Supplementary Appendix

Intraday Stochastic Volatility in Discrete Price Changes: the Dynamic Skellam Model

June 3, 2015

1 Monte Carlo likelihood and parameter estimation

We first present a short review of Monte Carlo techniques for likelihood evaluation for state space models and the efficient importance sampler (EIS) of Liesenfeld and Richard (2003) and Richard and Zhang (2007). Next, we present our bivariate extension of the numerically accelerated importance sampling techniques (NAIS) of Koopman, Lucas, and Scharth (2014).

NAIS : likelihood evaluation and importance sampling

The non-Gaussian nonlinear state space model in the main paper is introduced as

$$y_t \sim p(y_t|\boldsymbol{\theta}_t;\boldsymbol{\psi}), \qquad \boldsymbol{\theta}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_t, \qquad \boldsymbol{\alpha}_{t+1} \sim p_g(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t;\boldsymbol{\psi}), \qquad t = 1,\ldots,n,$$

where y_t is the observation, $\boldsymbol{\theta}_t$ is the unobserved signal vector composed of a constant vector $\boldsymbol{c}_t$ and a linear function of a fixed matrix $\boldsymbol{Z}_t$ and the dynamic, stochastic state vector $\boldsymbol{\alpha}_t$. The time index is t and we assume to have n observations for y_t. The fixed unknown parameter vector is $\boldsymbol{\psi}$ and will be estimated via the numerical optimization of the likelihood function that is evaluated by a simulation method as discussed below.

We assume that for given realizations of the signal $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_n')$ the observations $\boldsymbol{y} = (y_1, \ldots, y_n)'$ are conditionally independent. In vector notation, we have $\boldsymbol{\theta} = \boldsymbol{c} + \boldsymbol{Z}\boldsymbol{\alpha}$ with $\boldsymbol{c}' = (\boldsymbol{c}_1', \ldots, \boldsymbol{c}_n')$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_n')'$, and $\boldsymbol{Z}$ a block-diagonal matrix with blocks $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ on the leading diagonal. We can express the likelihood function by

$$L(\boldsymbol{y};\boldsymbol{\psi}) = \int p(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\theta} = \int p(\boldsymbol{y}|\boldsymbol{\theta};\boldsymbol{\psi})p_g(\boldsymbol{\theta};\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\theta}. \tag{1}$$

An analytical expression is not available for this high dimensional integral. In cases where the model is linear and Gaussian, the Kalman filter can be used for likelihood evaluation, signal extraction and forecasting. Here we rely on numerical integration techniques that need to be both practical and feasible. It is well established that we can use Monte Carlo simulation methods for the evaluation of (1); see Ripley (1987) for a general introduction. A naive Monte Carlo estimate of $L(\boldsymbol{y};\boldsymbol{\psi})$ is given by

$$\frac{1}{S}\sum_{k=1}^{M} p(\boldsymbol{y}|\boldsymbol{\theta}^{(k)};\boldsymbol{\psi}), \qquad \boldsymbol{\theta}^{(k)} \sim p_g(\boldsymbol{\theta};\boldsymbol{\psi}),$$

where S is the number of Monte Carlo replications and the simulated value of $\boldsymbol{\theta}^{(k)}$ is obtained by simulating the state vectors from the vector autoregressive process

$$\boldsymbol{\alpha}_{t+1} \;=\; \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \;\sim\; \mathrm{NID}(\boldsymbol{0}, \boldsymbol{Q}_t),$$

and with $\boldsymbol{\theta} = \boldsymbol{c} + \boldsymbol{Z}\boldsymbol{\alpha}$ for a given parameter vector $\boldsymbol{\psi}$. This Monte Carlo estimate is numerically highly inefficient since the simulated paths have no support from $\boldsymbol{y}$.

In various contributions in statistics and econometrics it is argued that (1) can be evaluated efficiently using the method of importance sampling; see, for example, Shephard and Pitt (1997), Durbin and Koopman (1997), Liesenfeld and Richard (2003) and Richard and Zhang (2007). For a feasible implementation of this method we require a Gaussian importance density $g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)$ from which the $\boldsymbol{\theta}$s are sampled conditional on the observation vector $\boldsymbol{y}$, where $\boldsymbol{\psi}^*$ denotes a fixed parameter vector, containing $\boldsymbol{\psi}$ as well as parameters $\tilde{\boldsymbol{\psi}}$ particular to the importance density $g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})$, i.e., $\boldsymbol{\psi}^* = (\boldsymbol{\psi}', \tilde{\boldsymbol{\psi}}')'$. Under the assumption that a numerically efficient device can be developed for sampling $\boldsymbol{\theta}$ from $g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)$, we can express the likelihood function (1) in terms of the importance density as

$$L(\boldsymbol{y};\boldsymbol{\psi}) = \int \frac{p(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)} g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)\,\mathrm{d}\boldsymbol{\theta}, \tag{2}$$

with the importance sampling estimate given by

$$\frac{1}{S}\sum_{k=1}^{S} \omega(\boldsymbol{y},\boldsymbol{\theta}^{(k)};\boldsymbol{\psi}^*), \qquad \omega(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi}^*) = \frac{p(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)}, \qquad \boldsymbol{\theta}^{(k)} \sim g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*), \tag{3}$$

where the number of simulations S should be sufficiently high and where $\boldsymbol{\theta}^{(k)}$ is drawn independently for $k = 1, \ldots, S$. In this framework we assume that $p_g(\boldsymbol{\theta};\boldsymbol{\psi}) = g(\boldsymbol{\theta};\boldsymbol{\psi})$, which implies that the marginal stochastic properties of $\boldsymbol{\theta}$ in the model are the same as in the importance sampling distribution. It follows immediately that

$$\omega(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi}^*) = \frac{p(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta};\boldsymbol{\psi})p_g(\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})g(\boldsymbol{\theta};\boldsymbol{\psi})/g(\boldsymbol{y};\boldsymbol{\psi}^*)} = g(\boldsymbol{y};\boldsymbol{\psi}^*)\,\frac{p(\boldsymbol{y}|\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})}, \tag{4}$$

see, for example, Durbin and Koopman (2012). The density $g(\boldsymbol{y};\boldsymbol{\psi}^*)$ can be taken as a scaling function since it does not depend on $\boldsymbol{\theta}$. The function $\omega(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi}^*)$ is usually referred to as the importance sampling weight function. If the variance of $\omega(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi}^*)$ exists, the estimate (3) is consistent for any $g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})$ and a central limit theorem applies; see Geweke (1989) and

Koopman, Shephard, and Creal (2009). We may expect that a well-behaved weight function leads to an efficient importance sampling estimate of the likelihood function.

NAIS : construction of the importance density

The key choice in selecting an importance density $g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)$ is numerical efficiency. We follow the predominant choice in the literature and opt for the Gaussian density; we construct $g(\,\cdot\,)$ efficiently using standard techniques such as regression analysis and the Kalman filter.

Several proposals for constructing a Gaussian $g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)$ have been developed. Shephard and Pitt (1997) and Durbin and Koopman (1997) determine the choice of $\tilde{\boldsymbol{\psi}}$ via a second order Taylor expansion of density $p(\boldsymbol{y}|\boldsymbol{\theta};\boldsymbol{\psi})$ around a $\boldsymbol{\theta}$ that is equal to the mode of $p(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi})$. The mode can be found by an iterative method involving the Kalman filter and the related smoother. Alternatively, in the EIS method of Liesenfeld and Richard (2003) and Richard and Zhang (2007), the appropriate Gaussian importance density is found by solving

$$\underset{\tilde{\boldsymbol{\psi}}_t}{\operatorname{argmin}} \int \lambda^2(y_t,\boldsymbol{\theta}_t;\boldsymbol{\psi}^*)\,\omega_t(y_t,\boldsymbol{\theta}_t;\boldsymbol{\psi}^*)\,g(\boldsymbol{\theta}_t|\boldsymbol{y};\boldsymbol{\psi}^*)\,\mathrm{d}\boldsymbol{\theta}_t, \tag{5}$$

for each $t = 1,\ldots,n$, with $\tilde{\boldsymbol{\psi}}' = (\tilde{\boldsymbol{\psi}}_1',\ldots,\tilde{\boldsymbol{\psi}}_n')$, $\boldsymbol{\psi}^{*\prime} = (\boldsymbol{\psi}',\tilde{\boldsymbol{\psi}}')$, and

$$\lambda(y_t,\boldsymbol{\theta}_t;\boldsymbol{\psi}^*) := \log\,\omega_t(y_t,\boldsymbol{\theta}_t;\boldsymbol{\psi}^*) := \log\,p(y_t|\boldsymbol{\theta}_t;\boldsymbol{\psi}) - \log\,g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t). \tag{6}$$

The importance density is effectively determined by the minimization of the variance of the log weight ω_t, for each t. Richard and Zhang (2007) evaluate the integral in (5) using importance sampling and perform its minimization via weighted least squares regression. Koopman, Lit, and Nguyen (2012) show that the EIS method can also fully rely on computationally efficient Kalman filter and smoothing methods. Their modification leads to a faster and efficient importance sampling method, especially for large state dimensions.

In a further development of EIS, Koopman et al. (2014) replace the evaluation of the integral in (5) by standard Gauss-Hermite quadrature methods. This results in a highly numerically efficient importance sampling technique, that can be augmented with easy-to-compute control variates to increase efficiency even further. They label their method numerically accelerated importance sampling (NAIS). The key to NAIS is the availability of analytic expressions for the marginal densities $g(\boldsymbol{\theta}_t|\boldsymbol{y};\boldsymbol{\psi}^*)$ given the Gaussian importance densities $g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})$ and a Gaussian marginal density $g(\boldsymbol{\theta};\boldsymbol{\psi}) = p_g(\boldsymbol{\theta};\boldsymbol{\psi})$. Although NAIS

was originally developed for a univariate signal $\theta_t \in \mathbb{R}$, the method can easily be extended to multiple dimensions; see Scharth (2012, Ch. 5) and the discussions in Koopman et al. (2014). Scharth (2012) proposes Halton sequences and quasi-Monte Carlo integration for the evaluation of high dimensional integrals. In the case of our dynamic Skellam model, the signal is only two-dimensional and hence we can still rely on Gauss-Hermite quadrature methods efficiently.

NAIS : bivariate numerically accelerated importance sampling

To facilitate the exposition, we express the Gaussian density as a kernel function in $\boldsymbol{\theta}_t$,

$$g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}}) = \prod_{t=1}^{n} g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t), \qquad g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t) = \exp\left(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_t - \frac{1}{2}\boldsymbol{\theta}_t'\boldsymbol{C}_t\boldsymbol{\theta}_t\right), \tag{7}$$

with scalar a_t, 2×1 vector $\boldsymbol{b}_t$, a symmetric 2×2 matrix $\boldsymbol{C}_t$, and bivariate $\boldsymbol{\theta}_t = (\theta_{1t}, \theta_{2t})'$. To ensure that $g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t)$ integrates to one, we set $a_t = -\log 2\pi + \frac{1}{2}\log|\boldsymbol{C}_t| - \frac{1}{2}\boldsymbol{b}_t'\boldsymbol{C}_t^{-1}\boldsymbol{b}_t$. We gather the five remaining parameters in $\boldsymbol{b}_t$ and $\boldsymbol{C}_t$ into the vector $\tilde{\boldsymbol{\psi}}_t$. NAIS obtains the importance sampling parameters $\tilde{\boldsymbol{\psi}}_t$ iteratively, starting from an initial guess $\tilde{\boldsymbol{\psi}}_t^{(0)}$, and updating it sequentially to $\tilde{\boldsymbol{\psi}}_t^{(k)}$ for $k = 1, 2, \ldots$, until convergence. Given $\tilde{\boldsymbol{\psi}}_t^{(k)}$, the next parameter vector $\tilde{\boldsymbol{\psi}}_t^{(k+1)}$ for the importance densities solves the EIS criterion

$$\underset{\tilde{\boldsymbol{\psi}}_t^{(k+1)}}{\operatorname{argmin}} \int\int \lambda^2(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^{*(k+1)}) \omega_t(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^{*(k)}) g(\boldsymbol{\theta}_t|\boldsymbol{y}; \boldsymbol{\psi}^{*(k)})\, d\theta_{1t}\, d\theta_{2t}, \tag{8}$$

where $\boldsymbol{\psi}^{*(k)}$ contains $\boldsymbol{\psi}$ and $\tilde{\boldsymbol{\psi}}^{(k)}$. The key to the implementation of NAIS is the availability of an analytical expression for the smoothing density $g(\boldsymbol{\theta}_t|\boldsymbol{y}; \boldsymbol{\psi}^{*(k)})$. In our case of Gaussian importance sampling distributions, we have

$$g(\boldsymbol{\theta}_t|\boldsymbol{y}; \boldsymbol{\psi}^{*(k)}) = \mathrm{N}(\widehat{\boldsymbol{\theta}}_t^{(k)}, \boldsymbol{V}_t^{(k)}) = \frac{1}{2\pi|\boldsymbol{V}_t^{(k)}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_t^{(k)})'(\boldsymbol{V}_t^{(k)})^{-1}(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_t^{(k)})\right), \tag{9}$$

where $\widehat{\boldsymbol{\theta}}_t^{(k)}$ and $\boldsymbol{V}_t^{(k)}$ are obtained from the Kalman filter and smoother, for given $\boldsymbol{\psi}^* = \boldsymbol{\psi}^{*(k)}$, applied to the linear Gaussian model $\boldsymbol{x}_t = \boldsymbol{\theta}_t + \boldsymbol{u}_t$ with disturbance $\boldsymbol{u}_t \sim \mathrm{N}(0, \boldsymbol{C}_t^{-1})$ and pseudo-observation $\boldsymbol{x}_t = \boldsymbol{C}_t^{-1}\boldsymbol{b}_t$, for $t = 1, \ldots, n$. It is straightforward to verify that the observation density $\prod_{t=1}^{n} g(\boldsymbol{x}_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t)$ is equivalent to $g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})$ in (7).

We numerically implement the minimization in (8) by the Gauss-Hermite quadrature

method; see, for example, Monahan (2001). For this purpose we define

$$\varphi(y_t, \boldsymbol{\theta}_t; \tilde{\boldsymbol{\psi}}_t^{(k+1)}, \boldsymbol{\psi}^{*(k)}) = \lambda^2(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^{*(k+1)})\omega_t(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^{*(k)}), \tag{10}$$

and we select a set of abscissae $\{z_i\}_{i=1}^M$ with associated Gauss-Hermite weights $h(z_i)$, for $i = 1, \dots M$. The numerical implementation of the minimization (8) becomes

$$\underset{\tilde{\boldsymbol{\psi}}_t^{(k+1)}}{\operatorname{argmin}} \sum_{i=1}^{M}\sum_{j=1}^{M} w_{ij} \cdot \varphi(y_t, \tilde{\boldsymbol{z}}_{ij,t}^{(k)}; \tilde{\boldsymbol{\psi}}_t^{(k+1)}, \boldsymbol{\psi}^{*(k)}), \tag{11}$$

with weight $w_{ij} = h(z_i)h(z_j)\exp(\frac{1}{2}z_i^2)\exp(\frac{1}{2}z_j^2)$ and $\tilde{\boldsymbol{z}}_{ij,t}^{(k)} = \widehat{\boldsymbol{\theta}}_t + \boldsymbol{F}_t^{(k)}\boldsymbol{z}_{ij}$, where the 2×2 square root matrix $\boldsymbol{F}_t^{(k)}$ is the result of the decomposition $\boldsymbol{V}_t^{(k)} = \boldsymbol{F}_t^{(k)}\boldsymbol{F}_t^{(k)\prime}$ and $\boldsymbol{z}_{ij} = (z_i\,,\, z_j)'$ for $i, j = 1, \dots, M$. In this implementation we have used the fact that $g(\tilde{\boldsymbol{z}}_{ij,t}^{(k)}|\boldsymbol{y}; \boldsymbol{\psi}^{*(k)}) \propto \exp(-\frac{1}{2}\boldsymbol{z}_{ij}'\boldsymbol{z}_{ij})$; see Koopman et al. (2014) and Scharth (2012, Ch. 5). The decomposition of $\boldsymbol{V}_t^{(k)}$ is needed because the joint set of M^2 abscissae $\boldsymbol{z}_{ij}$, for $i, j = 1, \dots, M$, is associated with the bivariate standard normal distribution.

We can express the minimization problem (11) as a standard weighted least squares computation applied to M^2 observations for the regression equation

$$\log p(y_t|\tilde{\boldsymbol{z}}_{ij,t}^{(k)}) \;\; = \;\; \text{constant} + \boldsymbol{\kappa}'\tilde{\boldsymbol{z}}_{ij,t}^{(k)} - \frac{1}{2}\boldsymbol{\xi}'\text{vech}(\tilde{\boldsymbol{z}}_{ij,t}^{(k)}\tilde{\boldsymbol{z}}_{ij,t}^{(k)\prime}) + \text{error}, \tag{12}$$

where $\boldsymbol{\kappa}$ and $\boldsymbol{\xi}$ are regression coefficient vectors and the regression weights are given by $w_{ij} \cdot \omega_t(y_t, \tilde{\boldsymbol{z}}_{ij,t}^{(k)}; \boldsymbol{\psi}^{*(k)}) \cdot g(\tilde{\boldsymbol{z}}_{ij,t}^{(k)}|\boldsymbol{y}; \boldsymbol{\psi}^{*(k)})$, and where $\text{vech}(\cdot)$ stacks elements of the upper triangular part of a symmetric matrix into a vector. The resulting weighted least squares estimates for $\boldsymbol{\kappa}$ and $\boldsymbol{\xi}$ yield the new values for $\boldsymbol{b}_t^{(k+1)}$ and $\text{vech}(\boldsymbol{C}_t^{(k+1)})$, respectively. Hence, new values for $\tilde{\boldsymbol{\psi}}_t^{(k+1)}$ are obtained for each $t = 1, \dots, n$. Using these new estimates, we can determine a new $g(\boldsymbol{\theta}_t|\boldsymbol{y}; \boldsymbol{\psi}^{*(k+1)})$ in (9) by constructing a new time series $\boldsymbol{x}_t$ and applying the Kalman filter and smoother to the linear Gaussian model given below (9). In this last step we obtain new values for $\widehat{\boldsymbol{\theta}}_t^{(k+1)}$ and $\boldsymbol{V}_t^{(k+1)}$, which we require in (9).

This procedure is iterated until convergence. Typically, we only need a small (< 10) number of iterations for the applications in this paper. We emphasize that the regression computations can be carried out in parallel over t, leading to a very efficient implementation.

NAIS : the algorithm

The minimum of (11) is obtained when $\log p(y_t|\boldsymbol{\theta}_t;\boldsymbol{\psi}) = \log g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t)$. Therefore we regress the log Gaussian density $\log g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t)$ as given by (7) on the log observation density $\log p(y_t|\boldsymbol{\theta}_t;\boldsymbol{\psi})$ by use of weighted least squares. The regression coefficient vector at time t, Ψ_t, consists of the intercept a_t, the individual components of the 2×1 vector $\boldsymbol{b}_t$ and the 2×2 matrix $\boldsymbol{C}_t$ at time t, i.e. $\Psi_t = (a_t, \boldsymbol{\kappa}', \boldsymbol{\xi}')'$. The optimum values $\widehat{\Psi}_t$ are obtained by applying the following iterative algorithm

(i) Find appropriate starting values for $\boldsymbol{\kappa}$ and $\boldsymbol{\xi}$ with $t = 1, ..., n$ and set $s = 1$ and $\Psi_t^{(s)} = (a_t, \boldsymbol{\kappa}', \boldsymbol{\xi}')'$. In most cases the algorithm is not very sensitive to starting values so $\boldsymbol{\kappa}$ consisting of ones and $\boldsymbol{C}_t(\boldsymbol{\xi})$ set to $\boldsymbol{I}_2$ suffices.

(ii) Construct the linear Gaussian state space model with observation equation $\boldsymbol{x}_t = \boldsymbol{\theta}_t + \boldsymbol{u}_t$ with disturbance $\boldsymbol{u}_t \sim \mathrm{N}(0, \boldsymbol{C}_t^{-1})$ and pseudo-observation $\boldsymbol{x}_t = \boldsymbol{C}_t^{-1}\boldsymbol{b}_t$, for $t = 1, \ldots, n$ and apply the Kalman filter and smoother to obtain $\widehat{\boldsymbol{\theta}}_t^{(k)}$ and $\boldsymbol{V}_t^{(k)}$ and use these to calculate $\tilde{\boldsymbol{z}}_{ij,t}^{(k)}$ as described below equation (11).

(iii) Minimize equation (11) by weighted least squares and obtain $\Psi_t^{(s+1)}$.

(iv) If $\sum_{t=1}^{n} ||\Psi_t^{(s+1)} - \Psi_t^{(s)}|| < \epsilon$, for some threshold value ϵ, the algorithm has converged and can be terminated. Otherwise, set $s = s + 1$ and go to step (ii).

Once the iterative algorithm has converged in step (iv), $\Psi_t^{(s+1)}, t = 1, \ldots, n$ represents the new importance density. The number of times the algorithm needs to be called before convergence depends on the model and the size of the dataset. Starting from init values the algorithm converges most of the time in 10 steps or less. For repeated analysis, values of $\widehat{\Psi}_t$ can be stored and used as starting values for the next call to the algorithm. The minimization of (11) can be carried out independently for all time points t and can therefore be done in parallel over t.

2 Intradaily time series of price changes in 2012

Table 1: Empirical distribution of price changes in multiples of tick size.

The table reports the empirical distribution (in percentage points) of tick price changes for the four stocks Walmart (WMT), Coca-Cola (KO), JPMorgan (JPM), and Caterpillar (CAT), in 2012. The majority of the observations are -1, 1 and 0, the distribution is close to symmetric and it centers around zero which validates the use of the MSKII(-1,1,0) distribution.

Company	≤ -4	-3	-2	-1	0	1	2	3	≥ 4
Wal-Mart Stores Inc. (WMT)	0.46	0.83	3.43	19.66	51.25	19.51	3.52	0.86	0.48
The Coca-Cola Company (KO)	0.25	0.44	2.09	18.11	58.31	17.90	2.20	0.45	0.25
JPMorgan Chase & Co. (JPM)	0.15	0.40	2.42	19.37	55.29	19.29	2.53	0.41	0.14
Caterpillar Inc. (CAT)	4.66	4.39	9.22	18.20	27.13	18.12	9.20	4.46	4.62

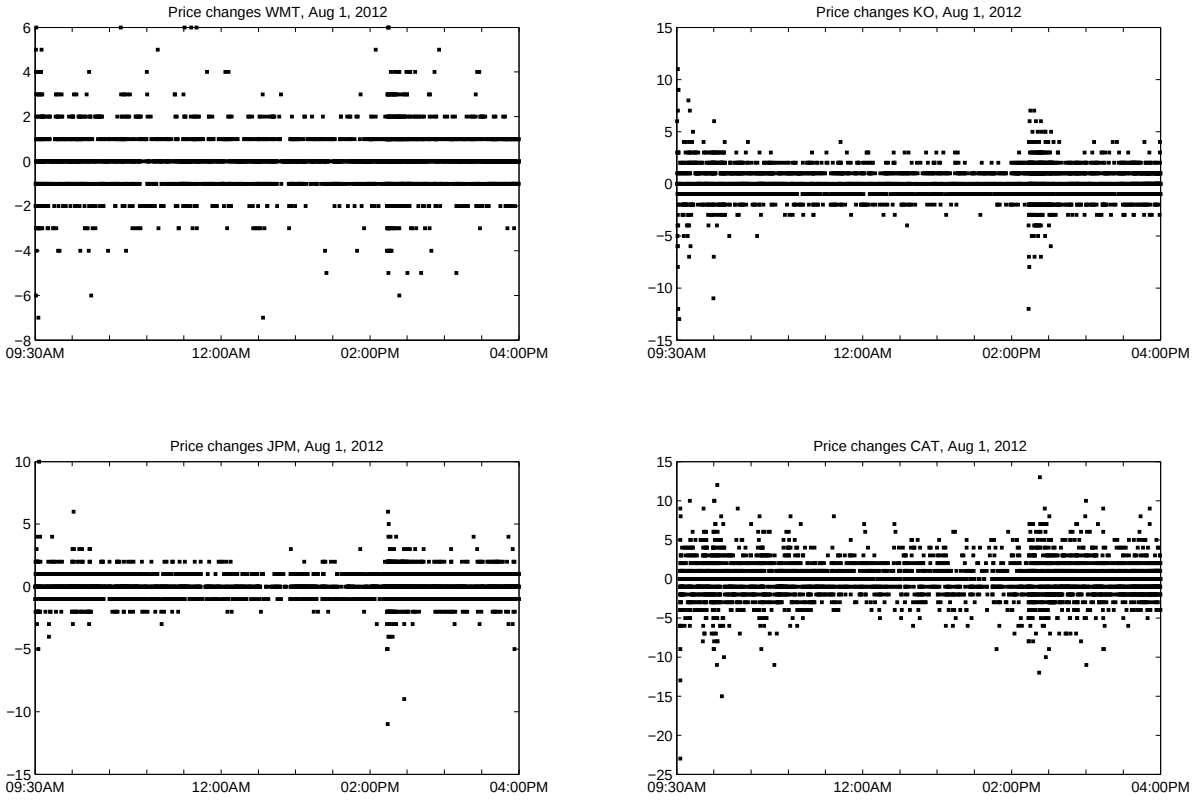

Figure 1: Price changes on August 1, 2012

The panels show the observed price changes for August 1, 2012 for the four stocks {WMT,KO,JPM,CAT}.

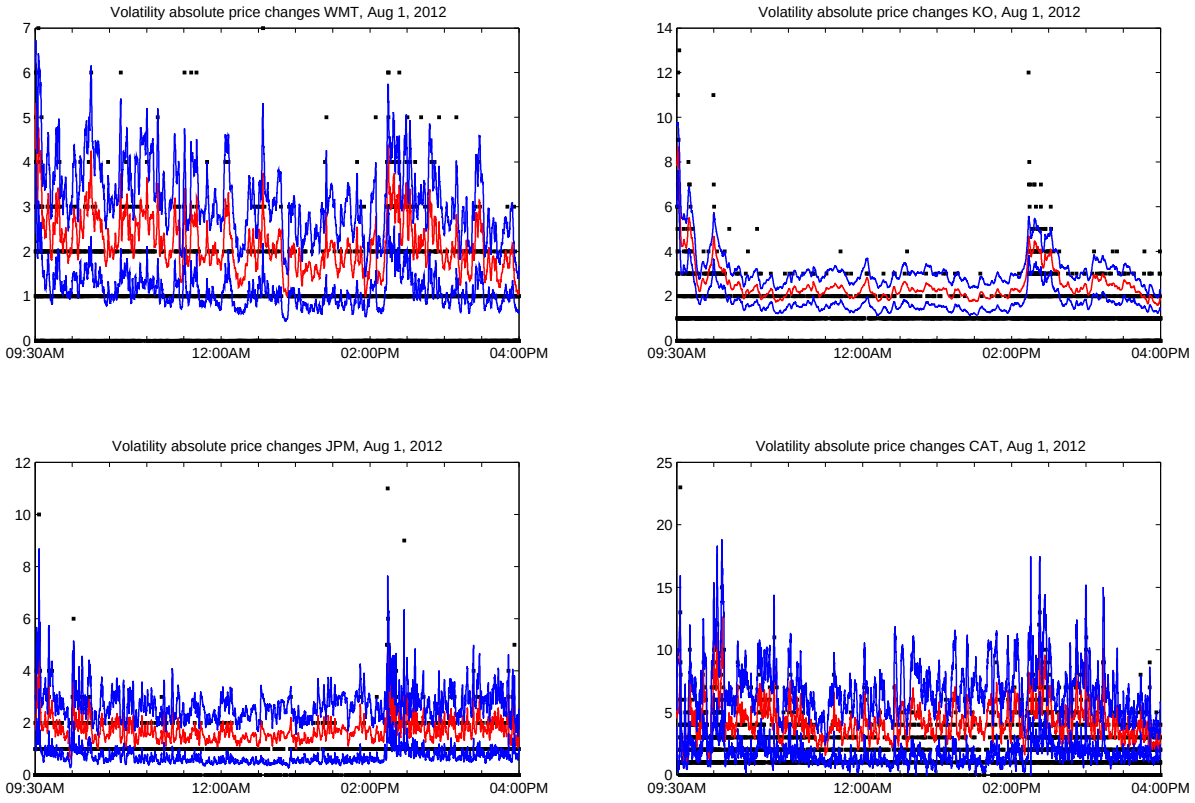

Figure 2: Absolute price changes on August 1, 2012 with volatility path estimate

The panels show the absolute values of observed price changes for August 1, 2012 for the four stocks {WMT,KO,JPM,CAT}. Furthermore, in each panel the estimate of $2 \times \sigma_t$ is presented together with its estimated 95% confidence interval.

References

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84(3)*, 669–684.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford: Oxford University Press.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–39.

Koopman, S. J., R. Lit, and T. M. Nguyen (2012). Modified efficient importance sampling using state space methods. Discussion paper, Tinbergen Institute.

Koopman, S. J., A. Lucas, and M. Scharth (2014). Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. *Journal of Business and Economic Statistics 33(1)*, 114–127.

Koopman, S. J., N. Shephard, and D. D. Creal (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics 149*, 2–11.

Liesenfeld, R. and J. F. Richard (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance 10*, 505–531.

Monahan, J. F. (2001). *Numerical methods of statistics*. Cambridge: Cambridge University Press.

Richard, J. F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 141*, 1385–1411.

Ripley, B. D. (1987). *Stochastic Simulation*. New York: Wiley.

Scharth, M. (2012). *Essays on Monte Carlo Methods for State Space Models*. Number 546 in Tinbergen Institute Research Series. Amsterdam: Thela Thesis and Tinbergen Institute.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84(3)*, 653–667.