Correlation between Noise and Returns: a Robust Integrated Variance Estimator

Stefano Peluso * Antonietta Mira † Pietro Muliere ‡

December 21, 2016

Abstract

Correlation between microstructure noise and latent financial logarithmic returns is an empirically relevant phenomenon with sound theoretical justification. With few notable exceptions, all integrated variance estimators proposed in the literature are not designed to explicitly handle such a dependence, or handle it only in special settings. In the present paper we provide an integrated variance estimator that is robust to correlated noise and returns. For this purpose, a generalization of the Forward Filtering Backward Sampling algorithm is proposed, to provide a sampling technique for a latent conditionally Gaussian random sequence. We apply our methodology to intraday Microsoft prices, and we compare it in simulation with established alternatives, showing an advantage in terms of bias and dispersion.

Keywords: Conditionally Gaussian Random Sequences; Forward Filtering and Backward Sampling; Integrated Variance; Kalman Filtering; State Space Models.

1 Introduction

Many statistical problems can be formulated as State Space models, where a latent stochastic process $\{\theta_t\}$ evolves in time with dynamics given by a transition equation $\theta_{t+1} = a_1(t)\theta_t + b_1(t)\epsilon_1(t+1)$, and the process is observed only through the noisy process $\{\xi_t\}$, corresponding, at time $t+1$, to $\xi_{t+1} = \tilde{A}_1(t)\theta_{t+1} + \tilde{B}_2(t)\epsilon_2(t+1)$, where $\epsilon_1(t)$ and $\epsilon_2(t)$ are Gaussian random variables and $a_1(t)$, $b_1(t)$, $\tilde{A}_1(t)$ and $\tilde{B}_1(t)$ are time-varying parameters. Kalman (1960) proposed the celebrated Kalman filtering algorithm as optimal solution, in mean square sense, to the *filtering problem*, that is the problem of estimating the unobservable θ_t by means of observations $\xi^t = \{\xi_1, \dots, \xi_t\}$. The Kalman filter is the starting point in Fruwirth-Schnatter (1994) and Carter and Kohn (1994) for an iterative sampling procedure, today commonly known as *Forward Filtering Backward Sampling* (FFBS), for obtaining posterior samples of $\{\theta_t\}$.

Liptser and Shiryayev (1972, 2001a,b) introduce the so-called *conditionally Gaussian random sequences*, whose main two features are: (a) model parameters depend in a measurable arbitrary way from past observations or from other random quantities, but once this dependence is realized,

*Corresponding author. University of Lugano, Swiss Finance Institute, E-mail: stefano.peluso@usi.ch. Via Giuseppe Buffi 13 CH-6904 Lugano. Tel.: +41 (0)58 666 44 95, Fax: +41 (0)58 666 46 47

†University of Lugano, Institute of Finance and InterDisciplinary Institute of Data Science, E-mail: antonietta.mira@usi.ch

‡Bocconi University of Milan, E-mail: pietro.muliere@unibocconi.it

the remaining randomness can be expressed in terms of Gaussian random variables, (b) correlation between ξ_t and θ_t is introduced through the presence, in both the transition and measurement equations, of common Brownian motions.

The Mixture Kalman filter of Chen and Liu (2000) and the Gibbs samplers of Sethuraman (1994) and Carter and Kohn (1996) are some relevant examples that, among other things, generalize Kalman filtering and posterior sampling of the latent stochastic process along direction (a) above. Other simulation techniques as the Extended Kalman filter of Gelb (1972), the Monte Carlo filter of Kitagawa (1987) and the Particle filter of Gordon et al. (1993) do not require the conditional Gaussianity, but are based on some form of approximation. Gelb (1972) provides a suboptimal solution to the filtering problem, by linearizing the transition equation. Kitagawa (1987) and Gordon et al. (1993) approximate the posterior distribution of the latent stochastic process through a weighted set of particles. de Jong and Shephard (1995) and Durbin and Koopman (2002) also developed FFBS algorithms. In particular, the methodology of de Jong and Shephard (1995) define the conditionally, linear Gaussian state space model in terms of a single source of error, and it is able to reproduce correlated shocks between the measurement equation and transition equation.

Harvey and Shepard (1996) and Sandmann and Koopman (1998) point out the empirical relevance of direction (b) for modeling the asymmetric behavior that is often found in stock prices, and Hull and White (1987) emphasize the role of correlation between observed prices and latent stochastic volatility, suggesting it can cause significant biases in financial option pricing, if neglected. Among others, Brandt and Kang (2004) and Jacquier et al. (2004) further study this phenomenon in the financial economic literature. Another concrete situation where neglecting the correlation in the two equations can be misleading, and this is the financial setting that motivates the development of the estimator proposed in the current paper, is in the estimation of the integrated volatility, when there is a dependence between microstructure noise and latent financial logarithmic returns, a phenomenon empirically found in Hansen and Lunde (2006) and theoretically justified in Diebold and Strasser (2013). Many integrated variance estimators proposed in the literature (Andersen et al. 2003; Ait-Sahalia et al. 2005; Zhang et al. 2005; Zhang 2006; Ait-Sahalia et al. 2010; Barndorff-Nielsen et al. 2011; Corsi et al. 2014; Peluso et al. 2014) are not designed to handle such a dependence, except for Barndorff-Nielsen et al. (2008), but only in the special setting of a linear model of endogeneity. To our knowledge, the only papers that deal with this endogenous noise are Kalnina and Linton (2008), which propose a robust version of Zhang et al. (2005), and the pre-averaging estimator of Jacod et al. (2009). Bandi and Russell (2011) despite assuming exogenous noise, still provides a good benchmark method, since empirically found to perform well even if the underlying assumptions are violated.

In the present paper we propose a realized variance estimator of the daily integrated volatility that is robust to the presence of correlation between microstructure noise and latent returns, generalizing the setting of Ait-Sahalia et al. (2010). For this purpose, we extend the FFBS algorithm from standard State Space models to the more general context of Liptser and Shiryayev (1972) in an exact form. The algorithm is presented in Section 3, after an introduction to conditionally Gaussian random sequences in Section 2. The motivating financial problem is studied in Section 4, and finally the conclusions are drawn in Section 5.

2 Conditionally Gaussian Random Sequences

In this Section we introduce the theoretical framework developed in Liptser and Shiryayev (1972) (see also Liptser and Shiryayev 2001a and Liptser and Shiryayev 2001b), with focus on the recursive equations of conditionally Gaussian random sequences for the solution to the filtering problem.

On a probability space $(\Omega, \mathcal{F}, P)$, the random sequence $\{\theta_t, \xi_t\}_t$, $t = 1, 2, \ldots$, with $\theta_t = (\theta_1(t), \ldots, \theta_k(t))$ and $\xi_t = (\xi_1(t), \ldots, \xi_l(t))$, defines the system of recursive equations

$$\theta_{t+1} = a_0(t,\omega) + a_1(t,\omega)\theta_t + b_1(t,\omega)\epsilon_1(t+1) + b_2(t,\omega)\epsilon_2(t+1) \quad (1)$$

$$\xi_{t+1} = A_0(t,\omega) + A_1(t,\omega)\theta_t + B_1(t,\omega)\epsilon_1(t+1) + B_2(t,\omega)\epsilon_2(t+1), \quad (2)$$

where $\epsilon_1(t) = (\epsilon_{1,1}(t), \ldots, \epsilon_{1,k}(t))$ and $\epsilon_2(t) = (\epsilon_{2,1}(t), \ldots, \epsilon_{2,l}(t))$ are independent Gaussian random variables with expected value $\mathbb{E}(\epsilon_{i,j}(t)) = 0$ and $\mathbb{E}(\epsilon_{i_1,j_1}(t)\epsilon_{i_2,j_2}(s)) = \delta(i_1, i_2)\delta(j_1, j_2)\delta(t, s)$, where

$$\delta(x, y) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$$

In the sequel, θ_t and ξ_t are, respectively, unobservable and observed vectors, with $\theta_0|\xi_0 \sim \Phi(m, \gamma)$, that is Gaussian with mean m and variance γ. $a_0(t,\omega)$ and $A_0(t,\omega)$ are vector functions, and $a_1(t,\omega)$, $A_1(t,\omega)$, $b_1(t,\omega)$, $b_2(t,\omega)$, $B_1(t,\omega)$ and $B_2(t,\omega)$ are matrix functions, square integrable and measurable at time t. All the vector and matrix functions at time t are collected in $D(t,\omega)$. In Liptser and Shiryayev (1972), $D(t,\omega)$ is assumed to be $\mathcal{F}_t^{\xi}$-measurable, where $\mathcal{F}_t^{\xi} = \sigma\{\omega : \xi_0, \ldots, \xi_t\}$ is the σ-algebra generated by the random variables $\xi_0, \ldots, \xi_t$. This assumption will be relaxed in Section 3, where measurability with respect to σ-algebras generated by other random variables will be considered. Denote $b \circ b = b_1 b_1^* + b_2 b_2^*$, $b \circ B = b_1 B_1^* + b_2 B_2^*$ and $B \circ B = B_1 B_1^* + B_2 B_2^*$ where X^* is the transposed matrix of X and $X^+ = Y^*(YY^*)^{-2}Y$ is the pseudo-inverse matrix of X, with Y such that $Y^*Y = X$. For ease of notation we suppress the dependence on ω.

Theorem 2.1. *Suppose that* $\mathbb{E}(||\theta_0||^2 + ||\xi_0||^2) < \infty$, $|(a_1(t,\omega))_{ij}| < L$ *and* $|(A_1(t,\omega))_{ij}| < L$, *where* L *is a positive constant. Then,* $\theta_t|\xi_0, \ldots, \xi_t \sim \Phi(m(t), \gamma(t))$, *where* $m(t)$ *and* $\gamma(t)$ *are determined from the recursive equations*

$$\begin{aligned} m(t+1) &= [a_0(t) + a_1(t)m(t)] + [b \circ B(t) + a_1(t)\gamma(t)A_1^*(t)] \\ &\quad \cdot [B \circ B(t) + A_1(t)\gamma(t)A_1^*(t)]^+ \cdot [\xi_{t+1} - A_0(t) - A_1(t)m(t)] \quad (3) \\ \gamma(t+1) &= [a_1(t)\gamma(t)a_1^*(t) + b \circ b(t)] - [b \circ B(t) + a_1(t)\gamma(t)A_1^*(t)] \\ &\quad \cdot [B \circ B(t) + A_1(t)\gamma(t)A_1^*(t)]^+ \cdot [b \circ B(t) + a_1(t)\gamma(t)A_1^*(t)]^* \quad (4) \end{aligned}$$

with the initial conditions $m(0) = m$ *and* $\gamma(0) = \gamma$.

Proof. See Liptser and Shiryayev (1972), Theorem 3.2. □

An important special case is when $D(t,\omega)$ is not random, but a deterministic function of time t. In this case, if the vector (θ_0, ξ_0) is Gaussian, the process (θ_t, ξ_t) will also be Gaussian, with non random covariance $\gamma(t)$. In this setting it is possible to reformulate the system of recursion equations (1) and (2) so that the dependence between ξ_t and θ_t is explicit and recover the Kalman filter as special case.

The random sequence $\{\theta_t, \xi_t\}_t$ is known as conditionally Gaussian since it follows a Gaussian distribution at any specific time t, conditional on the knowledge of $D(t,\omega)$ in the recursive equations. Note that this is not a restrictive assumption, since unconditionally the dependence in time and space is not necessarily linear (for instance when the distribution of $a_1(t)$ depends on θ_t) and the disturbances are location-scale mixture of Gaussian random variables. A wide class of continuous distributions may be constructed as location-scale mixture of Normal distributions, such as contaminated Normals, Student's t, Logistic, Laplace and Stable distributions. As specified in Marron and Wand (1992), one way of seeing that the class of Normal mixture densities is a very broad one comes from the fact that any density, even strongly multimodal and asymmetric, can be approximated arbitrarily well by a Normal mixture. This is a setting of interest in finance, where we often observe skewed return distributions (see, among others, Barndorff-Nielsen 1997 and Azzalini and Capitanio 2003). Furthermore, return distributions can be contaminated by outliers that are not easy to detect and correct for, and that can severely distort a non robust estimation methodology, causing for instance relevant consequences on asset allocation studies (Best and Grauer 1992). Finally, as pointed out in Engle and Smith (1999), multimodal distributions can model situations of regime switches, known to have a relevance in option pricing (see for instance Buffington and Elliott 2002) and mean-variance portfolio selection (Zhou and Yin 2003, among others).

3 Sampling Algorithm of the Latent Process

The system (1) and (2) can be reformulated to highlight the relation between ξ_t and θ_t, so that the sequence of the observations can be interpreted as a realization of a stochastic Markovian latent process with measurement noise:

$$\begin{aligned}\theta_{t+1} &= a_0(t,\omega) + a_1(t,\omega)\theta_t + b_1(t,\omega)\epsilon_1(t+1) + b_2(t,\omega)\epsilon_2(t+1) \quad (5)\\ \xi_{t+1} &= \tilde{A}_0(t,\omega) + \tilde{A}_1(t,\omega)\theta_{t+1}\\ &\quad +\tilde{B}_1(t,\omega)\epsilon_1(t+1) + \tilde{B}_2(t,\omega)\epsilon_2(t+1), \quad (6)\end{aligned}$$

where $a_0(t,\omega), a_1(t,\omega), b_1(t,\omega), b_2(t,\omega), \tilde{A}_0(t,\omega), \tilde{A}_1(t,\omega), \tilde{B}_1(t,\omega), \tilde{B}_2(t,\omega)$ are stored in $\tilde{D}(t,\omega)$. This alternative representation is much more common in the econometrics, financial and engineering literature, and it can be derived by the system (1) and (2) since, substituting (5) in (6), ξ_{t+1} can be written as

$$\begin{aligned}\xi_{t+1} &= \tilde{A}_0(t,\omega) + \tilde{A}_1(t,\omega)[a_0(t,\omega) + a_1(t,\omega)\theta_t + b_1(t,\omega)\epsilon_1(t+1)\\ &\quad +b_2(t,\omega)\epsilon_2(t+1)] + \tilde{B}_1(t,\omega)\epsilon_1(t+1) + \tilde{B}_2(t,\omega)\epsilon_2(t+1),\end{aligned}$$

clarifying that the relation between $D(t,\omega)$ and $\tilde{D}(t,\omega)$ is given by

$$\begin{cases} A_0(t) = \tilde{A}_0(t) + \tilde{A}_1(t)a_0(t)\\ A_1(t) = \tilde{A}_1(t)a_0(t)\\ B_1(t) = \tilde{A}_1(t)b_1(t) + \tilde{B}_1(t)\\ B_2(t) = \tilde{A}_1(t)b_2(t) + \tilde{B}_2(t). \end{cases} \quad (7)$$

Given the system (5)-(6), from Theorem 2.1 it follows that $\theta_t|\xi_1,\ldots,\xi_t \sim \Phi(m(t),\gamma(t))$, where $m(t)$ and $\gamma(t)$ are obtained by the recursive equations (3) and (4), but with $A_0(t)$, $A_1(t)$, $B_1(t)$ and $B_2(t)$

replaced by the respective right hand sides in (7). When $b_2(t) = 0$ and $\tilde{B}_1(t) = 0$ (or, equivalently, when $b_1(t) = 0$ and $\tilde{B}_2(t) = 0$) for all t, system (5)-(6) simplifies to

$$\begin{aligned} \theta_{t+1} &= a_0(t,\omega) + a_1(t,\omega)\theta_t + b_1(t,\omega)\epsilon_1(t+1) &\quad (8)\\ \xi_{t+1} &= \tilde{A}_0(t,\omega) + \tilde{A}_1(t,\omega)\theta_{t+1} + \tilde{B}_2(t,\omega)\epsilon_2(t+1), &\quad (9)\end{aligned}$$

for which the filtering problem can be solved through the Kalman filtering iterations:

$$\begin{aligned} m(t+1) &= [a_0(t) + a_1(t)m(t)] + [a_1(t)\gamma(t)A_1^*(t)] \\ &\quad \cdot [B_2^2(t) + A_1(t)\gamma(t)A_1^*(t)]^+ \cdot [\xi_{t+1} - A_0(t) - A_1(t)m(t)] \\ \gamma(t+1) &= [a_1(t)\gamma(t)a_1^*(t) + b_1^2(t)] - [a_1(t)\gamma(t)A_1^*(t)] \\ &\quad \cdot [B_2^2(t) + A_1(t)\gamma(t)A_1^*(t)]^+ \cdot [a_1(t)\gamma(t)A_1^*(t)]^* \end{aligned}$$

In the simplified setting of model (11)-(12), Fruwirth-Schnatter (1994) and Carter and Kohn (1994) introduce the Forward Filtering and Backward Sampling (FFBS) algorithm, to sample θ^T a posteriori from

$$p(\theta^T|\xi^T, \tilde{D}(1), \ldots, \tilde{D}(T)) \propto \prod_{t=1}^T \phi(V_t W_t, V_t),$$

where

$$\begin{aligned} V_t^{-1} &= A_1^*(t)(B_2^2(t))^+ A_1(t) + a_1^*(t)(b_1^2(t))^+ a_1(t) + \gamma^+(t) \\ W_t &= A_1(t)(B_2^2(t))^+(\xi_{t+1} - A_0(t)) + a_1^*(t)(b_1^2(t))^+(\theta_{t+1} - a_0(t)) + \gamma^+(t)m(t). \end{aligned}$$

Exploiting an extended factorization of the posterior density of θ, induced by the common Brownian motions, we derive a generalized version of the FFBS algorithm, to jointly sample

$$\theta_1, \ldots, \theta_T | \xi_1, \ldots, \xi_T, \tilde{D}(1), \ldots, \tilde{D}(T)$$

from the system (5)-(6) (an equivalent algorithm for the system (1)-(2) can also be formulated). For easier reference in the sequel, we denote this algorithm as G-FFBS.

Proposition 3.1. *Given ξ^T generated from model* (5)-(6)*, then*

$$p(\theta^T|\xi^T) \propto \prod_{t=1}^T \phi(V_t W_t, V_t),$$

where

$$\begin{aligned} V_t^{-1} &= (A_1(t) - B \circ b(t)(b \circ b)^+(t)a_1(t))^* \Sigma_t^+ (A_1(t) - B \circ b(t)(b \circ b)^+(t)a_1(t)) \\ &\quad + a_1^*(t)(b \circ b)^+(t)a_1(t) + \gamma^+(t) \\ W_t &= (A_1(t) - B \circ b(t)(b \circ b)^+(t)a_1(t))\Sigma_t^+(\xi_{t+1} - A_0(t) \\ &\quad - B \circ b(t)(b \circ b)^+(t)(\theta_{t+1} - a_0(t))) + a_1^*(t)(b \circ b)^+(t)(\theta_{t+1} - a_0(t)) + \gamma^+(t)m(t) \\ \Sigma_t &= B \circ B(t) - B \circ b(t)(b \circ b)^+(t) B \circ b^*(t). \end{aligned}$$

Proof. Using the notation $x^t = \{x_1, \ldots, x_t\}$ and suppressing the dependence on $\tilde{D}(1), \ldots, \tilde{D}(T)$, G-FFBS exploits the factorization

$$p(\theta^T|\xi^T) = \prod_{t=1}^{T} p(\theta_t|\theta_{t+1}, \ldots, \theta_T, \xi^T), \tag{10}$$

Noting that

$$\begin{aligned} p(\theta^T, \xi^T) &= \prod_{t=1}^{T} p(\theta_t, \xi_t|\theta_{t-1}) \\ &= \prod_{t=0}^{T-1} \phi\left\{ \begin{pmatrix} \xi_{t+1} \\ \theta_{t+1} \end{pmatrix}; \begin{pmatrix} A_0(t) + A_1(t)\theta_t \\ a_0(t) + a_1(t)\theta_t \end{pmatrix}, \begin{pmatrix} B \circ B(t) & b \circ B(t) \\ (b \circ B(t))^* & b \circ b(t) \end{pmatrix} \right\}, \end{aligned}$$

$\xi_{t+1}|\theta_{t+1}, \theta_t \sim \phi(\mu_t, \Sigma_t)$, where

$$\mu_t \quad = \quad A_0(t) + A_1(t)\theta_t + B \circ b(t)(b \circ b)^+(t)(\theta_{t+1} - a_0(t) - a_1(t)\theta_t)$$

Thus the factor $p(\theta_t|\theta_{t+1}, \ldots, \theta_T, \xi^T)$ in (10) can be expressed as

$$\begin{aligned} & p(\theta_t|\theta_{t+1}, \ldots, \theta_T, \xi^T) \propto p(\theta_t, \ldots, \theta_T, \xi_{t+1}, \ldots, \xi_T|\xi^t) \\ & = \prod_{i=t+1}^{T} p(\xi_i, \theta_i|\theta_{i-1}) \cdot p(\theta_t|\xi^t) \\ & \propto p(\xi_{t+1}|\theta_{t+1}, \theta_t) p(\theta_{t+1}|\theta_t) p(\theta_t|\xi^t) \\ & \propto \exp\left\{ -\frac{1}{2} \left((\xi_{t+1} - \mu_t)^* \Sigma_t^+ (\xi_{t+1} - \mu_t) \right) \right\} \\ & \quad \cdot \exp\left\{ -\frac{1}{2} \left((\theta_{t+1} - a_0(t) - a_1(t)\theta_t)^* (b \circ b)^+(t) (\theta_{t+1} - a_0(t) - a_1(t)\theta_t) \right) \right\} \\ & \quad \cdot \exp\left\{ -\frac{1}{2} \left((\theta_t - m(t))^* \gamma^+(t) (\theta_t - m(t)) \right) \right\} \\ & \propto \phi\Big(V_t W_t, V_t \Big) \end{aligned}$$

□

The proposed generalization over the traditional FFBS finds relevant empirical justification in the motivating example that will be discussed in Section 4. The algorithm requires a forward step in which the quantities of interest $m(t)$ and $\gamma(t)$ are computed following Theorem 2.1, and a backward step where the latent process is sampled according to the factorization in (10). In the traditional FFBS algorithm the factor at time t in (10) is proportional to $p(\theta_{t+1}|\theta_t)p(\theta_t|\xi^t)$, whilst in the proposed G-FFBS algorithm, there is the additional term $p(\xi_{t+1}|\theta_{t+1}, \theta_t)$, since the correlation between measurement and transition errors generates a conditional dependence between $\xi_{t+1}|\theta_{t+1}$ and θ_t. When $\tilde{B}_1(t) = 0$ and $b_2(t) = 0$ for any t or when $\tilde{B}_2(t) = 0$ and $b_1(t) = 0$, there is no correlation between the two errors, the conditional independence of the observations is restored, and G-FFBS reduces to FFBS.

For posterior inference on any function of the latent stochastic process $g(\theta^T)$, three cases can be distinguished: (i) When $\tilde{D}(t,\omega)$ is measurable at time t, (ii) $\tilde{D}(t,\omega)$ is unknown at time t but with known dynamics, (iii) $\tilde{D}(t,\omega)$ is unknown at time t and with unknown dynamics. In case (i), $\tilde{D}(t,\omega)$ is measurable at time t with respect to the σ-algebra generated by ξ^T or by some other observables, and any number of samples from $\theta^T|\xi^T$ can be obtained through the G-FFBS. In (ii) a simple procedure is to recursively estimate $\tilde{D}(t,\omega)$ with $\hat{D}(t,\omega)$, estimated by the known dynamics, and then use $\hat{D}(t,\omega)$ instead of $\tilde{D}(t,\omega)$ in the G-FFBS (see, for instance Smith and West 1983 and Campagnoli et al. 2001 for, respectively, a biometric and a financial application). When in (iii), $\tilde{D}(t,\omega)$ is unknown and can't be parametrically forecasted: a complete Bayesian model is specified, with prior $\pi(\tilde{D}(1,\omega),\ldots,\tilde{D}(T,\omega))$, and MCMC procedures are used to sample from the joint posterior $\mathbb{P}(\theta^T,\tilde{D}(1),\ldots,\tilde{D}(T)|\xi^T)$, by repeatedly sampling at each iteration

- $\mathbb{P}(\theta^T|\tilde{D}(1),\ldots,\tilde{D}(T),\xi^T)$,
- $\mathbb{P}(\tilde{D}(1),\ldots,\tilde{D}(T)|\theta^T,\xi^T) \propto \mathbb{P}(\theta^T,\xi^T|\tilde{D}(1),\ldots,\tilde{D}(T))\,\pi(\tilde{D}(1,\omega),\ldots,\tilde{D}(T,\omega))$.

The first step is executed through G-FFBS, and the whole algorithm is a Gibbs sampler (Geman and Geman 1984; Gelfand and Smith 1990) or a Metropolis-Hastings sampler (Metropolis et al. 1953; Hastings 1970), depending on $\pi(\tilde{D}(1,\omega),\ldots,\tilde{D}(T,\omega))$ being a conjugate prior or not.

4 Robust Integrated Variance Estimation

In this Section the developed sampling algorithm is applied to our motivating financial problem. Suppose that the logarithmic price of a given financial asset follows, within the trading day, the diffusion process

$$d\theta_t = c(t)dZ_t$$

where $c(t)$ is the instantaneous volatility and $\{Z_t\}_t$ is the standard Brownian motion. $IV = \int c^2(t)dt$ is known as *integrated variance* and is of interest as a measure of the true daily volatility. What we actually use for estimation is the discrete approximation of the continuous time process above: $\theta_{(t+1)/T} = \theta_{t/T} + c_{t/T}\sqrt{1/T}Z_t$, where T^{-1} is the discrete time interval between adjacent observations, $\theta_{t/T} - \theta_{(t-1)/T} = O_p(T^{-1/2})$ and Z_t standard Gaussian. IV is a latent quantity, usually estimated with the so-called *realized variance* $RV = \sum_{t=1}^{T}(\theta_{t/T} - \theta_{(t-1)/T})^2$, that is the sum of all intraday high frequency observed logarithmic returns. RV is a consistent and efficient estimator of IV (Andersen et al. 2003) when there is no microstructure noise, that is when $\theta_{t/T}$ for $t = 1,\ldots,T$ is directly observed. When microstructure noise is introduced, we observe $\xi_{t/T}$ instead of $\theta_{t/T}$, and the computable realized variance becomes $\tilde{RV} = \sum_{t=1}^{T}(\xi_{t/T} - \xi_{(t-1)/T})^2$. Note that we do not specify the continuous-time version of the measurement equation: the observed price relates to the latent price only through the microstructure noise, consequence of trades occurring at discrete times. Unfortunately, $\tilde{RV}$ loses the good properties of RV, since it is biased and inconsistent for the true integrated variance. As this problem arises mainly when the frequency of the observations approaches zero, it can be attenuated by sparse sampling, but this involves a loss of information because of the discarded data. Recently, some authors have followed the approach suggested by Ait-Sahalia et al. (2005) of sampling as often as possible and modeling the noise. In particular, a first consistent estimator of IV for financial data contaminated by microstructure noise has been proposed in Zhang et al. (2005) (whose order of convergence is improved in Zhang 2006), later followed by Barndorff-Nielsen et al. (2008), that propose a kernel-based estimator. There have been numerous extensions of the framework with noisy observations that account for additional empirically observed data irregularities, as asyncronicity of multivariate log prices, serially dependent microstructure noise, positivity of the estimator (see, for instance, Ait-Sahalia et al. 2010; Barndorff-Nielsen et al. 2011; Corsi et al. 2014; Peluso et al. 2014). Less attention has been posed on the dependence between microstructure noise and latent financial logarithmic returns, empirically found in Hansen and Lunde (2006). Also, common microstructure theories from financial economics literature induce a correlation between latent returns and microstructure noise (Diebold and Strasser 2013), and this correlation gives indication on the presence of uninformed trades, risk aversion and market makers learning speed. All the estimators mentioned above are not designed for such dependence, except for Barndorff-Nielsen et al. (2008), but only for a linear model of endogeneity. To our knowledge, the only papers that deal with this endogenous noise are Kalnina and Linton (2008), which propose a robust version of Zhang et al. (2005), and the pre-averaging estimator of Jacod et al. (2009). The kernel estimator of Bandi and Russell (2011) also shows robustness properties that justify its adoption in a setting with correlation between microstructure noise and latent returns.

The framework of conditionally Gaussian sequences, with the sampling algorithm introduced above, can be used to propose a new estimator of the integrated variance that is robust to the presence of correlation between microstructure noise and latent returns. Consider the bivariate system

$$\begin{aligned} \xi_{(t+1)/T} &= \theta_{(t+1)/T} + \tilde{B}_1(t)\epsilon_1(t+1) + \tilde{B}_2(t)\epsilon_2(t+1) && (11)\\ \theta_{(t+1)/T} &= \theta_{t/T} + b_1(t)\epsilon_1(t+1), && (12) \end{aligned}$$

in which $a_0(t) = \tilde{A}_0(t) = b_2(t) = 0$ and $a_1(t) = \tilde{A}_1(t) = 1$ for any t. Model (11)-(12) is completed by characterizing the prior distributions $\tilde{B}_1(t) \sim \phi(\mu_{B,t}, \sigma^2_{B,t})$, $b_1(t) \sim \phi(\mu_{b,t}, \sigma^2_{b,t})$ and finally $\tilde{B}_2(t) \sim IG(\alpha_{B,t}, \beta_{B,t})$. The correlation between microstructure noise and true returns is introduced through the random variable ϵ_1, present in both the equations of the system. Note that Hansen and Lunde (2006) found that microstructure noise and latent return are negatively correlated, and with a Gaussian prior on $B_1(t)$ it is possibile to center, a priori, this correlation on a negative value. Furthermore, Diebold and Strasser (2013) point out that a negative correlation appears most realistic, and that markets with no evidence of significant negative correlation are likely subject to an extraordinary microstructure effect such as high risk aversion.

The full conditional of θ^T is sampled with the G-FFBS. The forward step of the G-FFBS algorithm is performed through the following filtering iterations:

$$\begin{aligned} m(t+1) &= m(t) + \frac{b_1(t)B_1(t) + \gamma(t)}{B_1^2(t) + B_2^2(t) + \gamma(t)}(\xi_{(t+1)/T} - m(t)) \\ &= m(t) + \frac{b_1(t)(b_1(t) + \tilde{B}_1(t)) + \gamma(t)}{(b_1(t) + \tilde{B}_1(t))^2 + \tilde{B}_2^2(t) + \gamma(t)}(\xi_{(t+1)/T} - m(t)) && (13)\\ \gamma(t+1) &= (\gamma(t) + b_1^2(t)) - \frac{(b_1(t)B_1(t) + \gamma(t))^2}{B_1^2(t) + B_2^2(t) + \gamma(t)} \\ &= (\gamma(t) + b_1^2(t)) - \frac{(b_1(t)(b_1(t) + \tilde{B}_1) + \gamma(t))^2}{(b_1(t) + \tilde{B}_1(t))^2 + \tilde{B}_2^2(t) + \gamma(t)}. && (14) \end{aligned}$$

Note that if $\tilde{B}_1(t) = 0 \;\forall t$, the filtering iterations (13) and (14) simplify to the well known Kalman filter iterations (Kalman 1960). For the backward sampling step, $\theta^T|\xi^T$ are sampled from (10), where

$$\begin{aligned}
p(\theta_{t/T}|\theta_{(t+1)/T},\ldots,\theta_T,\xi^T) &\propto p(\xi_{(t+1)/T}|\theta_{(t+1)/T},\theta_{t/T})p(\theta_{(t+1)/T}|\theta_{t/T})p(\theta_{t/T}|\xi^t)\\
&= \phi\left(\xi_{(t+1)/T};\theta_{t/T}+\frac{B_1(t)}{b_1(t)}(\theta_{(t+1)/T}-\theta_{t/T}),B_2^2(t))\right)\\
&\quad \phi\left(\theta_{(t+1)/T};\theta_{t/T},b_1^2(t)\right)\phi\left(\theta_{t/T};m(t),\gamma(t)\right)\\
&\propto \phi(V_tW_t,V_t),
\end{aligned}$$

with

$$\begin{aligned}
V_t &= \left(\frac{\left(1-\frac{B_1(t)}{b_1(t)}\right)^2}{B_2^2(t)}+\frac{1}{b_1^2(t)}+\frac{1}{\gamma(t)}\right)^{-1}\\
&= \left(1-\frac{\left(1-\frac{B_1(t)}{b_1(t)}\right)^2 b_1^2(t)\gamma(t)+B_2^2(t)\gamma(t)}{\left(1-\frac{B_1(t)}{b_1(t)}\right)^2 b_1^2(t)\gamma(t)+B_2^2(t)\gamma(t)+B_2^2(t)b_1^2(t)}\right)\gamma(t)
\end{aligned}$$

$$\begin{aligned}
V_tW_t &= V_t\left[\frac{1-\frac{B_1(t)}{b_1(t)}}{B_2^2(t)}\left(\xi_{(t+1)/T}-\frac{B_1(t)}{b_1(t)}\theta_{(t+1)/T}\right)+\frac{\theta_{(t+1)/T}}{b_1^2(t)}+\frac{m(t)}{\gamma(t)}\right]\\
&= \left(1-\frac{\left(1-\frac{B_1(t)}{b_1(t)}\right)^2 b_1^2(t)\gamma(t)+B_2^2(t)\gamma(t)}{\left(1-\frac{B_1(t)}{b_1(t)}\right)^2 b_1^2(t)\gamma(t)+B_2^2(t)\gamma(t)+B_2^2(t)b_1^2(t)}\right)m(t)+\\
&\quad \frac{\left(1-\frac{B_1(t)}{b_1(t)}\right)\left(\frac{\xi_{(t+1)/T}}{\theta_{(t+1)/T}}-\frac{B_1(t)}{b_1(t)}\right)b_1^2(t)\gamma(t)+B_2^2(t)\gamma(t)}{\left(1-\frac{B_1(t)}{b_1(t)}\right)^2 b_1^2(t)\gamma(t)+B_2^2(t)\gamma(t)+B_2^2(t)b_1^2(t)}\theta_{(t+1)/T}.
\end{aligned}$$

Note that $B_1(t) = \tilde{B}_1(t) + b_1(t)$ and $B_2(t) = \tilde{B}_2(t)$. The correlation between transition and measurement error can be removed by fixing $\tilde{B}(t) = 0$. In this case, $B_1(t) = b_1(t)$ and, as expected, $V_t = \left(1-\frac{\gamma(t)}{b_1^2(t)+\gamma(t)}\right)\gamma(t)$ and $W_tV_t = \left(1-\frac{\gamma(t)}{b_1^2(t)+\gamma(t)}\right)m(t)+\frac{\gamma(t)}{b_1^2(t)+\gamma(t)}\theta_{(t+1)/T}$, as in the usual FFBS.

The difference between FFBS and G-FFBS can be crucial for the estimation of the latent stochastic process. For instance, for a correlation between microstructure noise and financial latent return of -0.10 (adopted in the simulation study of Kalnina and Linton 2008), a noise to signal ratio of 1.5 and an annualized latent variance of the transition error of 0.06, we simulate 500 trading days, with $T = 23400$ seconds per business day. The order of magnitude of the microstructure noise is controlled by $\tilde{B}_2$, the parameter linked to ϵ_2. For each day we compute the estimated quadratic variation for FFBS and G-FFBS, compared in Figure 1, where it is clear that neglecting the correlation has an impact on the inference of the latent process. As expected, the distance between the two methodologies widens as more correlation is introduced in the system (simulation results not shown for brevity).

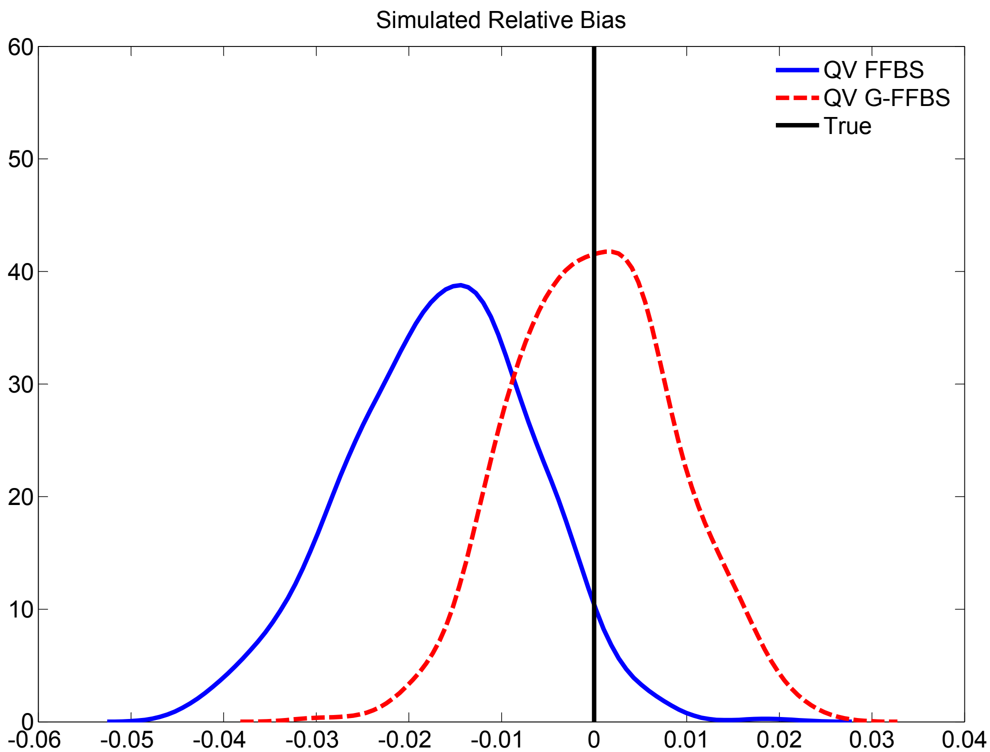

Figure 1: Relative Bias between Quadratic Variation simulated with FFBS and G-FFBS and the true latent value, over 500 trading days, when the correlation between microstructure noise and financial latent return is -0.10.

To sample from the remaining full conditional distributions, note that

$$\begin{pmatrix} \xi_{(t+1)/T} \\ \theta_{(t+1)/T} \end{pmatrix} |\theta_t, b_1(t), B_1(t), B_2(t) \sim \Phi \left\{ \begin{pmatrix} \theta_t \\ \theta_t \end{pmatrix}, \begin{pmatrix} B_1^2(t) + B_2^2(t) & b_1(t)B_1(t) \\ b_1(t)B_1(t) & b_1^2(t) \end{pmatrix} \right\},$$

and

$$\xi_{(t+1)/T}|\theta_{(t+1)/T}, \theta_t, b_1(t), B_1(t), B_2(t) \quad \sim \quad \phi\left(\theta_{t/T} + \frac{B_1(t)}{b_1(t)}(\theta_{(t+1)/T} - \theta_{t/T}), B_2^2(t)\right).$$

The full conditionals of $\tilde{B}_1(t)$ and $\tilde{B}_2^2(t)$ are in standard form:

$$
\begin{aligned}
&p\left(\tilde{B}_1(t)|\theta^T, \xi^T, \tilde{B}_2^T, b_1^T, \{\tilde{B}_1(s),\ s \neq t\}\right) \\
&\quad \propto p(\xi_{(t+1)/T}|\theta_{(t+1)/T}, \theta_{t/T}, \tilde{B}_1(t), \tilde{B}_2(t), b_1(t))p(\tilde{B}_1(t)) \\
&\quad \propto \frac{\phi\left\{\frac{\mu_{B,t}+\frac{\sigma^2_{B,t}}{b_1(t)\tilde{B}_2^2(t)}(\theta_{(t+1)/T}-\theta_{t/T})(\xi_{(t+1)/T}-\theta_{(t+1)/T})}{\sqrt{1+\frac{\sigma^2_{B,t}}{b_1^2(t)\tilde{B}_2^2(t)}(\theta_{(t+1)/T}-\theta_{t/T})^2}}, \sigma^2_{B,t}\right\}}{\sqrt{1+\frac{\sigma^2_{B,t}}{b_1^2(t)\tilde{B}_2^2(t)}(\theta_{(t+1)/T}-\theta_{t/T})^2}}
\end{aligned}
\tag{15}
$$

$$
\begin{aligned}
&p\left(\tilde{B}_2^2(t)|\theta^T, \xi^T, \tilde{B}_1^T, b_1^T, \{\tilde{B}_2(s),\ s \neq t\}\right) \\
&\quad \propto p(\xi_{(t+1)/T}|\theta_{(t+1)/T}, \theta_{t/T}, \tilde{B}_1(t), \tilde{B}_2^2(t), b_1(t))p(\tilde{B}_2^2(t)) \\
&\quad \propto IG\left\{\alpha_{B,t}+\frac{1}{2}, \beta_{B,t}+\frac{1}{2}\left(\xi_{(t+1)/T}-\theta_{(t+1)/T}-\frac{\tilde{B}_1(t)}{b_1(t)}(\theta_{(t+1)/T}-\theta_{t/T})\right)^2\right\}
\end{aligned}
\tag{16}
$$

On the other hand, we sample $b_1(t)$ with a *Hamiltonian* step (see Chapter 5 in Brooks et al. 2011 for an introduction to the algorithm). The motivation for using this algorithm is its capability to exploit the information in the gradient of the full conditional of $b_1(t)$ for a faster exploration of the parameter space, overcoming the random walk behavior of the Metropolis-Hastings step in a highly dimensional space. Please refer to the Appendix for the details on the Hamiltonian step. Note that, when there is no correlation (that is when $\tilde{B}_1(t) = 0$), the sampler can be reduced to the Gibbs algorithm in Peluso et al. (2014).

The output of the whole algorithm is a collection of samples

$$\{\theta^T_{(i)}, \tilde{B}^T_{1(i)}, \tilde{B}^T_{2(i)}, b^T_{1(i)}\}_{i=1}^M,$$

where M is the number of iterations of the MCMC scheme. Then the proposed estimator of the integrated variance is

$$\frac{1}{M-M_0}\sum_{i=M_0+1}^{M}\sum_{t=1}^{T}(\theta_{(t+1)/T,(i)}-\theta_{t,(i)})^2 \tag{17}$$

where $M_0 < M$ is the burn-in, that is the number of samples discarded at the beginning of the Markov chain.

We simulate 500 trading days, for $M = 1000$, $M_0 = 500$ and correlation -0.10, starting all the chains from strongly biased values. The hyperparameters are for $\mu_{B,t} = -1.48 \cdot 10^{-5}$, $\sigma^2_{B,t} = 1.53 \cdot 10^{-10}$, $\mu_{b,t} = 1.21 \cdot 10^{-4}$, $\sigma^2_{b,t} = 1.02 \cdot 10^{-8}$, $\alpha_{B,t} = 2.1$ and $\beta_{B,t} = 1.99 \cdot 10^{-8}$ for all t. Our methodology is compared with the estimators of Kalnina and Linton (2008), Bandi and Russell (2011) and Jacod et al. (2009) (for Jacod et al. 2009, both the adjusted and unadjusted estimators for small sample sizes are implemented). The method of Kalnina and Linton (2008) requires the choice of the tuning parameter K: we use $K = T^{2/3}$, since it performs well in the simulations in Kalnina and Linton (2008) and it is what the authors use in their empirical study. For the estimator proposed in Bandi and Russell (2011), the tuning parameters are chosen according to the rule of thumb proposed in Equation (26) of Bandi and Russell (2011), in simulation computed using the true values and in the application below to Microsoft Corporation, using the corresponding values in Table 1 of Bandi and Russell (2006). Finally, the tuning parameters of Jacod et al. (2009) are fixed, using their notation, to $k_n = 51$, $\theta = k_n/\sqrt{T}$ and $g(x) = x \wedge (1 - x)$, as in their simulation studies. The results are reported in Figure 2, there is a clear advantage for our methodology both in terms of bias and dispersion.

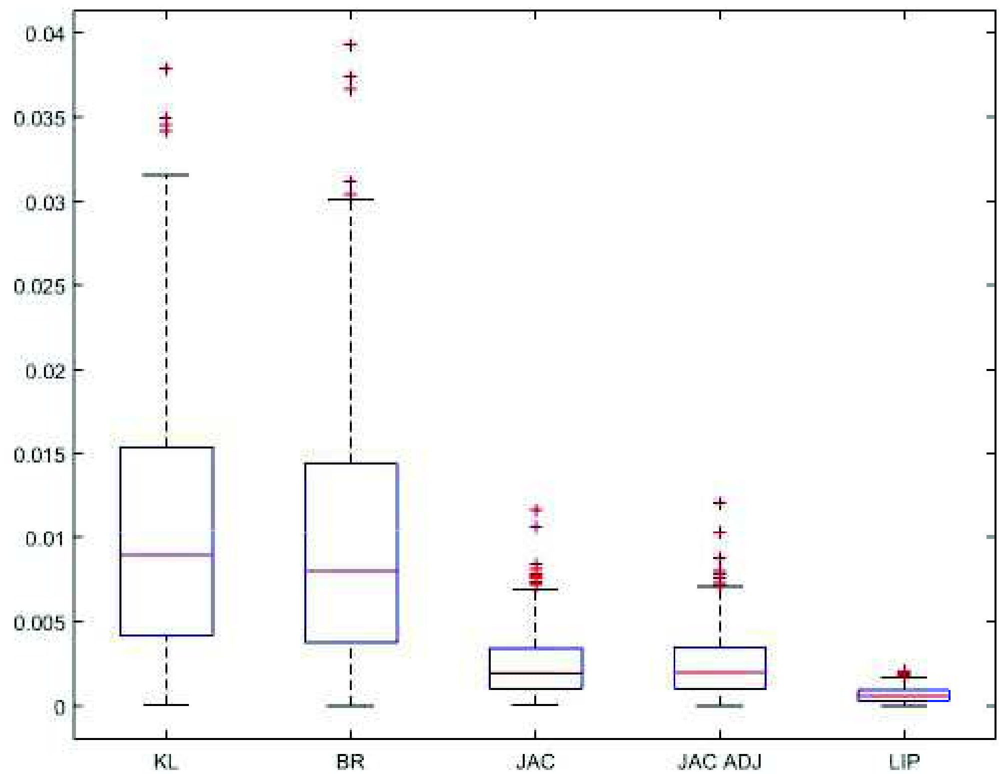

Figure 2: Absolute differences between estimated quadratic variations and the true latent value, for the methods in Kalnina and Linton (2008) (KL), Bandi and Russell (2011) (BR), Jacod et al. (2009) (JAC), the small sample adjusted estimator of Jacod et al. (2009) (JAC ADJ) and for our methodology (LIP). The correlation between microstructure noise and financial latent return is fixed to -0.10.

We also run the algorithm on 1 second frequency logarithmic prices of Microsoft Corporation, for the period April 1, 2007 - June 30, 2008, and the estimated annualized quadratic variations are reported in Figure 3.

5 Conclusions

Overwhelming evidence contrasts the independent microstructure noise assumption, in favour of market noise correlated with increments in the efficient price, with important implications for volatility estimation based on high frequency data (Hansen and Lunde 2006). Furthermore, such

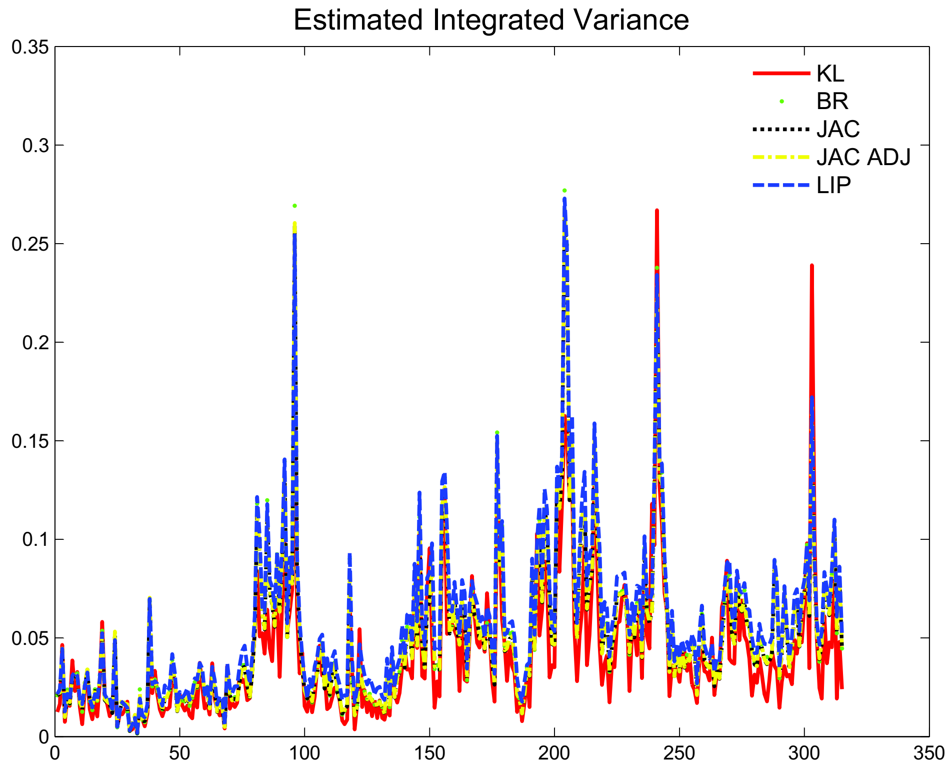

Figure 3: Integrated variance for Microsoft Corporation, estimated with the methods in Kalnina and Linton (2008) (KL), Bandi and Russell (2011) (BR), unadjusted (JAC) and adjusted (JAC ADJ) Jacod et al. (2009), and with our methodology (LIP), in the period April 1, 2007 - June 30, 2008.

a dependence naturally arises in common microstructure models, as discussed in depth in Diebold and Strasser (2013). On the other hand, with the notable exceptions of Barndorff-Nielsen et al. (2008), Kalnina and Linton (2008) and Zhang et al. (2005), several results in the literature analyze high-frequency volatility estimation assuming that the noise process is independent of the efficient price. In the present paper we use the theoretical framework of the conditionally Gaussian random sequences of Liptser and Shiryayev (1972, 2001a,b), to propose a new integrated variance estimator that is robust to the correlation between microstructure noise and latent returns. To this aim, we adopt a Bayesian perspective and we sample a posteriori the latent price process through a generalization of the Forward Filtering Backward Sampling algorithm of Fruwirth-Schnatter (1994) and Carter and Kohn (1994). An application to Microsoft 1-second frequency logarithmic prices is provided, and a simulation study shows an improved performance of our estimator in terms of bias and dispersion, relative to the alternatives in the literature. Our methodology can be implemented to other financial problems, for instance to generalize the framework of Barndorff-Nielsen (1997) to normal inverse Gaussian financial logarithmic returns with measurement error, or, following the approaches of Harvey et al. (1992) and Harvey et al. (1994), to ARCH and Stochastic Volatility models.

Acknowledgements

Stefano Peluso acknowledges support from the Swiss National Science Foundation (SNF), Grant No. P1TIP1_155031. Antonietta Mira also gratefully acknowledges the financial support by SNF.

Appendix: Hamiltonian step of Section 4

The Hamiltonian step is performed through the following substeps:

1. sample the auxiliary momentum variable $p\{1\}$ from $\Phi(0,1)$,
2. propose $b_1(t)^*$ from the Leapfrog algorithm. In details, fix $k\{1\}$ to the current value of $b_1(t)$. For step size ϵ and number of iterations L:

$$p\{1+\epsilon/2\} = p\{1\} - \frac{\epsilon}{2}\left[\frac{k\{1\}-\mu_{b,t}}{\sigma^2_{b,t}} + \frac{1}{\tilde{B}_2(t)^2}\left(\xi_{(t+1)/T} - \theta_{(t+1)/T} - \frac{\tilde{B}_1(t)}{k\{1\}}(\theta_{(t+1)/T}-\theta_{t/T})\right)\frac{\tilde{B}_1(t)}{k\{1\}^2}(\theta_{(t+1)/T}-\theta_{t/T})\right]$$

For $i = 1, \ldots, L-1$:

$$\begin{aligned}
k\{1+i\epsilon\} &= k\{1+(i-1)\epsilon\} + \epsilon p\{1+(i-1/2)\epsilon\} \\
p\{1+(i+1/2)\epsilon\} &= p\{1+(i-1/2)\epsilon\} - \epsilon\left[\frac{k\{1+i\epsilon\}-\mu_{b,t}}{\sigma^2_{b,t}} + \right. \\
&\left. \frac{1}{\tilde{B}_2(t)^2}\left(\xi_{(t+1)/T} - \theta_{(t+1)/T} - \frac{\tilde{B}_1(t)}{k\{1+i\epsilon\}}(\theta_{(t+1)/T}-\theta_{t/T})\right)\frac{\tilde{B}_1(t)}{k\{1+i\epsilon\}^2}(\theta_{(t+1)/T}-\theta_{t/T})\right]
\end{aligned}$$

Finally,

$$\begin{aligned}
k\{1+L\epsilon\} &= k\{1+(L-1)\epsilon\} + \epsilon p\{1+(L-1/2)\epsilon\} \\
p\{1+L\epsilon\} &= p\{1+(L-1/2)\epsilon\} - \frac{\epsilon}{2}\left[\frac{k\{1+L\epsilon\}-\mu_{b,t}}{\sigma^2_{b,t}} + \right. \\
&\left. \frac{1}{\tilde{B}_2(t)^2}\left(\xi_{(t+1)/T} - \theta_{(t+1)/T} - \frac{\tilde{B}_1(t)}{k\{1+L\epsilon\}}(\theta_{(t+1)/T}-\theta_{t/T})\right)\frac{\tilde{B}_1(t)}{k\{1+L\epsilon\}^2}(\theta_{(t+1)/T}-\theta_{t/T})\right]
\end{aligned}$$

and the proposed value is $b_1(t)^* = k\{1+L\epsilon\}$.

3. Evaluate potential and kinetic energies U and Z at proposed and current values:

$$\begin{aligned}
U(t) &\propto \frac{(b_1(t)-\mu_{b,t})^2}{2\sigma^2_{b,t}} + \frac{1}{2\tilde{B}_2(t)^2}\left[\xi_{(t+1)/T} - \theta_{(t+1)/T} - \frac{\tilde{B}_1(t)}{b_1(t)}(\theta_{(t+1)/T}-\theta_{t/T})\right]^2 \\
Z(t) &= \frac{1}{2}p\{1\}^2 \\
U(t)^* &\propto \frac{(b_1(t)^*-\mu_{b,t})^2}{2\sigma^2_{b,t}} + \frac{1}{2\tilde{B}_2(t)^2}\left[\xi_{(t+1)/T} - \theta_{(t+1)/T} - \frac{\tilde{B}_1(t)}{b_1(t)^*}(\theta_{(t+1)/T}-\theta_{t/T})\right]^2 \\
Z(t)^* &= \frac{1}{2}p\{1+L\epsilon\}^2
\end{aligned}$$

4. Accept $b_1(t)^*$ with probability

$$\min\left(1, \exp\{U(t) - U(t)^* + Z(t) - Z(t)^*\}\right).$$

References

Ait-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-Frequency Covariance Estimates with Noisy and Asynchronous Financial Data. *JASA*, 105:1504–1517.

Ait-Sahalia, Y., Mykland, P., and Zhang, L. (2005). How Often to Sample a Continuous-Time Process in the presence of Market Microstructure Noise. *The Review of Financial Studies*, 18:351–416.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003). Modeling and Foreasting Realized Volatility. *Econometrica*, 71:529–626.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society B*, 65:579–602.

Bandi, F. and Russell, J. (2006). Separating Microstructure Noise from Volatility. *Journal of Financial Economics.*

Bandi, F. and Russell, J. (2011). Market Microstructure Noise, Integrated Variance Estimators, and the Accuracy of Asymptotic Approximations. *Journal of Econometrics*, 160:145–159.

Barndorff-Nielsen, O. (1997). Normal inverse Gaussian distributions and Stochastic Volatility Modelling. *Scand. J. of Stat.*, 24:1–13.

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

Barndorff-Nielsen, O., Hansen, P., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149 – 169.

Best, M. and Grauer, R. (1992). Sensitivity analysis for Mean-Variance Portfolio Problems. *Management Science*, 37:980–989.

Brandt, M. and Kang, Q. (2004). On the relationship between the Conditional Mean and Volatility of Stock Returns: A Latent VAR Approach. *Journal of Financial Economics*, 72:217–257.

Brooks, S., Gelman, A., Jones, G., and Meng, X. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC; 1 edition (May 10, 2011), 6000 Broken Sound Parkway NW, Suite 300. Boca Raton, FL 33487-2742.

Buffington, J. and Elliott, R. (2002). American Options with Regime Switching. *International Journal of Theoretical and Applied Finance*, 05:497–514.

Campagnoli, P., Muliere, P., and Petrone, S. (2001). Generalized Dynamic Linear Models for Financial Time Series. *Applied Stochastic Models for Business and Industry*, 17:27–39.

Carter, C. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81:541–553.

Carter, C. and Kohn, R. (1996). Markov chain Monte Carlo in Conditionally Gaussian State Space Models. *Biometrika*, 83:589–601.

Chen, R. and Liu, J. (2000). Mixture Kalman Filters. *J. of Royal Stat. Soc. B*, 62:493–508.

Corsi, F., Peluso, S., and Audrino, F. (2014). Missing in asynchronicity: A kalman-em approach for multivariate realized covariance estimation. *Journal of Applied Econometrics. Forthcoming.*

de Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, 82:339–350.

Diebold, F. and Strasser, G. (2013). On the Correlation Structure of Microstructure Noise: A Financial Economic Approach. *Review of Economic Studies*, 80:1304–1337.

Durbin, J. and Koopman, S. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89:603–616.

Engle, R. and Smith, A. (1999). Stochastic Permanent Breaks. *Review of Economics and Statistics*, 81:553–574.

Fruwirth-Schnatter, S. (1994). Data Augmentation and Dynamic Linear Models. *J. of Time Series Analysis*, 15:183–202.

Gelb, A. (1972). *Applied Optimal Estimation*. The MIT Press.

Gelfand, A. and Smith, A. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85:398–409.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gordon, N., Salmond, D., and Smith, A. (1993). Novel Approach to Nonlinear Non-Gaussian State Space Models. *IEE Proc. F*, 40:107–113.

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

Harvey, A., Ruiz, E., and Sentana, E. (1992). Unobserved Component Time Series Models with Arch Disturbances. *Journal of Econometrics*, 52:129–157.

Harvey, A., Ruiz, E., and Shepard, N. (1994). Multivariate Stochastic Variance Models. *Review of Economic Studies*, 61:247–264.

Harvey, A. and Shepard, N. (1996). Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. *Journal of Business and Economic Statistics*, 14:429–434.

Hastings, W. (1970). Monte Carlo Sampling Methods Using Markov Chains and their Applications. *Biometrika*, 57:97–109.

Hull, J. and White, A. (1987). The Pricing of Options on Assets with Stochastic Volatility. *Journal of Finance*, 42:281–300.

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

Jacquier, E., Polson, N., and Rossi, P. (2004). Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors. *Journal of Econometrics*, 122:185–212.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45.

Kalnina, I. and Linton, O. (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics*, 147:47–59.

Kitagawa, G. (1987). Non-Gaussian State Space Modelling of Non-Stationary Time Series. *J. of American Stat. Ass.*, 82:1032–1063.

Liptser, R. and Shiryayev, A. (1972). Statistics of conditionally Gaussian random sequences. In *Sixth Berkley Symp. Math. Stat. and Prob.*, Univ. of Califonia Press.

Liptser, R. and Shiryayev, A. (2001a). *Statistics of Random Processes I. General Theory.* Springer-Verlag, New York.

Liptser, R. and Shiryayev, A. (2001b). *Statistics of Random Processes II. Applications.* Springer-Verlag, New York.

Marron, J. and Wand, M. (1992). Exact Mean Integrated Squared Error. *Annals of Stat.*, 20:712–736.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21:1087–1092.

Peluso, S., Corsi, F., and Mira, A. (2014). A Bayesian High-Frequency Estimator of the Multivariate Covariance of Noisy and Asynchronous Returns. *Journal of Financial Eonometrics*.

Sandmann, G. and Koopman, S. (1998). Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood. *Journal of Econometrics*, 87:271–301.

Sethuraman, J. (1994). A Constructive Definition of Dirichlet Priors. *Statistics Sinica*, 4:639–650.

Smith, A. and West, M. (1983). Monitoring Renal Transplants: An Application of the Multiprocess Kalman Filter. *Biometrics*, 39:867–878.

Zhang, L. (2006). Efficient Estimation of Stochastic Volatility using Noisy Observations: a Multi-Scale Approach. *Bernoulli*, 12:1019–1043.

Zhang, L., Myckland, P., and Ait-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, X. and Yin, G. (2003). Markowitz's Mean-Variance Portfolio Selection with Regime Switching: A Continuous-Time Model. *SIAM Journal on Control and Optimization*, 42:1466–1482.