	1566	16630	1502	154	19	2	0	0	-3	4	0.1583
20070618	T	0	0	3	17	133	1520	4148	1469	148	12	0	0	0	-4	3	0.1526
	B	0	0	0	13	136	1207	16483	1225	106	9	2	0	0	-3	4	0.1547
	A	0	0	0	7	118	1185	16622	1120	112	16	1	0	0	-3	4	0.1537
20070619	T	0	0	2	13	194	1783	4574	1853	181	20	0	0	0	-4	3	0.2103
	B	0	0	0	18	159	1380	16979	1502	136	11	1	0	0	-3	4	0.2178
	A	0	0	0	7	149	1400	17041	1427	140	18	4	0	0	-3	4	0.2180
20070620	T	0	1	8	61	408	2487	5261	2370	368	50	3	0	0	-6	4	0.4735
	B	0	2	6	59	341	2194	16168	2254	246	26	1	1	0	-5	6	0.4338
	A	0	2	5	34	317	2335	16145	2117	302	38	2	1	0	-7	6	0.4327
20070621	T	0	5	25	108	619	2724	5270	2696	661	116	25	6	0	-5	8	0.7131
	B	0	1	10	81	462	2558	15045	2752	444	63	3	4	0	-5	5	0.7327
	A	0	1	6	55	409	2807	14946	2621	464	100	10	4	0	-5	6	0.7349
20070622	T	0	13	19	125	575	2525	5363	2431	551	124	33	6	0	-7	7	0.6890
	B	0	8	34	79	489	2471	14718	2697	398	54	15	3	0	-6	9	0.7888
	A	0	1	16	78	449	2762	14579	2491	470	91	20	9	0	-5	10	0.7859
20070625	T	0	8	24	102	551	2631	5765	2590	560	105	17	4	0	-7	5	0.8449
	B	0	12	19	90	435	2495	15104	2609	412	74	13	1	0	-8	6	0.8479
	A	0	9	14	69	395	2706	15065	2436	436	111	14	9	0	-8	8	0.8427
20070626	T	0	5	20	118	653	2677	5549	2724	595	104	23	2	0	-5	5	0.7170
	B	0	4	20	109	501	2473	15462	2666	448	60	9	2	0	-6	6	0.7140
	A	0	4	6	68	479	2764	15368	2426	516	99	19	5	0	-6	6	0.7116
20070627	T	0	1	6	56	435	2610	5595	2546	542	76	11	1	0	-5	5	0.6057
	B	0	0	10	49	334	2348	15779	2461	403	43	7	2	0	-4	5	0.5414
	A	0	1	6	38	305	2512	15698	2366	415	83	10	2	0	-5	6	0.5477
20070628	T	1	8	24	110	455	2242	4555	2171	488	126	25	9	0	-46	6	0.5483
	B	1	11	26	80	428	2089	14457	2274	376	90	15	9	0	-47	8	0.5697
	A	1	4	16	89	363	2235	14557	2004	443	115	18	11	0	-48	6	0.5718
20070629	T	0	8	23	118	564	2486	5137	2505	568	99	16	6	0	-8	5	0.7242
	B	0	9	33	109	459	2416	14930	2511	422	95	23	7	0	-6	7	0.6679
	A	0	3	22	79	444	2612	14895	2326	501	107	12	13	0	-5	8	0.6656

Table 4: Some descriptive statistics for the high-frequency data.

9.1 Exact vs Approximate

First we will compare the two estimators, MC and MC*. We compute both estimators using three price series, transaction prices, ask quotes, and bid quotes, and six different orders for the Markov chain. This leads to 18 pairs of estimates for each of the 21 trading days. These estimates are given in Table 5.

The two estimators are empirically indistinguishable. So we will recommend the MC estimator, because its expression is given in closed form and therefore much faster to compute.

Next we compare the estimates across different values for k . We note that MC₁ (and in some cases both MC₁ and MC₂) can be quite different from those based on a larger k . We attribute the to inhomogeneity, that cannot be accounted for by small k .

		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$						
20070601	T	0.2395	0.2394	0.2558	0.2557	0.2393	0.2392	0.2313	0.2313	0.2327	0.2326	0.2142	0.2142
	B	0.2393	0.2393	0.2535	0.2534	0.2381	0.2381	0.2224	0.2224	0.2221	0.2223	0.2178	0.2176
	A	0.2397	0.2396	0.2584	0.2583	0.2307	0.2307	0.2210	0.2211	0.2120	0.2119	0.2054	0.2053
20070604	T	0.1663	0.1664	0.1784	0.1785	0.1686	0.1686	0.1652	0.1652	0.1757	0.1757	0.1828	0.1828
	B	0.1638	0.1638	0.1832	0.1832	0.1746	0.1746	0.1644	0.1644	0.1564	0.1564	0.1549	0.1550
	A	0.1681	0.1681	0.1943	0.1944	0.1685	0.1685	0.1823	0.1823	0.1734	0.1735	0.1739	0.1738
20070605	T	0.3299	0.3295	0.3157	0.3155	0.2900	0.2900	0.2882	0.2888	0.2938	0.2953	0.2831	0.2843
	B	0.2948	0.2948	0.3101	0.3102	0.2825	0.2826	0.2718	0.2721	0.2683	0.2684	0.2561	0.2564
	A	0.2946	0.2945	0.2968	0.2967	0.2731	0.2731	0.2791	0.2791	0.2737	0.2737	0.2607	0.2604
20070606	T	0.2877	0.2877	0.3088	0.3088	0.2856	0.2857	0.2928	0.2928	0.2846	0.2847	0.2814	0.2816
	B	0.2917	0.2917	0.2922	0.2921	0.2677	0.2677	0.2796	0.2795	0.2667	0.2667	0.2645	0.2645
	A	0.2876	0.2876	0.2949	0.2949	0.2585	0.2585	0.2705	0.2704	0.2768	0.2766	0.2924	0.2920
20070607	T	0.6197	0.6196	0.6581	0.6579	0.6053	0.6051	0.5798	0.5795	0.5666	0.5664	0.5836	0.5839
	B	0.6632	0.6632	0.6635	0.6635	0.6016	0.6017	0.5794	0.5795	0.5542	0.5544	0.5468	0.5467
	A	0.6677	0.6676	0.6722	0.6722	0.6009	0.6009	0.5665	0.5666	0.5780	0.5777	0.5832	0.5829

Table 5: A comparison of MC and MC* estimated with six different orders, k , using SPY trade and quote data for the the first five trading days of June, 2007.

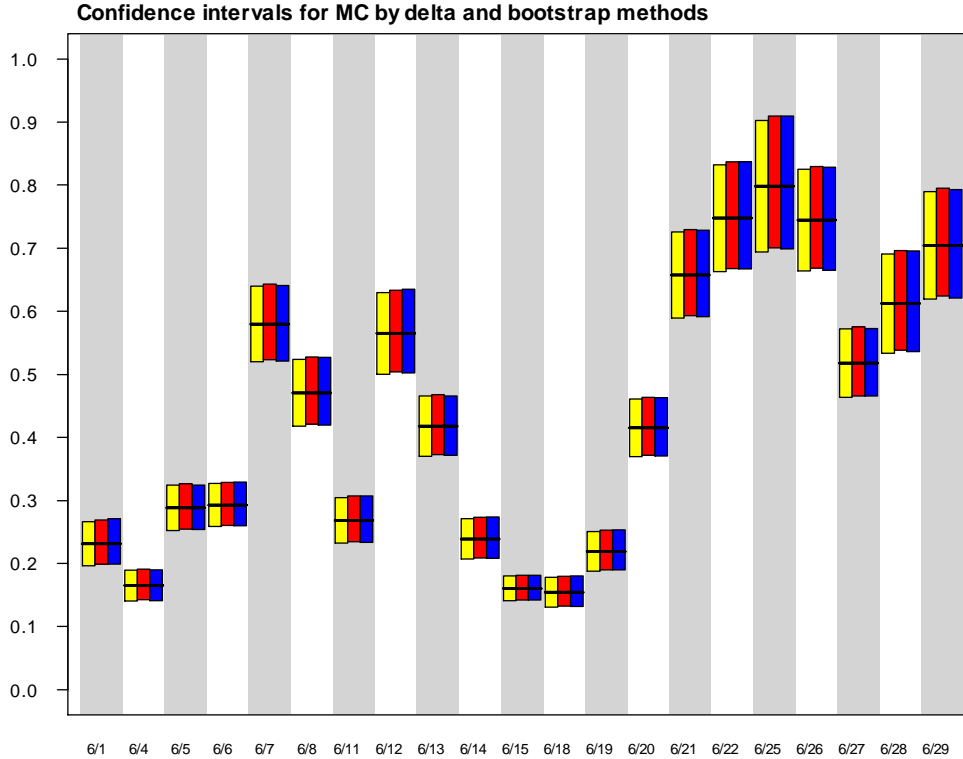


Figure 2: Three 95% confidence intervals (CI) for each daily estimate of MC_4 . The first two CIs are computed with the δ -method, using the central limit results for MC and $\log MC$, respectively. The last CI is based on the bootstrap with $B = 5,000$.

9.2 Confidence Interval by the δ -Method and the Bootstrap

We construct 95% confidence intervals for MC using the three methods we discussed in Section 4. These are the two confidence intervals based on the central limit theorems we obtained for MC and $\log(MC)$, which basically amounts to computing estimate of their asymptotic variance. The third confidence interval is constructed with the bootstrap, using $B = 5,000$ resamples. The confidence intervals for the estimator based on transaction data with $k = 4$ are plotted in Figure 2. We see a remarkable agreement between the three confidence intervals. So in situations, such as the present one, with thousands of intraday returns, we recommend to use one of the asymptotic methods for constructing confidence intervals, because these can be compute from analytical expressions.

Numerical values for the end point of the confidence intervals are given in a table (Table 7) in a separate appendix, both for the case $k = 3$ and $k = 4$. The confidence intervals for the estimator based on $k = 3$ are very similar to those for $k = 4$.

9.3 Sensitivity to Censoring

The jump on June 28, 2007, changes the transaction price from 151.00 to 150.54, corresponding to a log-return of -0.3051% , that would contribute about 0.093 to the quadratic variations. Figure 3 displays the MC_3 estimator using different state censoring schemes. The first estimator is the one presented earlier, where the jump on June 28th is omitted and infrequent states are aggregated into fewer states, specifically $\Delta X_{T_i} \in [+5, +10]$ cents are combined into a single state, and similarly for $\Delta X_{T_i} \in [-10, -5]$. The second estimator is computed with intraday returns, excluding the jump, but without any aggregation of the infrequent states. The third and last estimator is computed with all intraday returns, including the jump. The similarity across the estimators shows that the aggregations has little effect on the MC estimator, and by adding the squared jump to the first two estimates on 6/28 (as illustrated in Figure 3) make these two estimates quite similar to the third MC estimate that is computed with all intraday returns, including the jump.

10 Extensions and Related Estimators

10.1 Quadratic Form Expression

Our Markov chain estimator for the volatility of log-prices is given by

$$\text{MC}_k = \frac{n^2 \langle f, (2\hat{Z} - I)f \rangle_{\hat{\pi}}}{\sum_{i=1}^n X_i^2} = c_{\log} f' \{ \Lambda_{\hat{\pi}} (2\hat{Z} - I) \} f, \quad \text{with} \quad c_{\log} = \frac{n^2}{\sum X_i^2},$$

where k refers to the order of the Markov chain. First we note that

$$f' \{ \Lambda_{\hat{\pi}} (2\hat{Z} - I) \} f = f' \left\{ \frac{1}{2} \Lambda_{\hat{\pi}} (2\hat{Z} - I) + \frac{1}{2} (2\hat{Z}' - I) \Lambda_{\hat{\pi}} \right\} f = f' \{ \Lambda_{\hat{\pi}} \hat{Z} + \hat{Z}' \Lambda_{\hat{\pi}} - \Lambda_{\hat{\pi}} \} f.$$

Recall that $f_s = \mathbf{f}(\mathbf{x}_s)$ is the last element of the k -dimensional vector, \mathbf{x}_s . Thus, if we define the matrix

$$D_{i,j} = 1_{\{f_i = x_j\}}, \quad \text{for } i = 1, \dots, S^k \quad \text{and } j = 1, \dots, S,$$

then $f = Dx$. Consequently $\text{MC}_k = x' M x$, where

$$M = \frac{n^2}{\sum X_i^2} D' \left\{ \Lambda_{\hat{\pi}} \hat{Z} + \hat{Z}' \Lambda_{\hat{\pi}} - \Lambda_{\hat{\pi}} \right\} D,$$

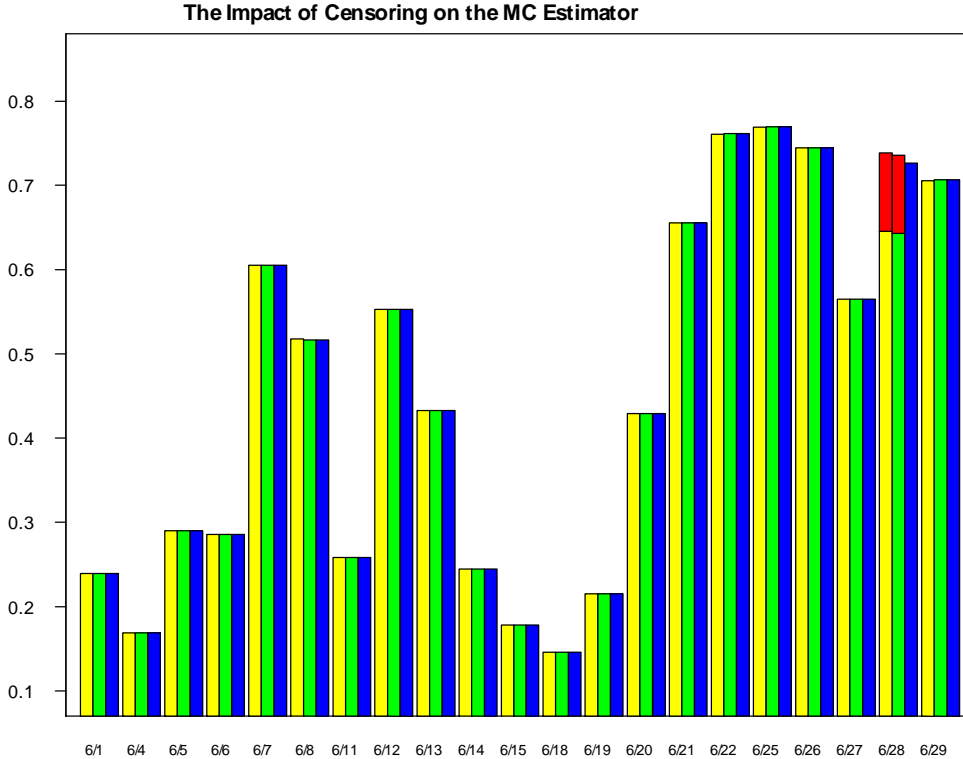


Figure 3: The MC estimator is barely influenced by reducing the state space by combining infrequent states into fewer states. Adding the jump on June 28th to the data increases the estimator by about the same amount as the jump’s contribution to the quadratic variation.

is a symmetric $S \times S$ matrix. We could, for example, study how $M = M_k$ varies as we increase the order of the Markov chain, k .

10.2 The Simplest Case ($S = 2$ and $k = 1$) and the Alternation Estimator

In this Section we consider the simplest case with two states and a Markov chain of order one. We consider the case where the state vector is $f = (\kappa, -\kappa)'$ so that $\Delta X_{T_i} = \pm\kappa$. The transition matrix is here

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix},$$

where $p, q < 1$ ensures that P is ergodic, and leads to the stationary distribution $\pi = \left(\frac{1-q}{2-p-q}, \frac{1-p}{2-p-q} \right)'$. Over a short interval of time, such as a day, is reasonable to assume that

the price process is drift-less, $E(\Delta X_{T_i}) = 0$. This leads to the requirement that $p = q$ so that $\pi = (\frac{1}{2}, \frac{1}{2})'$. The eigenvalues of P are then 1 and $\lambda = 2p - 1$ and it is simple to verify that⁴

$$\text{cov}(\Delta X_j, \Delta X_{j+h}) = \lambda^h \kappa^2. \quad (8)$$

It is well known that the RV $= \sum_{i=1}^n (\Delta X_{T_i})^2 = n\kappa^2$ suffers from autocorrelation in the returns, and here we note that $\{\Delta X_{T_i}\}$ is autocorrelated unless $\lambda \neq 0$. With $p = q$ it is convenient to write the transition matrix as

$$P = \frac{1}{2} \begin{pmatrix} 1 + \lambda & 1 - \lambda \\ 1 - \lambda & 1 + \lambda \end{pmatrix}.$$

The fundamental matrix is in this notation given by

$$Z = \frac{1}{1-\lambda} \begin{pmatrix} 1 - \lambda/2 & -\lambda/2 \\ -\lambda/2 & 1 - \lambda/2 \end{pmatrix}, \quad \text{so that} \quad \Lambda_\pi(2Z - I) = \frac{1}{1-\lambda} \begin{pmatrix} 1/2 & -\lambda/2 \\ -\lambda/2 & 1/2 \end{pmatrix}.$$

So the Markov-based estimator is simply given by

$$\text{MC}^\# = \langle f, (2\hat{Z} - I)f \rangle_{\hat{\pi}} = \frac{1 + \hat{\lambda}}{1 - \hat{\lambda}} n\kappa^2, \quad (9)$$

where $\hat{\lambda}$ is an estimator of λ . Not surprisingly is the RV $= n\kappa^2$ biased (unless $\lambda = 0$), so the scaling-factor, $\frac{1+\hat{\lambda}}{1-\hat{\lambda}}$, serves to offsets the bias that is induced by serial correlation in $\{\Delta X_{T_i}\}$.

10.2.1 The Alternation Estimator

Large (2006) was one of the first paper to propose an estimator that takes advantage of the discreteness of return data. He counts the number of consecutive increments in the same direction as *continuations*, c , and consecutive changes in opposite directions as *alternations*, a , so that $n = a + c$.⁵ The *alternation estimator* by Large (2006) is given by

$$\text{ALT} = \frac{c}{a} n\kappa^2.$$

We now show that the alternation estimator by Large (2006) is identical to $\text{MC}^\#$, with the implicit assumptions that $k = 1$ and the particular requirements that $S = 2$ with $f = (+\kappa, -\kappa)'$.

⁴An exercise in Campbell et al. (1997, p. 145) asks the reader to derive properties of returns in this model.

⁵A common terminology for c and a is concordant and discordant, respectively.

The maximum likelihood estimator of P , subject to the constraint that $p = q$ is

$$\hat{P} = \begin{pmatrix} \frac{c}{n} & \frac{a}{n} \\ \frac{a}{n} & \frac{c}{n} \end{pmatrix},$$

so that (by invariance) the maximum likelihood estimator of λ is $\hat{\lambda} = \det \hat{P} = \frac{c^2 - a^2}{n^2}$. The implication is that

$$\frac{1 + \hat{\lambda}}{1 - \hat{\lambda}} = \frac{n^2 + c^2 - a^2}{n^2 - c^2 + a^2} = \frac{2ac + 2c^2}{2ac + 2a^2} = \frac{c}{a},$$

which shows that ALT is a special case of the Markov estimator, (9). The noise permitted in Large's framework is limited by the requirement that $\{\Delta X_{T_i}\}$ has an AR(1) autocorrelation structure, see (8). The framework of the present paper allows for higher-order Markov chains and more than two states.

10.3 Markov Estimator by Eigenvalues and Eigenvectors

In the simple two-state Markov chain of order one the estimator of the quadratic variation is simply the realized variance scaled by a constant that depends on λ , which is the second eigenvalue of P , see (9). Thus an interesting question is whether the Markov estimator can be expressed in terms of the eigenvalues of P in general. The answer is yes when P is diagonalizable.

Let $\lambda_1 = 1 > \lambda_2 \geq \dots \lambda_S > -1$ be the eigenvalues of P and suppose that $P = V\Lambda W'$ where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_S)$. It can then be shown that

$$\text{MC}^\# = \langle \xi, \xi \rangle_\pi + 2 \sum_{s=2}^S \frac{\lambda_s}{1 - \lambda_s} \langle \xi, v_s w_s' \xi \rangle_\pi,$$

where $V = (v_1, \dots, v_S)$ and $W = (w_1, \dots, w_S)$. However, $\text{MC}^\# = \langle \xi, (2Z - I)\xi \rangle_\pi$ is simpler to compute and can be applied even in the case where P cannot be diagonalized.

11 Conclusion

In this paper we have established some general results concerning filtering semimartingales that are contaminated with measurement errors. We have shown that filtering with a theoretical information set, can eliminate the problems that are caused by market microstructure noise, under very weak assumptions. However, while the realized variance computed with returns of the filtered price produce a good estimator, the realized variance of the filtered

returns will not be a proper estimator of the quadratic variation. We have shown that the Markov chain framework offers a convenient and simple way to implement the filtering of high frequency data. Fortunately, the Markov chain estimator is easily adapted to estimate volatility of log-prices, which is usually the object of interest. The estimator does not appear to be hurt by the vast number of parameters that are estimated in higher-order Markov chains, and the estimator is surprisingly robust to inhomogeneity, and the presence of infrequent states, such as jumps.

The Markov chain estimator has many attractive features. It is very simple to compute and the same can be said about our estimator of the asymptotic variance. A choice variable is the order of the Markov chain, k . Information criteria may offer useful guidance about the choice for k , but $k = 3$ has worked well in all our simulations and empirical studies. Developing a data dependent procedure for selecting k is an interesting problem that we leave for future research. Further experience gained from applying the estimator to additional data, including other types of data, will offer useful guidance to this problem. Selecting k is similar to selecting the bandwidth parameter that is needed for the implementation of related estimators. Empirical methods for selecting the bandwidth are rather complex, and depend on preliminary estimates of features of the noise and the underlying process. Here, we have found the Markov chain estimator to be quite insensitive to the choice of k , once $k \geq 2$. In empirical situations where the estimator is found to be very sensitive to the choice for k , beyond the case $k = 1$, we recommend computing MC locally over time-intervals where inhomogeneity is less of an issue, and then combine the subsample estimates into an estimate for the whole period.

We have discussed that jumps can be dealt with by censoring large price changes, using an ad-hoc threshold. Price increments that are classified as jumps in this way can then be modelled separately as we discussed in Section 8.1.

It is our experience that the Markov chain estimator is remarkable resilient to inhomogeneity and the presence of jumps. In fact, removing the most extreme values does not affect the estimator much. We are only aware of one issue that can seriously affect the Markov chain estimator. The problem arises when the empirical transition matrix has an absorbing state, which happens when the last observed state is the only observation of that state. This phenomenon is increasingly relevant as we increase the order of the Markov chains. Fortunately, the remedy for this problem is very simple. By artificially extending the sam-

ple with k observations, as explained in Section 8.3, we can make the last observed state identical to the first observed state, which eliminates the possibility of \hat{P} having absorbing states.

There are several interesting extensions of the analysis presented here. For instance, the Markov chain filtered price may be used to compute other quantities, including properties of the implied noise process, or semi-variance, see Barndorff-Nielsen, Kinnebrock & Shephard (2008). Though we suspect that estimators of such quantities will be more sensitive to inhomogeneity than is the case for our estimator of the volatility. There are a number of ways that the Markov chain framework can be extended to estimate covolatility, that would enable the estimation of quadratic variation of multivariate processes. We will pursue this extensions in future research.

References

- Andersen, T., Dobrev, D. & Schaumburg, E. (2008), ‘Duration-based volatility estimation’, *working paper* .
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), ‘The distribution of realized stock return volatility’, *Journal of Financial Economics* **61**, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), ‘The distribution of exchange rate volatility’, *Journal of the American Statistical Association* **96**, 42–55. Correction published in 2003, volume 98, page 501.
- Bandi, F. M. & Russell, J. R. (2006), ‘Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations’, *working paper* .
- Bandi, F. M. & Russell, J. R. (2008), ‘Microstructure noise, realized variance, and optimal sampling’, *Review of Economic Studies* **75**, 339–369.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008a), ‘Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise’, *Econometrica* . Forthcoming.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008b), ‘Multivariate realised kernels: consistent positive semin-definite estimators of the covariation of equity prices with noise and non-synchronous trading’, *working paper* .
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008c), ‘Realised kernels in practice: Trades and quotes’, *forthcoming in Econometrics Journal* .
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008d), ‘Subsampling realised kernels’, *Journal of Econometrics* . Forthcoming.
- Barndorff-Nielsen, O. E., Kinnebrock, S. & Shephard, N. (2008), ‘Measuring downside risk - realised semi-variance’, *working paper* .
- Barndorff-Nielsen, O. E. & Shephard, N. (2002), ‘Econometric analysis of realised volatility and its use in estimating stochastic volatility models’, *Journal of the Royal Statistical Society, Series B* **64**, 253–280.
- Barndorff-Nielsen, O. E. & Shephard, N. (2004), ‘Power and bipower variation with stochastic volatility and jumps (with discussion)’, *Journal of Financial Econometrics* **2**, 1–48.

- Barndorff-Nielsen, O. E. & Shephard, N. (2005), How accurate is the asymptotic approximation to the distribution of realised volatility?, in D. W. K. Andrews & J. H. Stock, eds, 'Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg', Cambridge University Press, Cambridge, pp. 306–331.
- Basawa, I. & Green, T. (1990), 'Asymptotic bootstrap validity for finite Markov chains.', *Communications in Statistics: Theory and Methods* **19**(4), 1493–1510.
- Brémaud, P. (1999), *Markov chains: Gibbs fields, Monte Carlo Simulation, and Queues*, Springer.
- Campbell, J. Y., Lo, A. W. & Mackinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.
- Christensen, K., Oomen, R. C. & Podolskij, M. (2008), 'Realised quantile-based estimation of the integrated variance', *working paper* .
- Christensen, K. & Podolskij, M. (2007), 'Asymptotic theory for range-based estimation of integrated volatility of a continuous semi-martingale', *Journal of Econometrics* **141**, 323–349.
- Delattre, S. & Jacod, J. (1997), 'A central limit theorem for normalized functions of the increments of a diffusion process in the presence of round off errors', *Bernoulli* **3**, 1–28.
- Derman, C. (1956), 'Some asymptotic distribution theory for markov chains with a denumerable number of states', *Biometrika* **43**(3/4), 285–294.
- Duflo, M. (1997), *Random Iterative Models*, Springer.
- Goncalves, S. & Meddahi, N. (2006), 'Box-cox transforms for realized volatility'. Unpublished working paper.
- Goncalves, S. & Meddahi, N. (2008), 'Bootstrapping realized volatility', *Econometrica* . Forthcoming.
- Griffin, J. & Oomen, R. (2006), 'Sampling returns for realized variance calculations: tick time or transaction time?', *Forthcoming in Econometric Reviews* .
- Hansen, P. R., Large, J. & Lunde, A. (2008), 'Moving average-based estimators of integrated variance', *Econometric Reviews* . forthcoming.
- Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–218. The 2005 Invited Address with Comments and Rejoinder.
- Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2008), 'Microstructure noise in the continuous case: The pre-averaging approach', *working paper* .
- Kemeny, J. & Snell, J. (1976), *Finite Markov Chains*, Springer.
- Large, J. (2006), 'Estimating quadratic variation when quoted prices jump by a constant increment', *Journal of Econometrics* . Forthcoming.
- Li, Y. & Mykland, P. (2007), 'Are volatility estimators robust to modelling assumptions?', *Bernoulli* **13**, 601–622.
- Li, Y. & Mykland, P. A. (2006), 'Determining the volatility of a price process in the presence of rounding errors'. Working paper.
- Maheu, J. M. & McCurdy, T. H. (2002), 'Nonlinear features of realised FX volatility', *Review of Economics and Statistics* **84**, 668–681.
- Meddahi, N. (2002), 'A theoretical comparison between integrated and realized volatilities', *Journal of Applied Econometrics* **17**, 479–508.
- Mykland, P. A. & Zhang, L. (2008), 'Inference for continuous semimartingales observed at high frequency: A general approach', *working paper* .
- Phillips, P. C. & Yu, J. (2008), 'Information loss in volatility measurement with flat price trading', *working paper* .

- Russell, J. R. & Engle, R. F. (2006), ‘A discrete-state continuous-time model of financial transaction prices and times’, *Journal of Business and Economic Statistics* .
- Schweitzer, P. (1968), ‘Perturbation theory and finite markov chains’, *Journal of Applied Probability* **5**, 401–413.
- Shao, J. & Tu, D. (1995), *The Jackknife and Bootstrap*, Springer.
- Zhang, L. (2006), ‘Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach’, *Bernoulli* **12**, 1019–1043.
- Zhang, L., Mykland, P. A. & Aït-Sahalia, Y. (2005), ‘A tale of two time scales: Determining integrated volatility with noisy high frequency data’, *Journal of the American Statistical Association* **100**, 1394–1411.
- Zhou, B. (1996), ‘High-frequency data and volatility in foreign-exchange rates’, *Journal of Business and Economic Statistics* **14**, 45–52.
- Zhou, B. (1998), Parametric and nonparametric volatility measurement, in C. L. Dunis & B. Zhou, eds, ‘Nonlinear Modelling of High Frequency Financial Time Series’, John Wiley Sons Ltd, chapter 6, pp. 109–123.

Appendix of Proofs

Proof of Lemma 1. We need to show that $\mathbb{E} |\mathbb{E}(U_{t+h}|\mathcal{G}_t) - \mathbb{E}(U_t)| \rightarrow 0$ when $h \rightarrow \infty$.

$$\begin{aligned} \mathbb{E} |\mathbb{E}(U_{t+h}|\mathcal{G}_t) - \mathbb{E}(U_{t+h})| &= \int |\mathbb{E}(U_{t+h}|d\omega) - \mathbb{E}(U_{t+h})| \mu(d\omega \in \mathcal{G}_t) \\ &= \int \left| \int U_{t+h}(\omega') (\Pr(U_{t+h} \in d\omega')|d\omega) \right. \\ &\quad \left. - \Pr(U_{t+h} \in d\omega') \right) \mu(d\omega') | \mu(d\omega \in \mathcal{G}_t) \leq \mathbb{E}(|U_{t+h}|)\phi(h), \end{aligned}$$

which converges to 0 as $h \rightarrow \infty$, by the definition of ϕ -mixing. \square

Proof of Theorem 1. By the martingale property $\mathbb{E}(M_{t+h}|\mathcal{G}_t) = M_t$ we have,

$$\mathbb{E}(X_{t+h}|\mathcal{G}_t) = M_t + \mathbb{E}(\text{FV}_{t+h}|\mathcal{G}_t) + \mathbb{E}(U_{t+h}|\mathcal{G}_t),$$

where the last term is assumed to be continuous finite variation process. Now complete the proof by showing that $\mathbb{E}(\text{FV}_{t+h}|\mathcal{G}_t)$ is a continuous finite variation process.

Let $0 = T_0 < \dots < T_n = T$ be a partition of $[0, T]$. We need to show that

$$\sum_{i=2}^n |\mathbb{E}(\text{FV}_{T_i}|\mathcal{G}_{T_{i-1}}) - \mathbb{E}(\text{FV}_{T_{i-1}}|\mathcal{G}_{T_{i-2}})| < \infty,$$

when $\sup_{1 \leq i \leq n} |T_i - T_{i-1}| \rightarrow 0$. First we note that

$$\begin{aligned} |\mathbb{E}(\text{FV}_{T_i}|\mathcal{G}_{T_{i-1}}) - \mathbb{E}(\text{FV}_{T_{i-1}}|\mathcal{G}_{T_{i-2}})| &\leq |\mathbb{E}(\text{FV}_{T_i}|\mathcal{G}_{T_{i-1}}) - \text{FV}_{T_{i-1}}| + |\text{FV}_{T_{i-1}} - \text{FV}_{T_{i-2}}| \\ &\quad + |\text{FV}_{T_{i-2}} - \mathbb{E}(\text{FV}_{T_{i-1}}|\mathcal{G}_{T_{i-2}})|. \end{aligned}$$

By summing over i we get:

$$\sum_{i=2}^n |\mathbb{E}(\text{FV}_{T_i}|\mathcal{G}_{T_{i-1}}) - \mathbb{E}(\text{FV}_{T_{i-1}}|\mathcal{G}_{T_{i-2}})| \leq 2 \sum_{i=1}^n \mathbb{E}(|\text{FV}_{T_i} - \text{FV}_{T_{i-1}}| | \mathcal{G}_{T_{i-1}})$$

$$+ \sum_{i=1}^n |\text{FV}_{T_i} - \text{FV}_{T_{i-1}}|.$$

The last term is finite, since FV_t has finite variation, and if $\sum_{i=1}^n \mathbb{E}(|\text{FV}_{T_i} - \text{FV}_{T_{i-1}}| | \mathcal{G}_{T_{i-1}}) \not\leq \infty$ would imply $\mathbb{E}(\sum_{i=1}^n |\text{FV}_{T_i} - \text{FV}_{T_{i-1}}|) \not\leq \infty$ which contradicts the fact that FV_t has locally integrable variation. Next we address continuity.

$$|\mathbb{E}(\text{FV}_{t+\epsilon+h} | \mathcal{G}_{t+\epsilon}) - \mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_t)| \leq |\mathbb{E}(\text{FV}_{t+\epsilon+h} | \mathcal{G}_{t+\epsilon+h}) - \mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_{t+\epsilon+h})| + |\mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_{t+\epsilon+h}) - \mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_{t+h})|,$$

where the first term of the right hand side is bounded by

$$|\mathbb{E}(\text{FV}_{t+\epsilon+h} | \mathcal{G}_{t+\epsilon+h}) - \mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_{t+\epsilon+h})| \leq \mathbb{E}(|\text{FV}_{t+h+\epsilon} - \text{FV}_{t+h}| | \mathcal{G}_{t+h+\epsilon}),$$

which vanishes because $\lim_{\epsilon \rightarrow 0} |\text{FV}_{t+h+\epsilon} - \text{FV}_{t+h}| = 0$ by continuity of FV_t . For the second term we have $|\mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_{t+\epsilon+h}) - \mathbb{E}(\text{FV}_{t+h} | \mathcal{G}_{t+h})| \rightarrow 0$ by continuity of \mathcal{G}_t at $t+h$. \square

Proposition A.1 *Some properties of $Z^{(h)} = I + \sum_{j=1}^h (P^j - \Pi)$, when P generates an ergodic Markov chain.*

(i) $(P - \Pi)^j = P^j - \Pi$ so that $Z^{(h)} = I + \sum_{j=1}^h (P - \Pi)^j$.

(ii) $Z = \lim_{h \rightarrow \infty} Z^{(h)} = (I - P + \Pi)^{-1}$.

(iii) $Z\iota = \iota$, $\pi'Z = \pi'$, and $PZ = ZP = Z - I + \Pi$.

Proof. We prove (i) by induction. The identity is obvious for $j = 1$. Now suppose that the identity holds for j . Then

$$(P - \Pi)^{j+1} = (P - \Pi)(P^j - \Pi) = P^{j+1} - \Pi P^j + \Pi^2 - P\Pi = P^{j+1} - \Pi,$$

where the last identity follows from $\Pi P^j = \Pi^2 = P\Pi = \Pi$.

(ii) Since the chain is ergodic we have $\|P - \Pi\| < 1$. By the Perron-Frobenius theorem, P^h converges geometrically to Π with rate given by the second largest eigenvalue of P , namely $\|P^h - \Pi\| = O(|\lambda_2|^h)$. The series $I + \sum_{j=1}^{\infty} (P^j - \Pi) = I + \sum_{j=1}^{\infty} (P - \Pi)^j$ is therefore absolutely convergent with $I + \sum_{j=1}^{\infty} (P - \Pi)^j = (I - (P - \Pi))^{-1} = Z$.

(iii) $P^j\iota = \iota$ and $\pi'P^j = \pi'$ for any $j \in \mathbb{N}$; and $\Pi\iota = \iota$ and $\pi'\Pi = \pi'$, so that have $(P^j - \Pi)\iota = \pi'(P^j - \Pi) = 0$. The first two results follow from $Z = I + \sum_{j=1}^{\infty} (P^j - \Pi)$. Next, $PZ = ZP = P + \sum_{j=1}^{\infty} (P^{j+1} - \Pi)$ and

$$P + \sum_{j=1}^{\infty} (P^{j+1} - \Pi) = P + \sum_{j=1}^{\infty} (P^j - \Pi) - P + \Pi = Z - I + \Pi.$$

■

Proof of Lemma 2. We have $\mathbb{E}(\Delta X_{T_{i+h}} | \Delta X_{T_i} = x_r) = (P^h x)_r = e_r' P^h x$ so that for $(\Delta X_{T_i}, \Delta X_{T_{i-1}}) = (x_r, x_s)$ we have

$$y^{(h)}(\mathbf{x}_r, \mathbf{x}_s) = f_s + \sum_{l=1}^h \mathbb{E}(\Delta X_{T_{i+l}} | \Delta \mathcal{X}_{T_i} = \mathbf{x}_s) - \sum_{l=1}^h \mathbb{E}(\Delta X_{T_{i+l-1}} | \Delta \mathcal{X}_{T_{i-1}} = \mathbf{x}_r)$$

$$\begin{aligned}
&= \left[e'_s + e'_s \sum_{l=1}^h P^l - e'_r \sum_{l=1}^h P^l \right] f \\
&= \left[e'_s + e'_s \sum_{l=1}^h (P^l - \Pi) - e'_r \sum_{l=1}^h (P^l - \Pi) \right] f \\
&= \left[e'_r + e'_s \{I + \sum_{l=1}^h (P^l - \Pi)\} - e'_r \{I + \sum_{l=1}^h (P^l - \Pi)\} \right] f \\
&= \left[e'_r + e'_s Z^{(h)} - e'_r Z^{(h)} \right] x,
\end{aligned}$$

and the result follows. \square

Proof of Theorem 2. Since ΔX_{T_i} is ergodic, then so is the overlapping chain $(\Delta X_{T_i}, \dots, \Delta X_{T_{i-k}})$. So by the ergodic theorem, we simply need to compute the expected value, as this will be the limit almost surely. The stationary distribution of $(\Delta X_{T_i}, \dots, \Delta X_{T_{i-k}})$ is given by

$$\Pr \{ (\Delta X_{T_{i+1}}, \Delta \mathcal{X}_{T_i}) = (x_s, \mathbf{x}_r) \} = \pi_r P_{r,s}.$$

Now we will show that $\text{RV}_F^{(h)} \xrightarrow{a.s.} \langle \xi, (2Z^{(h)} - I)\xi \rangle_\pi + 2\langle (I - Z^{(h)})\xi, (P^{h+1} - \Pi)\xi \rangle_\pi$. To simplify our expressions, we write $\Delta Y_{T_i} = y^{(h)}(\Delta \mathcal{X}_{T_{i-1}}, \Delta \mathcal{X}_{T_i})$, and recall that $y^{(h)}(\mathbf{x}_r, \mathbf{x}_s) = e'_r(I - Z^{(h)})f + e'_s Z^{(h)}f$, such that

$$\begin{aligned}
\mathbb{E}\{(\Delta Y_{T_i})^2\} &= \sum_{r,s} \pi_r P_{r,s} \{y^{(h)}(\mathbf{x}_r, \mathbf{x}_s)\}^2 \\
&= f'(I - Z^{(h)})' \sum_{r,s} \pi_r P_{r,s} e_r e'_r (I - Z^{(h)})f \\
&\quad + 2f'(I - Z^{(h)})' \sum_{r,s} \pi_r P_{r,s} e_r e'_s Z^{(h)}f + f' Z^{(h)'} \sum_{r,s} \pi_r P_{r,s} e_s e'_s Z^{(h)}f.
\end{aligned}$$

The identities

$$\sum_{r,s} \pi_r P_{r,s} e_r e'_r = \sum_{r,s} \pi_r P_{r,s} e_s e'_s = \Lambda_\pi, \quad \text{and} \quad \sum_{r,s} \pi_r P_{r,s} e_r e'_s = \Lambda_\pi P,$$

simplifies our expression to

$$\begin{aligned}
\mathbb{E}\{(\Delta Y_{T_i})^2\} &= \langle (I - Z^{(h)})f, (I - Z^{(h)})f \rangle_\pi + 2\langle (I - Z^{(h)})f, PZ^{(h)}f \rangle_\pi + \langle Z^{(h)}f, Z^{(h)}f \rangle_\pi \\
&= \langle (I - Z^{(h)})f, (I - 2Z^{(h)})f \rangle_\pi + \langle f, Z^{(h)}f \rangle_\pi + 2\langle (I - Z^{(h)})f, PZ^{(h)}f \rangle_\pi \\
&= \langle f, Z^{(h)}f \rangle_\pi + \langle (I - Z^{(h)})f, (I + 2(P - I)Z^{(h)})f \rangle_\pi.
\end{aligned}$$

Next we note that

$$PZ^{(h)} = P + \sum_{l=1}^h (P^{l+1} - \Pi) = \Pi + \sum_{l=1}^{h+1} (P^l - \Pi) = \Pi + Z^{(h+1)} - I,$$

and that $\langle (I - Z^{(h)})f, \Pi f \rangle_\pi = f'(I - Z^{(h)})' \Lambda_\pi \pi' f = f'(I - Z^{(h)})' \pi \pi' f = 0$ since

$$\pi' Z^{(h)} = \pi' + \pi' \sum_{l=1}^h (P - \Pi)^l = \pi'.$$

So

$$\begin{aligned}
\mathbb{E}\{(\Delta Y_{T_i})^2\} &= \langle f, Z^{(h)}f \rangle_\pi + \langle (I - Z^{(h)})f, (I + 2(\Pi - I + Z^{(h+1)} - Z^{(h)}))f \rangle_\pi \\
&= \langle f, Z^{(h)}f \rangle_\pi + \langle (I - Z^{(h)})f, (-I + 2Z^{(h+1)} - 2Z^{(h)})f \rangle_\pi \\
&= \langle f, (2Z^{(h)} - I)f \rangle_\pi + \langle (I - Z^{(h)})f, 2(Z^{(h+1)} - Z^{(h)})f \rangle_\pi,
\end{aligned}$$

and since $Z^{(h+1)} - Z^{(h)} = P^{h+1} - \Pi$, we have established that $\text{RV}_F^{(h)} \xrightarrow{a.s.} \langle \xi, (2Z^{(h)} - I)\xi \rangle_\pi + 2\langle (I - Z^{(h)})\xi, (P^{h+1} - \Pi)\xi \rangle_\pi$.

Finally, $\text{RV}_F = \lim_{h \rightarrow \infty} \text{RV}_F^{(h)}$, follows from $\|Z^{(h)} - Z\| = \|\sum_{l>h} P^l - \Pi\| = O(|\lambda_2|^h)$, where λ_2 is the second largest eigenvalue of P , so that $Z^{(h)}$ converges to Z geometrically. \square

Lemma A.1 *Asymptotic distribution of the empirical stationary distribution.*

Let $n_{s,\cdot} = \sum_{i=1}^n 1_{\{\Delta X_{T_i}=x_s\}}$. Then

$$\sqrt{n} \left\{ \left(\frac{n_{1,\cdot}}{n}, \dots, \frac{n_{S,\cdot}}{n} \right)' - \pi \right\} \xrightarrow{d} \text{N}(0, \Lambda_\pi Z + Z' \Lambda_\pi - \pi \pi' - \Lambda_\pi).$$

Proof of Lemma A.1 By a Cramér-Wold argument it follows by Proposition 1 that the limiting distribution is Gaussian, because any linear combination of $\left(\frac{n_{1,\cdot}}{n}, \dots, \frac{n_{S,\cdot}}{n}\right)$ can be expressed as a function $\sqrt{n}\Delta X_{T_i}$ using the indicator functions, $1_{\{\Delta X_{T_i}=x_s\}}$, $s = 1, \dots, S$. So we simply need to derive the asymptotic variance. Denote the asymptotic variance matrix by Σ . For the s -th diagonal element, we use $g(\sqrt{n}\Delta X_{T_i}) = 1_{\{\Delta X_{T_i}=x_s\}}$, so that the state vector is the s -th unit vector, $g = e_s$. By Proposition 1 we obtain the asymptotic variance

$$\Sigma_{s,s} = \langle e_s, (2Z - I - \Pi)e_s \rangle_\pi = 2\pi_s Z_{s,s} - \pi_s(1 + \pi_s).$$

Next, for $r \neq s$, we consider $g(\sqrt{n}\Delta X_{T_i}) = 1_{\{\Delta X_{T_i}=x_r\}} + 1_{\{\Delta X_{T_i}=x_s\}}$, which leads to $g = e_r + e_s$, and the asymptotic variance,

$$\langle e_r + e_s, (2Z - I - \Pi)(e_r + e_s) \rangle_\pi.$$

Similarly with $g(\sqrt{n}\Delta X_{T_i}) = 1_{\{\Delta X_{T_i}=x_r\}} - 1_{\{\Delta X_{T_i}=x_s\}}$ we obtain the asymptotic variance

$$\langle e_r - e_s, (2Z - I - \Pi)(e_r - e_s) \rangle_\pi.$$

We now obtain the covariances by the identity $\text{cov}(X, Y) = \frac{1}{4} \{\text{var}(X + Y) - \text{var}(X - Y)\}$, so that

$$\begin{aligned}
\Sigma_{r,s} &= \frac{1}{4} \{ \langle e_r + e_s, (2Z - I - \Pi)(e_r + e_s) \rangle_\pi - \langle e_r - e_s, (2Z - I - \Pi)(e_r - e_s) \rangle_\pi \} \\
&= \frac{1}{4} \{ 2\langle e_r, (2Z - I - \Pi)e_s \rangle_\pi + 2\langle e_s, (2Z - I - \Pi)e_r \rangle_\pi \} \\
&= \frac{1}{4} \{ 2\langle e_r, (2Z - \Pi)e_s \rangle_\pi + 2\langle e_s, (2Z - \Pi)e_r \rangle_\pi \} \\
&= \pi_r Z_{r,s} + \pi_s Z_{s,r} - \pi_r \pi_s.
\end{aligned}$$

This completes the proof. \square

We have the following results for $n_{r,s}$, which is used extensively in our proofs.

Proposition A.2 *Let*

$$\Sigma_{n_{r,s}, n_{u,v}} = \lim_{n \rightarrow \infty} \text{cov} \left(\frac{n_{r,s}}{\sqrt{n}}, \frac{n_{u,v}}{\sqrt{n}} \right) \quad \text{and} \quad \Sigma_{n_{r,s}, n_{u,\cdot}} = \lim_{n \rightarrow \infty} \text{cov} \left(\frac{n_{r,s}}{\sqrt{n}}, \frac{n_{u,\cdot}}{\sqrt{n}} \right).$$

Then

$$\Sigma_{n_{r,s}, n_{u,v}} = P_{r,s} P_{u,v} (\pi_r Z_{s,u} + \pi_u Z_{v,r} - 3\pi_r \pi_u) + \pi_r P_{r,s} \delta_{r,u} \delta_{s,v}, \quad (\text{A.1})$$

$$\Sigma_{n_{r,s}, n_{u,\cdot}} = \pi_r P_{r,s} Z_{s,u} - 2\pi_r \pi_u P_{r,s} + \pi_u P_{r,s} Z_{u,r}. \quad (\text{A.2})$$

The results is due to Derman (1956). Here we state and prove the results with a modern notation, such as the fundamental matrix, Z .

Proof of Proposition A.2. We have

$$\begin{aligned} \frac{\text{cov}(n_{r,s}, n_{u,v})}{n} &= n^{-1} \sum_{i < j} \pi_r P_{r,s} (P^{j-i-1})_{s,u} P_{u,v} + \pi_r P_{r,s} \delta_{r,u} \delta_{s,v} \\ &\quad + n^{-1} \sum_{j < i} \pi_u P_{u,v} (P^{j-i-1})_{v,r} P_{r,s} - n \pi_r P_{r,s} \pi_u P_{u,v} \\ &= \pi_r P_{r,s} \sum_{l=0}^{n-2} \frac{n-l-1}{n} (P^l - \Pi)_{s,u} P_{u,v} + \pi_r P_{r,s} \delta_{r,u} \delta_{s,v} \\ &\quad + \pi_u P_{u,v} \sum_{l=0}^{n-2} \frac{n-l-1}{n} (P^l - \Pi)_{v,r} P_{r,s} - 3\pi_r P_{r,s} \pi_u P_{u,v} \\ &\rightarrow \pi_r P_{r,s} Z_{s,u} P_{u,v} + \pi_u P_{u,v} Z_{v,r} P_{r,s} + \pi_r P_{r,s} \delta_{r,u} \delta_{s,v} - 3\pi_r P_{r,s} \pi_u P_{u,v}, \end{aligned}$$

as $n \rightarrow \infty$. This proves (A.1). Next, summing (A.1) over v yields

$$\begin{aligned} &\sum_v P_{r,s} P_{u,v} (\pi_r Z_{s,u} + \pi_u Z_{v,r} - 3\pi_r \pi_u) + \pi_r P_{r,s} \delta_{r,u} \delta_{s,v} \\ &= \pi_r P_{r,s} Z_{s,u} + \pi_u P_{r,s} \sum_v P_{u,v} Z_{v,r} - 3\pi_r \pi_u P_{r,s} + \pi_r P_{r,s} \delta_{r,u} \end{aligned}$$

and the identity $\sum_v P_{u,v} Z_{v,r} = (PZ)_{u,r} = (Z - I + \Pi)_{u,r}$ leads to

$$\pi_r P_{r,s} Z_{s,u} + \pi_u P_{r,s} Z_{u,r} - \pi_u P_{r,s} \delta_{r,u} + \pi_u \pi_r P_{r,s} - 3\pi_r \pi_u P_{r,s} + \pi_r P_{r,s} \delta_{r,u},$$

which simplifies to (A.2). \square

Proof of Theorem 3. The Gaussian limit follows from Proposition 1 that also yields the expression, $\langle g, (2ZZ - \Pi Z - I)g \rangle_{\pi_Z}$, for the asymptotic variance. To obtain the expression for Σ_{RV_F} that is stated in the Theorem, we note that the variance of $\text{RV}_F = \sum_{r,s=1}^{S^k} n_{r,s} y_{(r,s)}^2$ is given from the variance-covariance of $(n_{r,s})_{r,s=1,\dots,S^k}$ that are denoted by $\Sigma_{n_{r,s}, n_{u,v}}$. The expression for $\Sigma_{n_{r,s}, n_{u,v}}$ are derived in Proposition A.2. \square

Proof of Proposition 2. We know that $\sqrt{n} \left\{ \left(\frac{n_{r,s}}{n}, \frac{n_{r,\cdot}}{n}, \frac{n_{u,v}}{n}, \frac{n_{u,\cdot}}{n} \right)' - \left(\frac{P_{r,s}}{\pi_r}, \pi_r, \frac{P_{u,v}}{\pi_r}, \pi_u \right)' \right\} \xrightarrow{d} N(0, A)$, where the elements of A is given from Proposition A.2. Since $\hat{P}_{r,s} = \frac{n_{r,s}}{n} / \frac{n_{r,\cdot}}{n}$, we can obtain the asymptotic variance by the delta method. Applying the mapping $(x, y, z, w)' \mapsto \left(\frac{x}{y}, \frac{z}{x} \right)'$ to $\left(\frac{n_{r,s}}{n}, \frac{n_{r,\cdot}}{n}, \frac{n_{u,v}}{n}, \frac{n_{u,\cdot}}{n} \right)$, leads to the following gradient:

$$\nabla' = \begin{pmatrix} \pi_r^{-1} & -\pi_r^{-1} P_{r,s} & 0 & 0 \\ 0 & 0 & \pi_u^{-1} & -\pi_u^{-1} P_{u,v} \end{pmatrix}'.$$

The asymptotic covariance of $\hat{P}_{r,s} = \frac{n_{r,s}}{n_{r,\cdot}}$ and $\hat{P}_{u,v} = \frac{n_{u,v}}{n_{u,\cdot}}$ is given as the upper-right (or lower-left) element of $\nabla' A \nabla$, which is given by

$$\begin{aligned} (\Sigma P)_{(r,s)(u,v)} &= (\pi_r^{-1}, -\pi_r^{-1} P_{r,s}) \begin{pmatrix} \Sigma_{n_{r,s}, n_{u,v}} & \Sigma_{n_{r,s}, n_{u,\cdot}} \\ \Sigma_{n_{r,\cdot}, n_{u,v}} & \Sigma_{n_{r,\cdot}, n_{u,\cdot}} \end{pmatrix} \begin{pmatrix} \pi_u^{-1} \\ -\pi_u^{-1} P_{u,v} \end{pmatrix} \\ &= \frac{1}{\pi_r \pi_u} (1, -P_{r,s}) \begin{pmatrix} \Sigma_{n_{r,s}, n_{u,v}} & \Sigma_{n_{r,s}, n_{u,\cdot}} \\ \Sigma_{n_{r,\cdot}, n_{u,v}} & \Sigma_{n_{r,\cdot}, n_{u,\cdot}} \end{pmatrix} \begin{pmatrix} 1 \\ -P_{u,v} \end{pmatrix}. \end{aligned}$$

Next, we substitute the expressions we determined in Proposition A.2. This yields

$$\begin{aligned} &= \frac{1}{\pi_r \pi_u} \{P_{r,s} P_{u,v} (\pi_r Z_{s,u} + \pi_u Z_{v,r} - 3\pi_r \pi_u) + \pi_r P_{r,s} \delta_{r,u} \delta_{s,v} \\ &\quad - P_{u,v} (\pi_r P_{r,s} Z_{s,u} - 2\pi_r \pi_u P_{r,s} + \pi_u P_{r,s} Z_{u,r}) \\ &\quad - P_{r,s} (\pi_u P_{u,v} Z_{v,r} - 2\pi_u \pi_r P_{u,v} + \pi_r P_{u,v} Z_{r,u}) \\ &\quad + P_{r,s} P_{u,v} (\pi_r Z_{r,u} + \pi_u Z_{u,r} - \pi_r \pi_u - \delta_{r,u} \pi_r)\} \\ &= \frac{P_{r,s} \delta_{r,u} \delta_{s,v}}{\pi_r} + \frac{P_{r,s} P_{u,v}}{\pi_r \pi_u} \{ \pi_r Z_{s,u} + \pi_u Z_{v,r} - 3\pi_r \pi_u \\ &\quad - (\pi_r Z_{s,u} - 2\pi_r \pi_u + \pi_u Z_{u,r}) \\ &\quad - (\pi_u Z_{v,r} - 2\pi_u \pi_r + \pi_r Z_{r,u}) \\ &\quad + \pi_r Z_{r,u} + \pi_u Z_{u,r} - \pi_r \pi_u - \delta_{r,u} \pi_r \} \\ &= \frac{P_{r,s} \delta_{r,u} \delta_{s,v}}{\pi_r} + \frac{P_{r,s} P_{u,v}}{\pi_r \pi_u} \{-\delta_{r,u} \pi_r\} = \frac{P_{r,s}}{\pi_r} \delta_{r,u} (\delta_{s,v} - P_{r,v}). \end{aligned}$$

So we immediately see that $(\Sigma P)_{(r,s)(u,v)} = 0$ if $r \neq u$, $(\Sigma P)_{(r,s)(r,v)} = -\frac{1}{\pi_r} P_{r,s} P_{r,v}$, if $s \neq v$, and that $(\Sigma P)_{(r,s)(r,s)} = \frac{1}{\pi_r} P_{r,s} (1 - P_{r,s})$. This completes the proof. \square

Proof of Lemma 3. Schweitzer (1968) perturbs finite Markov chains and computes, in particular, the derivatives of the stationary distribution and of the fundamental matrix with respect to P ,

$$\frac{\partial \pi_v}{\partial P_{r,s}} = \pi_r Z_{s,v} \quad \text{and} \quad \frac{\partial Z_{u,v}}{\partial P_{r,s}} = Z_{u,r} Z_{s,v} - \pi_r (Z^2)_{s,v}.$$

So we have,

$$\frac{\partial \langle \xi, Z \xi \rangle_\pi}{\partial P_{k,l}} = \sum_i (Z \xi)_i \xi_i \frac{\partial \pi_i}{\partial P_{k,l}} + \sum_i \sum_j \pi_i \xi_i \xi_j \frac{\partial Z_{i,j}}{\partial P_{k,l}}.$$

The first term can we expressed as

$$\sum_i (Z \xi)_i \xi_i \frac{\partial \pi_i}{\partial P_{k,l}} = \pi_k \sum_i (Z \xi)_i Z_{l,i} \xi_i = \pi_k [Z \Lambda_\xi Z \xi]_l,$$

and the second term can be written as

$$\begin{aligned} \sum_i \sum_j \pi_i \xi_i \xi_j \frac{\partial Z_{i,j}}{\partial P_{k,l}} &= \sum_{i,j} \pi_i \xi_i \xi_j Z_{i,k} Z_{l,j} - \sum_{i,j} \pi_i \xi_i \pi_k (Z^2)_{l,j} \xi_j \\ &= \langle \xi, Z_{\cdot,k} \rangle_\pi (Z \xi)_l - \pi_k (Z^2 \xi)_l \langle \pi, \xi \rangle. \end{aligned}$$

Next we find that $\frac{\partial \langle \xi, \xi \rangle_\pi}{\partial P_{k,l}} = \pi_k \sum_i Z_{l,i} \xi_i^2 = \pi_k (Z\xi^2)_l$. Adding up the three terms and defining $\mu = \langle \pi, \xi \rangle$, we have

$$\begin{aligned} \frac{\partial \langle \xi, (2Z - I)\xi \rangle_\pi}{\partial P_{k,l}} &= 2\{\pi_k [Z\Lambda_\xi Z\xi]_l + (Z\xi)_l \langle \xi, Z_{\cdot k} \rangle_\pi - \pi_k \mu (Z^2\xi)_l\} - \pi_k (Z\xi^2)_l \\ &= [\pi_k 2Z\Lambda_\xi Z\xi + 2\xi' \Lambda_\pi Z_{\cdot k} Z\xi - 2\pi_k \mu Z Z\xi - \pi_k Z\Lambda_\xi \xi]_l \\ &= [Z \{ \Lambda_\xi (2I - Z^{-1}) - 2\mu I \} Z\xi \pi_k + 2Z\xi \xi' \Lambda_\pi Z_{\cdot k}]_l, \end{aligned}$$

and the results follows from $2Z - I = (2I - Z^{-1})Z = \{2I - (I - P + \Pi)\}Z$, and picking the l -th row and k -th column of the matrix stated in the Lemma. \square

Proof of Corollary 1. To simplify notation set $\Xi_{r,s} = \partial \langle \xi, (2Z - I)\xi \rangle_\pi / \partial P_{r,s}$. From Proposition 2 we find

$$\begin{aligned} \Sigma_{\text{MC}} &= \sum_{r,s,u,v} \Xi_{r,s} (\Sigma P)_{(r,s)(u,v)} \Xi_{u,v} = \sum_{r,s,v} \Xi_{r,s} \left\{ \frac{1}{\pi_r} (\Lambda_{P_{r,\cdot}} - P'_{r,\cdot} P_{r,\cdot})_{s,v} \right\} \Xi_{r,v} \\ &= \sum_{r,s,v} \frac{1}{\pi_r} \Xi_{r,\cdot} (\Lambda_{P_{r,\cdot}} - P'_{r,\cdot} P_{r,\cdot}) \Xi_{r,\cdot}, \end{aligned}$$

and the result follows from the expression derived in Lemma 3. \square

Proof of Theorem 5. We first prove that $\text{RV}_{\hat{F}} = \text{MC}^\#$ when the first and last observed state coincide. In this event it follows that $\hat{\pi}_r = \sum_s n_{r,s}/n$, so that $n_{r,s} = \hat{\pi}_r \hat{P}_{r,s} n$, and hence

$$\text{RV}_{\hat{F}} = \sum_{r,s} n_{r,s} \hat{y}_{(r,s)}^2 = \sum_{r,s} \hat{\pi}_r \hat{P}_{r,s} n \left\{ e'_r (I - \hat{Z})f + e'_s \hat{Z}f \right\}^2 = \sum_{r,s} \hat{\pi}_r \hat{P}_{r,s} \left\{ e'_r (I - \hat{Z})f + e'_s \hat{Z}f \right\}^2.$$

Following the same steps as those used in the proof of Theorem 2, we find (with $h = \infty$) that $\text{RV}_{\hat{F}} = \langle \xi, (2\hat{Z} - I)\xi \rangle_{\hat{\pi}} = \text{MC}^\#$. Next, we turn to the general situation, where we establish the asymptotic equivalence of $\text{RV}_{\hat{F}}$ and $\text{MC}^\#$ by relating both estimators to a third estimator. The third estimator is defined by $\widetilde{\text{MC}} = \xi' \Lambda_{\tilde{\pi}} (2\tilde{Z} - I)\xi$, where $\tilde{\pi}$ and \tilde{Z} are associated with the transition matrix, \tilde{P} . Here \tilde{P} is the maximum likelihood estimator of P , that is compute with the extended sample

$$(x_1, x_2, \dots, x_{n-1}, x_n, x_1, \dots, x_k).$$

In comparison, \hat{P} is computed with the sample $(x_1, x_2, \dots, x_{n-1}, x_n)$. Artificially adding a finite number of observations implies that $\hat{P} - \tilde{P} = O_p(n^{-1})$. The two estimators are given by $g(\hat{P})$ and $g(\tilde{P})$ for some continuous function g , and we derived the differential $\partial g(P)/\partial P$ in Lemma 3 from which it follows that $\text{MC}^\# - \widetilde{\text{MC}} = O_p(n^{-1})$. Finally we show that $\text{RV}_{\hat{F}} - \widetilde{\text{MC}} = O_p(n^{-1})$. Since the first and last observed state in the extended sample are identical, we have $\widetilde{\text{MC}} = \text{RV}_{\tilde{F}} = \sum_{r,s} \tilde{n}_{r,s} \tilde{y}_{(r,s)}^2$, where $\tilde{y}_{(r,s)} = e'_r (I - \tilde{Z})f + e'_s \tilde{Z}f$ and $\tilde{n}_{r,s}$ is the number of transitions from r to s in the extended sample. Since $f = n^{-1/2}\xi$ we have $\tilde{y}_{(r,s)}^2 = O_p(n^{-1})$ so that

$$\text{RV}_{\hat{F}} = \sum_{r,s} \tilde{n}_{r,s} \tilde{y}_{(r,s)}^2 - \sum_{r,s} n_{r,s} \tilde{y}_{(r,s)}^2 = O_p(n^{-1}).$$

So the proof is completed by showing $\sum_{r,s} n_{r,s} (\tilde{y}_{(r,s)}^2 - y_{(r,s)}^2) = O_p(n^{-1})$. First we note that $\{e'_r(I - \tilde{Z}) + e'_s \tilde{Z}\} - \{e'_r(I - \hat{Z}) + e'_s \hat{Z}\} = O_p(n^{-1})$ because $Z = Z(P)$ is continuous differentiable in P . Since $G(x) = xx'$ is also “smooth” we have $G\{e'_r(I - \tilde{Z}) + e'_s \tilde{Z}\} - G\{e'_r(I - \hat{Z}) + e'_s \hat{Z}\} = O_p(n^{-1})$. It now follows that

$$\begin{aligned} \sum_{r,s} n_{r,s} (\tilde{y}_{(r,s)}^2 - y_{(r,s)}^2) &= \sum_{r,s} n_{r,s} f' \left[G\{e'_r(I - \tilde{Z}) + e'_s \tilde{Z}\} - G\{e'_r(I - \hat{Z}) + e'_s \hat{Z}\} \right] f \\ &= \sum_{r,s} \frac{n_{r,s}}{n} \xi' \left[G\{e'_r(I - \tilde{Z}) + e'_s \tilde{Z}\} - G\{e'_r(I - \hat{Z}) + e'_s \hat{Z}\} \right] \xi, \end{aligned}$$

is $O_p(n^{-1})$, which completes the proof. \square

Proof of Corollary 2. Follows by the delta method. \square

Proof of Theorem 6. To simplify notation, write “ f ” as short for “ $\int_{T_k}^{T_k+h}$ ”. First notice that

$$\Pr(X_{T_k+h} = j | X_{T_k} = i) = \left[\mathbb{E} \left\{ e^{A \int \lambda_u du} | \mathcal{F}_{T_k} \right\} \right]_{i,j},$$

so by Bayes’ formula we have for $i \neq j$,

$$\begin{aligned} \Pr(X_{T_k+h} = j | X_{T_k} = i, X_{T_k+h} \neq i) &= \frac{\Pr(X_{T_k+h} = j, X_{T_k+h} \neq i | X_{T_k} = i)}{\Pr(X_{T_k+h} \neq i | X_{T_k} = i)} \\ &= \frac{\Pr(X_{T_k+h} = j \neq i | X_{T_k} = i)}{1 - \Pr(X_{T_k+h} = i | X_{T_k} = i)} \\ &= \frac{[\mathbb{E} \{ e^{A \int \lambda_u du} | \mathcal{F}_{T_k} \}]_{i,j}}{1 - [\mathbb{E} \{ e^{A \int \lambda_u du} | \mathcal{F}_{T_k} \}]_{i,i}}. \end{aligned}$$

Since $\mathbb{E} \{ e^{A \int \lambda_u du} | \mathcal{F}_{T_k} \} = I + \lambda_{T_k} A + o(h)$, we have (for $i \neq j$)

$$\Pr(X_{T_k+h} = j | X_{T_k} = i, X_{T_k+h} \neq i) = \frac{[I + \lambda_{T_k} A + o(h)]_{i,j \neq i}}{[\lambda_{T_k} A + o(h)]_{i,i}} \rightarrow \frac{A_{i,j}}{A_{i,i}},$$

as $h \rightarrow 0$. Consider two observations at time T_k and T_{k+1} , and recall that $\Pr(T_{k+1} - T_k \in dh | \mathcal{F}_{T_k}) = \mathbb{E} \{ \mu_{T_k+h} e^{-\int \mu_u du} | \mathcal{F}_{T_k} \}$. So that

$$\Pr(X_{T_{k+1}} = j | X_{T_k} = i, X_{T_k+h} \neq i) = \frac{[P(T_k, T_{k+1})]_{i,j}}{1 - [P(T_k, T_{k+1})]_{i,i}},$$

where $P(T_k, T_{k+1}) = \mathbb{E}(\int_0^\infty e^{A \int \lambda_u du} \mu_{T_k+h} e^{-\int \mu_u du} dh | \mathcal{F}_{T_k})$. Next we use that $e^{A \int \lambda_u du} = \sum_{k=0}^\infty \frac{A^k}{k!} (\int \lambda_u du)^k$, so that

$$\int_0^\infty \frac{(\int \lambda_u du)^k}{k!} \mu_{T_k+h} e^{-\int \mu_u du} dh = \int_0^\infty \lambda_{T_k+h} \frac{(\int \lambda_u du)^{k-1}}{(k-1)!} e^{-\int \mu_u du} dh,$$

and

$$\left\| \int_0^\infty e^{A \int \lambda_u du} \mu_{T_k+h} e^{-\int \mu_u du} dh - I - A \int_0^\infty \lambda_{T_k+h} e^{-\int \mu_u du} dh \right\|$$

$$\leq \sum_{k \geq 2} \frac{\|A\|^k}{(k-1)!} \int_0^\infty \lambda_{T_k+h} (f \lambda_u du)^{k-1} e^{-f \mu_u du} dh$$

Having assumed that $\frac{\lambda_t(n)}{\mu_t(n)} \rightarrow 0$ uniformly, we know that there exists $C_n \rightarrow \infty$, such that $\mu_t(n) \geq C_n \lambda_t(n)$.

$$\begin{aligned} & \sum_{k \geq 2} \frac{\|A\|^k}{(k-1)!} \int_0^\infty \lambda_{T_k+h} (f \lambda_u du)^{k-1} e^{-f \mu_u du} dh \\ & \leq \sum_{k \geq 2} \frac{\|A\|^k}{(k-1)!} \int_0^\infty \lambda_{T_k+h} (f \lambda_u du)^{k-1} e^{-C_n f \lambda_u du} dh \\ & = \sum_{k \geq 2} \frac{\|A\|^k}{(k-1)!} \int_0^\infty v^{k-1} e^{-C_n v} dv = \sum_{k \geq 2} \frac{\|A\|^k}{C_n^k} = \frac{\|A\|^2}{C_n^2 - C_n \|A\|} = O\left(\frac{1}{C_n^2}\right). \end{aligned}$$

Thus, $\int_0^\infty e^{A f \lambda_u du} \mu_{T_k+h} e^{-f \mu_u du} dh = I + A \int_0^\infty \lambda_{T_k+h} e^{-f \mu_u du} dh + o\left(\frac{1}{C_n}\right)$, and taking expected value we have

$$P(T_k, T_{k+1}) = I + AE\left(\int_0^\infty \lambda_{T_k+h} e^{-f \mu_u du} dh | \mathcal{F}_{T_k}\right) + o\left(\frac{1}{C_n}\right).$$

Finally we can conclude that

$$\Pr(X_{T_{k+1}} = j | X_{T_k} = i, X_{T_k+h} \neq i) = \frac{[I + AE\{\int_0^\infty \lambda_{T_k+h} e^{-f \mu_u du} dh | \mathcal{F}_{T_k}\} + o(\frac{1}{C_n})]_{i,j}}{1 - [I + AE\{\int_0^\infty \lambda_{T_k+h} e^{-f \mu_u du} dh | \mathcal{F}_{T_k}\} + o(\frac{1}{C_n})]_{i,i}} \rightarrow \frac{A_{i,j}}{A_{i,i}}.$$

This completes the proof. \square

12 Separate Appendix (Not-for-Publication)

		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$						
20070601	T	0.2395	0.2394	0.2558	0.2557	0.2393	0.2392	0.2313	0.2313	0.2327	0.2326	0.2142	0.2142
	B	0.2393	0.2393	0.2535	0.2534	0.2381	0.2381	0.2224	0.2224	0.2221	0.2223	0.2178	0.2176
	A	0.2397	0.2396	0.2584	0.2583	0.2307	0.2307	0.2210	0.2211	0.2120	0.2119	0.2054	0.2053
20070604	T	0.1663	0.1664	0.1784	0.1785	0.1686	0.1686	0.1652	0.1652	0.1757	0.1757	0.1828	0.1828
	B	0.1638	0.1638	0.1832	0.1832	0.1746	0.1746	0.1644	0.1644	0.1564	0.1564	0.1549	0.1550
	A	0.1681	0.1681	0.1943	0.1944	0.1685	0.1685	0.1823	0.1823	0.1734	0.1735	0.1739	0.1738
20070605	T	0.3299	0.3295	0.3157	0.3155	0.2900	0.2900	0.2882	0.2888	0.2938	0.2953	0.2831	0.2843
	B	0.2948	0.2948	0.3101	0.3102	0.2825	0.2826	0.2718	0.2721	0.2683	0.2684	0.2561	0.2564
	A	0.2946	0.2945	0.2968	0.2967	0.2731	0.2731	0.2791	0.2791	0.2737	0.2737	0.2607	0.2604
20070606	T	0.2877	0.2877	0.3088	0.3088	0.2856	0.2857	0.2928	0.2928	0.2846	0.2847	0.2814	0.2816
	B	0.2917	0.2917	0.2922	0.2921	0.2677	0.2677	0.2796	0.2795	0.2667	0.2667	0.2645	0.2645
	A	0.2876	0.2876	0.2949	0.2949	0.2585	0.2585	0.2705	0.2704	0.2768	0.2766	0.2924	0.2920
20070607	T	0.6197	0.6196	0.6581	0.6579	0.6053	0.6051	0.5798	0.5795	0.5666	0.5664	0.5836	0.5839
	B	0.6632	0.6632	0.6635	0.6635	0.6016	0.6017	0.5794	0.5795	0.5542	0.5544	0.5468	0.5467
	A	0.6677	0.6676	0.6722	0.6722	0.6009	0.6009	0.5665	0.5666	0.5780	0.5777	0.5832	0.5829
20070608	T	0.5166	0.5173	0.5288	0.5288	0.5179	0.5177	0.4710	0.4707	0.4647	0.4672	0.5039	0.5061
	B	0.5581	0.5584	0.5745	0.5746	0.5143	0.5141	0.4934	0.4926	0.4838	0.4815	0.4986	0.4989
	A	0.5310	0.5312	0.5610	0.5612	0.4967	0.4968	0.4861	0.4862	0.4740	0.4736	0.4842	0.4839
20070611	T	0.2452	0.2453	0.2585	0.2587	0.2582	0.2583	0.2685	0.2688	0.2833	0.2837	0.2869	0.2872
	B	0.2508	0.2508	0.2725	0.2725	0.2554	0.2555	0.2708	0.2706	0.2743	0.2744	0.2982	0.2986
	A	0.2419	0.2420	0.2768	0.2768	0.2564	0.2564	0.2657	0.2658	0.2866	0.2867	0.2915	0.2916
20070612	T	0.4627	0.4629	0.5500	0.5501	0.5526	0.5528	0.5649	0.5654	0.5699	0.5708	0.5951	0.5958
	B	0.4921	0.4922	0.5500	0.5501	0.5636	0.5638	0.5763	0.5764	0.5638	0.5638	0.5533	0.5528
	A	0.4820	0.4822	0.5557	0.5557	0.5577	0.5577	0.5912	0.5908	0.5958	0.5953	0.6178	0.6181
20070613	T	0.4609	0.4612	0.4785	0.4785	0.4324	0.4322	0.4178	0.4172	0.3959	0.3954	0.4046	0.4049
	B	0.4841	0.4842	0.4863	0.4865	0.4235	0.4237	0.4059	0.4060	0.4053	0.4055	0.4025	0.4022
	A	0.4827	0.4829	0.4977	0.4978	0.4572	0.4574	0.4253	0.4254	0.4090	0.4091	0.4080	0.4085
20070614	T	0.2496	0.2496	0.2644	0.2645	0.2447	0.2448	0.2391	0.2392	0.2342	0.2344	0.2377	0.2379
	B	0.2550	0.2551	0.2626	0.2626	0.2362	0.2362	0.2291	0.2292	0.2186	0.2188	0.2139	0.2139
	A	0.2484	0.2484	0.2653	0.2654	0.2313	0.2313	0.2270	0.2270	0.2079	0.2080	0.2106	0.2107
20070615	T	0.1831	0.1831	0.1917	0.1916	0.1777	0.1777	0.1609	0.1609	0.1598	0.1599	0.1512	0.1515
	B	0.1725	0.1724	0.1735	0.1735	0.1514	0.1514	0.1479	0.1479	0.1377	0.1377	0.1444	0.1445
	A	0.1750	0.1751	0.1716	0.1716	0.1596	0.1596	0.1582	0.1582	0.1475	0.1472	0.1555	0.1556

Table continued on next page...

		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$						
20070618	T	0.1445	0.1445	0.1604	0.1604	0.1458	0.1458	0.1547	0.1547	0.1599	0.1599	0.1584	0.1584
	B	0.1542	0.1542	0.1532	0.1532	0.1490	0.1491	0.1540	0.1535	0.1580	0.1583	0.1607	0.1609
	A	0.1521	0.1521	0.1494	0.1494	0.1447	0.1447	0.1544	0.1543	0.1639	0.1637	0.1636	0.1635
20070619	T	0.1842	0.1842	0.2200	0.2200	0.2154	0.2156	0.2192	0.2196	0.2178	0.2180	0.2188	0.2191
	B	0.1897	0.1897	0.2279	0.2277	0.2157	0.2156	0.2230	0.2229	0.2158	0.2161	0.2219	0.2222
	A	0.1968	0.1968	0.2277	0.2276	0.2215	0.2214	0.2145	0.2144	0.2251	0.2251	0.2180	0.2177
20070620	T	0.3815	0.3819	0.4367	0.4373	0.4289	0.4294	0.4151	0.4155	0.4377	0.4382	0.4094	0.4101
	B	0.4025	0.4029	0.4397	0.4401	0.4310	0.4313	0.4417	0.4416	0.4476	0.4478	0.4642	0.4641
	A	0.4014	0.4017	0.4470	0.4473	0.4354	0.4354	0.4367	0.4360	0.4528	0.4522	0.4512	0.4506
20070621	T	0.5986	0.5993	0.6622	0.6630	0.6556	0.6567	0.6577	0.6587	0.6680	0.6691	0.6804	0.6816
	B	0.6315	0.6318	0.6780	0.6783	0.6715	0.6717	0.6581	0.6584	0.6777	0.6782	0.6630	0.6632
	A	0.6269	0.6271	0.6847	0.6850	0.6819	0.6821	0.6825	0.6826	0.6883	0.6884	0.7190	0.7191
20070622	T	0.6940	0.6947	0.7669	0.7685	0.7608	0.7621	0.7478	0.7485	0.7604	0.7606	0.7789	0.7780
	B	0.6956	0.6962	0.7908	0.7918	0.7717	0.7727	0.7758	0.7757	0.7702	0.7716	0.7588	0.7615
	A	0.7020	0.7026	0.7650	0.7654	0.7470	0.7474	0.7474	0.7461	0.7536	0.7531	0.7823	0.7841
20070625	T	0.6179	0.6190	0.7452	0.7466	0.7694	0.7709	0.7982	0.7992	0.8371	0.8372	0.8241	0.8248
	B	0.6360	0.6378	0.7245	0.7270	0.7471	0.7470	0.7494	0.7491	0.7682	0.7682	0.8481	0.8494
	A	0.6370	0.6390	0.7401	0.7427	0.7382	0.7383	0.8119	0.8117	0.8098	0.8102	0.7896	0.7895
20070626	T	0.6933	0.6935	0.8068	0.8069	0.7447	0.7448	0.7446	0.7446	0.7094	0.7096	0.7058	0.7050
	B	0.7097	0.7099	0.7499	0.7501	0.7409	0.7410	0.7605	0.7602	0.7355	0.7353	0.7071	0.7070
	A	0.7004	0.7006	0.8169	0.8171	0.7619	0.7617	0.7422	0.7417	0.7224	0.7222	0.7556	0.7552
20070627	T	0.5035	0.5039	0.5673	0.5676	0.5648	0.5653	0.5177	0.5168	0.5048	0.5030	0.5018	0.5003
	B	0.5165	0.5168	0.5571	0.5574	0.5488	0.5492	0.5376	0.5379	0.4884	0.4887	0.4964	0.4965
	A	0.5171	0.5175	0.5546	0.5550	0.5527	0.5531	0.5311	0.5314	0.5118	0.5118	0.5144	0.5141
20070628	T	0.5927	0.5928	0.6433	0.6434	0.6457	0.6463	0.6123	0.6120	0.6167	0.6183	0.5425	0.5431
	B	0.6018	0.6029	0.6658	0.6651	0.6227	0.6210	0.5750	0.5767	0.5433	0.5433	0.4809	0.4833
	A	0.6203	0.6202	0.6919	0.6917	0.6388	0.6403	0.6106	0.6115	0.5830	0.5858	0.5746	0.5916
20070629	T	0.6442	0.6451	0.7258	0.7262	0.7055	0.7066	0.7044	0.7040	0.6841	0.6839	0.6902	0.6910
	B	0.6680	0.6691	0.7215	0.7224	0.6855	0.6854	0.6782	0.6794	0.6536	0.6546	0.6541	0.6557
	A	0.6704	0.6720	0.7149	0.7167	0.7027	0.7047	0.6899	0.6911	0.6642	0.6647	0.6664	0.6679

Table 6: A comparison of MC and MC* for 21 trading days of high-frequency data, using six different orders for the Markov chain being estimator.

Date	$k = 3$							$k = 4$						
	MC	Delta	LogDelta		Bootstrap		MC	Delta	LogDelta		Bootstrap			
20070601	0.239	0.209	0.269	0.211	0.271	0.210	0.272	0.231	0.197	0.266	0.199	0.269	0.199	0.271
20070604	0.169	0.146	0.192	0.147	0.193	0.147	0.193	0.165	0.141	0.190	0.143	0.191	0.141	0.190
20070605	0.290	0.258	0.322	0.260	0.324	0.260	0.322	0.288	0.252	0.324	0.254	0.327	0.254	0.324
20070606	0.286	0.257	0.314	0.259	0.315	0.259	0.315	0.293	0.259	0.327	0.261	0.329	0.260	0.329
20070607	0.605	0.549	0.662	0.552	0.664	0.552	0.664	0.580	0.520	0.639	0.523	0.642	0.521	0.641
20070608	0.518	0.463	0.573	0.465	0.576	0.465	0.576	0.471	0.418	0.524	0.421	0.527	0.419	0.527
20070611	0.258	0.228	0.288	0.230	0.290	0.229	0.292	0.269	0.232	0.305	0.235	0.307	0.234	0.307
20070612	0.553	0.496	0.609	0.499	0.612	0.498	0.614	0.565	0.500	0.629	0.504	0.633	0.502	0.635
20070613	0.432	0.390	0.475	0.392	0.477	0.391	0.476	0.418	0.370	0.465	0.373	0.468	0.372	0.466
20070614	0.245	0.217	0.272	0.219	0.274	0.218	0.273	0.239	0.207	0.271	0.209	0.273	0.209	0.274
20070615	0.178	0.158	0.198	0.159	0.199	0.158	0.198	0.161	0.141	0.180	0.143	0.182	0.143	0.181
20070618	0.146	0.127	0.164	0.128	0.166	0.128	0.165	0.155	0.131	0.178	0.133	0.180	0.132	0.180
20070619	0.215	0.189	0.242	0.191	0.243	0.190	0.243	0.219	0.188	0.250	0.190	0.253	0.190	0.253
20070620	0.429	0.385	0.473	0.387	0.475	0.386	0.473	0.415	0.369	0.461	0.372	0.463	0.371	0.463
20070621	0.656	0.593	0.718	0.596	0.721	0.596	0.722	0.658	0.589	0.726	0.593	0.730	0.591	0.728
20070622	0.761	0.683	0.839	0.687	0.843	0.684	0.839	0.748	0.663	0.832	0.668	0.837	0.668	0.837
20070625	0.769	0.684	0.854	0.689	0.859	0.687	0.858	0.798	0.694	0.903	0.700	0.910	0.699	0.910
20070626	0.745	0.677	0.812	0.680	0.815	0.679	0.813	0.745	0.664	0.825	0.668	0.830	0.665	0.829
20070627	0.565	0.511	0.618	0.514	0.621	0.514	0.620	0.518	0.463	0.572	0.466	0.575	0.465	0.573
20070628	0.646	0.571	0.721	0.575	0.725	0.572	0.724	0.612	0.534	0.691	0.538	0.696	0.536	0.696
20070629	0.706	0.636	0.775	0.639	0.779	0.639	0.777	0.704	0.619	0.790	0.624	0.795	0.621	0.793

Table 7: A comparison of 95% confidence intervals for MC using three different methods: The δ -method applied to MC and $\log(\text{MC})$ and the bootstrap method, here computed with $B = 5,000$ resamples. We present results for $k = 3$ and $k = 4$. The point estimates are listed in bold font and the upper end of the 95% confidence intervals are identified by a gray background. We note that the three types of confidence intervals are very similar in all cases.