



Duisenberg school of finance – Tinbergen Institute Discussion Paper

**TI 14-032 /IV/ DSF 73**

## **The Dynamic Skellam Model with Applications**

*Siem Jan Koopman<sup>1,2</sup>*

*Rutger Lit<sup>1</sup>*

*André Lucas<sup>1</sup>*

<sup>1</sup> *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute,*

<sup>2</sup> *CREATES, Aarhus University, Denmark.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

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

*Siem Jan Koopman<sup>(a,b)</sup>, Rutger Lit<sup>(a)</sup> and André Lucas<sup>(a)</sup>*

<sup>(a)</sup>VU University Amsterdam and Tinbergen Institute

<sup>(b)</sup>CREATES, Aarhus University

May 25, 2015

## **Abstract**

We introduce a dynamic Skellam model that measures stochastic volatility from high-frequency tick-by-tick discrete stock price changes. The likelihood function for our model is analytically intractable and requires Monte Carlo integration methods for its numerical evaluation. The proposed methodology is applied to tick-by-tick data of four stocks traded on the New York Stock Exchange. We require fast simulation methods for likelihood evaluation since the number of observations per series per day varies from 1000 to 10,000. Complexities in the intraday dynamics of volatility and in the frequency of trades without price impact require further non-trivial adjustments to the dynamic Skellam model. In-sample residual diagnostics and goodness-of-fit statistics show that the final model provides a good fit to the data. An extensive forecasting study of intraday volatility shows that the dynamic modified Skellam model provides accurate forecasts compared to alternative modeling approaches.

*Key Words:* non-Gaussian time series models; volatility models; importance sampling; numerical integration; high-frequency data; discrete price changes.

---

\*We thank István Barra, Asger Lunde and Albert J. Menkveld for their comments on an earlier draft. We also thank the two referees, the Associate Editor and the Editor whose many insightful suggestions have helped us to reshape and improve the paper considerably. Lit and Lucas acknowledge the financial support of the Dutch National Science Foundation (NWO grant VICI453-09-005). Koopman acknowledges support from CREATES, Aarhus University, Denmark, funded by the Danish National Research Foundation, (DNRF78).

# 1 Introduction

Stochastic volatility is typically associated with the time-varying variance in time series of daily continuously compounded rates of financial returns; for a review of the relevant literature, see Shephard (2005). The availability of high-frequency intraday trade information has moved the focus towards the estimation of volatility using realized measures such as realized volatility and realized kernels; see the seminal contributions of Barndorff-Nielsen and Shephard (2001, 2002), Andersen, Bollerslev, Diebold, and Labys (2001) and Hansen and Lunde (2006). Recent research has moved beyond the use of high-frequency data for obtaining daily observations of (realized) variances to the actual modeling of high-frequency price changes themselves at the intraday level. For example, Barndorff-Nielsen, Pollard, and Shephard (2012) and Shephard and Yang (2015) formulate continuous-time stochastic processes and design econometric models based on integer-valued Lévy processes using Skellam distributed random variables. Price changes of a stock are measured on a grid of one dollar cent and hence the tick-by-tick price change can be treated as a Skellam distributed random variable that takes values in  $\mathbb{Z}$ . Also Hansen, Horel, Lunde, and Archakov (2015) study the discrete nature of high-frequency price changes and explore their dynamic properties by formulating a stochastic Markov-chain process.

In our current study we develop a new statistical model that is empirically relevant for the discrete time series of tick-by-tick financial data. Such data enjoy the increasing interest of government regulators as well as industry participants given their potential impact on the stability of financial markets. Our new model has three important features that are needed to capture typical intraday properties of the data. First, the model builds on a dynamic modified Skellam distribution to make the model congruent with the realized data that consist of discrete-valued tick-size price changes defined on the set of integers  $\mathbb{Z}$ . Second, our modified Skellam distribution features a doubly dynamic variance parameter. The variance is allowed to be different over the course of a trading day due to intraday seasonal patterns, which we capture by including a spline function over the time of day. On top of this, we also allow for autoregressive intraday stochastic volatility dynamics to capture any remaining volatility dynamics over the course of the trading day that cannot be attributed to seasonal patterns. Third, our data requires a careful treatment of small price changes of the order

of 0, 1, or -1 dollar cents. For this purpose, we modify the dynamic Skellam distribution by allowing for a probability mass transfer between these different price change realizations. The probability mass transfer needs to vary over time as well because the data reveal that trades with a zero price-change are not spread evenly across the trading day. The resulting new model with these three features embedded performs well in terms of fit, diagnostics, and forecasting power compared to a range of alternative models.

Our model stands in a much longer tradition of dynamic models for count data. Early contributions regarding the dynamic modeling of count data in  $\mathbb{N}$  are reviewed in Durbin and Koopman (2012, Ch. 9). An example is the contribution of Jorgensen, Lundbye-Christensen, Song, and Sun (1999), who propose to model Poisson counts by a state space model driven by a latent gamma Markov process. The Skellam distribution is a natural extension to this literature, as it was originally introduced as the difference of two Poisson random variables; see Irwin (1937) and Skellam (1946). However it is not immediate clear how the treatment of Jorgensen et al. (1999) can be extended for the difference of Poisson random variables as it requires an analytical expression of a conditional distribution for a gamma variable given a Skellam variable. Other related initial work is presented by Rydberg and Shephard (2003) who propose a dynamic model for data in  $\mathbb{Z}$  by decomposing stock price movements into activity, direction of moves, and size of the moves. A very different approach to observations in  $\mathbb{Z}$  is related to integer autoregressive (INAR) models. Barreto-Souza and Bourguignon (2013), Zhang, Wang, and Zhu (2009), Freeland (2010), Kachour and Truquet (2010), Alzaid and Omair (2014) and Andersson and Karlis (2014) all propose extensions to the INAR model to enable the treatment of variables in  $\mathbb{Z}$ . These models are relatively simple to analyze as closed form expressions for the likelihood are available. However, a major drawback of these models in our current context is their lack of flexibility to incorporate missing observations and to allow for a time-varying variance process. Most related to our work is the contribution of Shahtahmassebi (2011) and Shahtahmassebi and Moyeed (2014) who adopt the Skellam distribution to analyze time series data in  $\mathbb{Z}$  within a Bayesian framework, whereas we use simulated maximum likelihood methods. However, their work does not treat the specific features of intraday financial price changes such as intraday seasonality, long stretches of missing values, and the time-varying modifications for the Skellam distribution. All these features are key for our current analysis of the empirical data. In addition, our new dynamic

modified Skellam distribution may also provide a useful flexible modeling framework in other empirical settings.

Our data consist of tick-by-tick discrete price changes for 4 stocks traded on the New York Stock Exchange (NYSE). For each second, there is either a trade or a missing value, such that the methodology needs to be able to account for possibly many missing values efficiently. Our state space framework for the dynamic modified Skellam model meets this requirement and can handle long time series that consist of a mix of observations and missing values. The number of zeros in the data does not appear to match the prediction by the standard Skellam distribution as it fails to pass various residual diagnostic tests. We therefore introduce a modified Skellam distribution that allows for a time-varying probability mass transfer and obtain a zero-deflated or zero-inflated Skellam model. This appropriately modified Skellam model passes the diagnostic tests and is successful in our forecasting exercise when compared to alternative models.

The new dynamic modified Skellam model has an intractable likelihood function. We therefore reformulate the model in terms of a nonlinear non-Gaussian state space model and estimate the static parameters by means of simulated maximum likelihood and importance sampling methods. In particular, we apply the numerically accelerated importance sampling (NAIS) methods of Koopman, Lucas, and Scharth (2014) which is an extension of the efficient importance sampling (EIS) method of Liesenfeld and Richard (2003) and Richard and Zhang (2007). The NAIS methodology obtains the parameters of the importance sampling distribution using Gauss-Hermite quadrature rather than simulation, and is applicable for high-dimensional state vectors. In Supplementary Appendix D we provide the details of how the NAIS methodology can be implemented to accommodate for both a time-varying mean and variance. Long time series can pose particular efficiency problems for importance sampling methods; see Robert and Casella (2004, §3.3) and Cappé, Moulines, and Ryden (2005, §6.1 and 9.1). However, we find that the dynamic Skellam model can be efficiently treated using the NAIS methodology for time series as long as 23,400 observations. The presented diagnostic tests show that the importance sampling weights are well-behaved in almost all cases.

The remainder of this paper is organized as follows. We present the new dynamic modified Skellam model in Section 2 and explain how it can be cast into a nonlinear non-Gaussian

state space form. Section 3 applies the dynamic Skellam model to four stocks, traded on NYSE, for all trading days in the year 2012. This section also contains information on model fit, diagnostic checks and forecasting performance. Section 4 concludes.

## 2 The dynamic Skellam model

Consider a variable  $Y_t$  that only takes integer values, that is  $Y_t \in \mathbb{Z}$ . Our aim is to analyze a time series of realizations for  $Y_t$  denoted by  $y_1, \dots, y_n$  where  $n$  is the length of the time series. We consider the Skellam distribution for  $Y_t$ , propose a novel modification of the Skellam distribution, and specify dynamic processes for the mean and variance.

### 2.1 The Skellam distribution

The probability mass function (pmf) of a Skellam distributed random variable  $Y_t \in \mathbb{Z}$  with parameters  $\mathbb{E}(Y_t) = \mu \in \mathbb{R}$  and  $\text{Var}(Y_t) = \sigma^2 \in \mathbb{R}^+$  is defined as  $\Pr(Y_t = y_t) = p(y_t; \mu, \sigma^2)$ , with

$$p(y_t; \mu, \sigma^2) = \exp(-\sigma^2) \left( \frac{\sigma^2 + \mu}{\sigma^2 - \mu} \right)^{y_t/2} I_{|y_t|}(\sqrt{\sigma^4 - \mu^2}), \quad (1)$$

where  $I_{|y_t|}(\cdot)$  is the modified Bessel function of order  $|y_t|$ ; see Abramowitz and Stegun (1972). The Skellam distribution was originally derived from the difference of two Poisson distributions; see Irwin (1937) and Skellam (1946). We then have  $\mu = \lambda_1 - \lambda_2$  and  $\sigma^2 = \lambda_1 + \lambda_2$ , where  $\lambda_1$  and  $\lambda_2$  are the intensities of the two underlying Poisson distributions; see also Alzaid and Omair (2010). Karlis and Ntzoufras (2009) show that the underlying Poisson assumption can be dispensed with and that the Skellam distribution can also be considered by itself as an interesting distribution defined on integers.

The Skellam distribution is right-skewed for  $\mu > 0$ , left-skewed for  $\mu < 0$ , and symmetric for  $\mu = 0$ . If  $\mu = 0$ , the Skellam pmf simplifies to

$$p(y_t; 0, \sigma^2) = \exp(-\sigma^2) I_{|y_t|}(\sigma^2). \quad (2)$$

In the upper panels of Figure 1 we present examples of Skellam distributions for a range of values for  $\mu$  and  $\sigma^2$ . The excess kurtosis of the Skellam distribution is  $1/\sigma^2$  and the

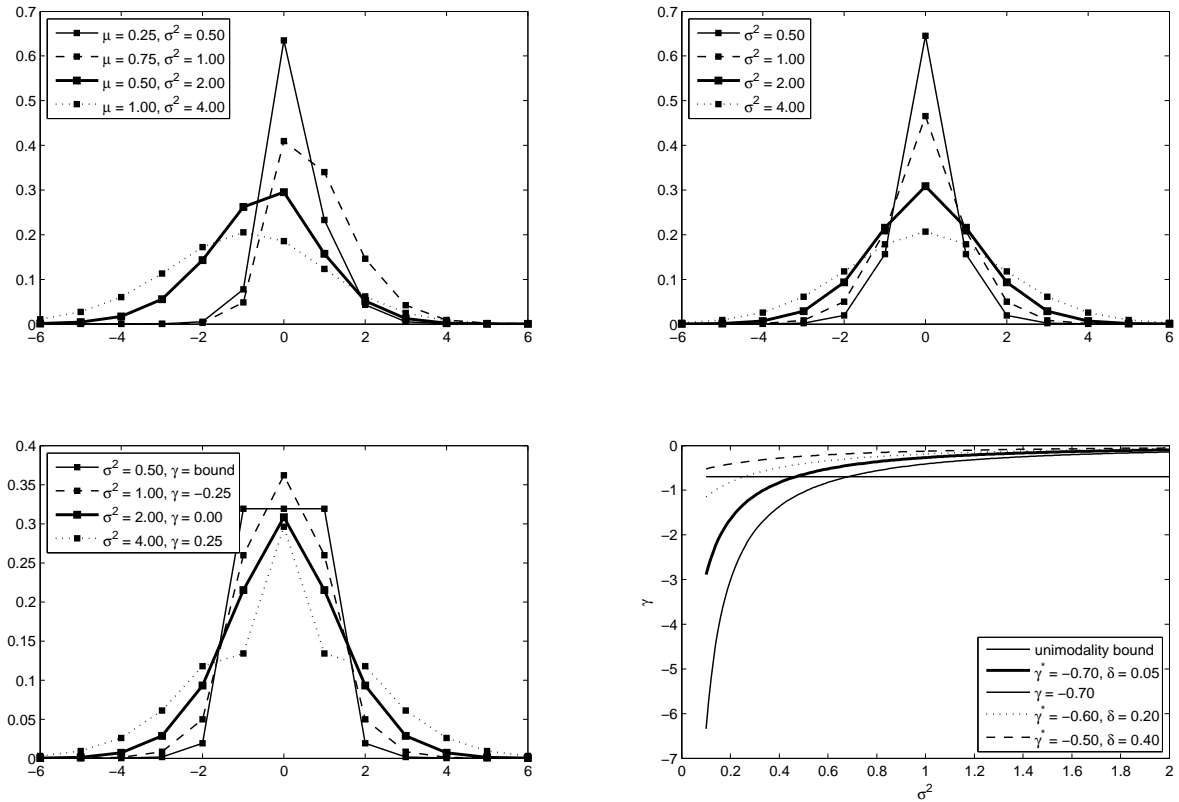


Figure 1: Skellam distributions and properties

Panel 1: Skellam distribution examples with pmf (1) for several combinations of  $\mu$  and  $\sigma^2$ . Panel 2: zero-mean ( $\mu = 0$ ) Skellam distribution (2) examples for several combinations of  $\sigma^2$ . Panel 3: MSKII( $-1, 1, 0; \mu, \sigma^2, \gamma$ ) distribution examples with pmf (3) for  $\mu = 0$  and several combinations of  $\sigma^2$  and  $\gamma_t$ . The distributions provide discrete support: the connecting lines are drawn for clarity and do not indicate continuity. Panel 4: unimodality bound and parameterized zero deflation bounds.

Gaussian distribution is a limiting case of the Skellam distribution; see Johnson, Kotz, and Kemp (1992) and references therein.

## 2.2 The modified Skellam distribution

The upper panels of Figure 1 reveal that the Skellam distribution is highly peaked at zero for low values of  $\sigma^2$ . This particular feature does not match the high-frequency tick-by-tick discrete stock price data well in our empirical application. To accommodate some more flexible patterns, we propose a modification of the Skellam distribution to compensate for the over- or under-representation of specific integers. For example, in our empirical application the standard Skellam distribution over-predicts the occurrence of 0s and under-predicts the occurrence of  $\pm 1$ s.



The first obvious modification of the Skellam distribution is the zero-altered Skellam distribution of Karlis and Ntzoufras (2006, 2009). Although they originally propose a modified Skellam distribution with a higher (zero-inflated) probability of observing  $Y_t = 0$ , their method can easily be adapted to accommodate a lower (zero-deflated) probability of observing  $Y_t = 0$ . To obtain a zero-deflated Skellam distribution, we transfer probability mass from  $Y_t = 0$  to  $Y_t \neq 0$ . We refer to this distribution as the modified Skellam distribution of type I (MSKI). More details of MSKI are presented in Appendix A.

The obvious consequence of redistributing the probability mass for  $Y_t = 0$  to all remaining integers is that the tails of the distribution inflate or deflate. The effect on the tails may be undesirable and we may want to accommodate for it by a further modification of MSKI. Our new proposed modified Skellam distribution of type II transfers probability mass from one specific integer to two other integers, that is, from  $Y_t = k$  to  $Y_t = i$  and  $Y_t = j$ , for the case of  $k$ -deflation, and the other way around for  $k$ -inflation, with  $i, j, k \in \mathbb{Z}$ . In this way, the probability mass at the remaining integers remains unchanged. The  $\text{MSKII}(i, j, k; \mu, \sigma^2, \gamma)$  distribution is defined by its pmf

$$p_{II}(y_t; i, j, k, \mu, \sigma^2, \gamma) = \begin{cases} P_{y_t}, & \text{for } y_t \notin \{i, j, k\}, \\ (1 - \gamma)P_i, & \text{for } y_t = i, \\ (1 - \gamma)P_j, & \text{for } y_t = j, \\ \gamma P_i + \gamma P_j + P_k, & \text{for } y_t = k, \end{cases} \quad (3)$$

where  $P_q = p(q; \mu, \sigma^2)$  is defined in equation (1) and  $q \in \mathbb{Z}$ , and with coefficient  $\gamma \in (-P_k/(P_i + P_j), 1)$ . The sign of the coefficient  $\gamma$  determines whether we inflate  $P_k$  (positive) or whether we deflate this probability (negative). For  $\gamma = 0$ , we recover the original Skellam distribution defined in (1). The lower bound of  $\gamma$  follows directly from the last equation in (3) since  $\gamma P_i + \gamma P_j + P_k \geq 0$  implies  $\gamma \geq -P_k/(P_i + P_j)$ . The mean and variance of the  $\text{MSKII}(i, j, k; \mu, \sigma^2, \gamma)$  distribution are given by

$$\begin{aligned} \mathbb{E}(Y_t) &= \mu_{II} = \mu - \gamma(i \cdot P_i + j \cdot P_j) + k \cdot \gamma(P_i + P_j), \\ \text{Var}(Y_t) &= \sigma_{II}^2 = \sigma^2 + \mu^2 + \gamma P_i(k^2 - i^2) + \gamma P_j(k^2 - j^2) - \mu_{II}^2, \end{aligned} \quad (4)$$

respectively, see Appendix B for derivations. For  $\gamma = 0$ , we clearly have  $\mu_{II} = \mu$  and

$\sigma_{II}^2 = \sigma^2$ . Given the data in our empirical application below, the  $\text{MSKII}(-1, 1, 0; 0, \sigma^2, \gamma)$  distribution will prove to be of particular interest.

If  $\gamma$  is sufficiently negative, the  $\text{MSKII}(i, j, k; \mu, \sigma^2, \gamma)$  distribution may become bimodal which can be undesirable in specific applications and for estimation purposes. However, we can formulate a stricter lower bound on  $\gamma$  to enforce unimodality. In particular, to ensure unimodality for the  $\text{MSKII}(-1, 1, 0; \mu, \sigma^2, \gamma)$  distribution under zero deflation we require  $P_0 > P_{-1}$  and  $P_0 > P_1$ , such that the lower bound  $\underline{\gamma}(\mu, \sigma^2)$  for  $\gamma$  is given by

$$\underline{\gamma}(\mu, \sigma^2) = (\min(P_{-1}, P_1) - P_0) / (\min(P_{-1}, P_1) + P_1 + P_{-1}). \quad (5)$$

The probability  $P_q$  is a function of  $\mu$  and  $\sigma^2$  for all  $q \in \mathbb{Z}$ . In Panel 3 of Figure 1 we present  $\text{MSKII}(-1, 1, 0; \mu, \sigma^2, \gamma)$  distributions for  $\mu = 0$  and different values of  $\sigma^2$  and  $\gamma$ . The figure reveals the effect of  $\gamma$  on the peakedness of the distribution. Panel 4 of Figure 1 presents examples of unimodal bounds  $\underline{\gamma}(\mu, \sigma^2)$  for  $\mu = 0$  and for different values of  $\sigma^2$ . We can select different model specifications to enforce  $\gamma$  to lie in the unimodality range; see section 3.2.

### 2.3 The Skellam model with dynamic mean and variance

Consider an observed time series for  $y_t \in \mathbb{Z}$  with  $t = 1, \dots, n$  where  $n$  is the time series length. The possible serial dependence in the time series  $y_1, \dots, y_n$  can be analyzed on the basis of a Skellam model with dynamic stochastic processes for the mean  $\mu_t$  and the variance  $\sigma_t^2$ . The dynamic MSKII model can be specified by

$$Y_t | \mu_t, \sigma_t^2 \sim \text{MSKII}(-1, 1, 0; \mu_t, \sigma_t^2, \gamma_t), \quad t = 1, \dots, n, \quad (6)$$

where  $\gamma_t = \gamma(\mu_t, \sigma_t^2)$  is the time-varying coefficient  $\gamma$  in (3) and is a function of  $\mu_t$  and  $\sigma_t^2$ . Hence we assume that the serial dependence in  $Y_t$  is accounted for by the time variation in  $\mu_t$  and  $\sigma_t^2$  only. In other words, conditional on  $\mu_t$  and  $\sigma_t^2$ ,  $Y_t$  is not subject to other dynamic processes. We model the dynamics of  $\mu_t$  and  $\sigma_t^2$  by a (possibly) nonlinear transformation of

an autoregressive process,

$$\begin{pmatrix} \mu_t \\ \sigma_t^2 \end{pmatrix} = s(\boldsymbol{\theta}_t), \quad \boldsymbol{\theta}_t = \mathbf{c}_t + \mathbf{Z}_t \boldsymbol{\alpha}_t, \quad (7)$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{d}_t + \mathbf{T}_t \boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \text{NID}(\mathbf{0}, \mathbf{Q}_t), \quad (8)$$

for  $t = 1, \dots, n$ , where vector  $s(\cdot)$  is referred to as the link function,  $\boldsymbol{\theta}_t \in \mathbb{R}^{r \times 1}$  is the signal vector, with  $r = 2$ ,  $\boldsymbol{\alpha}_t \in \mathbb{R}^{m \times 1}$  is the state vector,  $\mathbf{c}_t \in \mathbb{R}^{r \times 1}$  is a scalar intercept,  $\mathbf{d}_t \in \mathbb{R}^{m \times 1}$  is a vector of intercepts,  $\mathbf{Z}_t \in \mathbb{R}^{r \times m}$  is a matrix of coefficients,  $\mathbf{T}_t \in \mathbb{R}^{m \times m}$  is a transition matrix, and the disturbances  $\boldsymbol{\eta}_t$  are normally and independently distributed (NID) with mean zero and variance matrix  $\mathbf{Q}_t \in \mathbb{R}^{m \times m}$ . The vectors  $\mathbf{c}_t$ ,  $\mathbf{d}_t$  and matrices  $\mathbf{Z}_t$ ,  $\mathbf{T}_t$ ,  $\mathbf{Q}_t$  are typically constant but possibly time-varying in a deterministic manner. Typical examples of link functions  $s(\cdot)$  are the exponential function (to ensure positivity) and the scaled logistic function (to preserve lower and upper bounds). When the link function  $s(\cdot)$  directly requires the state vector  $\boldsymbol{\alpha}_t$  as an argument, we simply set  $r = m$ ,  $\mathbf{c}_t = \mathbf{0}$ , and  $\mathbf{Z}_t = \mathbf{I}_m$ . For an application with an observation distribution that only requires a time-varying mean or variance, we have a univariate signal and  $r = 1$ . The initial conditions for the elements of the state vector  $\boldsymbol{\alpha}_1$  depend on their dynamic properties. The variance matrix  $\mathbf{Q}_t$  is possibly positive semi-definite and hence the vector  $\boldsymbol{\eta}_t$  may contain zeros.

The model specified in equations (7)–(8) allows for a wide variety of dynamic patterns in  $\mu_t$  and  $\sigma_t^2$ , including autoregressive moving average dynamics, time-varying seasonal and cyclical patterns, deterministic and stochastic trends, and their combinations. Regression and intervention effects can be added to the signal as well. More details of their formulations in the form of (8) are provided in Durbin and Koopman (2012, Ch. 3). The dynamic Skellam model as specified above falls within the class of non-Gaussian nonlinear state space models which can be represented as

$$y_t \sim p(y_t | \boldsymbol{\theta}_t; \boldsymbol{\psi}), \quad \boldsymbol{\theta}_t = \mathbf{c}_t + \mathbf{Z}_t \boldsymbol{\alpha}_t, \quad \boldsymbol{\alpha}_{t+1} \sim p_g(\boldsymbol{\alpha}_{t+1} | \boldsymbol{\alpha}_t; \boldsymbol{\psi}), \quad t = 1, \dots, n, \quad (9)$$

with  $\boldsymbol{\alpha}_1 \sim p_g(\boldsymbol{\alpha}_1; \boldsymbol{\psi})$ , where  $\boldsymbol{\psi}$  is an unknown and fixed parameter vector gathering all the parameters in  $\mathbf{c}_t$ ,  $\mathbf{Z}_t$ ,  $\mathbf{d}_t$ ,  $\mathbf{T}_t$ , and  $\mathbf{Q}_t$ , and possibly in the signal function  $s(\cdot)$ . The

observation density  $p(y_t|\boldsymbol{\theta}_t; \boldsymbol{\psi})$  refers to the dynamic (possibly modified) Skellam distribution from Section 2 with signal  $\boldsymbol{\theta}_t$  representing the dynamic mean  $\mu_t$  and/or variance  $\sigma_t^2$ . The updating Gaussian state density  $p_g(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t; \boldsymbol{\psi})$  refers to the linear Markov process (8), and  $p_g(\boldsymbol{\alpha}_1; \boldsymbol{\psi})$  represents the initial condition for  $\boldsymbol{\alpha}_1$ . We assume that for given realizations of the signal  $\boldsymbol{\theta}' = (\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_n)$  the observations  $\mathbf{y} = (y_1, \dots, y_n)'$  are conditionally independent, and also write  $\boldsymbol{\theta} = \mathbf{c} + \mathbf{Z}\boldsymbol{\alpha}$  with  $\mathbf{c}' = (\mathbf{c}'_1, \dots, \mathbf{c}'_n)$ ,  $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_n)'$ , and  $\mathbf{Z}$  a block-diagonal matrix with blocks  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  on the leading diagonal. The joint conditional density for all observations and the marginal density for all states can now be written as

$$p(\mathbf{y}|\boldsymbol{\theta}; \boldsymbol{\psi}) = \prod_{t=1}^n p(y_t|\boldsymbol{\theta}_t; \boldsymbol{\psi}), \quad p_g(\boldsymbol{\alpha}; \boldsymbol{\psi}) = p_g(\boldsymbol{\alpha}_1; \boldsymbol{\psi}) \prod_{t=2}^n p_g(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}; \boldsymbol{\psi}), \quad (10)$$

respectively. Given the linear dependence of  $\boldsymbol{\theta}$  on  $\boldsymbol{\alpha}$ , the density  $p_g(\boldsymbol{\theta}; \boldsymbol{\psi})$  can be constructed directly from  $p_g(\boldsymbol{\alpha}; \boldsymbol{\psi})$ .

The state space representation implied by equations (9) or (10) for the dynamic Skellam model allows us to build on a well developed framework for the parameter estimation of  $\boldsymbol{\psi}$ , for the signal extraction of  $\boldsymbol{\theta}$  and the filtering and smoothing of  $\boldsymbol{\alpha}$ ; see Durbin and Koopman (2012) for a textbook treatment. As for all non-Gaussian nonlinear state space models, the main complication for the dynamic Skellam model is that the likelihood function  $\int p(\mathbf{y}|\boldsymbol{\theta}; \boldsymbol{\psi})p_g(\boldsymbol{\alpha}; \boldsymbol{\psi}) d\boldsymbol{\alpha}$  is analytically intractable. We therefore adopt the method of Monte Carlo maximum likelihood for parameter estimation, but also for signal extraction. In particular, we apply the numerically accelerated importance sampling (NAIS) method of Koopman et al. (2014) and show that it can efficiently handle long univariate time series (large  $n$ ). If we require a time-varying  $\mu_t$  or  $\sigma_t^2$ , that is a univariate signal,  $r = 1$ , we can apply the NAIS method of Koopman et al. (2014) without extensions. For handling both a time-varying mean  $\mu_t$  and variance  $\sigma_t^2$ , we have developed a bivariate extension of the NAIS methodology available in the Supplementary Appendix D. In our empirical study below we find we can set  $\mu_t = 0$ , such that we only consider a stochastic time-varying variance  $\sigma_t^2$ .

### 3 Analysis of high-frequency Skellam price changes

We study the dynamic properties of intraday high-frequency U.S. stock price changes listed at the New York Stock Exchange using our new dynamic Skellam model. High-frequency changes in stock prices evolve as positive and negative integer multiples of a fixed tick size. The tick size of stock prices at the NYSE is \$0.01, irrespective of the level of the stock price. This contrasts with other exchanges where tick sizes may increase with the price level of the traded instrument. For example, a sufficiently liquid stock with a price of \$4.00 rarely faces price jumps higher than 4 ticks, that is a 1% price change. On the other hand, a 4 tick price jump for a stock priced at \$100.00 represents a price change of only 0.04% and occurs much more frequently.

Rather than aggregating the data to one-minute or five-minute intervals, we analyze stock price changes on a second-by-second basis within a single trading day. As a consequence, all series have the same length of  $n = 23,400$  (6.5 hours  $\times$  3600) seconds with many missing values. By explicitly considering missing values in our analysis we take account of the duration between consecutive trades. Since there is more active trading at the beginning and end of a trading day, the number of missing values also varies throughout the day. We exploit Kalman filter and smoothing methods to handle missing values. Descriptive statistics for the data are reported and discussed below.

We analyze the intraday prices using the dynamic Skellam model as developed above. In accordance with other analyses of high-frequency stock returns, the sample mean in price changes for a sufficiently large sample size is typically close to zero; see, for example, Andersen and Bollerslev (1997). Hence we set  $\mu_t = 0$  and focus on the modeling of stochastic volatility  $\sigma_t^2$ . This yields a univariate signal ( $r = 1$ ) in our state space representation of the model as discussed in Section 2.3.

#### 3.1 Data

We use data from the trades and quotes (TAQ) database of the New York Stock Exchange at a one-second frequency. The data consist of the prices of four different stocks traded over the entire year 2012. We select companies from different industries and with different trade intensities. We analyze the tick-by-tick data without the “odd-lots” that represent

Table 1: Descriptive statistics of the four selected stocks for all trading days in 2012 combined as one sample. The table reports data characteristics of tick changes between 9:30am and 4:00pm. We report the “opening price” at 9:30 am (OP) January 1, 2012, the “closing price” at 16:00 pm (CP) December 31, 2012, the total number of trades in 2012 (#Trades), the percentage of zero price changes (%0), the percentage of  $-1, 1$  price changes ( $\% \pm 1$ ), variance (V), skewness (S), kurtosis (K) and the largest up tick (Max) and down tick (Min).

Company	OP	CP	#Trades	%0	$\% \pm 1$	V	S	K	Max	Min
Wal-Mart Stores Inc.	59.98	68.27	647,707	51.25	39.17	1.07	-0.01	13.59	19	-21
Coca-Cola Company	70.40	36.27	679,556	58.31	36.01	0.75	-0.00	15.65	19	-19
JPMorgan Chase	34.10	44.00	1,029,957	55.29	38.66	0.72	-0.01	7.96	15	-16
Caterpillar Inc.	93.43	89.57	792,829	27.13	36.32	4.82	-0.00	8.84	32	-32

trades with volumes less than 100 and are not recorded on the consolidated tape; see the discussion in O’Hara, Yao, and Ye (2014). The data require standard pre-processing. For a review of high-frequency data cleaning procedures; see for example Falkenberry (2002). We apply the cleaning algorithm of Brownlees and Gallo (2006) after applying a rudimentary filter corresponding to the cleaning steps P1, P2, P3 and T1, T2, T3 of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008, p. 8). Descriptive statistics are presented in Table 1.

The large difference in opening price and closing price for Coca-Cola Company is due to a 2:1 stock split on August 13, 2012. The number of trades ranges from almost 650,000 to more than a million over 2012. At the same time, the column “%0” in Table 1 shows that many trades do not result in a price change: the percentage of zeros ranges from 27% for Caterpillar to 58% for Coca-Cola. We can conclude from the “%0” and “ $\% \pm 1$ ” columns that the majority of trades only induce a maximum price change of  $\pm 1$ . A full breakdown of the empirical distribution of tick-size price changes is provided in the Supplementary Appendix E. The correct handling of zero price change trades is challenging for two reasons. First, zero price changes are not randomly distributed over the trading day. A Wald-Wolfowitz runs test, see Bradley (1968, Ch. 12), strongly rejects the null hypothesis of zeros following a random sequence throughout the trading day. The largest  $p$ -value of the runs test is  $8.73 \times 10^{-6}$  out of the 1000 days under consideration (4 stocks  $\times$  250 trading days in 2012). Second, long streaks of zeros and/or missing values occur regularly during slow trading periods of the day. This leads to a low volatility in price changes. Although the majority of observations within a trading day are either missing or are equal to  $-1, 0$  and  $1$ , large price changes (or jumps) do occur as indicated by the “Max” and “Min” columns in Table 1. Also the reported yearly sample variance and kurtosis for each stock reflect sufficient variation in

the tick-by-tick stock price changes. The challenge for our statistical dynamic model is to address all of these salient features appropriately.

### 3.2 Dynamic Skellam model with Intraday Stochastic Volatility

We consider the conditional observation density (6) with pmf (3). The standard Skellam model is a special case with  $\gamma_t = 0$ . The model specification for the dynamic variance, or the stochastic volatility, is based on the link function with  $r = 1$  given by

$$\sigma_t^2 = s(\boldsymbol{\theta}_t) = \exp(\theta_t), \quad t = 1, \dots, n, \quad (11)$$

where scalar  $\theta_t$  represents log-volatility. The dynamic signal process accommodates the salient features of intraday volatility by the following decomposition:

$$\theta_t = c + s_t + \alpha_t, \quad \alpha_{t+1} = \phi\alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_{\eta,t}^2), \quad (12)$$

for  $t = 1, \dots, n$ , where the constant  $c$  represents the overall daily log-volatility,  $s_t$  reflects the seasonal variation in intraday volatility, and the autoregressive component  $\alpha_t$  captures the local clustering of high and low price changes throughout the day. The constant and seasonal effects are treated as fixed and deterministic. The dynamic component  $\alpha_t$  is assumed stationary ( $|\phi| < 1$ ) and is driven by the disturbance or innovation  $\eta_t$ . We assume  $\eta_t$  is normally and independently distributed with mean zero and a time-varying variance. The time-varying variance is specified as a fixed function of time and reflects scheduled news announcements that may lead to relatively large price adjustments.

The seasonality in volatility is typically due to the high trading intensity at the beginning and end of the trading day, and the low intensity during the lunch break. A parsimonious specification for the seasonal effect is obtained by using a spline function that can interpolate different levels of volatility smoothly over the time-of-day. In particular, we let  $s_t$  be an intraday zero-sum regression spline function that we can represent as

$$s_t = \boldsymbol{\beta}' \tilde{\mathbf{W}}_t, \quad t = 1, \dots, n, \quad \sum_{t=1}^n s_t = 0, \quad (13)$$

where  $\beta$  is a  $K \times 1$  vector of parameters associated with the location of  $K + 1$  spline knots and  $\tilde{\mathbf{W}}_t$  is the  $t$ -th column of the zero sum interpolation weight matrix  $\tilde{\mathbf{W}}$  as constructed in Harvey and Koopman (1993); see also Poirier (1973). The zero-sum spline implies a restriction ( $K + 1$  knots,  $K$  parameters) to ensure the identification of the constant  $c$ . For our data set, a sharp decrease in volatility takes place in the first half hour (09:30-10:00) of many trading days. Furthermore, the lunch break and close of the market are key events. Therefore, we set  $K = 3$  and choose the knot positions at  $\{09:30, 10:00, 12:30, 16:00\}$ . Many variations around these knot locations have been considered but do not significantly affect the results reported below.

The variance of the innovations for the stationary component  $\alpha_t$  is time-varying to account for increased volatility due to special news announcements during the trading day. Many of such news announcements are released at pre-set time periods, such as 08:30, 10:00, and other; see Andersen, Bollerslev, Diebold, and Vega (2003). The effect of the news announcement before the opening of the market at 09:30 is captured by the first knot of the spline  $s_t$ . The possible effect of, say, a 10:00 news announcement, however, is harder to accommodate by the spline or AR(1) process only. For this purpose we introduce a separate parameter to model a (possible) temporary jump in volatility between 10:00 and 10:01. We do so by defining the indicator variable  $\tau_S(t) = 1$  for  $t = 1800, \dots, 1860$  (corresponding to the first minute after 10am), and zero otherwise, thus increasing the variance of  $\eta_t$  from  $\sigma_\eta^2$  to  $\sigma_\eta^2 + \sigma_{\eta,S}^2$  during this period, where  $\sigma_{\eta,S}^2 > 0$ . An increase of the variance for  $\eta_t$  allows  $\alpha_{t+1}$  to vary more than in other time periods.

We ensure unimodality of the  $\text{MSKII}(-1, 1, 0; 0, \sigma_t^2, \gamma_t)$  distribution under zero deflation via a parsimonious re-parameterization as follows. We introduce the coefficients  $-1 < \gamma^* < 1$  and  $\delta > 0$ . Then we determine  $\gamma_t$  as  $\gamma_t = \gamma^*$ , when  $\gamma^* \geq 0$ , and  $\gamma_t = -\gamma^* \times \underline{\gamma}(0, \sigma_t^2 + \delta)$ , when  $\gamma^* < 0$ , since  $\underline{\gamma}(0, \cdot) < 0$ . The coefficient  $\delta$  ensures a left-horizontal shift from the lower bound of  $\underline{\gamma}(0, \sigma_t^2)$  in order to avoid potential numerical issues for its limit as  $\sigma_t^2 \rightarrow 0$ ; see Panel 4 of Figure 1. The condition of unimodality stabilizes some numerical issues in likelihood evaluation since the construction of an importance density for bimodal distributions is rather challenging; see the discussion in Durbin and Koopman (2012, p. 253).



### 3.3 Parameter estimation results

The parameter vector for our dynamic Skellam model is given by  $\boldsymbol{\psi} = (\phi, \sigma_\eta, \sigma_{\eta,S}, c, \delta, \gamma^*, \boldsymbol{\beta}')'$ . The log-likelihood function is computed by the NAIS algorithm of Koopman et al. (2014); see Appendix D for the details. The log-likelihood is maximized for each trading day and stock using a quasi-Newton optimization method based on the numerical evaluation of the score with respect to  $\boldsymbol{\psi}$ . In NAIS, we require the evaluation of a Gauss-Hermite polynomial and base it on  $M = 12$  abscissae points. Higher values of  $M$  does not lead to more accurate results. The actual likelihood evaluation in NAIS is based on  $S = 100$  simulations with common random numbers during the optimization. The average optimizing time for one trading day ( $K = 3$ , 9 parameters,  $n = 23,400$ ) is between 5 and 15 minutes. Computations are performed on a i7-2600, 3.40 GHz desktop PC using four cores. Appendix C provides some further simulation evidence of the estimation procedure and its time requirements.

The parameter vector is estimated for each stock and each trading day in 2012. Given the large number of estimates, we provide a graphical presentation in Figure 2. In particular, we present the parameter estimates of  $\phi$ ,  $\sigma_\eta$ ,  $c$ , and  $\gamma^*$ . The estimates vary from day to day and characterize the intraday dynamics of price changes for that specific day. We have between 2500 and 4000 observations available for the estimation of  $\boldsymbol{\psi}$  on daily basis; see Table 1. It also allows us to carry out a meaningful forecasting study in Section 3.7.

The top row in Figure 2 shows the estimates of  $\phi$ . Overall, the estimates indicate a high degree of persistence of the autoregressive process  $\alpha_t$ . The average estimate of  $\phi$  over all trading days of 2012 exceeds 0.94 for each stock. Some individual days exhibit a  $\phi$  estimate that is clearly below the average. It indicates that the cubic spline  $c + s_t$  already captures most of the information for that specific day. We investigate the individual contribution of the spline versus the autoregressive component in Section 3.6 in more detail.

The second row reveals how the daily estimate of the volatility of the autoregressive component varies over time. Volatility levels appear to be somewhat higher in February and/or August for most stocks.

The third row shows the daily estimates of the constant  $c$ . For Walmart, the time series of  $c$  estimates shows a steady increase of the overall average daily volatility level during the year. For Coca-Cola, the structural break in the daily estimates of  $c$  in Augustus 13,

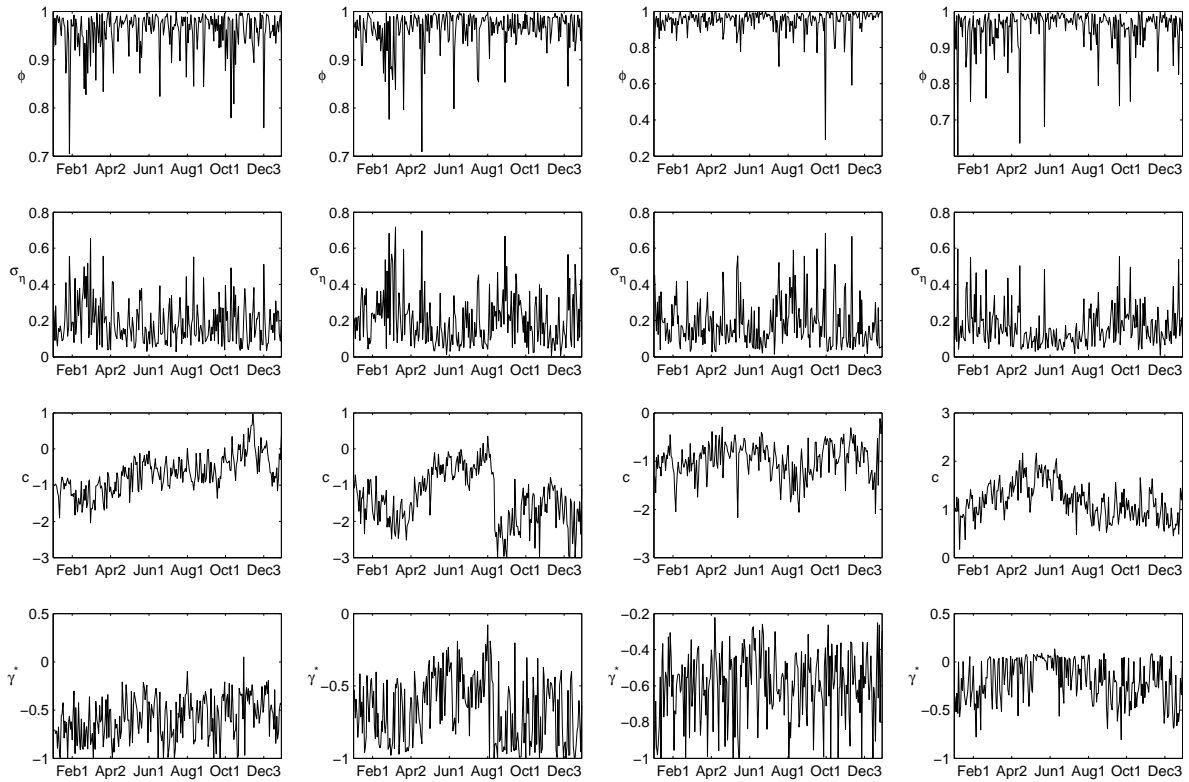


Figure 2: Maximum likelihood estimates of  $\psi$

The figure shows the maximum likelihood estimates of the first four elements of  $\psi$  where each column correspond to one of the four stocks in the order WMT, KO, JPM and CAT and the rows represent the parameter estimates in the order  $(\phi, \sigma_\eta, c, \gamma^*)$ .

2012, clearly coincides with the 2:1 stock split on that day. The constants  $c$  naturally play a dominant role in the overall level of daily log-volatility. As such, they may be compared to alternative estimates of integrated volatility based on high-frequency data. Interestingly, the time series correlations over all trading days in 2012 of our estimates of  $c$  with the logged realized volatility (RV) measure as estimated using the algorithm of Aït-Sahalia, Mykland, and Zhang (2011), based on 5-minute intervals, are high. The correlations are 0.90, 0.88, 0.67, and 0.93 for Walmart, Coca-Cola, JPMorgan, and Caterpillar, respectively.

The bottom panels in Figure 2 show the parameter estimates of  $\gamma^*$ . The estimates of  $\gamma^*$  are typically highly statistically significant, which indicates that our modification of the standard Skellam distribution is empirically relevant. For all stocks the 0-deflated model ( $\gamma^* < 0$ ) is clearly preferred. Only for CAT we have that some periods are subject to 0-inflation. CAT has the largest stock price compared to the others stocks, resulting in a larger

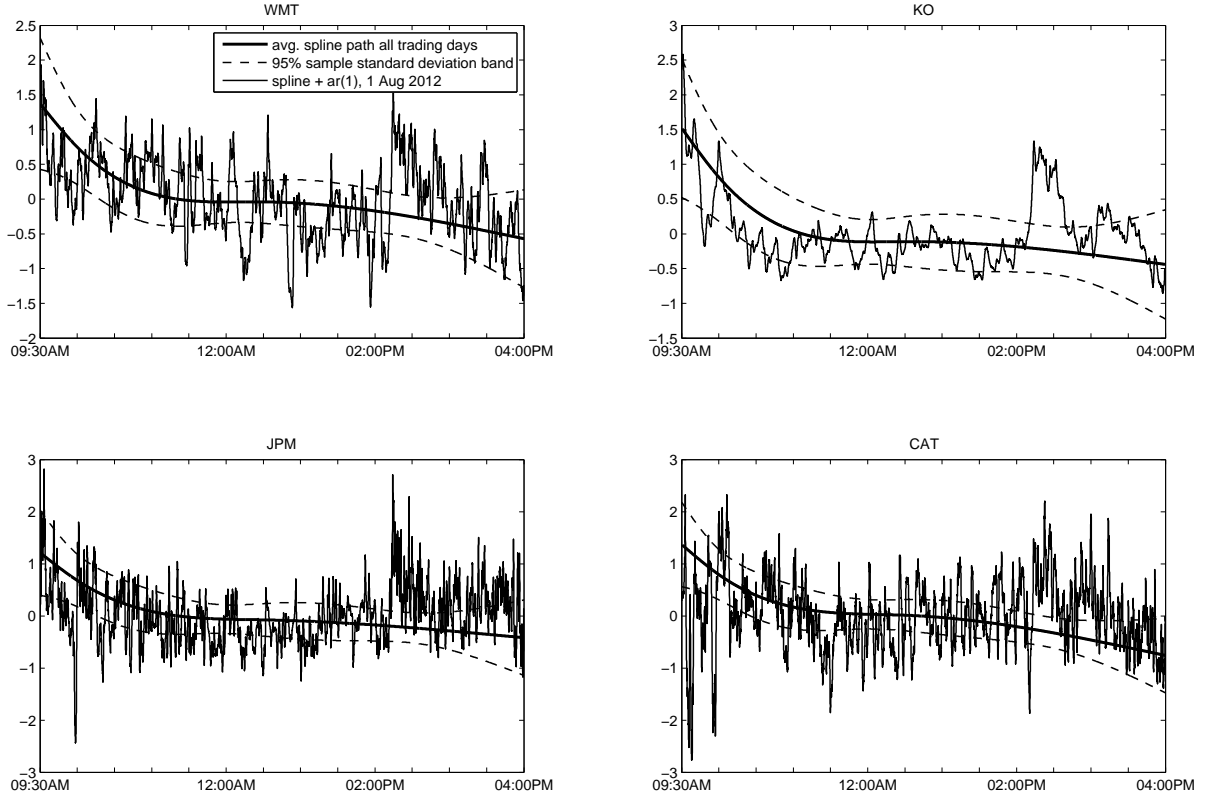


Figure 3: Average spline path

The figure shows the time series average of the zero sum spline  $s_t$  and a 95% confidence band based on all trading days of 2012. For Aug 1, 2012, it also shows the value of  $s_t + \alpha_t$ .

value of  $\sigma_t^2$  on average. A larger value of  $\sigma_t^2$  comes with a lower predicted probability of 0s, such that zero inflation rather than deflation becomes more relevant for CAT compared to the other stocks. Our type II modified Skellam model also outperforms the standard zero-deflation type I modification of the Skellam model of Karlis and Ntzoufras (2006, 2009), which is why we do not report the latter here.

### 3.4 Signal extraction

Figure 3 presents the time series average of our estimated zero sum cubic spline  $s_t$ , with corresponding 95% confidence bands. Instead of the commonly found volatility U-shape, we only find increased levels of volatility at the start, but not at the end of the average trading day in 2012.

To highlight the possible departures of the the fitted signal from the average spline level across all days, we also present the estimates of the spline plus the autoregressive

component  $(s_t + \alpha_t)$  for one specific day (August 1, 2012) in Figure 3. We find that for each of the four stocks the intraday volatility pattern is close to the overall average spline pattern. At the same time, we observe that particularly the autoregressive component picks up substantial temporary departures from the average level within the day. The size and patterns of the departures vary per stock and per day. For some stocks, departures appear relatively short-lived; see, for example, Caterpillar and JPMorgan. For other stocks, such as Walmart and Coca-Cola, departures are much more persistent. These patterns reveal why the autoregressive component  $\alpha_t$  contributes to the model specification and why it is statistically significant. In Section 3.7 we also verify whether  $\alpha_t$  leads to more precise forecasts of the magnitude of price changes for the next day.

### 3.5 Goodness-of-fit

To assess the model fit and the statistical contribution of the autoregressive component  $\alpha_t$ , we consider three different model specifications. All three specifications are based on the modified type II Skellam distribution but differ in the composition of the log-volatility signal:

- (i) Model  $\mathcal{A}$ : the static type II modified Skellam model with  $\mu_t = 0$  and static  $\sigma_t^2 = \exp(c)$ . The parameter vector is given by  $\boldsymbol{\psi} = (c, \delta, \gamma^*)'$ .
- (ii) Model  $\mathcal{B}$ : the spline-based model with  $\mu_t = 0$  and time-varying  $\sigma_t^2 = \exp(c + s_t)$ , where  $s_t$  is the zero sum cubic spline specified in (13). The parameter vector is given by  $\boldsymbol{\psi} = (c, \delta, \gamma^*, \boldsymbol{\beta}')$ .
- (iii) Model  $\mathcal{C}$ : the complete model with  $\mu_t = 0$  and  $\sigma_t^2 = \exp(c + s_t + \alpha_t)$  as in (12). The parameter vector is given in Section 3.3.

For each model specification, the parameter vector is estimated by maximum likelihood using NAIS. Figure 4 presents the log-likelihood differences (times 2) between Model  $\mathcal{B}$  and Model  $\mathcal{C}$  only, because the log-likelihood differences with respect to Model  $\mathcal{A}$  are all much larger. For almost all stocks and days, the differences between the maximized log-likelihood values are large and statistically significant. In most cases the differences are so large that also in terms of model selection criteria, such as the Akaike information criterion, model  $\mathcal{C}$  is strongly preferred over model  $\mathcal{B}$ .

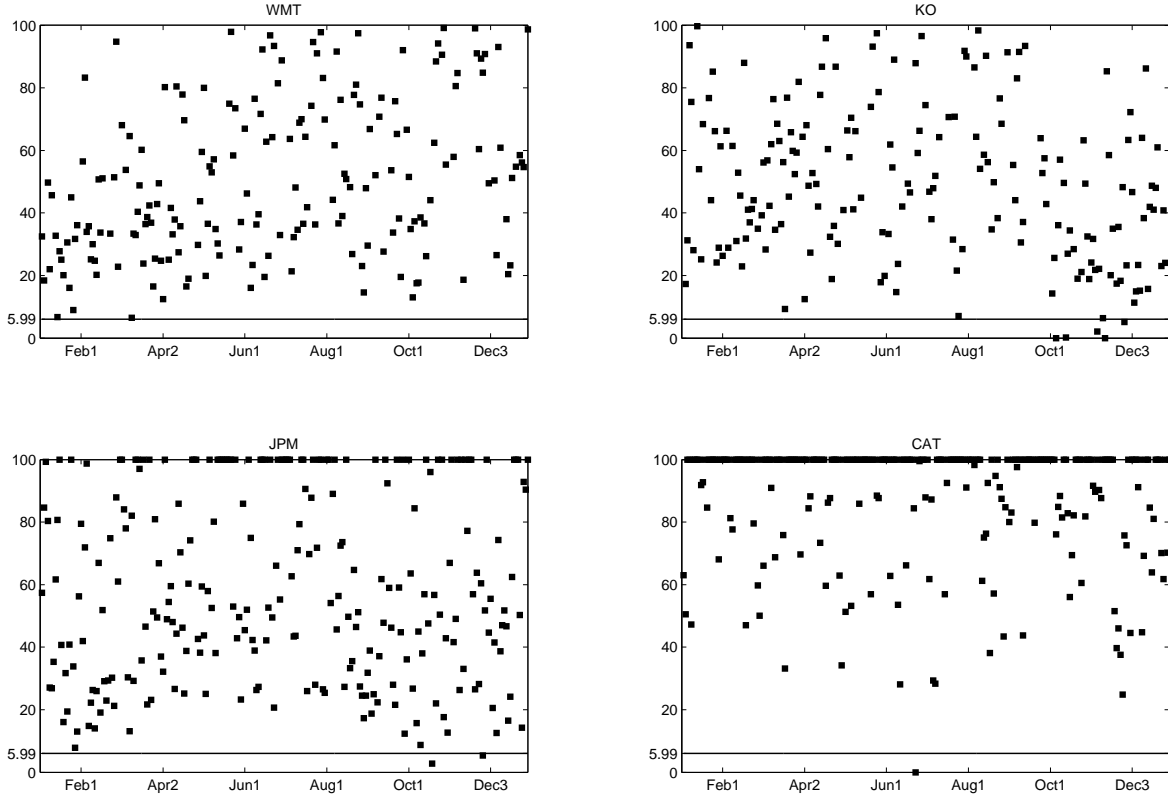


Figure 4: Log-likelihood ratios for Model  $\mathcal{B}$  (with spline) and Model  $\mathcal{C}$  (spline plus autoregressive).

Each panel is for a stock and presents the log-likelihood differences (times 2) for all days in 2012. A dot indicates the log-likelihood ratio values for a specific day in 2012 between a model with only a constant and a spline  $c + s_t$ , Model  $\mathcal{B}$ , and a model with spline and autoregressive component  $c + s_t + \alpha_t$ , Model  $\mathcal{C}$ . The horizontal line indicates the 5% critical value for the  $\chi^2(2)$  distribution corresponding to hypothesis  $H_0: \phi = 0, \sigma_\eta = 0$ . The differences are capped at 100 for visualization purposes.

### 3.6 Diagnostic checking

#### Variance of importance sampling weights

The estimation results from Section 3.3 rely on importance sampling methods. The log importance sampling weights can be used for diagnostic checking purposes. When the sample variance of the importance weights is high, likelihood calculations and signal extraction may change substantially when a different simulation sample is used. Geweke (1989) argues that importance sampling methods should only be used if the variance of the importance weights is known to exist. Robert and Casella (2004) provide examples of importance samplers that do not meet this condition and cases where this leads to biased results.

For the data at hand, we find that sample variances of the importance sampling weights are generally low, typically smaller than 1. To verify more formally whether the variances

of the importance weights exist, we follow Koopman, Shephard, and Creal (2009). Using maximum likelihood, they estimate the shape parameter  $\xi$  and the scale parameter  $\beta$  of a generalized Pareto distribution for the largest 1% to 50% out of 100,000 importance sampling weights. If the null hypothesis  $H_0 : \xi \leq 1/2$  cannot be rejected, they conclude that the variance of the importance sampling weights is finite and that results can be trusted.

## Pearson residuals

Diagnostic tests can also be based on the standardized Pearson residuals as given by

$$e_t = \frac{y_t - \mathbb{E}(y_t|y_{1:t-1})}{\sqrt{\text{Var}(y_t|y_{1:t-1})}}, \quad t = 1, \dots, n, \quad (14)$$

where  $y_{1:t-1}$  is the set of past Skellam returns  $\{y_1, \dots, y_{t-1}\}$ , and  $\mathbb{E}(y_t|y_{1:t-1})$  and  $\text{Var}(y_t|y_{1:t-1})$  are the one-step ahead observation forecast and its variance. Both of these depend on the filtered estimate of the scale parameter  $\mathbb{E}(\sigma_t^2|y_{1:t-1})$ . The importance sampling methods used for estimation can also be used for filtering and forecasting, albeit at a substantial computational cost given the large time series length  $n$ . However, for diagnostic checking purposes these computations only need to be performed once. We therefore regard the extra computation time as acceptable. An alternative is the use of nonlinear filtering methods such as the particle filter. The Pearson residuals  $e_t$ , for  $t = 1, \dots, n$ , of a correctly specified model have mean zero and unit variance, and both  $e_t$  and  $e_t^2$  should be serially uncorrelated. These properties can be verified by a number of diagnostic tests.

## Forecast distribution tests

Once the one-step-ahead predicted estimates of  $\sigma_t^2$ , for  $t = 1, \dots, n$ , are obtained we can test the distributional assumptions of the model. In particular, we test whether our dynamic modified Skellam model assigns the correct probabilities to the observations. We follow Jung, Kukuk, and Liesenfeld (2006) and draw a uniform random variable  $\tilde{u}_t$  on the interval  $[P(x_t \leq y_t - 1|y_{1:t-1}), P(x_t \leq y_t|y_{1:t-1})]$ . For a correctly specified model, the random draws  $\tilde{u}_t$ , for  $t = 1, \dots, n$ , are serially independent and uniformly distributed on the interval  $[0, 1]$ . The variable  $\tilde{u}_t$  can be transformed to a standard normal variable:  $e_t^* = F_N^{-1}(\tilde{u}_t)$ , where  $F_N^{-1}$  is the inverse normal distribution function. The transformed residuals  $e_t^*$  are also standard

normally distributed, and both  $e_t^*$  and  $(e_t^*)^2$  are serially uncorrelated, when the model is correctly specified.

### Diagnostic testing results

We apply the above diagnostic tests to our MSKII( $-1, 1, 0; 0, \sigma_t^2, \gamma(\sigma_t^2)$ ) model, Model  $\mathcal{C}$ . We benchmark the results against the two alternative specifications, Models  $\mathcal{A}$  and  $\mathcal{B}$ . We select the first trading day of every even month and present the corresponding diagnostic test results for this day in Table 2.

Table 2 shows that except for the single case of Caterpillar on Dec 03, 2012, the null hypothesis of a finite variance of the importance sampling weights is never rejected. The results also clearly support that allowing for intraday dynamics in  $\sigma_t^2$  is important. The static model  $\mathcal{A}$  is uniformly rejected based on all versions of the Ljung-Box test statistics. Interestingly, the results for the spline-based model  $\mathcal{B}$  and the dynamic model  $\mathcal{C}$  appear to be more similar. Based on autocorrelations in the levels of  $e_t$  or  $e_t^*$  the two models perform very similar, with a slight advantage for model  $\mathcal{C}$ . However, the dynamic model is much more adequate in filtering out the serial dependence in the second order moments, as revealed by the test results for  $e_t^2$  and  $e_t^{*2}$ . Whereas model  $\mathcal{B}$  has unacceptable diagnostics for most stocks and days, the diagnostic tests for model  $\mathcal{C}$  are mostly insignificant. We conclude that the autoregressive intraday component present in our new dynamic modified Skellam model is key to the good performance of the model. It results in a better performance than the commonly used intraday spline-based model.

### 3.7 Forecasting study

To verify the performance of the new model further, we perform a forecasting study for all 21 trading days in June 2012 in which we compare our dynamic modified Skellam model to four alternative methods. We focus on the prediction of volatility for each model by evaluating the probability of absolute price tick changes  $X_{t+1} = |Y_{t+1}|$ , for intraday times  $t = \tau, \dots, n - 1$ , for each day. The pmf of  $X_t$  is given by

$$p_{|II|}(X_t = x_t; \sigma_t^2, \gamma_t) = \begin{cases} p_{II}(Y_t = 0; -1, 1, 0, 0, \sigma_t^2, \gamma_t), & \text{for } x_t = 0, \\ 2 \cdot p_{II}(Y_t = x_t; -1, 1, 0, 0, \sigma_t^2, \gamma_t), & \text{for } x_t \geq 1. \end{cases} \quad (15)$$





The five considered models have in common that they all derive probabilities according to the type II modified Skellam distribution. They differ in the way the Skellam parameters  $\sigma_{t+1}^2$  and  $\gamma_{t+1}$  are obtained. Models  $\mathcal{A}, \mathcal{B}, \mathcal{C}$  are the parametric models as listed in Section 3.5. Models  $\mathcal{D}$  and  $\mathcal{E}$  are nonparametric benchmarks that are specified as follows.

- (iv) Model  $\mathcal{D}$ : we estimate  $\sigma_{t+1}^2$  using the sample variance using all observations in a rolling window of the past 900 seconds. We set  $\gamma_t = 0$ , such that the model collapses to the standard Skellam model.
- (v) Model  $\mathcal{E}$ : both  $\sigma_{t+1}^2$  and  $\gamma_{t+1}$  are obtained non-parametrically from the data. Define the empirical probability of a zero as  $\hat{P}_0$  and  $\hat{\sigma}_{t+1}^2$  as obtained under model  $\mathcal{D}$ . We then solve two equations for two unknowns, namely

$$\hat{\sigma}_{t+1}^2 = \sigma_{t+1}^2 - 2\gamma_{t+1}P_1, \quad (16)$$

$$\hat{P}_0 = P_0 + 2\gamma_{t+1}P_1, \quad (17)$$

where equations (16) and (17) follow from equations (4) and (3), respectively. By the substitution of (17) into (16), we obtain  $\hat{\sigma}_{t+1}^2 = \sigma_{t+1}^2 - \hat{P}_0 + P_0$  which we solve numerically for  $\sigma_{t+1}^2$  using a binary search algorithm. The resulting  $\sigma_{t+1}^2$  is substituted into (16) to obtain  $\gamma_{t+1}$ .

We emphasize that Models  $\mathcal{A}, \mathcal{B}, \mathcal{C}$  use the subsequent estimated parameter vectors from the day before. Further extensions can be obtained by considering a forecasting model for the daily estimates of  $\psi$ ; for instance, see Diebold and Li (2006). Even without these modifications, the forecasting experiment already produces some clear advantages of the new dynamic Skellam model, Model  $\mathcal{C}$ . For all models and all trading days, we start our forecast evaluation after a burn-in period of  $\tau = 60$  seconds. Models  $\mathcal{D}$  and  $\mathcal{E}$  subsequently extend the burn-in window to 900 seconds, after which the forecasts are updated using a rolling window. The results are presented in Table 3.

The performance of the models is first assessed in terms of an out-of-sample probabilistic loss function LOGL, which can be classified as a proper scoring rule; see Winkler (1969).  $\text{LOGL}_h$  sums the log probabilities for Model  $h \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}\}$  using the model's predictive pmf and the realized absolute tick-size change  $x_{t+1}$ . A loss of zero indicates that the absolute

Table 3: The table presents the total log loss (LOGL) of the 21 trading days of June 2012. The losses are based on the forecasting study presented in Section 3.7. The DM statistic represents the Diebold and Mariano (1995) statistic which is asymptotically distributed as a standard normal random variable and hence rejects the null hypothesis of equal predictive accuracy at the 5% level of significance in favour of Model  $\mathcal{C}$  if the DM test statistic is smaller than  $-1.65$ .

Model	Wal Mart (WMT)		Coca-Cola (KO)		JPMorgan (JPM)		Caterpillar (CAT)	
	LOGL	DM	LOGL	DM	LOGL	DM	LOGL	DM
$\mathcal{A}$	-57846	-25.18	-58754	-22.24	-96479	-31.43	-128170	-40.81
$\mathcal{B}$	-56595	-20.03	-57283	-18.91	-94611	-26.97	-124351	-35.27
$\mathcal{C}$	-55221	—	-55993	—	-92943	—	-121218	—
$\mathcal{D}$	-55715	-7.61	-56612	-8.06	-93860	-11.40	-121325	-1.20
$\mathcal{E}$	-55907	-9.58	-57147	-12.21	-93729	-9.94	-121901	-6.32

tick-size change  $x_{t+1}$  was perfectly predicted by the model. The log loss differences can also be compared between models using the Diebold Mariano (DM) test statistic; see Diebold and Mariano (1995). The DM statistic is asymptotically normally distributed under the null hypothesis of equal predictive accuracy. We take Model  $\mathcal{C}$  as our benchmark in the computation of the Diebold Mariano statistics.

Table 3 shows that the forecasts based on Model  $\mathcal{C}$  have always the lowest log loss. The new fully dynamic type II modification of the Skellam model clearly outperforms its static (Model  $\mathcal{A}$ ) and spline-based (Model  $\mathcal{B}$ ) counterparts, as well as the non-parametric zero-inflation model, Model  $\mathcal{E}$ . Using a one-sided test, Model  $\mathcal{C}$  also significantly outperforms the nonparametric benchmark Model  $\mathcal{D}$  for 3 out of the 4 stocks. Only for Caterpillar, the two models cannot be distinguished in a statistically significant manner. However, the excellent forecasting performance of Model  $\mathcal{C}$  remains despite its use of the estimate of the constant, spline, and autoregressive parameters of the day before. We emphasize that the parameter estimates are not recursively updated during the day. Models  $\mathcal{D}$  and  $\mathcal{E}$ , by contrast, do not rely on any parameter estimates from the previous day.

## 4 Conclusions

We have modeled tick-by-tick discrete price changes for U.S. stocks listed on the New York Stock Exchange. The analysis of high-frequency data attracts ever more attention from both government regulators and the financial industry. Hence the understanding of the dynamics in high-frequency data has become important. We have shown that the empirical analysis of

high-frequency tick-by-tick data can be based on modifications and dynamic extensions of the Skellam distribution. Our type II modified Skellam distribution features a dynamic variance parameter, and a dynamic transfer of probability mass to accommodate the non-standard properties of the data in terms of the occurrence of zero-price-changes. These features of our model are needed to have a stable importance sampling estimation procedure, a good in-sample fit, an adequate diagnostic performance, and an accurate out-of-sample forecasting performance, in comparison to a number of relevant benchmark models. We conclude that the new dynamic modified Skellam model provides a flexible modeling framework that can be effectively employed to capture the dynamics in high-frequency tick-by-tick data with many missing entries. Since the model produces intraday patterns of high-frequency volatility dynamics, it may provide an interesting and complementary perspective to the literature on nonparametric realized volatility measures and realized kernels which are proposed by Barndorff-Nielsen and Shephard (2001, 2002) and Andersen et al. (2001).

## References

- Abramowitz, M. and I. A. Stegun (1972). *Handbook of mathematical functions*. New York: Dover publications.
- Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics* 160(1), 160–175.
- Alzaid, A. and M. A. Omair (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society* 33(1), 17–45.
- Alzaid, A. A. and M. A. Omair (2014). Poisson difference integer valued autoregressive model of order one. *Bulletin of the Malaysian Mathematical Sciences Society* (forthcoming).
- Andersen, T. G. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115–158.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *J. American Statistical Association* 96(453), 42–55.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review* 93(1), 38–62.

- Andersson, J. and D. Karlis (2014). A parametric time series model with covariates for integers in  $z$ . *Statistical Modelling* 14(2), 135–156.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Realised kernels in practice: Trades and quotes. *Econometrics Journal* 4, 1–33.
- Barndorff-Nielsen, O. E., D. G. Pollard, and N. Shephard (2012). Integer-valued levy processes and low latency financial econometrics. *Quantitative Finance* 12(4), 587–605.
- Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. Royal Statistical Society B* 63(2), 167–241.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. Royal Statistical Society B* 64(2), 253–280.
- Barreto-Souza, W. and M. Bourguignon (2013). A skew true inar(1) process with application. Discussion paper, arXiv/1306.0156.
- Bradley, J. V. (1968). *Distribution-Free Statistical Tests*. New Jersey: Prentice-Hall.
- Brownlees, C. T. and G. M. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis* 51, 2232–2245.
- Cappé, O., E. Moulines, and T. Ryden (2005). *Inference in Hidden Markov Models*. New York: Springer.
- Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–265.
- 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.
- Falkenberry, T. N. (2002). High frequency data filtering. Technical report, Tick Data.
- Freeland, R. K. (2010). True integer value time series. *AStA-Advances in Statistical Analysis* 94, 217–229.

- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57, 1317–39.
- Hansen, P. R., G. Horel, A. Lunde, and I. Archakov (2015). A markov chain estimator of multivariate volatility from high frequency data. Discussion paper, CREATES, Aarhus University.
- Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise (with discussion). *J. Business and Economic Statist.* 24, 127–161.
- Harvey, A. C. and S. J. Koopman (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association* 88(424), 1228–1236.
- Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same poisson distribution. *J. Royal Statistical Society A* 100(3), 415–416.
- Johnson, N., S. Kotz, and A. W. Kemp (1992). *Univariate discrete distributions*. New York: Wiley.
- Jorgensen, B., S. Lundbye-Christensen, P. Song, and L. Sun (1999). A state space model for multivariate longitudinal count data. *Biometrika* 86(1), 169–181.
- Jung, R. C., M. Kukuk, and R. Liesenfeld (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics & Data Analysis* 51(4), 2350–2364.
- Kachour, M. and L. Truquet (2010). A p-order signed integer-valued autoregressive (sinar(p)) model. *Journal of Time Series Analysis* 32, 223–236.
- Karlis, D. and I. Ntzoufras (2006). Bayesian analysis of the differences of count data. *Statistics in medicine* 25(11), 1885–1905.
- Karlis, D. and I. Ntzoufras (2009). Bayesian modelling of football outcomes: using the Skellam’s distribution for the goal difference. *IMA Journal of Management Mathematics* 20, 133–145.
- Koopman, S. J., R. Lit, and A. Lucas (2015). Intraday stock price dependence using dynamic discrete copula distributions. Discussion paper, Tinbergen Institute.
- 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.
- O’Hara, M., C. Yao, and M. Ye (2014). What’s not there: Odd lots and market data. *Journal of Finance* 69(5), 2199–2236.
- Poirier, D. J. (1973). Piecewise regression using cubic spline. *Journal of the American Statistical Association* 68(343), 515–524.
- 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.
- Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer.
- Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics* 1(1), 2–25.
- 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.
- Shahtahmassebi, G. (2011). *Bayesian Modelling of Ultra High-Frequency Financial Data*. Doctoral thesis, Research with Plymouth University. University of Plymouth.
- Shahtahmassebi, G. and R. Moyeed (2014). Bayesian modelling of integer data using the generalised poisson difference distribution. *International Journal of Statistics and Probability* 3, 24–35.
- Shephard, N. (2005). *Stochastic volatility: Selected Readings*. New York: Oxford University Press.
- Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84(3), 653–667.
- Shephard, N. and J. J. Yang (2015). Continuous time analysis of fleeting discrete price moves. Discussion paper, Harvard University.

- Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society* 109(3), 296.
- Winkler, R. L. (1969). Scoring rules and the evaluation of probability assessors. *Journal of the American Statistical Association* 64(327), 1073–1078.
- Zhang, H., D. Wang, and F. Zhu (2009). Inference for inar(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference* 140, 667–683.

## Appendices

### A Modified Skellam distribution of type I

The MSKI distribution in which probability mass is transferred from  $Y_t \neq 0$  to  $Y_t = 0$  or vice versa is defined by its pmf

$$p_I(y_t; \mu, \sigma^2, \gamma) = \begin{cases} (1 - \gamma)p(Y_t = y_t; \mu, \sigma^2), & \text{for } y_t \neq 0, \\ \gamma + (1 - \gamma)p(Y_t = 0; \mu, \sigma^2), & \text{for } y_t = 0, \end{cases}$$

where  $P_q = p(q; \mu, \sigma^2)$  as defined in equation (1),  $q \in \mathbb{Z}$ , and  $\gamma \in (\frac{P_0}{P_0-1}, 1)$ . For  $\gamma = 0$  we recover the Skellam distribution as defined in (1) and for  $\gamma = \frac{P_0}{P_0-1}$  we have the lower bound  $P_0 = 0$ . If unimodality is required the zero deflation should be bounded as  $\gamma \in (\frac{\min(P_{-1}, P_1) - P_0}{1 + \min(P_{-1}, P_1) - P_0}, 1)$  which ensures  $P_0 \geq \min(P_{-1}, P_1)$ . The mean and variance of the MSKI distribution are  $\mathbb{E}(Y_t) = (1 - \gamma)\mu$  and  $\mathbb{V}\text{ar}(Y_t) = (1 - \gamma)\sigma^2 + \gamma(1 - \gamma)\mu^2$  which follows from

$$\mathbb{V}\text{ar}(Y_t) = (1 - \gamma) \sum_{x=-\infty}^{\infty} x^2 p(Y_t = x; \mu, \sigma^2) - (1 - \gamma)^2 \left[ \sum_{x=-\infty}^{\infty} x p(Y_t = x; \mu, \sigma^2) \right]^2,$$

with  $\sum_{x=-\infty}^{\infty} x^2 p(Y_t = x; \mu, \sigma^2) = \sigma^2 + \mu^2$  being the second moment of the Skellam distribution of (1). The inflation/deflation of probability mass to non-zero values of  $Y_t$  can be achieved in a similar way.

## B Moments of the MSKII( $i, j, k$ ) distribution

Let  $\mu$  and  $\sigma^2$  denote the mean and variance of the standard (non-deflated) Skellam distribution. The mean of the MSKII( $i, j, k, \mu, \sigma^2, \gamma$ ) distribution is given by

$$\begin{aligned}
\mathbb{E}(Y_t) &= \sum_{x \in \mathbb{Z}} x p_{II}(Y_t = x; \mu, \sigma^2, \gamma) \\
&= \left[ \sum_{x \in \mathbb{Z} \setminus \{i, j, k\}} x p(Y_t = x; \mu, \sigma^2) \right] + i(1 - \gamma)P_i + j(1 - \gamma)P_j + k(\gamma P_i + \gamma P_j + P_k) \\
&= \left[ \sum_{x \in \mathbb{Z} \setminus \{i, j, k\}} x p(Y_t = x; \mu, \sigma^2) \right] + iP_i + jP_j + kP_k - i\gamma P_i - j\gamma P_j + k\gamma P_i + k\gamma P_j \\
&= \mu - i\gamma P_i - j\gamma P_j + k\gamma P_i + k\gamma P_j,
\end{aligned} \tag{B.1}$$

which is equal to the first equation of (4). The second moment of the MSKII( $i, j, k, \mu, \sigma^2, \gamma$ ) distribution is given by

$$\begin{aligned}
\mathbb{E}(Y_t^2) &= \sum_{x \in \mathbb{Z}} x^2 p_{II}(Y_t = x; \mu, \sigma^2, \gamma) \\
&= \left[ \sum_{x \in \mathbb{Z} \setminus \{i, j, k\}} x^2 p(Y_t = x; \mu, \sigma^2) \right] + i^2(1 - \gamma)P_i + j^2(1 - \gamma)P_j + k^2(\gamma P_i + \gamma P_j + P_k) \\
&= \left[ \sum_{x \in \mathbb{Z} \setminus \{i, j, k\}} x^2 p(Y_t = x; \mu, \sigma^2) \right] + i^2 P_i + j^2 P_j + k^2 P_k \\
&\quad - i^2 \gamma P_i - j^2 \gamma P_j + k^2 \gamma P_i + k^2 \gamma P_j \\
&= \sigma^2 + \mu - i^2 \gamma P_i - j^2 \gamma P_j + k^2 \gamma P_i + k^2 \gamma P_j.
\end{aligned} \tag{B.2}$$

Combining (B.1) and (B.2) leads to the variance of the MSKII( $i, j, k$ ) distribution as presented in the second equation of (4).

## C Simulation study

We conduct a simulation study to verify the performance of the importance sampling estimation methodology explained in the Online Appendix D in combination with the Skellam model as presented in (12). The case of zero inflation, zero deflation and zero neutral is covered in this study. We assume that the Skellam model of (12) is the true data generating process and we simulate time series of Skellam variables with length  $n = 23,400$  which is



Table C.1: Simulation results of the zero-altered Skellam model

This table reports simulation averages of maximum likelihood estimates of the static parameters for the dynamic Skellam model of Section 3. The simulation averages are calculated with  $R = 100$  replications of time series with length  $n = 23,400$ . The true parameter values are in the table above the simulated values. Standard deviations of the estimates over the Monte Carlo simulations are in parentheses. The column  $t(s)$  denotes the average computation time (in seconds) for finding the maximum of the log likelihood function. Computations are carried out on a i7-2600, 3.40 GHz desktop PC using four cores.

	$\phi$	$\sigma_\eta$	$c$	$\gamma^*$	$\beta_1$	$\beta_2$	$t(s)$
true	0.99	0.05	-0.30	0.00	1.00	-0.40	
	0.987	0.055	-0.298	-0.024	1.005	-0.400	356.24
	(0.007)	(0.022)	(0.065)	(0.082)	(0.131)	(0.064)	
true	0.95	0.15	0.10	-0.50	1.00	-0.40	
	0.944	0.154	0.101	-0.498	0.997	-0.395	271.71
	(0.022)	(0.046)	(0.059)	(0.140)	(0.110)	(0.055)	
true	0.95	0.15	0.10	0.25	1.00	-0.40	
	0.945	0.150	0.104	0.252	0.996	-0.396	269.58
	(0.030)	(0.054)	(0.056)	(0.028)	(0.107)	(0.054)	

equal to the length of the tick price change series in the application of this paper. To incorporate missing values in the simulated data sets we denote  $P_{NaN}$  which is the probability of no trade at time  $t$ . We set  $P_{NaN} = 0.85$  at 09:30 and 16:00 and  $P_{NaN} = 0.95$  at 13:00. Every  $P_{NaN}$  between the time points 09:30–13:00 and 13:00–16:00 is determined by two triangles with the hypotenuses connecting  $P_{NaN} = 0.95$  in the middle of the day and  $P_{NaN} = 0.85$  at the beginning and end of the day. With the probability of a missing value over the day, missing values are randomly positioned at time points with the idea that the probability of a missing values is highest when trading activity is lowest. We refer to, for example, Koopman, Lit, and Lucas (2015) for graphs of trading patterns. For this simulation study, we obtain an average of 2000-2500 simulated trades out of 23,400 which is just below average.

The simulated data comes from a slightly more parsimonious model specification than (12). We set  $\delta = 0.30$ ,  $\sigma_{\eta,S} = 0$  and the vector of hyper parameters has dimension 6 and is given by

$$\boldsymbol{\psi}_{sim} = (\phi, \sigma_\eta, c, \gamma^*, \boldsymbol{\beta}', )',$$

where the elements of the  $2 \times 1$  vector  $\boldsymbol{\beta}$  correspond to a zero sum spline with spline knots placed at  $\{09:30, 12:30, 16:00\}$ . We present the estimation results in Table C.1.

Given that we are estimating a non-Gaussian state space model for a time series of length of  $n = 23,400$ , our estimation procedure is generally fast with optimizing times of only a couple of minutes. We also note that our methodology in combination with the novel Skellam model is able to estimate the parameter vector  $\boldsymbol{\psi}$  with high precision. Finally, the model is able to distinguish both zero-inflation and zero-deflation situations accurately. The results of this simulation study provide confidence for applying the Skellam model to real data sets.

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

*Siem Jan Koopman, Rutger Lit, André Lucas*

## D 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 et al. (2014).

### NAIS : likelihood evaluation and importance sampling

We can express the likelihood function for the non-Gaussian nonlinear state space model (9) as

$$L(\mathbf{y}; \boldsymbol{\psi}) = \int p(\mathbf{y}, \boldsymbol{\theta}; \boldsymbol{\psi}) d\boldsymbol{\theta} = \int p(\mathbf{y}|\boldsymbol{\theta}; \boldsymbol{\psi})p_g(\boldsymbol{\theta}; \boldsymbol{\psi}) d\boldsymbol{\theta}. \quad (\text{D.1})$$

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 (D.1); see Ripley (1987) for a general introduction. A naive Monte Carlo estimate of  $L(\mathbf{y}; \boldsymbol{\psi})$  is given by

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

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

In various contributions in statistics and econometrics it is argued that (D.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  $g(\boldsymbol{\theta}|\mathbf{y}; \boldsymbol{\psi}^*)$  from which the  $\boldsymbol{\theta}$ s are sampled conditional on the observation vector  $\mathbf{y}$ , where  $\boldsymbol{\psi}^*$  denotes a fixed parameter vector, containing  $\boldsymbol{\psi}$  as well as parameters  $\tilde{\boldsymbol{\psi}}$  particular to the importance density  $g(\mathbf{y}|\boldsymbol{\theta}; \tilde{\boldsymbol{\psi}})$ , 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  $g(\boldsymbol{\theta}|\mathbf{y}; \boldsymbol{\psi}^*)$ , we can express the likelihood function (D.1) in terms of the importance density as

$$L(\mathbf{y}; \boldsymbol{\psi}) = \int \frac{p(\mathbf{y}, \boldsymbol{\theta}; \boldsymbol{\psi})}{g(\boldsymbol{\theta}|\mathbf{y}; \boldsymbol{\psi}^*)} g(\boldsymbol{\theta}|\mathbf{y}; \boldsymbol{\psi}^*) d\boldsymbol{\theta}, \quad (\text{D.2})$$

with the importance sampling estimate given by

$$\frac{1}{S} \sum_{k=1}^S \omega(\mathbf{y}, \boldsymbol{\theta}^{(k)}; \boldsymbol{\psi}^*), \quad \omega(\mathbf{y}, \boldsymbol{\theta}; \boldsymbol{\psi}^*) = \frac{p(\mathbf{y}, \boldsymbol{\theta}; \boldsymbol{\psi})}{g(\boldsymbol{\theta}|\mathbf{y}; \boldsymbol{\psi}^*)}, \quad \boldsymbol{\theta}^{(k)} \sim g(\boldsymbol{\theta}|\mathbf{y}; \boldsymbol{\psi}^*), \quad (\text{D.3})$$

where the number of simulations  $S$  should be sufficiently high and where  $\boldsymbol{\theta}^{(k)}$  is drawn independently for  $k = 1, \dots, S$ . 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

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

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

$$\operatorname{argmin}_{\tilde{\boldsymbol{\psi}}_t} \int \lambda^2(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^*) \omega_t(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^*) g(\boldsymbol{\theta}_t | \mathbf{y}; \boldsymbol{\psi}^*) d\boldsymbol{\theta}_t, \quad (\text{D.5})$$

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

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

The importance density is effectively determined by the minimization of the variance of the log weight  $\omega_t$ , for each  $t$ . Richard and Zhang (2007) evaluate the integral in (D.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 (D.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  $g(\boldsymbol{\theta}_t | \mathbf{y}; \boldsymbol{\psi}^*)$  given the Gaussian importance densities  $g(\mathbf{y} | \boldsymbol{\theta}; \tilde{\boldsymbol{\psi}})$  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  $\boldsymbol{\theta}_t$ ,

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

with scalar  $a_t$ ,  $2 \times 1$  vector  $\mathbf{b}_t$ , a symmetric  $2 \times 2$  matrix  $\mathbf{C}_t$ , 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  $a_t = -\log 2\pi + \frac{1}{2} \log |\mathbf{C}_t| - \frac{1}{2} \mathbf{b}'_t \mathbf{C}_t^{-1} \mathbf{b}_t$ . We gather the five remaining parameters in  $\mathbf{b}_t$  and  $\mathbf{C}_t$  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  $k = 1, 2, \dots$ , 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

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

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  $g(\boldsymbol{\theta}_t | \mathbf{y}; \boldsymbol{\psi}^{*(k)})$ . In our case of Gaussian importance sampling distributions, we have

$$g(\boldsymbol{\theta}_t | \mathbf{y}; \boldsymbol{\psi}^{*(k)}) = \text{N}(\hat{\boldsymbol{\theta}}_t^{(k)}, \mathbf{V}_t^{(k)}) = \frac{1}{2\pi |\mathbf{V}_t^{(k)}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_t^{(k)})' (\mathbf{V}_t^{(k)})^{-1} (\boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_t^{(k)})\right), \quad (\text{D.9})$$

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

We numerically implement the minimization in (D.8) by the Gauss-Hermite quadrature method; see, for example, Monahan (2001). For this purpose we define

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

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 (D.8) becomes

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

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

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

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

where  $\boldsymbol{\kappa}$  and  $\boldsymbol{\xi}$  are regression coefficient vectors and the regression weights are given by

$w_{ij} \cdot \omega_t(y_t, \tilde{\mathbf{z}}_{ij,t}^{(k)}; \boldsymbol{\psi}^{*(k)}) \cdot g(\tilde{\mathbf{z}}_{ij,t}^{(k)} | \mathbf{y}; \boldsymbol{\psi}^{*(k)})$ , 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  $\mathbf{b}_t^{(k+1)}$  and  $\text{vech}(\mathbf{C}_t^{(k+1)})$ , 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  $g(\boldsymbol{\theta}_t | \mathbf{y}; \boldsymbol{\psi}^{*(k+1)})$  in (D.9) by constructing a new time series  $\mathbf{x}_t$  and applying the Kalman filter and smoother to the linear Gaussian model given below (D.9). In this last step we obtain new values for  $\hat{\boldsymbol{\theta}}_t^{(k+1)}$  and  $\mathbf{V}_t^{(k+1)}$ , which we require in (D.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 (D.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 (D.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  $t$ ,  $\Psi_t$ , consists of the intercept  $a_t$ , the individual components of the  $2 \times 1$  vector  $\mathbf{b}_t$  and the  $2 \times 2$  matrix  $\mathbf{C}_t$  at time  $t$ , i.e.  $\Psi_t = (a_t, \boldsymbol{\kappa}', \boldsymbol{\xi}')'$ . The optimum values  $\hat{\Psi}_t$  are obtained by applying the following iterative algorithm

- (i) Find appropriate starting values for  $\boldsymbol{\kappa}$  and  $\boldsymbol{\xi}$  with  $t = 1, \dots, n$  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  $\mathbf{C}_t(\boldsymbol{\xi})$  set to  $\mathbf{I}_2$  suffices.
- (ii) Construct the linear Gaussian state space model with observation equation  $\mathbf{x}_t = \boldsymbol{\theta}_t + \mathbf{u}_t$  with disturbance  $\mathbf{u}_t \sim N(0, \mathbf{C}_t^{-1})$  and pseudo-observation  $\mathbf{x}_t = \mathbf{C}_t^{-1} \mathbf{b}_t$ , for  $t = 1, \dots, n$  and apply the Kalman filter and smoother to obtain  $\hat{\boldsymbol{\theta}}_t^{(k)}$  and  $\mathbf{V}_t^{(k)}$  and use these to calculate  $\tilde{\mathbf{z}}_{ij,t}^{(k)}$  as described below equation (D.11).
- (iii) Minimize equation (D.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  $\epsilon$ , 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),  $\Psi_t^{(s+1)}, t = 1, \dots, n$  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  $\hat{\Psi}_t$  can be stored and used as starting values for the next call to the algorithm. The minimization of (D.11) can be carried out independently for all time points  $t$  and can therefore be done in parallel over  $t$ .

# E Intradaily time series of price changes in 2012

Table E.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 presented in (3).

Company	$\leq -4$	-3	-2	-1	0	1	2	3	$\geq 4$
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

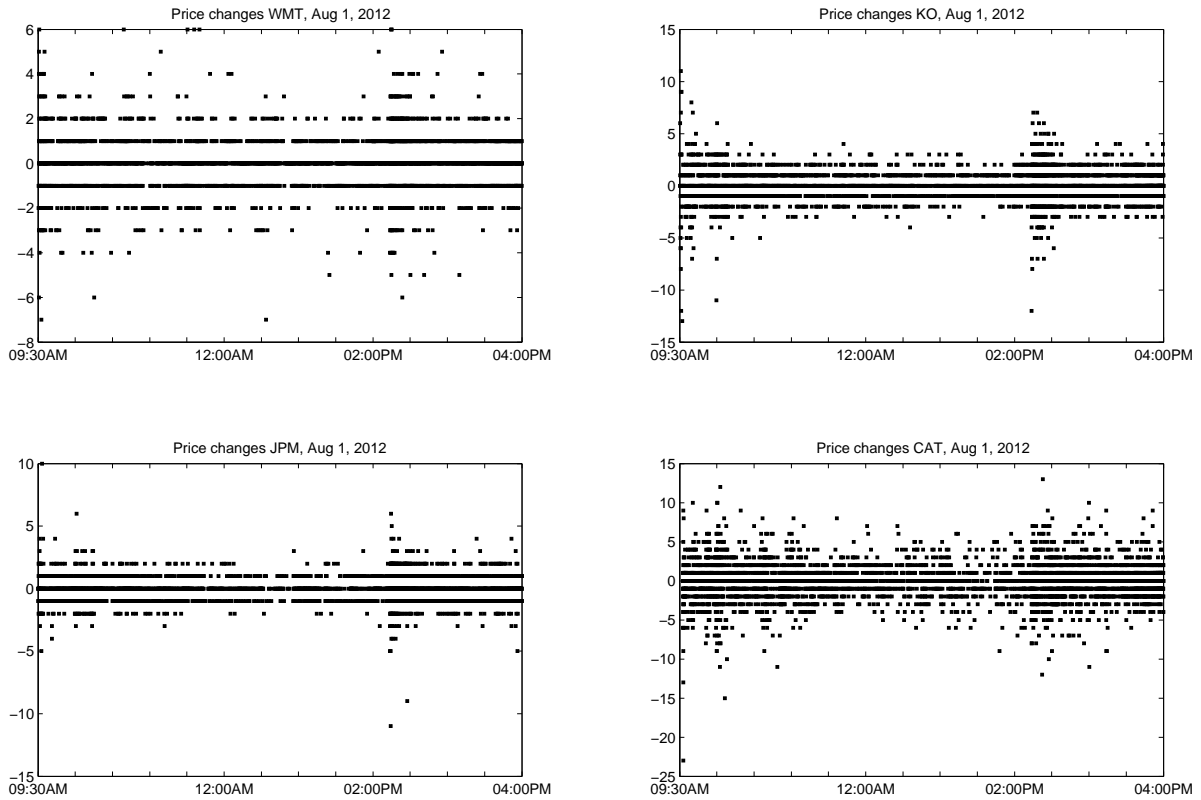


Figure E.1: Price changes on August 1, 2012

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

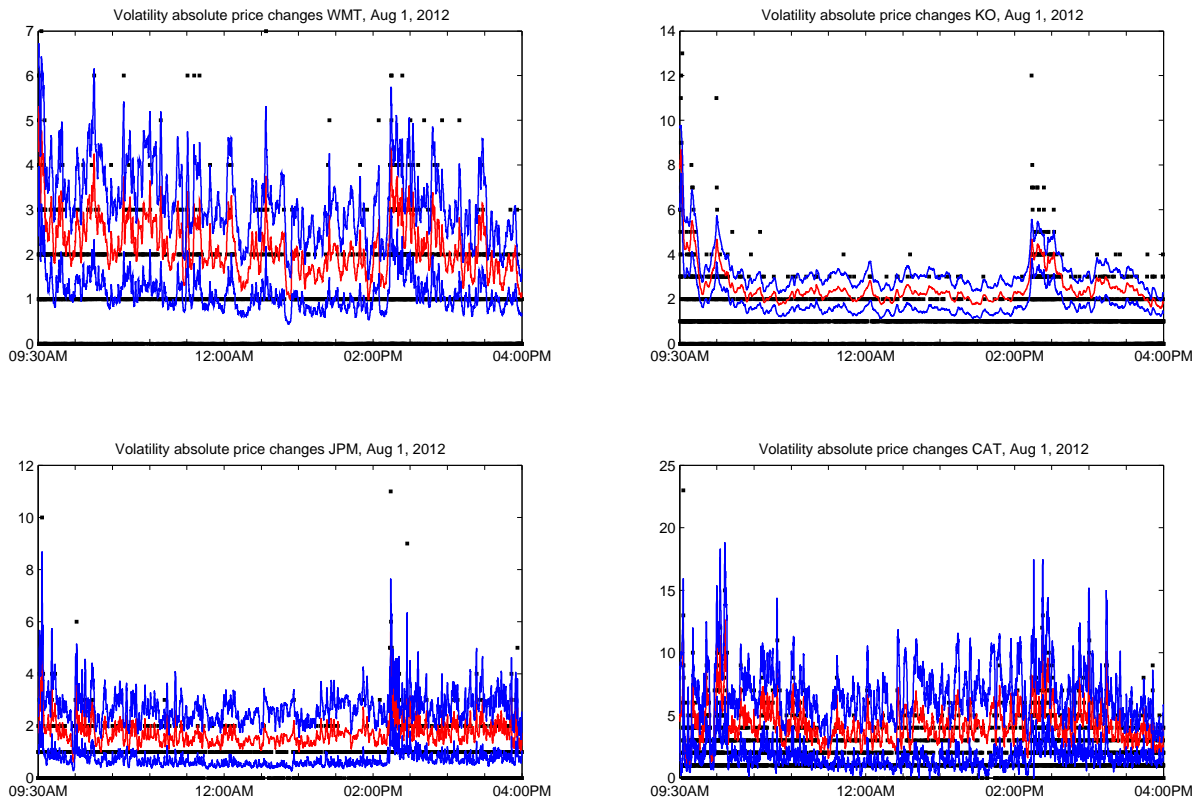


Figure E.2: Absolute price changes on August 1, 2012 with volatility path estimate

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.