

Barra, Istvan; Koopman, Siem Jan

Working Paper

Bayesian Dynamic Modeling of High-Frequency Integer Price Changes

Tinbergen Institute Discussion Paper, No. 16-028/III

Provided in Cooperation with:

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Barra, Istvan; Koopman, Siem Jan (2016) : Bayesian Dynamic Modeling of High-Frequency Integer Price Changes, Tinbergen Institute Discussion Paper, No. 16-028/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/145335>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TI 2016-028/III
Tinbergen Institute Discussion Paper



Bayesian Dynamic Modeling of High-Frequency Integer Price Changes

Istvan Barra

Siem Jan Koopman

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.

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

Bayesian Dynamic Modeling of High-Frequency Integer Price Changes*

István Barra^(a) and *Siem Jan Koopman* ^(a,b)

^(a) Vrije Universiteit Amsterdam and Tinbergen Institute

^(b) CREATES, Aarhus University

Abstract

We investigate high-frequency volatility models for analyzing intra-day tick by tick stock price changes using Bayesian estimation procedures. Our key interest is the extraction of intra-day volatility patterns from high-frequency integer price changes. We account for the discrete nature of the data via two different approaches: ordered probit models and discrete distributions. We allow for stochastic volatility by modeling the variance as a stochastic function of time, with intra-day periodic patterns. We consider distributions with heavy tails to address occurrences of jumps in tick by tick discrete prices changes. In particular, we introduce a dynamic version of the negative binomial difference model with stochastic volatility. For each model we develop a Markov chain Monte Carlo estimation method that takes advantage of auxiliary mixture representations to facilitate the numerical implementation. This new modeling framework is illustrated by means of tick by tick data for several stocks from the NYSE and for different periods. Different models are compared with each other based on predictive likelihoods. We find evidence in favor of our preferred dynamic negative binomial difference model.

Keywords: Bayesian inference; discrete distributions; high-frequency dynamics; Markov chain Monte Carlo; stochastic volatility.

*Corresponding author: S.J. Koopman, Department of Econometrics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Emails : barra.istvan@gmail.com s.j.koopman@vu.nl IB thanks Dutch National Science Foundation (NWO) for financial support. SJK acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. We both thank the European Union 7th Framework Programme (FP7-SSH/2007-2013, grant agreement 320270 - SYRTO) for financial support. We are indebted to Lennart Hoogerheide, Rutger Lit, André Lucas and Mike Pitt for their help and support in this research project and to Rudolf Frühwirth for providing the C code for auxiliary mixture sampling. This version: April 19, 2016

1 Introduction

High-frequency price changes observed at stock, futures and commodity markets can typically not be regarded as continuous variables. In most electronic markets, the smallest possible price difference is set by the regulator or the trading platform. Here we develop and investigate dynamic models for high-frequency integer price changes that take the discreteness of prices into account. We explore the dynamic properties of integer time series observations. In particular, we are interested in the stochastic volatility dynamics of price changes within intra-daily time intervals. This information can be used for the timely identification of changes in volatility and to obtain more accurate estimates of integrated volatility.

In the current literature on high-frequency returns, price discreteness is typically neglected. However, the discreteness can have an impact on the distribution of price changes and on its volatility; see, for example, Security and Exchange Commission Report (2012), Chakravarty et al. (2004) and Ronen and Weaver (2001). Those assets that have prices with a spread of almost always equal to one tick are defined as large tick assets; see, Eisler et al. (2012). These large tick assets are especially affected by the discreteness through the effect of different quoting strategies on these assets; see the discussions in Chordia and Subrahmanyam (1995) and Cordella and Foucault (1999). Also the effect of liquidity on large tick assets can be substantial as it is documented by O'Hara et al. (2014) and Ye and Yao (2014). Many large tick assets exist on most US exchange markets as the tick size is set to only one penny for stocks with a price greater than 1\$ by the Security and Exchange Commission in Rule 612 of the Regulation National Market System. Hence almost all low price stocks are large tick assets. Moreover, many future contracts are not decimalized for example, five-years U.S Treasury Note futures and EUR/USD futures fall into this category (see Dayri and Rosenbaum (2013)).

The relevance of discreteness and its effect on the analysis of price changes have been the motivation to develop models that account for integer prices. Similar to the case of continuous returns, we are primarily interested in the extraction of volatility from discrete price changes. We consider different dynamic model specifications for the high-frequency integer price changes with a focus on the modeling and extraction of stochastic volatility. We have encountered the studies of Müller and Czado (2006) and Stefanos (2015) who propose ordered probit models with time-varying variance specifications. We adopt their modeling approaches as a reference and also use their treatments of Bayesian estimation. The main novelty of our study is the specification of a new model for tick by tick price changes based on the discrete negative binomial distribution which we shall refer to shortly as the Δ NB distribution. The properties of this distribution are explored in detail in our study. In particular, the heavy tail properties are emphasized. In our analysis, we adopt the Δ NB distribution conditional on a Gaussian latent state vector

process which represent the components of the stochastic volatility process. The volatility process accounts for the periodic pattern in high-frequency volatility due to intra-day seasonal effects such as the opening, lunch and closing hours. Our Bayesian modeling approach provides a flexible and unified framework to fit the observed tick by tick price changes. The ΔNB properties closely mimic the empirical stylized properties of trade by trade price changes. Hence we will argue that the ΔNB model with stochastic volatility is an attractive alternative to models based on the Skellam distribution as suggested earlier; see Koopman et al. (2014). We further decompose the unobserved log volatility into intra-daily periodic and transient volatility components. We propose a Bayesian estimation procedure using standard Gibbs sampling methods. Our procedure is based on data augmentation and auxiliary mixtures; it extends the auxiliary mixture sampling procedure proposed by Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter et al. (2009). The procedures are implemented in a computationally efficient manner.

In our empirical study we consider six stocks from the NYSE in a volatile week in October 2008 and a calmer week in April 2010. We compare the in-sample and out-of-sample fits of four different model specifications: ordered probit model based on the normal and Student's t distributions, the Skellam distribution and the ΔNB model. We compare the models in terms of Bayesian information criterion and predictive likelihoods. We find that the ΔNB model is favored for stocks with a relatively low tick size and in periods of more volatility.

Our study is related to different strands in the econometrics literature. Modeling discrete price changes with static Skellam and ΔNB distributions has been introduced by Alzaid and Omair (2010) and Barndorff-Nielsen et al. (2012). The dynamic specification of the Skellam distribution and its (non-Bayesian) statistical treatment have been explored by Koopman et al. (2014). Furthermore, our study is related to Bayesian treatments of stochastic volatility models for continuous returns; see, for example, Chib et al. (2002), Kim et al. (1998), Omori et al. (2007) and, more recently, Stroud and Johannes (2014). We extend this literature on trade by trade price changes by explicitly accounting for prices discreteness and heavy tails of the tick by tick return distribution. These extensions are explored in other contexts in Engle (2000), Czado and Haug (2010), Dahlhaus and Neddermeyer (2014) and Rydberg and Shephard (2003).

The remainder is organized as follows. In Section 2 we review different dynamic model specifications for high-frequency integer price changes. We give most attention to the introduction of the dynamic ΔNB distribution. Section 3 develops a Bayesian estimation procedure based on Gibbs sampling, mainly for the ΔNB case of which the Skellam is a special case. In Section 4 we present the details of our empirical study including a description of our dataset, the data cleaning procedure, the presentation of our estimation results and a discussion of our overall empirical findings. Section 5 concludes.

2 Dynamic models for discrete price changes

In this section we first review the modeling of integer valued variables using ordered probit models based on normal and Student's t distributions with stochastic volatility. Next we introduce the dynamic negative binomial difference (Δ NB) model with stochastic volatility and discuss its features. The dynamic Skellam model is a special case of Δ NB.

2.1 Ordered normal stochastic volatility model

In econometrics, the ordered probit model is typically used for the modeling of ordinal variables. But we can also adopt the ordered probit model in a natural way for the modeling of discrete price changes. In this approach we effectively round a realization from a continuous distribution to its nearest integer. The continuous distribution can be subject to stochastic volatility; this extension is relatively straightforward. Let r_t^* be the continuous return which is rounded to $r_t = k$ when $r_t^* \in [k - 0.5, k + 0.5)$. We observe r_t and we regard r_t^* as a latent variable. By neglecting the discreteness of r_t during the estimation procedure, we clearly would distort the measurement of the scaling or variation of r_t^* . Therefore we need to take account of the rounding of r_t by specifying an ordered probit model with rounding thresholds $[k - 0.5, k + 0.5)$. We assume that the underlying distribution for r_t^* is subject to stochastic volatility. We obtain the following specification

$$r_t = k, \quad \text{with probability } \Phi\left(\frac{k + 0.5}{\exp(h_t/2)}\right) - \Phi\left(\frac{k - 0.5}{\exp(h_t/2)}\right), \quad \text{for } k \in \mathbb{Z}, \quad (1)$$

for $t = 1, \dots, T$, where h_t is the logarithm of the time varying stochastic variance for the standard normal distribution with cumulative density function $\Phi(\cdot)$ for the latent variable r_t^* . Similar ordered probit specifications with stochastic volatility are introduced by Müller and Czado (2006) and Stefanos (2015). The dynamic model specification for h_t is given by

$$h_t = \mu_h + x_t, \quad x_{t+1} = \varphi x_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \quad (2)$$

for $t = 1, \dots, T$, where μ_h is the unconditional mean of the log volatility of the continuous returns, x_t is a zero mean autoregressive process (AR) of order one, that is AR(1), with φ as the persistence parameter for the log volatility process and σ_η^2 as the variance of the Gaussian disturbance term η_t . The mean μ_h represents the daily log volatility and the autoregressive process x_t captures the changes in log volatility due to firm specific or market information experienced during the day. The latent variable x_t is specified as an AR(1) process with zero mean; this restriction is enforced to allow for the identification of μ_h .

The basic model specification (1) - (2) accounts for the discreteness of the prices via

the ordered probit specification and for intra-day volatility clustering via the possibly persistent dynamic process of x_t . The model captures the salient empirical features of high-frequency trade by trade price changes. Another stylized fact of intra-day price changes is the seasonality pattern in the volatility process. In particular, the volatility at the opening minutes of the trading day is high, during the lunch-hour it is lowest, and at the closing minutes it is increasing somewhat. We can account for such an intra-day volatility pattern by including a cubic spline in the log volatility specification, that is

$$h_t = \mu_h + s_t + x_t, \quad E(s_t) = 0, \quad (3)$$

where s_t is a normalized spline function with its unconditional mean equal to zero. This specification implies a decomposition of the variance of the continuous return distribution r_t^* into a deterministic daily seasonal pattern s_t and a stochastically time varying signal x_t . We use intradaily cubic spline function, constructed from $K + 1$ piecewise cubic polynomials, to capture the daily seasonality. We adopt the representation of Poirier (1973) where the periodic cubic spline s_t is based on K knots and the regression equation

$$s_t = w_t \beta \quad (4)$$

where w_t is a $1 \times K$ weight vector and β is a $K \times 1$ vector which contains values of the spline function at the K knots. Further details about the spline and the Poirier representation are presented in Appendix B. For alternative treatments of intra-daily seasonality, we refer to Bos (2008), Stroud and Johannes (2014) and Weinberg et al. (2007).

The model can be modified and extended in several ways. First, we can account for the market microstructure noise observed in tick by tick returns (see for example, Aït-Sahalia et al. (2011) and Griffin and Oomen (2008)) by including an autoregressive moving average (ARMA) process in the specification of the mean of r_t^* . In a similar way, we can facilitate the incorporation of explanatory variables such as market imbalance which can also have predictive power. Second, to include predetermined announcement effects, we can include regression effects in the specification as proposed in Stroud and Johannes (2014). Third, it is possible that the unconditional mean μ_h of the volatility of price changes is time varying. For example, we may expect that for larger price stocks the volatility is higher and therefore the volatility is not properly scaled when the price has changed. The time-varying conditional mean of the volatility can be easily incorporated in the model, by specifying a random walk dynamics for μ_h , which would allow for smooth changes in the mean over time. For our current purposes below we can rely on the specification as given by equation (3).

2.2 Ordered t stochastic volatility model

It is well documented in the financial econometrics literature that asset prices are subject to jumps; see, for example, Aït-Sahalia et al. (2012). However, the ordered normal specification, as we have introduced it above, does not deliver sufficiently heavy tails in its asset price distribution to accommodate the jumps that are typically observed in high-frequency returns. To account for the jumps more appropriately, we can consider a heavy tailed distribution instead of the normal distribution. In this way we can assign probability mass to the infrequently large jumps in asset returns. An obvious choice for a heavy tailed distribution is the Student's t -distribution which would imply the following specification,

$$r_t = k, \quad \text{with probability } \mathcal{T}\left(\frac{k+0.5}{\exp(h_t/2)}, \nu\right) - \mathcal{T}\left(\frac{k-0.5}{\exp(h_t/2)}, \nu\right), \quad \text{for } k \in \mathbb{Z}, \quad (5)$$

which effectively replaces model equation (1), where $\mathcal{T}(\cdot, \nu)$ is the cumulative density function of the Student's t -distribution with ν as the degrees of freedom parameter. The model specification for h_t is provided by equation (2) or (3).

The parameter vector of this model specification is denoted by ψ and includes the degrees of freedom ν , the unconditional mean of log volatility μ_h , the volatility persistence coefficient φ , the variance of the log volatility disturbance σ_η^2 , and the unknown vector β in (4) with values of the spline at its knot positions. In case of the normal ordered probit specification, we can rely on the same parameters but without ν . The estimation procedure for these unknown parameters in the ordered probit model specifications are carried out by standard Bayesian simulation methods for which the details are provided in Appendix C.

2.3 Dynamic Δ NB model

Positive integer variables can alternatively be modeled directly via discrete distributions such as the Poisson or the negative binomial, see Johnson et al. (2005). These well-known distributions only provide support to positive integers. When modeling price differences, we also need to allow for negative integers. For example, in this case, the Skellam distribution can be considered, see Skellam (1946). The specification of these distributions can be extended to stochastic volatility model straightforwardly. However, the analysis and estimation based on such models are more intricate. In this context, Alzaid and Omair (2010) advocates the use of the Skellam distribution based on the difference of two Poisson random variables. Barndorff-Nielsen et al. (2012) introduces the negative binomial difference (Δ NB) distribution which have fatter tails compared to the Skellam distribution. Next we review the Δ NB distribution and its properties. We

further introduce a dynamic version of the ΔNB model from which the dynamic Skellam model is a special case.

The ΔNB distribution is implied by the construction of the difference of two negative binomial random variables which we denote by NB^+ and NB^- where the variables have number of failures λ^+ and λ^- , respectively, and failure rates ν^+ and ν^- , respectively. We denote the ΔNB variable as the random variable R and is simply defined as

$$R = NB^+ - NB^-. \quad (6)$$

We then assume that R is distributed as

$$R \sim \Delta\text{NB}(\lambda^+, \nu^+, \lambda^-, \nu^-), \quad (7)$$

where ΔNB is the difference negative binomial distribution with probability mass function given by

$$f_{\Delta\text{NB}}(r; \lambda^+, \nu^+, \lambda^-, \nu^-) = m \times \begin{cases} d^+ \times F(\nu^+ + r, \nu^-, r + 1; \tilde{\lambda}^+ \tilde{\lambda}^-), & \text{if } r \geq 0, \\ d^- \times F(\nu^+, \nu^- - r, -r + 1; \tilde{\lambda}^+ \tilde{\lambda}^-), & \text{if } r < 0, \end{cases}$$

where $m = (\tilde{\nu}^+)^{\nu^+} (\tilde{\nu}^-)^{\nu^-}$, $d^{[s]} = (\tilde{\lambda}^{[s]})^r (\nu^{[s]})_r / r!$,

$$\tilde{\nu}^{[s]} = \frac{\nu^{[s]}}{\lambda^{[s]} + \nu^{[s]}}, \quad \tilde{\lambda}^{[s]} = \frac{\lambda^{[s]}}{\lambda^{[s]} + \nu^{[s]}},$$

for $[s] = +, -$, and with the hypergeometric function

$$F(a, b, c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!},$$

where $(x)_n$ is the Pochhammer symbol of falling factorial and is defined as

$$(x)_n = x(x-1)(x-2)\cdots(x-n+1) = \frac{\Gamma(x+1)}{\Gamma(x-n+1)}. \quad (8)$$

More details about the ΔNB distribution, its probability mass function and properties are provided by Barndorff-Nielsen et al. (2012). For example, the ΔNB distribution has the following first and second moments

$$\mathbb{E}(R) = \lambda^+ - \lambda^-, \quad \text{Var}(R) = \lambda^+ \left(1 + \frac{\lambda^+}{\nu^+}\right) + \lambda^- \left(1 + \frac{\lambda^-}{\nu^-}\right). \quad (9)$$

The variables ν^+ , ν^- , λ^+ and λ^- are treated typically as unknown coefficients.

An important special case is the zero mean ΔNB distribution which is obtained when

$\lambda = \lambda^+ = \lambda^-$ and $\nu = \nu^+ = \nu^-$. The probability mass function for the corresponding random variable R is given by

$$f_0(r; \lambda, \nu) = \left(\frac{\nu}{\lambda + \nu}\right)^{2\nu} \left(\frac{\lambda}{\lambda + \nu}\right)^{|r|} \frac{\Gamma(\nu + |r|)}{\Gamma(\nu)\Gamma(|r| + 1)} F\left(\nu + |r|, \nu, |r| + 1; \left(\frac{\lambda}{\lambda + \nu}\right)^2\right).$$

In this case we have obtained a zero mean random variable R with its variance given by

$$\text{Var}(R) = 2\lambda \left(1 + \frac{\lambda}{\nu}\right). \quad (10)$$

We denote the distribution for the zero mean random variable R by $\Delta\text{NB}(\lambda, \nu)$. This random variable R can alternatively be considered as being generated from a compound Poisson process, that is

$$R = \sum_{i=1}^N M_i, \quad (11)$$

where random variable N is generated by the Poisson distribution with intensity

$$\lambda \times (z_1 + z_2), \quad z_1, z_2 \sim \text{Ga}(\nu, \nu) \quad (12)$$

with $\text{Ga}(\nu, \nu)$ being the Gamma distribution, having its shape and scale both equal to ν , and where indicator variable M_i is generated as

$$M_i = \begin{cases} 1, & \text{with probability } P(M_i = 1) = z_1 / (z_1 + z_2), \\ -1, & \text{with probability } P(M_i = -1) = z_2 / (z_1 + z_2). \end{cases} \quad (13)$$

We will use this representation of a zero mean ΔNB variable for developments below.

In the empirical analyses of this study, we adopt the zero inflated versions of the ΔNB distributions, because empirically we observe a clear overrepresentation of trade by trade price changes that are equal to zero. The number of these zero price changes are especially high for the more liquid stocks. This is due to the available volumes on best bid and ask prices which are relatively much higher. Hence the price impact of one trade is much lower as a result. The zero inflated version is accomplished by the specification of the random variable R_0 as

$$r_0 = \begin{cases} r, & \text{with probability } (1 - \gamma)f_{\Delta\text{NB}}(r; \lambda^+, \nu^+, \lambda^-, \nu^-), \\ 0, & \text{with probability } \gamma + (1 - \gamma)f_{\Delta\text{NB}}(0; \lambda^+, \nu^+, \lambda^-, \nu^-), \end{cases}$$

where $f_{\Delta\text{NB}}(r; \lambda^+, \nu^+, \lambda^-, \nu^-)$ is the probability mass function for r and $0 < \gamma < 1$ is treated as a fixed and unknown coefficient. We denote the zero inflated ΔNB probability mass function with f_0 .

The dynamic specifications of the Δ NB distributions can be obtained by letting the variables $\nu^{[s]}$ and/or $\lambda^{[s]}$ be time-varying random variables, for $[s] = +, -$. We opt to have a time-varying $\lambda^{[s]}$ since it is more natural for an intensity than a degrees of freedom parameter to vary over time. We restrict our analysis to the zero inflated zero mean Δ NB distribution $f_0(r_t; \lambda_t, \nu)$ which means we use zero inflation and we assume that the degree of freedom parameters for positive and negative price changes are the same and $\lambda_t = \lambda_t^+ = \lambda_t^-$. Taking the above considerations into account, the dynamic Δ NB model can be specified as above but with

$$\lambda_t = \exp(h_t),$$

where h_t is specified as in equation (2) or (3). Hence this dynamic specification is similar to the one described in Section 2.1 for the ordered normal specification.

2.4 Dynamic Skellam model

The dynamic Δ NB model embeds the dynamic Skellam model as considered by Koopman et al. (2014). It is obtained as the limiting case of letting ν go to infinity, that is $\nu \rightarrow \infty$; for a derivation and further details, see Appendix A.

3 Bayesian estimation procedures

Bayesian estimation procedures for the ordered normal and ordered Student's t stochastic volatility models are discussed by Müller and Czado (2006) and Stefanos (2015); their procedures, with some details, are presented in Appendix C.

Here we develop a Bayesian estimation procedure for observations y_t , with $t = 1, \dots, T$, coming from the dynamic Δ NB model. We provide the details of the procedure and discuss its computational implementation. Our reference dynamic Δ NB model is given by

$$\begin{aligned} y_t &\sim f_0(y_t; \lambda_t, \nu), & \lambda_t &= \exp h_t, \\ h_t &= \mu_h + s_t + x_t, & s_t &= w_t \beta, & x_{t+1} &= \varphi x_t + \eta_t, \end{aligned} \tag{14}$$

where $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$, for $t = 1, \dots, T$. The details of the model are discussed in Section 2. The variable parameters ν , μ_h , β , φ and σ_η^2 are static while x_t is a latent variable that is modeled as a stationary autoregressive process. The intra-day seasonal effect s_t is represented by a Poirier spline; see Appendix B.

Our proposed Bayesian estimation procedure aims to estimate all static variables jointly with the time-varying signal h_1, \dots, h_T for the dynamic Δ NB model. It is based on Gibbs sampling, data augmentation and auxiliary mixture sampling methods which

are developed by Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter et al. (2009). At each time point t , for $t = 1, \dots, T$, we introduce a set of latent auxiliary variables to facilitate the derivation of conditional distributions. By introducing these auxiliary variables we are able to specify the model as a linear state space model with non-Gaussian observation disturbances. Moreover using an auxiliary mixture sampling procedure, we can even obtain conditionally an approximating linear Gaussian state space model. In such a setting, we can exploit highly efficient Kalman filtering and smoothing procedures for the sampling of many full paths for the dynamic latent variables. These ingredients are key for a computational feasible implementation of our estimation process.

3.1 Data augmentation: our latent auxiliary variables

We use the following auxiliary variables for the data augmentation. We define N_t as the sum of NB^+ and NB^- , the gamma mixing variables z_{t1} and z_{t2} . Moreover conditional on z_{t1} , z_{t2} and the intensity λ_t , we can interpret N_t as a Poisson process on $[0, 1]$ with intensity $(z_{t1} + z_{t2})\lambda_t$ based on the result in equation (12). We can introduce the latent arrival time of the N_t -th jump of the Poisson process τ_{t2} and the arrival time between the N_t -th and $N_t + 1$ -th jump of the process τ_{t1} for every $t = 1, \dots, T$. The interarrival time τ_{t1} can be assumed to come from an exponential distribution with intensity $(z_{t1} + z_{t2})\lambda_t$ while the N_t th arrival time can be treated as the gamma distributed variable with density function $\text{Ga}(N_t, (z_{t1} + z_{t2})\lambda_t)$. We have

$$\tau_{t1} = \frac{\xi_{t1}}{(z_{t1} + z_{t2})\lambda_t}, \quad \xi_{t1} \sim \text{Exp}(1) \quad (15)$$

$$\tau_{t2} = \frac{\xi_{t2}}{(z_{t1} + z_{t2})\lambda_t}, \quad \xi_{t2} \sim \text{Ga}(N_t, 1), \quad (16)$$

where we can treat ξ_{t1} and ξ_{t2} as auxiliary variables. By taking the logarithm of the equations and substituting the definition of $\log \lambda_t$ from equation (3), we can rewrite the above equations as

$$-\log \tau_{t1} = \log(z_{t1} + z_{t2}) + \mu_h + s_t + x_t + \xi_{t1}^*, \quad \xi_{t1}^* = -\log \xi_{t1} \quad (17)$$

$$-\log \tau_{t2} = \log(z_{t1} + z_{t2}) + \mu_h + s_t + x_t + \xi_{t2}^*, \quad \xi_{t2}^* = -\log \xi_{t2}. \quad (18)$$

These equations are linear in the state vector, which would facilitate the use of Kalman filtering. However, the error terms ξ_{t1}^* and ξ_{t2}^* are non-normal. We can adopt solutions as in Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter et al. (2009) where the gamma and exponential distributions are approximated by normal mixture distributions.

In particular, we can specify the approximations as

$$f_{\xi^*}(x; N_t) \approx \sum_{i=1}^{C(N_t)} \omega_i(N_t) \phi(x, m_i(N_t), v_i(N_t)), \quad (19)$$

where $C(N_t)$ is the number of mixture components at time t , for $t = 1, \dots, T$, $\omega_i(N_t)$ is the weight, and $\phi(x, m, v)$ is the normal density for variable x with mean m and variance v . These approximations remain to depend on N_t because the log gamma distribution is not canonical and it has different shapes for different values of N_t .

3.2 Mixture indicators for obtaining conditional linear model

Conditional on $N, z_1, z_2, \tau_1, \tau_2$ and $C = \{c_{tj}, t = 1, \dots, T, j = 1, \dots, \min(N_t + 1, 2)\}$ we can write the following state space form

$$\underbrace{\tilde{y}_t}_{\min(N_t+1,2) \times 1} = \underbrace{\begin{bmatrix} 1 & w_t & 1 \\ 1 & w_t & 1 \end{bmatrix}}_{\min(N_t+1,2) \times (K+2)} \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_t \end{bmatrix}}_{(K+2) \times 1} + \underbrace{\varepsilon_t}_{\min(N_t+1,2) \times 1}, \quad \varepsilon_t \sim \mathcal{N}(0, H_t) \quad (20)$$

$$\alpha_{t+1} = \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_{t+1} \end{bmatrix}}_{(K+2) \times 1} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & I_K & 0 \\ 0 & 0 & \varphi \end{bmatrix}}_{(K+2) \times (K+2)} \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_t \end{bmatrix}}_{(K+2) \times 1} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \eta_{t+1} \end{bmatrix}}_{(K+2) \times 1}, \quad \eta_{t+1} \sim \mathcal{N}(0, \sigma_\eta^2) \quad (21)$$

(22)

where

$$\underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_1 \end{bmatrix}}_{(K+2) \times 1} \sim \mathcal{N} \left(\underbrace{\begin{bmatrix} \mu_0 \\ \beta_0 \\ 0 \end{bmatrix}}_{(K+2) \times 1}, \underbrace{\begin{bmatrix} \sigma_\mu^2 & 0 & 0 \\ 0 & \sigma_\beta^2 I_K & 0 \\ 0 & 0 & \sigma_\eta^2 / (1 - \varphi^2) \end{bmatrix}}_{(K+2) \times (K+2)} \right) \quad (23)$$

$H_t = \text{diag}(v_{c_{t1}}^2(1), v_{c_{t2}}^2(N_t))$ and

$$\underbrace{\tilde{y}_t}_{\min(N_t+1,2) \times 1} = \begin{pmatrix} -\log \tau_{t1} - m_{c_{t1}}(1) - \log(z_{t1} + z_{t2}) \\ -\log \tau_{t2} - m_{c_{t2}}(N_t) - \log(z_{t1} + z_{t2}) \end{pmatrix} \quad (24)$$

Using the mixture of normal approximation of ξ_{t1}^* and ξ_{t2}^* , allows us to build an efficient Gibbs sampling procedure where we can sample the latent state paths in one block, efficiently using Kalman filtering and smoothing techniques.

3.3 The sampling of event times N_t

The remaining challenge is the sampling of N_t as all the other full conditionals are standard. We notice that conditional on z_{t1} , z_{t2} and the intensity λ_t , the N_t 's are independent over time. We have

$$p(N|\gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, z_1, z_2, y) = \prod_{t=1}^T p(N_t|\gamma, \lambda_t, z_{t1}, z_{t2}, y_t), \quad (25)$$

where the t element vectors (v_1, \dots, v_t) containing time dependent variables for all time time periods, are denoted by v , the variable without a subscript. For a given time index t , we can draw N_t from a discrete distribution with

$$\begin{aligned} p(N_t|\gamma, \lambda_t, z_{t1}, z_{t2}, y_t) &= \frac{p(N_t, y_t|\gamma, \lambda_t, z_{t1}, z_{t2})}{p(y_t|\gamma, \lambda_t, z_{t1}, z_{t2})} \\ &= \frac{p(y_t|N_t, \gamma, \lambda_t, z_{t1}, z_{t2})p(N_t|\gamma, \lambda_t, z_{t1}, z_{t2})}{p(y_t|\gamma, \lambda_t, z_{t1}, z_{t2})} \\ &= [\gamma \mathbf{1}_{\{y_t=0\}} + (1-\gamma)p(y_t|N_t, \lambda_t, z_{t1}, z_{t2})] \\ &\times \frac{p(N_t|\gamma, \lambda_t, z_{t1}, z_{t2})}{p(y_t|\gamma, \lambda_t, z_{t1}, z_{t2})} \end{aligned} \quad (26)$$

The denominator in equation (26) is a Skellam distribution with intensity $\lambda_t z_{t1}$ and $\lambda_t z_{t2}$. We can calculate probability

$$p(y_t|N_t, \lambda_t, z_{t1}, z_{t2}) \quad (27)$$

using the results from equation (12) condition on λ_t , z_{t1} and z_{t2} , y_t is distributed as a marked Poisson process with marks given by

$$M_i = \begin{cases} 1, & \text{with } P(M_i = 1) = \frac{z_{t1}}{z_{t1} + z_{t2}} \\ -1, & \text{with } P(M_i = -1) = \frac{z_{t2}}{z_{t1} + z_{t2}} \end{cases}, \quad (28)$$

which implies that we can represent y_t as $\sum_{i=0}^{N_t} M_i$.

$$p(y_t|N_t, \lambda_t, z_{t1}, z_{t2}) = \begin{cases} 0, & \text{if } y_t > N_t \text{ or } |y_t| \bmod 2 \neq |N_t| \bmod 2 \\ \binom{N_t}{\frac{N_t+y_t}{2}} \left(\frac{z_{t1}}{z_{t1} + z_{t2}}\right)^{\frac{N_t+y_t}{2}} \left(\frac{z_{t2}}{z_{t1} + z_{t2}}\right)^{\frac{N_t-y_t}{2}}, & \text{otherwise} \end{cases} \quad (29)$$

Conditional on z_{t1} , z_{t2} and λ_t , N_t is a realization of a Poisson process on $[0, 1]$ with intensity $(z_{t1} + z_{t2})\lambda_t$, hence the probability $p(N_t|\gamma, \lambda_t, z_{t1}, z_{t2})$ is a Poisson random variable with intensity equal to $\lambda_t(z_{t1} + z_{t2})$. We can draw N_t parallel over $t = 1, \dots, T$ by drawing

a uniform random variable $u_t \sim U[0, 1]$ and

$$N_t = \min \left\{ n : u_t \leq \sum_{i=0}^n p(i|\gamma, \lambda_t, z_{t1}, z_{t2}, y_t) \right\} \quad (30)$$

3.4 Markov chain Monte Carlo algorithm

The complete MCMC algorithm is outlined below. Various details of the MCMC steps are presented in Appendix D. In an algorithmic style, the MCMC steps are given as follows.

1. Initialize $\mu_h, \varphi, \sigma_\eta^2, \gamma, \nu, C, \tau, N, z_1, z_2, s$ and x
2. Generate $\varphi, \sigma_\eta^2, \mu_h, s$ and x from $p(\varphi, \sigma_\eta^2, \mu_h, s, x|\gamma, \nu, C, \tau, N, z_1, z_2, s, y)$
 - (a) Draw φ, σ_η^2 from $p(\varphi, \sigma_\eta^2|\gamma, \nu, C, \tau, N, z_1, z_2, s, y)$
 - (b) Draw μ_h, s and x from $p(\mu_h, s, x|\varphi, \sigma_\eta^2, \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$
3. Generate γ from $p(\gamma|\nu, \mu_h, \varphi, \sigma_\eta^2, x, C, \tau, N, z_1, z_2, s, y)$
4. Generate $C, \tau, N, z_1, z_2, \nu$ from $p(C, \tau, N, z_1, z_2, \nu|\gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$
 - (a) Draw ν from $p(\nu|\gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$
 - (b) Draw z_1, z_2 from $p(z_1, z_2|\nu, \gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$
 - (c) Draw N from $p(N|z_1, z_2, \nu, \gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$
 - (d) Draw τ from $p(\tau|N, z_1, z_2, \nu, \gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$
 - (e) Draw C from $p(C|\tau, N, z_1, z_2, \nu, \gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$
5. Go to 2

To validate our estimation procedure for the dynamic Skellam and Δ NB models we simulate 20,000 observation and apply our MCMC procedure with 100,000 replications, in a single experiment. Our true parameters are chosen as $\mu = -1.7$, $\varphi = 0.97$, $\sigma_\eta = 0.02$, $\gamma = 0.001$ and $\nu = 15$ which are close to those estimated from real data in our empirical study of Section 4. In Table 1 we summarize the results of our estimates and their corresponding highest posterior density (HPD) regions. The results indicate that in our stylized setting, the algorithm can estimate the parameters accurately since the true parameters are within the HPD regions based on the estimates. The posterior distributions of the parameters for the Δ NB model are presented in Figure 1; those for the dynamic Skellam model are presented in Appendix D. The most atypical posterior distributions are displayed for the autoregressive coefficient φ and for the state variance σ_η^2 . Hence we may conclude that these parameters are the most challenging to estimate.

Table 1: Estimation results from a dynamic Skellam and Δ NB model based on 20,000 observations and 100,000 iterations from which 20,000 used as a burn in sample. The 95 % HPD regions are in brackets.

	True	Skellam	Δ NB		True	Skellam	Δ NB
μ	-1.7	-1.72 [-1.797,-1.642]	-1.726 [-1.804,-1.651]	β_1	1.13	1.139 [0.884,1.392]	1.128 [0.875,1.38]
φ	0.97	0.973 [0.965,0.979]	0.975 [0.969,0.981]	β_2	-0.29	-0.306 [-0.453,-0.158]	-0.297 [-0.448,-0.151]
σ^2	0.02	0.018 [0.013,0.023]	0.015 [0.011,0.02]	β_3	-0.80	-0.801 [-0.943,-0.657]	-0.793 [-0.933,-0.65]
γ	0.001	0.005 [0,0.017]	0.003 [0,0.01]	β_4	0.09	0.091 [-0.052,0.23]	0.099 [-0.04,0.24]
ν	15		12.191 [8,16.4]				

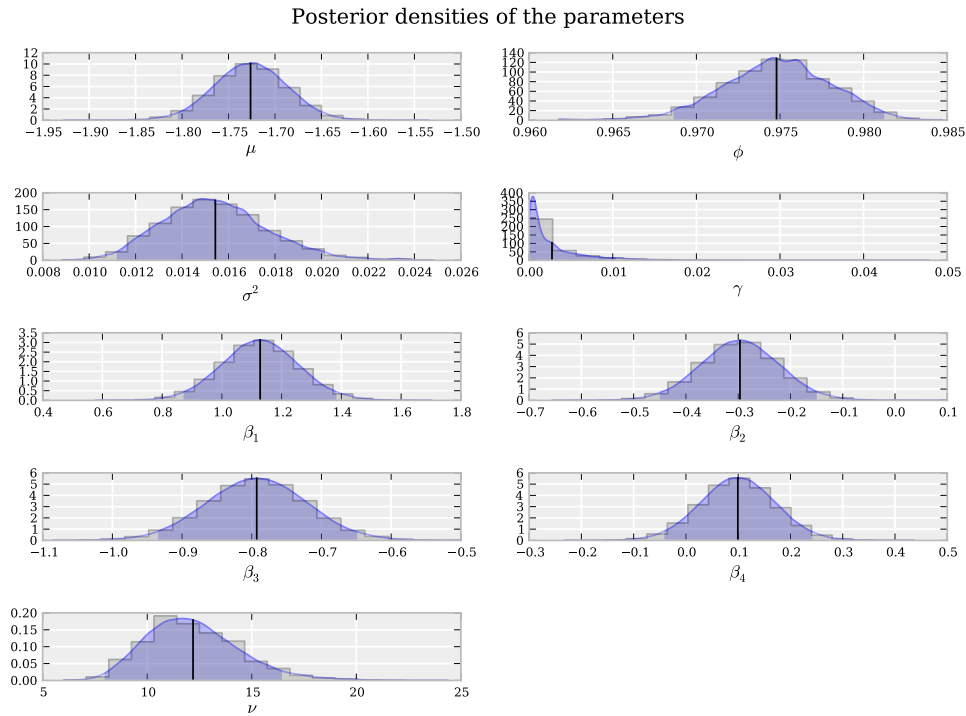


Figure 1: Posterior distributions of the parameters from a dynamic Δ NB model based on 20000 observations and 100000 iterations from which 20000 used as a burn in sample. Each picture shows the histogram of the posterior draws the kernel density estimate of the posterior distribution, the HPD region and the posterior mean. The true parameters are $\mu = -1.7$, $\varphi = 0.97$, $\sigma_\eta = 0.02$, $\gamma = 0.001$ and $\nu = 15$.

4 Empirical study

In this section we present and discuss the empirical findings from our analyses concerning tick by tick price changes for six different stocks traded at the NYSE, for two different periods. We consider two model classes and two models for each class. The first set consists of the ordered probit models with normal and Student's t stochastic volatility. The second set includes the dynamic Skellam and dynamic Δ NB models. The analyses include in-sample and out-of-sample marginal likelihood comparison of the models. Our aims of the empirical study is twofold. First, the usefulness of the Δ NB model on a challenging dataset is investigated. In particular, we validate our estimation procedure and reveal possible shortcomings in the estimation of the parameters in the Δ NB model. Second, we intend to find out what the differences are when the considered models are based on heavy-tailed distributions (ordered t and Δ NB models) or not (ordered normal and dynamic Skellam models). Also, we compare the different model classes: ordered model versus integer distribution model.

4.1 Data

We have access to the Thomson Reuters Sirca dataset that contains all trades and quotes with millisecond time stamps for all stocks listed at NYSE. We have collected the data for Alcoa (AA), Coca-Cola (KO), International Business Machines (IBM), J.P. Morgan (JPM), Ford (F) and Xerox (XRX). These stocks differ in liquidity and in price magnitude. In our study we concentrate on two weeks of price changes: the first week of October 2008 and the last week of April 2010. These weeks exhibit different market sentiments and volatility characteristics. The month of October 2008 is in the middle of the 2008 financial crises with record high volatilities and some markets experienced their worst weeks in October 2008 since 1929. The month of April 2010 is a much calmer month with low volatilities.

The cleaning process of the data consists of a number of filtering steps that are similar to the procedures described in Boudt et al. (2012), Barndorff-Nielsen et al. (2008) and Brownlees and Gallo (2006). First, we remove the quotes-only entries which is a large portion of the data. By excluding the quotes we lose around 70 – 90% of the data. In the next step, we delete the trades with missing or zero prices or volumes. We also restrict our analysis to the trading period. The fourth step is to aggregate the trades which have the same time stamp. We take the trades with the last sequence number when there are multiple trades at the same millisecond. We regard the last price as the price that we can observe with a millisecond resolution. Finally, we treat outliers by following the rules as suggested by Barndorff-Nielsen et al. (2008). We delete trades with prices smaller than the bid-price minus the bid-ask spread and higher than the ask-price plus the bid-ask

spread. Tables 2 and 3 present the descriptive statistics for our resulting data from the 3rd to 10th October 2008 and from the 23rd to 30th April 2010, respectively. A more detailed account of the cleaning process can be found in Tables 10 and 11 in Appendix E. We treat the periods from the 3rd to 9th October 2008 and from the 23rd to 29th April 2010 as the in-sample periods. The two out-of-sample periods are 10 October 2008 and 30 April 2010.

Table 2: Descriptive statistics of the data from 3rd to 10th October 2008. Column In displays the statistics on the in-sample period from 3rd to 9th October 2008, while the column Out shows the descriptives for the out-of-sample period 10th October. We show the number of observations (Num.obs), average price (Avg. price), mean price change (Mean), standard deviation of price changes (Std), minimum and max integer price changes (Min,Max) and the percentage of zeros in the sample (% Zeros).

	AA		F		IBM	
	In	Out	In	Out	In	Out
Num. obs	64 807	14 385	32 756	14 313	68 002	20 800
Avg. price	16.75	11.574	3.077	2.112	96.796	87.583
Mean	-0.007	-0.004	-0.007	0	-0.02	-0.004
Std	1.63	2.126	0.745	0.601	6.831	7.09
Min	-33	-51	-18	-10	-197	-105
Max	38	39	21	9	186	140
% Zeros	48.76	48.76	77.08	77.08	39.9	39.9
	JPM		KO		XRX	
	In	Out	In	Out	In	Out
Num. obs	142 867	43 230	70 356	25 036	26 020	8 623
Avg. price	42.773	38.889	49.203	41.875	9.049	7.768
Mean	-0.009	0.012	-0.012	0.005	-0.006	0.004
Std	2.368	2.779	1.758	2.734	0.816	1.285
Min	-48	-40	-33	-50	-17	-17
Max	74	55	30	63	19	12
% Zeros	43.78	43.78	34.39	34.39	54.98	54.98

4.2 Estimation results

We start our analyses with the dynamic Skellam and Δ NB models for all considered stocks in the periods from 3rd to 9th October 2008 and from 23rd to 29th April 2010. In this study, after some initial experimentation, we use the following prior distributions

$$\mu_h \sim \mathcal{N}(0, 1), \quad \beta_i \sim \mathcal{N}(0, 1), \quad \frac{\varphi + 1}{2} \sim \mathcal{B}(20, 1.5),$$

$$\sigma_\eta^2 \sim \mathcal{IG}(2.5, 0.025), \quad \gamma \sim \mathcal{B}(1.7, 10), \quad \nu \sim \mathcal{DU}(2, 128),$$

for $i = 1, \dots, K$, where \mathcal{N} is the normal, \mathcal{B} is the beta, \mathcal{IG} is the inverse gamma, and \mathcal{DU} is the discrete uniform distribution. In the MCMC procedure, we draw 100,000 samples

Table 3: Descriptive statistics of the data from 23rd to 30th April 2010. Column In displays the statistics on the in-sample period from 23rd to 29th October 2008, while the column Out shows the descriptives for the out-of-sample period 30th October. We show the number of observations (Num.obs), average price (Avg. price), mean price change (Mean), standard deviation of price change (Std), minimum and max integer price change (Min,Max) and the percentage of zeros in the sample (% Zeros).

	AA		F		IBM	
	In	Out	In	Out	In	Out
Num. obs	27 550	4 883	63 241	9 894	43 606	8 587
Avg. price	13.749	13.519	13.734	13.231	130.176	129.575
Mean	-0.001	-0.006	-0.001	-0.006	0.001	-0.019
Std	0.468	0.502	0.448	0.454	1.424	1.371
Min	-3	-2	-5	-2	-22	-15
Max	3	2	4	3	24	9
% Zeros	75.02	75.02	79.73	79.73	51.93	51.93
	JPM		KO		XRX	
	In	Out	In	Out	In	Out
Num. obs	101 045	21 443	34 469	6 073	36 332	4 326
Avg. price	43.702	42.854	53.628	53.732	11.164	11.025
Mean	-0.001	-0.007	-0.003	-0.006	0	-0.007
Std	0.615	0.638	0.647	0.696	0.494	0.459
Min	-5	-10	-9	-5	-9	-2
Max	5	5	7	5	7	3
% Zeros	68.73	68.73	65.09	65.09	79.29	79.29

from the Markov chain and disregard the first 20,000 draws as burn-in samples. The results of parameter estimation for the 2008 and 2010 data periods are reported in Tables 4, 5, 6 and Tables 7, 8, 9 respectively.

The unconditional mean volatility differ across stocks and time periods. The unconditional mean of the latent state is higher for stocks with higher price and it is higher in the more volatile periods in 2008. These results are consistent with intuition but we should not take strong conclusions from these findings. For example, we cannot compare the means between models as they have somewhat different meanings in different model specifications. The estimated AR(1) coefficients for the different series range from 0.88 to 0.99. This finding suggests persistent dynamic volatility behaviour within a trading day, even after accounting for the intra-day seasonal pattern in volatility. However, by comparing the two different periods, we find that the transient volatility is less persistent in the more volatile crises period. We only included the zero inflation specification for the Δ NB and dynamic Skellam distributions when additional flexibility appears to be needed in the observation density. This flexibility has been required for higher price stocks and during the more volatile periods. In case of the April 2010 period we used the zero inflation only for IBM, while in the October 2008 period we included the zero inflation for all stocks expect for the two lowest price stocks F and XRX. The estimates for the zero

Table 4: Estimation results from a dynamic Skellam and Δ NB model for Alcoa (AA) and Ford (F) during the period from 3rd to 9th October 2008. The posterior mean estimates are based on 100,000 iterations (20,000 used as burn-in). The 95 % HPD regions are in brackets. MaxIneff and minESS are maximum inefficiency among parameters and minimum effective sample size, respectively.

	AA				F			
	Skellam	Ord Norm	Δ NB	Ord t	Skellam	Ord Norm	Δ NB	Ord t
μ	-0.179	0.403	-0.264	0.394	-1.874	-1.561	-1.864	-1.53
	[-0.238,-0.12]	[0.339,0.466]	[-0.323,-0.204]	[0.321,0.467]	[-1.944,-1.803]	[-1.619,-1.501]	[-1.935,-1.792]	[-1.587,-1.473]
ϕ	0.927	0.914	0.941	0.941	0.938	0.881	0.945	0.892
	[0.921,0.934]	[0.906,0.922]	[0.933,0.947]	[0.933,0.951]	[0.929,0.949]	[0.869,0.895]	[0.936,0.954]	[0.88,0.905]
σ^2	0.211	0.32	0.126	0.188	0.114	0.286	0.093	0.231
	[0.194,0.23]	[0.287,0.353]	[0.109,0.148]	[0.157,0.226]	[0.096,0.136]	[0.252,0.323]	[0.077,0.107]	[0.203,0.259]
γ	0.247	0.277	0.243	0.287				
	[0.238,0.255]	[0.268,0.284]	[0.233,0.252]	[0.279,0.295]				
β_1	0.434	0.454	0.372	0.47	0.296	0.303	0.289	0.296
	[0.327,0.543]	[0.342,0.569]	[0.272,0.477]	[0.343,0.596]	[0.158,0.436]	[0.187,0.418]	[0.151,0.432]	[0.182,0.41]
β_2	-0.184	-0.2	-0.151	-0.217	-0.116	-0.12	-0.114	-0.119
	[-0.272,-0.097]	[-0.292,-0.109]	[-0.236,-0.07]	[-0.32,-0.116]	[-0.221,-0.008]	[-0.209,-0.033]	[-0.223,-0.005]	[-0.204,-0.03]
ν			8.684	16.828			14.389	98.252
			[6.4,11.2]	[10.8,24.2]			[10.4,18.2]	[60.4, 128]
maxIneff	2 820.98	1 175.6	8 166.65	9 218.06	1 756.28	221.98	1 022.51	362.84
minESS	28.36	68.05	9.8	8.68	45.55	360.39	78.24	220.49

Table 5: Estimation results from a dynamic Skellam and Δ NB model for International Business Machines (IBM) and JP Morgan (JPM) during the period from 3rd to 9th October 2008. The posterior mean estimates are based on 100,000 iterations (20,000 used as burn-in). The 95 % HPD regions are in brackets. MaxIneff and minESS are maximum inefficiency among parameters and minimum effective sample size, respectively.

	IBM				JPM			
	Skellam	Ord Norm	Δ NB	Ord t	Skellam	Ord Norm	Δ NB	Ord t
μ	1.939	2.708	1.252	2.73	0.236	0.913	0.233	0.952
	[1.865,2.013]	[2.633,2.783]	[1.203,1.299]	[2.657,2.806]	[0.191,0.28]	[0.866,0.961]	[0.191,0.277]	[0.906,0.999]
ϕ	0.882	0.894	0.936	0.898	0.898	0.903	0.904	0.908
	[0.874,0.891]	[0.888,0.901]	[0.931,0.941]	[0.892,0.905]	[0.895,0.903]	[0.898,0.907]	[0.9,0.91]	[0.903,0.912]
σ^2	0.84	0.767	0.12	0.698	0.449	0.496	0.386	0.437
	[0.757,0.915]	[0.71,0.821]	[0.108,0.131]	[0.649,0.749]	[0.424,0.474]	[0.466,0.521]	[0.356,0.404]	[0.409,0.462]
γ	0.279	0.295	0.3	0.299	0.198	0.233	0.204	0.241
	[0.274,0.286]	[0.29,0.301]	[0.295,0.305]	[0.294,0.304]	[0.191,0.204]	[0.228,0.239]	[0.198,0.21]	[0.236,0.247]
β_1	0.283	0.307	0.208	0.318	0.357	0.382	0.343	0.379
	[0.161,0.414]	[0.171,0.441]	[0.121,0.297]	[0.183,0.45]	[0.284,0.432]	[0.303,0.463]	[0.271,0.414]	[0.299,0.459]
β_2	0.075	0.069	0.02	0.056	0.011	0.023	0.015	0.019
	[-0.035,0.183]	[-0.045,0.186]	[-0.055,0.097]	[-0.06,0.171]	[-0.056,0.076]	[-0.052,0.094]	[-0.049,0.081]	[-0.053,0.091]
ν			2	82.281			88.472	124.04
			[2,2]	[44.9, 127.5]			[76.4,98.8]	[116.5, 128]
maxIneff	16 517.03	953.63	2 661.61	2 571.53	6 351.47	2 862.84	12 600.8	6 550.42
minESS	4.84	83.89	30.06	31.11	12.6	27.94	6.35	12.21

Table 6: Estimation results from a dynamic Skellam and Δ NB model for Coca-Cola (KO) and Xerox (XRX) during the period from 3rd to 9th October 2008. The posterior mean estimates are based on 100,000 iterations (20,000 used as burn-in). The 95 % HPD regions are in brackets. MaxIneff and minESS are maximum inefficiency among parameters and minimum effective sample size, respectively.

	KO				XRX			
	Skellam	Ord Norm	Δ NB	Ord t	Skellam	Ord Norm	Δ NB	Ord t
μ	0.182 [0.141,0.224]	0.864 [0.82,0.908]	0.146 [0.106,0.189]	0.861 [0.816,0.906]	-1.42 [-1.474,-1.364]	-1.155 [-1.202,-1.107]	-1.419 [-1.474,-1.364]	-1.332 [-1.405,-1.26]
ϕ	0.938 [0.932,0.944]	0.94 [0.934,0.945]	0.942 [0.937,0.948]	0.943 [0.938,0.949]	0.925 [0.909,0.94]	0.749 [0.718,0.779]	0.94 [0.928,0.952]	0.96 [0.949,0.97]
σ^2	0.081 [0.072,0.09]	0.087 [0.077,0.098]	0.068 [0.06,0.076]	0.079 [0.068,0.089]	0.073 [0.056,0.091]	0.503 [0.44,0.574]	0.049 [0.036,0.061]	0.036 [0.025,0.048]
γ	0.104 [0.097,0.11]	0.144 [0.139,0.15]	0.102 [0.095,0.109]	0.145 [0.14,0.152]				
β_1	0.569 [0.501,0.639]	0.611 [0.54,0.685]	0.542 [0.475,0.61]	0.613 [0.539,0.687]	0.565 [0.452,0.679]	0.601 [0.506,0.696]	0.535 [0.421,0.649]	0.607 [0.468,0.75]
β_2	-0.209 [-0.278,-0.141]	-0.235 [-0.307,-0.162]	-0.195 [-0.261,-0.128]	-0.238 [-0.311,-0.165]	-0.143 [-0.23,-0.055]	-0.148 [-0.222,-0.076]	-0.132 [-0.22,-0.041]	-0.151 [-0.26,-0.041]
ν			34.682 [28.6,41.8]	122.006 [110.9, 128]			8.791 [6.2,12]	5.046 [4.7,5.5]
maxIneff	1 295.71	199.34	2 013.99	277.22	1 131.81	543.75	1 604.94	400.61
minESS	61.74	401.32	39.72	288.58	70.68	147.13	49.85	199.69

Table 7: Estimation results from a dynamic Skellam and Δ NB model for Alcoa (AA) and Ford (F) during the period from 23rd to 29th April 2010. The posterior mean estimates are based on 100,000 iterations (20,000 used as burn-in). The 95 % HPD regions are in brackets. MaxIneff and minESS are maximum inefficiency among parameters and minimum effective sample size, respectively.

	AA				F			
	Skellam	Ord Norm	Δ NB	Ord t	Skellam	Ord Norm	Δ NB	Ord t
μ	-2.23 [-2.29,-2.17]	-1.915 [-1.96,-1.869]	-2.227 [-2.288,-2.167]	-1.898 [-1.943,-1.854]	-2.397 [-2.442,-2.351]	-2.037 [-2.07,-2.005]	-2.393 [-2.436,-2.348]	-2.005 [-2.036,-1.973]
ϕ	0.956 [0.944,0.968]	0.9 [0.881,0.921]	0.958 [0.947,0.971]	0.917 [0.898,0.933]	0.942 [0.933,0.951]	0.888 [0.876,0.901]	0.944 [0.936,0.953]	0.906 [0.894,0.918]
σ^2	0.029 [0.02,0.04]	0.081 [0.061,0.105]	0.027 [0.018,0.039]	0.058 [0.043,0.075]	0.061 [0.051,0.078]	0.124 [0.105,0.142]	0.057 [0.046,0.068]	0.086 [0.072,0.101]
γ								
β_1	0.037 [-0.052,0.13]	0.038 [-0.032,0.106]	0.037 [-0.056,0.13]	0.038 [-0.032,0.107]	0.148 [0.089,0.207]	0.131 [0.089,0.175]	0.149 [0.09,0.206]	0.128 [0.086,0.171]
β_2	-0.041 [-0.138,0.057]	-0.033 [-0.107,0.04]	-0.041 [-0.137,0.061]	-0.034 [-0.107,0.04]	-0.188 [-0.259,-0.115]	-0.161 [-0.213,-0.107]	-0.188 [-0.26,-0.115]	-0.157 [-0.208,-0.103]
ν			20.367 [15,25.8]	114.925 [93, 128]			27.436 [21.4,33.8]	121.529 [109.9, 128]
maxIneff	3 243.85	326.86	1 399.33	200.48	2 297.71	278.35	3 321.8	279.19
minESS	24.66	244.76	57.17	399.04	34.82	287.41	24.08	286.55

Table 8: Estimation results from a dynamic Skellam and Δ NB model for International Business Machines (IBM) and JP Morgan (JPM) during the period from 23rd to 29th April 2010. The posterior mean estimates are based on 100,000 iterations (20,000 used as burn-in). The 95 % HPD regions are in brackets. MaxIneff and minESS are maximum inefficiency among parameters and minimum effective sample size, respectively.

	IBM				JPM			
	Skellam	Ord Norm	Δ NB	Ord t	Skellam	Ord Norm	Δ NB	Ord t
μ	-0.088	0.511	-0.216	0.423	-1.674	-1.408	-1.673	-1.505
	[-0.16,-0.014]	[0.444,0.579]	[-0.296,-0.135]	[0.312,0.535]	[-1.716,-1.632]	[-1.43,-1.386]	[-1.716,-1.631]	[-1.546,-1.464]
ϕ	0.974	0.947	0.983	0.991	0.992	0.872	0.993	0.988
	[0.966,0.981]	[0.934,0.96]	[0.978,0.989]	[0.988,0.994]	[0.99,0.994]	[0.858,0.887]	[0.991,0.994]	[0.985,0.991]
σ^2	0.026	0.076	0.011	0.007	0.002	0.114	0.002	0.004
	[0.018,0.035]	[0.05,0.097]	[0.006,0.015]	[0.004,0.011]	[0.002,0.003]	[0.098,0.129]	[0.002,0.003]	[0.003,0.006]
γ	0.286	0.312	0.269	0.316				
	[0.276,0.297]	[0.303,0.322]	[0.257,0.283]	[0.308,0.325]				
β_1	0.473	0.488	0.424	0.547	0.195	0.244	0.195	0.229
	[0.357,0.591]	[0.385,0.589]	[0.312,0.546]	[0.375,0.722]	[0.124,0.266]	[0.207,0.281]	[0.121,0.266]	[0.16,0.295]
β_2	0.206	0.244	0.181	0.209	0.029	0.022	0.029	0.014
	[0.084,0.327]	[0.14,0.348]	[0.059,0.302]	[0.02,0.397]	[-0.039,0.1]	[-0.013,0.057]	[-0.043,0.098]	[-0.051,0.079]
ν			6.241	9.35			36.288	9.476
			[4.4,8]	[8,10.7]			[29.6,43.8]	[8.8,10.1]
maxIneff	4 024.03	907.22	2 445.44	469.68	2 963.88	621.05	625.7	275.89
minESS	19.88	88.18	32.71	170.33	26.99	128.81	127.86	289.97

Table 9: Estimation results from a dynamic Skellam and Δ NB model for Coca-Cola (KO) and Xerox (XRX) during the period from 23rd to 29th April 2010. The posterior mean estimates are based on 100,000 iterations (20,000 used as burn-in). The 95 % HPD regions are in brackets. MaxIneff and minESS are maximum inefficiency among parameters and minimum effective sample size, respectively.

	KO				XRX			
	Skellam	Ord Norm	Δ NB	Ord t	Skellam	Ord Norm	Δ NB	Ord t
μ	-1.636	-1.364	-1.638	-1.438	-2.334	-2.006	-2.326	-1.946
	[-1.693,-1.581]	[-1.403,-1.325]	[-1.694,-1.582]	[-1.493,-1.384]	[-2.393,-2.275]	[-2.05,-1.962]	[-2.386,-2.266]	[-1.989,-1.903]
ϕ	0.98	0.852	0.98	0.96	0.943	0.838	0.948	0.884
	[0.973,0.987]	[0.826,0.877]	[0.973,0.987]	[0.943,0.974]	[0.929,0.959]	[0.817,0.862]	[0.934,0.962]	[0.862,0.906]
σ^2	0.007	0.144	0.007	0.021	0.059	0.231	0.05	0.129
	[0.004,0.01]	[0.117,0.176]	[0.004,0.01]	[0.011,0.033]	[0.037,0.076]	[0.188,0.274]	[0.035,0.072]	[0.097,0.161]
γ								
β_1	0.355	0.421	0.352	0.406	0.647	0.603	0.641	0.587
	[0.268,0.443]	[0.36,0.481]	[0.264,0.441]	[0.325,0.485]	[0.553,0.739]	[0.536,0.671]	[0.548,0.733]	[0.518,0.656]
β_2	0.067	0.061	0.069	0.05	-0.457	-0.411	-0.452	-0.397
	[-0.032,0.164]	[-0.006,0.126]	[-0.031,0.166]	[-0.039,0.136]	[-0.545,-0.367]	[-0.475,-0.348]	[-0.541,-0.362]	[-0.462,-0.334]
ν			22.404	11.375			16.886	106.091
			[16.6,27.8]	[9.2,13.7]			[12.4,22.4]	[73.9, 128]
maxIneff	1 252.48	269.82	812.64	817.67	2 675.36	607.26	5 104.53	644.36
minESS	63.87	296.5	98.44	97.84	29.9	131.74	15.67	124.15

inflation parameter γ ranges from 0.1 to 0.3. The degrees of freedom parameter ν for the ΔNB distribution is estimated as a higher value during the more quiet 2010 period which suggests that the distribution of the tick by tick price change is closer to a thin tailed distribution during such periods. In addition, we have found that the estimated degrees of freedom parameter is a lower value for stocks with a higher average price.

From a more technical perspective, our study has revealed that the parameters of our ΔNB modeling framework mix relatively slowly. This may indicate that our procedure can be rather inefficient. However, it turns out that the troublesome parameters are in all cases the persistence parameter of the volatility process, φ , and the volatility of volatility, σ_η . It is well established and documented that these coefficients are not easy to estimate as they have not a direct impact on the observations as such; see the discussions in Kim et al. (1998) and Stroud and Johannes (2014)). Furthermore, our empirical study is faced with some challenging numerical problems. First, we should emphasize that for some stock we analyze almost 100,000 observations, and the shortest time series has around 30,000 observations. Such a long time series will typically lead us to a slow mixing process in a Bayesian MCMC estimation process because the full conditional distributions are highly informative. Hence the role of some parameters, specifically those in the state equation, in the estimation process is rather weak. However, we do not conclude that we cannot estimate such parameters accurately. Our simulated experiment in the previous section has shown that our algorithm is successful. It just requires more numerical efforts to obtain accurate results. Second, we have anticipated that parameter estimation for the dynamic Skellam and ΔNB models is less numerically efficient and overall more challenging when compared to the estimation for ordered normal and ordered t models. Parameter estimation for the discrete distribution models requires more auxiliary variables and the analysis is based on additional conditional statements.

On the basis of the output of our MCMC estimation procedure, we obtain the estimates for the latent volatility variable h_t but we can also decompose these estimates into the corresponding components of h_t , these are μ_h , s_t and x_t ; see equation (3). Figure 2 presents the intra-day, tick by tick Coca Cola price changes and its estimated components s_t and x_t for the logarithm of volatility h_t in the Skellam model, from 23rd to 29th April 2010. We notice that apart from the pronounced intra-day seasonality in volatility, many volatility changes occur within a trading day. Some of these volatility changes may have been sparked by news announcements while others may have occurred as the result of the trading process.

4.3 In-sample comparison

It is widely established in Bayesian studies that the computation of sequential Bayes factors (BF) is infeasible in this framework as it requires sequential parameter estimation.

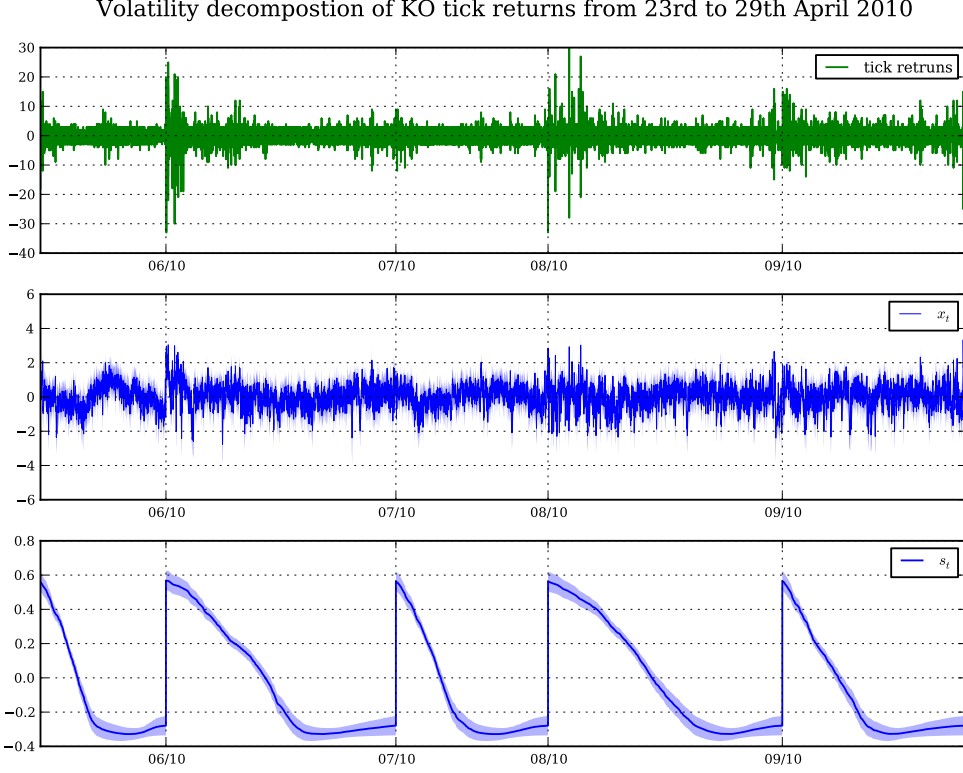


Figure 2: Decomposition of log volatility in the dynamic Skellam model for KO from 23rd to 29th April 2010.

The sequential estimation of the parameters in our model is computationally prohibitive given the very high time dimensions. To provide some comparative assessments of the four models that we have considered in our study, we follow Stroud and Johannes (2014) and calculate Bayesian Information Criteria (BIC) for model \mathcal{M} as

$$BIC_T(\mathcal{M}) = -2 \sum_{t=1}^T \log p(y_t | \hat{\theta}, \mathcal{M}) + d_i \log T \quad (31)$$

where $p(y_t | \theta, \mathcal{M})$ can be calculated by means of a particle filter and $\hat{\theta}$ is the posterior mean of the parameters. The implementation of the particle filter for all considered models is rather straightforward given the provided details of the models in Section 2. The BIC gives an asymptotic approximation to the Bayes factor by

$$BIC_T(\mathcal{M}_i) - BIC_T(\mathcal{M}_j) \approx -2 \log BF_{i,j}.$$

We will use this approximation for our sequential model comparisons.

Figures 3 and Figure 5 present the in-sample Bayes factors for the periods from 23rd to 29th October 2008 and from 3rd to 9th April 2010, respectively. These graphs are rather insightful as they indicate that for all stocks, no evidence in favor of the ΔNB model can be detected for the 2008 period. In the 2010 period, only the IBM stock favors the ΔNB

distribution. In 2008 on the lower price stocks AA, F and XRX, the ordered t model seems to provide the best fit. In 2010 the ordered normal model performs the best on the lower priced stocks, while the high priced stocks are treated more successfully with fat tailed distributions such as the ordered t or the Δ NB distribution models. Furthermore, we may conclude from the sequential Bayes factor results that the ordered t and Δ NB model tends to be favored in case sudden big jumps in volatility have occurred. Such large to extreme realizations of price changes, possibly leading to a prolonged period of high volatility, suggest the need of the Δ NB model. These findings are consistent with the intuition that for time varying volatility models, the identification of parameters determining the tail behaviour requires extreme or excessive observations in combination with low volatility.

4.4 Out-of-sample comparisons

The performances of the dynamic Skellam and Δ NB models can also be compared in terms of predictive likelihoods. The one-step-ahead predictive likelihood for model \mathcal{M} is

$$\begin{aligned} p(y_{t+1}|y_{1:t}, \mathcal{M}) &= \\ &\int \int p(y_{t+1}|y_{1:t}, x_{t+1}, \theta, \mathcal{M})p(x_{t+1}, \theta|y_{1:t}, \mathcal{M})dx_{t+1}d\theta = \\ &\int \int p(y_{t+1}|y_{1:t}, x_{t+1}, \theta, \mathcal{M})p(x_{t+1}|\theta, y_{1:t}, \mathcal{M})p(\theta|y_{1:t}, \mathcal{M})dx_{t+1}d\theta. \end{aligned} \quad (32)$$

Generally, the h -step-ahead predictive likelihood can be decomposed to the sum of one-step-ahead predictive likelihoods via

$$\begin{aligned} p(y_{t+1:t+h}|y_{1:t}, \mathcal{M}) &= \prod_{i=1}^h p(y_{t+i}|y_{1:t+i-1}, \mathcal{M}) = \prod_{i=1}^h \int \int p(y_{t+i}|y_{1:t+i-1}, x_{t+i}, \theta, \mathcal{M}). \\ &\times p(x_{t+i}|\theta, y_{1:t+i-1}, \mathcal{M})p(\theta|y_{1:t+i-1}, \mathcal{M})dx_{t+i}.d\theta \end{aligned} \quad (33)$$

These results suggest that we require the computation of $p(\theta|y_{1:t+i-1}, m)$, for $i = 1, 2, \dots$, that is the posterior of the parameters using sequentially increasing data samples. It requires the MCMC procedure to be repeated as many times as we have number of out-of-sample observations. In our application, for each stock and each model, it implies several thousands of MCMC replications for a predictive analysis of a single out-of-sample day. This exercise is computationally not practical or even infeasible. However, we may be able to rely on the approximation

$$\begin{aligned} p(y_{t+1:t+h}|y_{1:t}, \mathcal{M}) &\approx \prod_{i=1}^h \int \int p(y_{t+i}|y_{1:t+i-1}, x_{t+i}, \theta, \mathcal{M}) \\ &\times p(x_{t+i}|\theta, y_{1:t+i-1}, \mathcal{M})p(\theta|y_{1:t}, \mathcal{M})dx_{t+i}d\theta. \end{aligned} \quad (34)$$

This approximation is based on the notion that, after observing a considerable amount of data, that is for t sufficiently large, the posterior distribution of the static parameters should not change much and hence $p(\theta|y_{1:t+i-1}, \mathcal{M}) \approx p(\theta|y_{1:t}, \mathcal{M})$.

Based on this approximation, we carry out the following exercise. From our MCMC output we obtain a sample of posterior distributions based on the in-sample observations. For each parameter draws from the posterior distribution we estimate the likelihood using the particle filter for the out-of-sample period.

Figures 4 and Figure 6 present the out-of-sample sequential predictive Bayes factors for the 10th October 2008 and 30th April 2010, respectively. On the 10th October 2008, the dynamic Skellam model is preferred for IBM and JPM, the ordered normal model is preferred for AA, F and KO while the Δ NB model is the preferred model for XRX. The dynamic Skellam model for IBM and JPM is consistently winning for both the in-sample and out-of-sample periods. On 30th April 2010 the ordered t model performs the best for AA, F, JPM ,KO and XRX, while the Δ NB model is the best for IBM. The different models appear only to be consistent in terms of the in-sample and out-of-sample performances for the high price stocks.

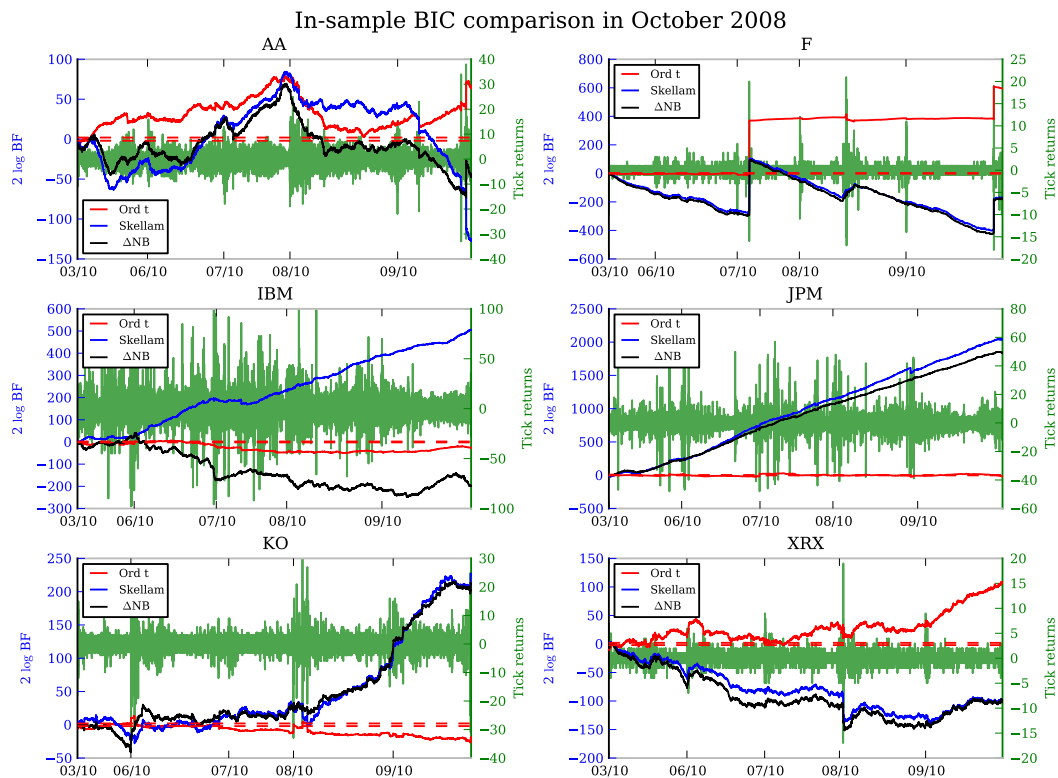


Figure 3: Sequential Bayes factors approximation based on BIC on data from 3rd to 9th October 2008.

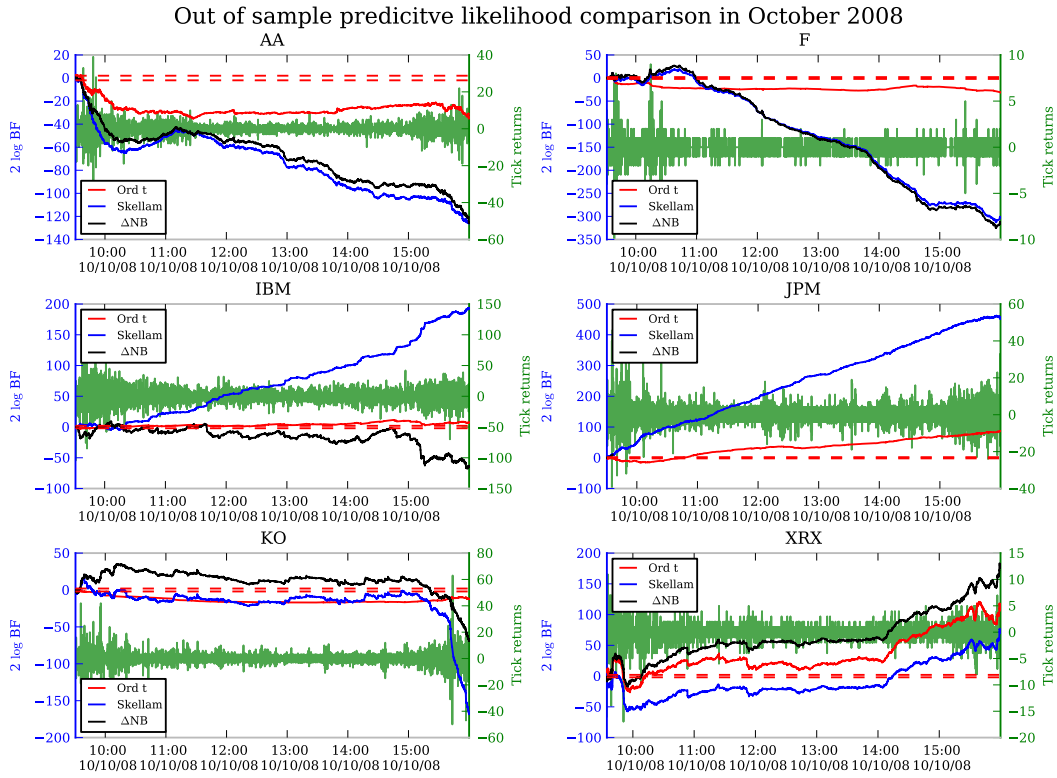


Figure 4: Sequential predictive Bayes factors on 10th October 2008.

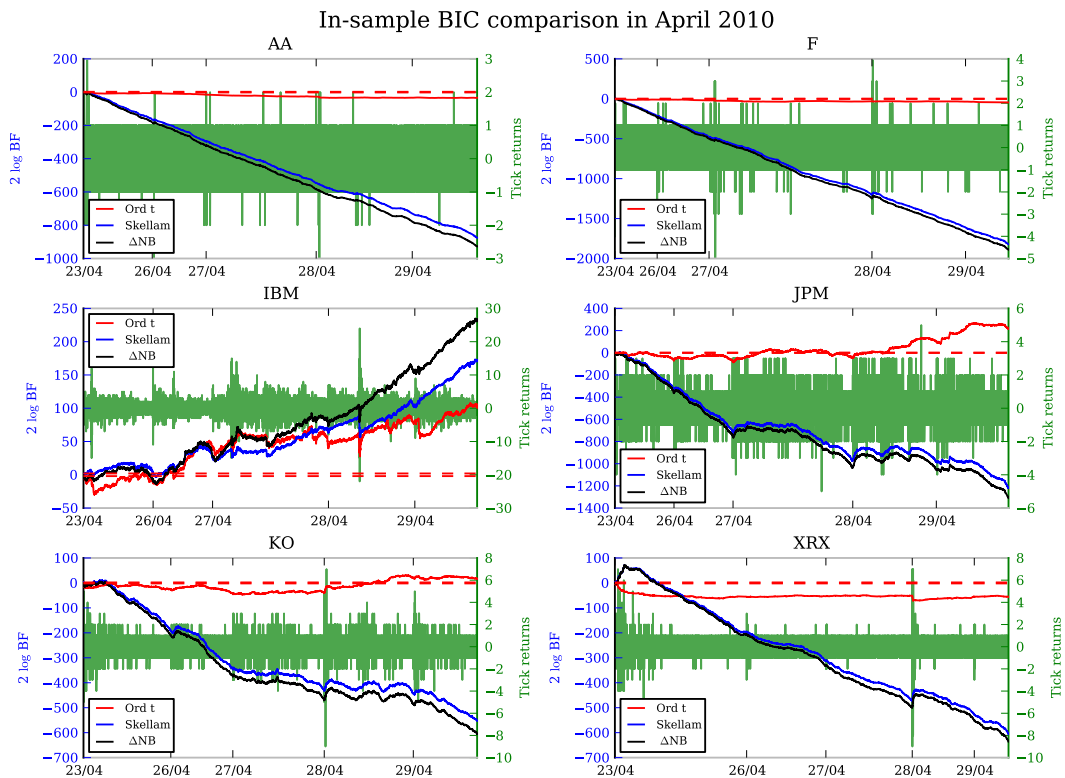


Figure 5: Sequential Bayes factors approximation based on BIC on data from 23rd to 29th April 2010.

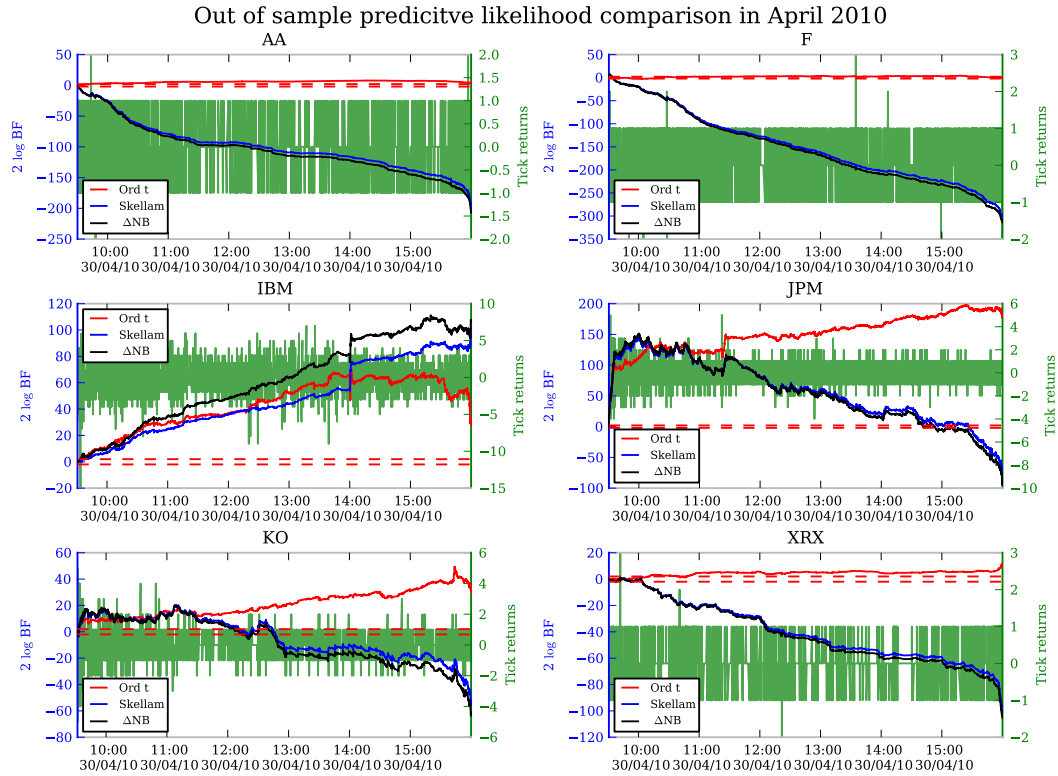


Figure 6: Sequential predictive Bayes factors on 30th April 2010.

5 Conclusion

We have reviewed and introduced dynamic models for high-frequency integer price changes. In particular, we have introduced the dynamic negative binomial difference model, referred to as the Δ NB model. We have developed a Markov chain Monte Carlo procedure (based on Gibbs sampling) for the Bayesian estimation of parameters in the dynamic Skellam and Δ NB models. Furthermore, we have demonstrated our estimation procedures for simulated data and for real data consisting of tick by tick prices from NYSE stocks. We have compared the in-sample and out-of-sample performances of the different models.

References

- Aït-Sahalia, Y., J. Jacod, and J. Li (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics* 168, 207–222.
- Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics* 160, 160–175.
- Alzaid, A. and M. A. Omair (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Science Society* 33, 17–45.

- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Realized Kernels in Practice: Trades and Quotes. *Econometrics Journal* 4, 1–32.
- Barndorff-Nielsen, O. E., D. G. Pollard, and N. Shephard (2012). Integer-valued Levy process and low latency financial econometrics. Working Paper.
- Bos, C. (2008). Model-based estimation of high frequency jump diffusions with microstructure noise and stochastic volatility. TI Discussion Paper.
- Boudt, K., J. Cornelissen, and S. Payseur (2012). Highfrequency: Toolkit for the analysis of Highfrequency financial data in R. Working paper.
- Brownlees, C. and G. Gallo (2006). Financial econometrics analysis at ultra-high frequency: Data handling concerns. *Computational Statistics and Data Analysis* 51, 2232–2245.
- Chakravarty, S., R. A. Wood, and R. A. V. Ness (2004). Decimals and liquidity: A study of the NYSE. *Journal of Financial Research* 27, 75–94.
- Chib, S., F. Nardari, and N. Shephard (2002). Markov Chain Monte Carlo for stochastic volatility models. *Journal of Econometrics* 108, 281–316.
- Chordia, T. and A. Subrahmanyam (1995). Market making, the tick size and payment-for-order-flow: Theory and evidence. *Journal of Business* 68, 543–576.
- Cordella, T. and T. Foucault (1999). Minimum price variations, time priority and quote dynamics. *Journal of Financial Intermediation* 8, 141–173.
- Czado, C. and S. Haug (2010). An ACD-ECOGARCH(1,1) model. *Journal of Financial Econometrics* 8, 335–344.
- Dahlhaus, R. and J. C. Neddermeyer (2014). Online Spot Volatility-Estimation and Decomposition with Nonlinear Market Microstructure Noise Models. *Journal of Financial Econometrics* 12, 174–212.
- Dayri, K. and M. Rosenbaum (2013). Large tick assets: implicit spread and optimal tick size. Working paper.
- Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89, 603–616.
- Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (Second Edition ed.). Oxford University Press.

- Eisler, Z., J. P. Bouchaud, and J. Kockelkoren (2012). The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance* 12, 1395–1419.
- Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica* 68, 1–22.
- Frühwirth-Schnatter, S., R. Frühwirth, L. Held, and H. Rue (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing* 19, 479–492.
- Frühwirth-Schnatter, S. and H. Wagner (2006). Auxiliary mixture sampling for parameter-driven models of time series of small counts with applications to state space modeling. *Biometrika* 93, 827–841.
- Griffin, J. and R. Oomen (2008). Sampling retruns for realized varaiance calculations: tick time or transaction time? *Econometric Reviews* 27, 230–253.
- Johnson, N. L., A. W. Kemp, and S. Kotz (2005). *Univariate Discrete Distributions* (Third Edition ed.). John Wiley and Sons.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.
- Koopman, S. J., R. Lit, and A. Lucas (2014). The Dynamic Skellam Model with Applications. TI Discussion Paper 14-032.
- Müller, G. J. and C. Czado (2006). Stochastic volatility models for ordinal valued time series with application to finance. Working paper.
- Münnix, M. C., R. Schäfer, and T. Guhr (2010). Impact of the tick-size on financial returns and correlations. *Physica A: Statistical Mechanics and its Applications* 389(21), 4828–4843.
- O’Hara, M., G. Saar, and Z. Zhong (2014). Relative Tick Size and the Trading Environment. Working paper.
- Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: fast likelihood inference. *Journal of Econometrics* 140, 425–449.
- Poirier, D. J. (1973). Piecewise Regression Using Cubic Splines. *Journal of the American Statistical Association* 68, 515–524.
- Roberts, G. O. and J. S. Rosenthal (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* 18, 349–367.

- Ronen, T. and D. G. Weaver (2001). "teenies" anyone? *Journal of Financial Markets* 4, 231–260.
- Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics* 1, 2–25.
- SEC (2012). Report to Congress on Decimalization. US Securities and Exchange Commission report.
- 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.
- Stefanos, D. (2015). Bayesian inference for ordinal-response state space mixed models with stochastic volatility. Working paper.
- Stroud, J. R. and M. S. Johannes (2014). Bayesian modeling and forecasting of 24-hour high-frequency volatility: A case study of the financial crisis. Working paper.
- Weinberg, J., L. D. Brown, and R. S. J (2007). Bayesian forecasting of an inhomogenous Poisson process with application to call center data. *Journal of the American Statistical Association* 102, 1185–1199.
- Ye, M. and C. Yao (2014). Tick Size Constrains, Market Structure, and Liquidity. Working paper.

A Negative Binomial distribution

Different parametrization of the NB distribution

$$f(k; \nu, p) = \frac{\Gamma(\nu + k)}{\Gamma(\nu)\Gamma(k + 1)} p^k (1 - p)^\nu \quad (\text{A1})$$

Using

$$\lambda = \nu \frac{p}{1 - p} \Rightarrow p = \frac{\lambda}{\lambda + \nu} \quad (\text{A2})$$

$$f(k; \lambda, \nu) = \frac{\Gamma(\nu + k)}{\Gamma(\nu)\Gamma(k + 1)} \left(\frac{\lambda}{\nu + \lambda} \right)^k \left(\frac{\nu}{\nu + \lambda} \right)^\nu \quad (\text{A3})$$

Mean

$$\mu = \lambda \quad (\text{A4})$$

Variance

$$\sigma^2 = \lambda \left(1 + \frac{\lambda}{\nu} \right) \quad (\text{A5})$$

Dispersion index

$$\frac{\sigma^2}{\mu} \left(1 + \frac{\lambda}{\nu} \right) \quad (\text{A6})$$

The NB distribution is over dispersed and which means that there are more intervals with low counts and more intervals with high counts, compared to a Poisson distribution. As we increase ν we get back to the Poisson case.

The Poisson distribution can be obtained from the NB distribution as follows

$$\begin{aligned} \lim_{\nu \rightarrow \infty} f(k; \lambda, \nu) &= \frac{\lambda^k}{k!} \lim_{\nu \rightarrow \infty} \frac{\Gamma(\nu + k)}{\Gamma(\nu)(\nu + \lambda)^k} \frac{1}{\left(1 + \frac{\lambda}{\nu}\right)^\nu} = \frac{\lambda^k}{k!} \lim_{\nu \rightarrow \infty} \frac{(\nu + k - 1) \dots \nu}{(\nu + \lambda)^k} \frac{1}{\left(1 + \frac{\lambda}{\nu}\right)^\nu} \\ &= \frac{\lambda^k}{k!} \cdot 1 \cdot \frac{1}{e^\lambda} = \text{Poi}(\lambda) \end{aligned} \quad (\text{A7})$$

The NB distribution $Y \sim NB(\lambda, \nu)$ can be written as a Poisson-Gamma mixture or Poisson distribution with Gamma heterogeneity where the Gamma heterogeneity has mean 1.

$$Y \sim \text{Poi}(\lambda U) \quad \text{where} \quad U \sim \text{Ga}(\nu, \nu), \quad (\text{A8})$$

where we use the $\text{Ga}(\alpha, \beta)$ is given by

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad (\text{A9})$$

$$\begin{aligned}
f(k; \lambda, \nu) &= \int_0^{\infty} f_{\text{Poisson}}(k; \lambda u) f_{\text{Gamma}}(u; \nu, \nu) du \\
&= \int_0^{\infty} \frac{(\lambda u)^k e^{-\lambda u}}{k!} \frac{\nu^\nu u^\nu e^{-\nu u}}{\Gamma(\nu)} du \\
&= \frac{\lambda^k \nu^\nu}{k! \Gamma(\nu)} \int_0^{\infty} e^{-(\lambda+\nu)u} u^{k+\nu-1} du
\end{aligned}$$

Substituting $(\lambda + \nu)u = s$ we get

$$\begin{aligned}
&= \frac{\lambda^k \nu^\nu}{k! \Gamma(\nu)} \int_0^{\infty} e^{-s} \frac{s^{k+\nu-1}}{(\lambda + \nu)^{k+\nu-1}} \frac{1}{(\lambda + \nu)} ds \\
&= \frac{\lambda^k \nu^\nu}{k! \Gamma(\nu)} \frac{1}{(\lambda + \nu)^{k+\nu}} \int_0^{\infty} e^{-s} s^{k+\nu-1} ds \\
&= \frac{\lambda^k \nu^\nu}{k! \Gamma(\nu)} \frac{\Gamma(k + \nu)}{(\lambda + \nu)^{k+\nu}} \\
&= \frac{\Gamma(\nu + k)}{\Gamma(\nu) \Gamma(k + 1)} \left(\frac{\lambda}{\nu + \lambda} \right)^k \left(\frac{\nu}{\nu + \lambda} \right)^\nu
\end{aligned} \tag{A10}$$

B Daily volatility patterns

We want to approximate the function $f : \mathbb{R} \rightarrow \mathbb{R}$ with a continuous function which is built up from piecewise polynomials of degree at most three. Let the set $\Delta = \{k_0, \dots, k_K\}$ denote the set of knots k_j $j = 0, \dots, K$. Δ is some times called a mesh on $[k_0, k_K]$. Let $y = \{y_0, \dots, y_K\}$ where $y_j = f(x_j)$. We denote a cubic spline on Δ interpolating to y as $S_\Delta(x)$. $S_\Delta(x)$ has to satisfy

1. $S_\Delta(x) \in C^2[k_0, k_K]$
2. $S_\Delta(x)$ coincides with a polynomial of degree at most three on the intervals $[k_{j-1}, k_j]$ for $j = 0, \dots, K$.
3. $S_\Delta(x) = y_j$ for $j = 0, \dots, K$.

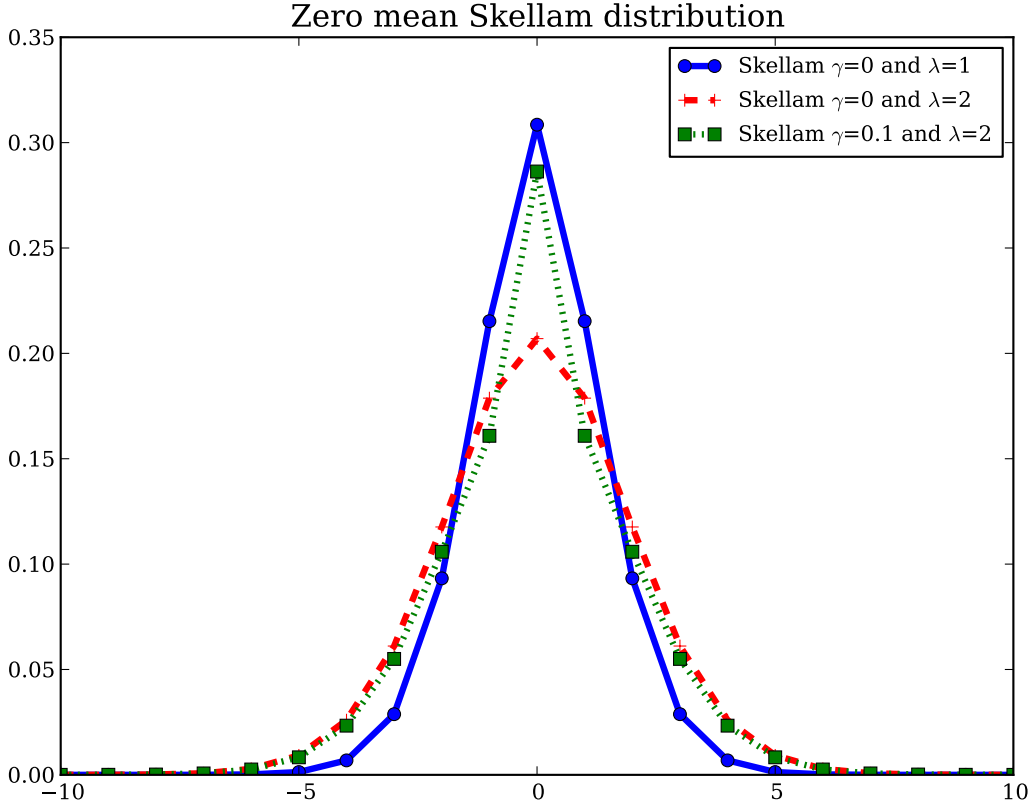


Figure 7: The picture shows the Skellam distribution with different parameters

Using the 2 we know that the $S''_{\Delta}(x)$ is a linear function on $[k_{j-1}, k_j]$ which means that we can write $S''_{\Delta}(x)$ as

$$S''_{\Delta}(x) = \left[\frac{k_j - x}{h_j} \right] M_{j-1} + \left[\frac{x - k_{j-1}}{h_j} \right] M_j \quad \text{for } x \in [k_{j-1}, k_j] \quad (\text{A11})$$

where $M_j = S''_{\Delta}(k_j)$ and $h_j = k_j - k_{j-1}$. Integrating $S''_{\Delta}(x)$ and solving the integrating for the two integrating constants (using $S_{\Delta}(x) = y_j$) Poirier (1973) shows that we get

$$S'_{\Delta}(x) = \left[\frac{h_j}{6} - \frac{(k_j - x)^2}{2h_j} \right] M_{j-1} + \left[\frac{(x - k_{j-1})^2}{2h_j} - \frac{h_j}{6} \right] M_j + \frac{y_j - y_{j-1}}{h_j} \quad \text{for } x \in [k_{j-1}, k_j] \quad (\text{A12})$$

$$\begin{aligned} S_{\Delta}(x) &= \frac{k_j - x}{6h_j} [(k_j - x)^2 - h_j^2] M_{j-1} + \frac{x - k_{j-1}}{6h_j} [(x - k_{j-1})^2 - h_j^2] M_j \\ &+ \left[\frac{k_j - x}{h_j} \right] y_{j-1} + \left[\frac{x - k_{j-1}}{h_j} \right] y_j \quad \text{for } x \in [k_{j-1}, k_j] \end{aligned} \quad (\text{A13})$$

In the above expression only M_j for $j = 0, \dots, K$ are unknown. We can use the continuity restrictions which enforce continuity at the knots by requiring that the derivatives are

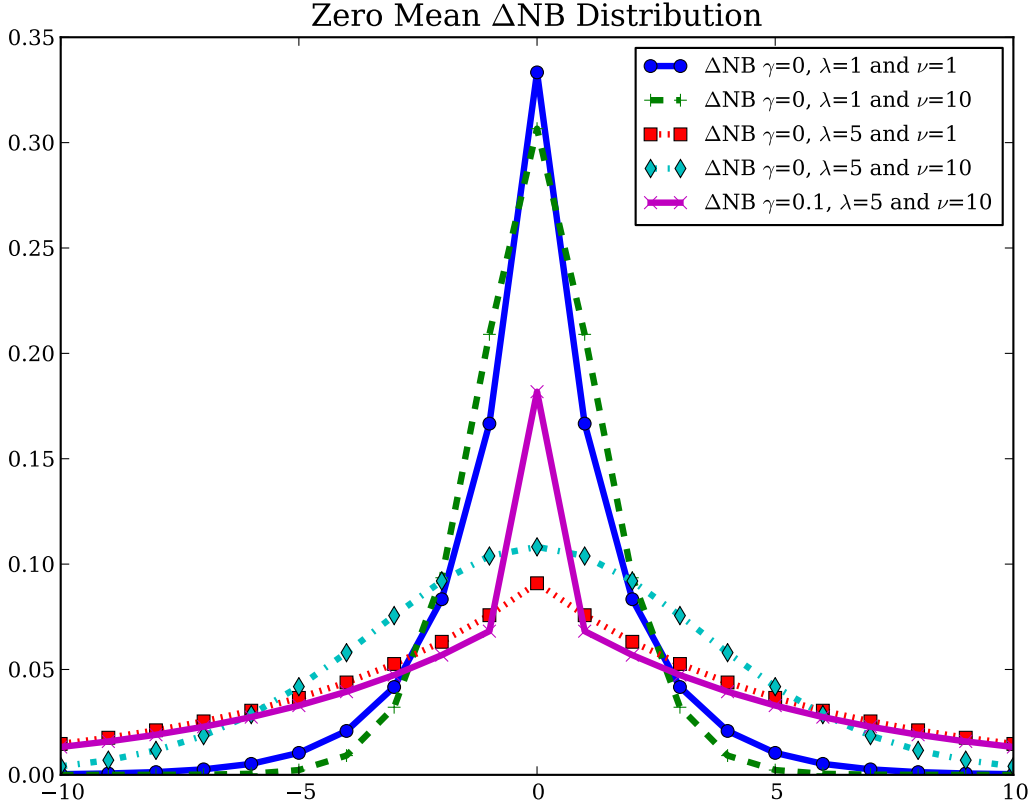


Figure 8: The picture shows the Δ NB distribution with different parameters

equal at the knots k_j for $j = 1, \dots, K - 1$

$$S'_{\Delta}(k_j^-) = h_j M_{j-1}/6 + h_j M_j/3 + (y_j - y_{j-1})/h_j \quad (\text{A14})$$

$$S'_{\Delta}(k_j^+) = -h_{j+1} M_j/3 - h_{j+1} M_{j+1}/6 + (y_{j+1} - y_j)/h_{j+1} \quad (\text{A15})$$

which yields $K - 1$ conditions

$$(1 - \lambda_j)M_{j-1} + 2M_j + \lambda_j M_{j+1} = \frac{6y_{j-1}}{h_j(h_j + h_{j+1})} - \frac{6y_j}{h_j h_{j+1}} + \frac{6y_{j+1}}{h_{j+1}(h_j + h_{j+1})} \quad (\text{A16})$$

where

$$\lambda_j = \frac{h_{j+1}}{h_j + h_{j+1}} \quad (\text{A17})$$

Using two end conditions we have $K + 1$ unknowns and $K + 1$ equations and we can solve the linear equation system for M_j . Using the $M_0 = \pi_0 M_1$ and $M_K = \pi_K M_{K-1}$ end

conditions we can write

$$\underbrace{\Lambda}_{(K+1) \times (K+1)} = \begin{bmatrix} 2 & -2\pi_0 & 0 & \dots & 0 & 0 & 0 \\ 1-\lambda_1 & 2 & \lambda_1 & \dots & 0 & 0 & 0 \\ 0 & 1-\lambda_2 & 2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & \lambda_{K-2} & 0 \\ 0 & 0 & 0 & \dots & 1-\lambda_{K-1} & 2 & \lambda_{K-1} \\ 0 & 0 & 0 & \dots & 0 & -2\pi_K & 2 \end{bmatrix} \quad (\text{A18})$$

$$\Theta = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{6}{h_1(h_1+h_2)} & -\frac{6}{h_1h_2} & \frac{6}{h_2(h_1+h_2)} & \dots & 0 & 0 & 0 \\ 0 & \frac{6}{h_2(h_2+h_3)} & -\frac{6}{h_2h_3} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\frac{6}{h_{K-2}h_{K-1}} & \frac{6}{h_{K-1}(h_{K-2}+h_{K-1})} & 0 \\ 0 & 0 & 0 & \dots & \frac{6}{h_{K-1}(h_{K-1}+h_K)} & -\frac{6}{h_{K-1}h_K} & \frac{6}{h_K(h_{K-1}+h_K)} \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}$$

$$\underbrace{m}_{(K+1) \times 1} = \begin{bmatrix} M_0 \\ M_1 \\ \vdots \\ M_{K-1} \\ M_K \end{bmatrix} \quad (\text{A19})$$

$$\underbrace{y}_{(K+1) \times 1} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{K-1} \\ y_K \end{bmatrix} \quad (\text{A20})$$

The linear equation system is given by

$$\Lambda m = \Theta y \quad (\text{A21})$$

and the solution is

$$m = \Lambda^{-1} \Theta y \quad (\text{A22})$$

Using this result and equation (A13) we can calculate

$$\underbrace{S_{\Delta}(\xi)}_{N \times 1} = \begin{bmatrix} S_{\Delta}(\xi_1) \\ S_{\Delta}(\xi_2) \\ \vdots \\ S_{\Delta}(\xi_{N-1}) \\ S_{\Delta}(\xi_N) \end{bmatrix} \quad (\text{A23})$$

Lets denote P the $N \times (K+1)$ matrix where i th row $i = 1, \dots, N$ given that $k_{j-1} \leq \xi \leq k_j$ can be written as

$$\underbrace{p_i}_{1 \times (K+1)} = \left[\underbrace{0, \dots, 0}_{\text{first } j-2}, \frac{k_j - \xi_i}{6h_j} [(k_j - \xi_i)^2 - h_j^2], \frac{\xi_i - k_{j-1}}{6h_j} [(\xi_i - k_{j-1})^2 - h_j^2], \underbrace{0, \dots, 0}_{\text{last } K+1-j} \right]$$

Moreover denote Q the $N \times (K+1)$ matrix where i th row $i = 1, \dots, N$ given that $k_{j-1} \leq \xi \leq k_j$ can be written as

$$\underbrace{q_i}_{1 \times (K+1)} = \left[\underbrace{0, \dots, 0}_{\text{first } j-2}, \frac{k_j - \xi_i}{h_j}, \frac{\xi_i - k_{j-1}}{h_j}, \underbrace{0, \dots, 0}_{\text{last } K+1-j} \right] \quad (\text{A24})$$

Now using (A13) and (A22) we get

$$S_{\Delta}(\xi) = Pm + Qy = P\Lambda^{-1}\Theta y + Qy = (P\Lambda^{-1}\Theta + Q)y = \underbrace{W}_{N \times (K+1)} \underbrace{y}_{(K+1) \times 1} \quad (\text{A25})$$

where

$$W = P\Lambda^{-1}\Theta + Q \quad (\text{A26})$$

In practical situations we might only know the knots but we don't know we observe the spline values with error. In this case we have

$$s = S_{\Delta}(\xi) + \varepsilon = Wy + \varepsilon, \quad (\text{A27})$$

where

$$\underbrace{s}_{N \times 1} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{N-1} \\ s_N \end{bmatrix} \quad (\text{A28})$$

and

$$\underbrace{\varepsilon}_{N \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{N-1} \\ \varepsilon_N \end{bmatrix} \quad (\text{A29})$$

with

$$E(\varepsilon) = 0 \quad \text{and} \quad E(\varepsilon\varepsilon') = \sigma^2 I \quad (\text{A30})$$

Notice that after fixing the knots we only have to estimate the value of the spline at he knots and this determines the whole shape of the spline. We cab do this by simple OLS

$$\hat{y} = (W^\top W)^{-1} W^\top s \quad (\text{A31})$$

For identification reasons we want

$$\sum_{j:\text{unique}\xi_j} S_\Delta(\xi_j) = \sum_{j:\text{unique}\xi_j} w_j y = w^* y = 0 \quad (\text{A32})$$

where w_i is the i th row of W and

$$\underbrace{w^*}_{1 \times (K+1)} = \sum_{j:\text{unique}\xi_j} w_j \quad (\text{A33})$$

The restriction can be enforced by one of the elements of y . This ensures that $E(s_t) = 0$ so s_t and μ_h can be identified. If we drop y_K we can substitute

$$y_K = - \sum_{i=0}^{K-1} (w_i^*/w_K^*) y_i \quad (\text{A34})$$

where w_i^* is the i th element of w^* . Substituting this into

$$\begin{aligned} \sum_{j:\text{unique}\xi_j} S_\Delta(\xi_j) &= \sum_{j:\text{unique}\xi_j} w_j y = \sum_{j:\text{unique}\xi_j} \sum_{i=0}^K w_{ji} y_i = \sum_{j:\text{unique}\xi_j} \sum_{i=0}^{K-1} w_{ji} y_i - w_{jK} \sum_{i=0}^{K-1} (w_i^*/w_K^*) y_i \\ &= \sum_{j:\text{unique}\xi_j} \sum_{i=0}^{K-1} (w_{ji} - w_{jK} w_i^*/w_K^*) y_i = \sum_{i=0}^{K-1} \sum_{j:\text{unique}\xi_j} (w_{ji} - w_{jK} w_i^*/w_K^*) y_i \\ &= \sum_{i=0}^{K-1} (w_i^* - w_K^* w_i^*/w_K^*) y_i = \sum_{i=0}^{K-1} (w_i^* - w_i^*) y_i = 0 \end{aligned} \quad (\text{A35})$$

Lets partition W in the following way

$$\underbrace{W}_{N \times (K+1)} = \left[\underbrace{W_{-K}}_{N \times K} : \underbrace{W_K}_{N \times 1} \right] \quad (\text{A36})$$

where W_{-K} is equal to the first K columns of W and W_K is the K th column of W . Moreover

$$\underbrace{w^*}_{1 \times (K+1)} = \left[\underbrace{w_{-K}^*}_{1 \times K} : \underbrace{w_K^*}_{1 \times 1} \right] \quad (\text{A37})$$

We can define

$$\underbrace{\widetilde{W}}_{N \times K} = \underbrace{W_{-K}}_{N \times K} - \frac{1}{w_K^*} \underbrace{W_K}_{N \times 1} \underbrace{w_{-K}^*}_{1 \times K} \quad (\text{A38})$$

and we have

$$s = S_\Delta(\xi) + \varepsilon = \underbrace{\widetilde{W}}_{N \times K} \underbrace{\tilde{y}}_{K \times 1} + \varepsilon. \quad (\text{A39})$$

C MCMC estimation of the ordered t-SV model

In this section, the t element vectors (v_1, \dots, v_t) containing time dependent variables for all time time periods, are denoted by v , the variable without a subscript.

C.1 Generating the parameters $x, \mu_h, \varphi, \sigma_\eta^2$ (Step 2)

Notice that conditional on $C = \{c_t, t = 1, \dots, T\}$, r_t^* we have

$$2 \log r_t^* = \mu + s_t + x_t + \log \lambda_t + m_{c_t} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, v_{c_t}^2) \quad (\text{A40})$$

which implies the following following state space form

$$\tilde{y}_t = \underbrace{\begin{bmatrix} 1 & w_t & 1 \end{bmatrix}}_{1 \times (K+2)} \underbrace{\begin{bmatrix} \mu \\ \beta \\ x_t \end{bmatrix}}_{(K+2) \times 1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, v_{c_t}^2) \quad (\text{A41})$$

$$\alpha_{t+1} = \underbrace{\begin{bmatrix} \mu \\ \beta \\ x_{t+1} \end{bmatrix}}_{(K+2) \times 1} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & I_K & 0 \\ 0 & 0 & \varphi \end{bmatrix}}_{(K+2) \times (K+2)} \underbrace{\begin{bmatrix} \mu \\ \beta \\ x_t \end{bmatrix}}_{(K+2) \times 1} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \eta_{t+1} \end{bmatrix}}_{(K+2) \times 1}, \quad \eta_{t+1} \sim \mathcal{N}(0, \sigma_\eta^2) \quad (\text{A42})$$

where

$$\underbrace{\begin{bmatrix} \mu \\ \beta \\ x_1 \end{bmatrix}}_{(K+2) \times 1} \sim \mathcal{N} \left(\underbrace{\begin{bmatrix} \mu_0 \\ \beta_0 \\ 0 \end{bmatrix}}_{(K+2) \times 1}, \underbrace{\begin{bmatrix} \sigma_\mu^2 & 0 & 0 \\ 0 & \sigma_\beta^2 I_K & 0 \\ 0 & 0 & \sigma_\eta^2 / (1 - \varphi^2) \end{bmatrix}}_{(K+2) \times (K+2)} \right), \quad (\text{A43})$$

and

$$\tilde{y}_t = 2 \log r_t^* - \log \lambda_t - m_{r_{t1}} \quad (\text{A44})$$

First we draw φ, σ_η^2 from $p(\varphi, \sigma_\eta^2 | \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$. Notice that

$$p(\varphi, \sigma_\eta^2 | \gamma, \nu, C, \tau, N, z_1, z_2, s, y) = p(\varphi, \sigma_\eta^2 | \tilde{y}_t, C, N) \propto p(\tilde{y}_t | \varphi, \sigma_\eta^2, C, N) p(\varphi) p(\sigma_\eta^2), \quad (\text{A45})$$

where \tilde{y}_t is defined above in equation (A67). The likelihood can be evaluated using standard Kalman filtering and prediction error decomposition (see e.g, Durbin and Koopman (2012)) taking advantage of fact that conditional on the auxiliary variables we have a linear Gaussian state space form given by equation (A64),(A65), (A66) and (A67). We draw from the posterior using an adaptive random walk Metropolis-Hastings step proposed by Roberts and Rosenthal (2009). Conditional on φ, σ_η^2 we draw μ_h, s and x from $p(\mu_h, s, x | \varphi, \sigma_\eta^2, \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$, which is done simulating from the smoothed state density of the linear Gaussian state space model given by (A41),(A42), (A43) and (A44). We use the simulation smoother proposed by Durbin and Koopman (2002).

C.2 Generating γ (Step 3)

$$p(\gamma | \nu, \mu, \varphi, \sigma_\eta^2, x, s, C, y, r_t^*) = p(\gamma | \nu, h, y) \quad (\text{A46})$$

because given ν, h and y , the variables $C, \varphi, \sigma_\eta^2, r_t^*$ are redundant.

$$p(\gamma | \nu, h, y) \propto p(y | \gamma, \nu, h) p(\gamma | \nu, h) = p(y | \gamma, \nu, h) p(\gamma) \quad (\text{A47})$$

as γ is independent from ν and h .

$$\begin{aligned}
p(y|\gamma, \nu, h)p(\gamma) &= \prod_{t=1}^T \left\{ \gamma \mathbf{1}_{\{y_t=0\}} + (1-\gamma) \right. \\
&\times \left. \left[\mathcal{T}\left(\frac{y_t+0.5}{\exp(h_t/2)}, \nu\right) - \mathcal{T}\left(\frac{y_t-0.5}{\exp(h_t/2)}, \nu\right) \right] \right\} \frac{\gamma^{a-1}(1-\gamma)^{b-1}}{\text{B}(a, b)} \\
&\propto \prod_{t=1}^T \left\{ \gamma^a(1-\gamma)^{b-1} \mathbf{1}_{\{y_t=0\}} + \gamma^{a-1}(1-\gamma)^b \right. \\
&\times \left. \left[\mathcal{T}\left(\frac{y_t+0.5}{\exp(h_t/2)}, \nu\right) - \mathcal{T}\left(\frac{y_t-0.5}{\exp(h_t/2)}, \nu\right) \right] \right\},
\end{aligned}$$

where $\mathcal{T}(\cdot, \nu)$ is the Student's t density function with mean zero scale one and degree of freedom parameter ν . We sample from this posterior using an adaptive random walk Metropolis-Hastings sampler by Roberts and Rosenthal (2009).

C.3 Generating r^*

$$p(r^*|\gamma, \nu, \mu, \varphi, \sigma_\eta^2, x, s, C, \lambda, y) = p(r^*|\gamma, h, \lambda, y) = \prod_{t=1}^T p(r_t^*|\gamma, h_t, \lambda_t, y_t) \quad (\text{A48})$$

Using the law of total probability

$$\begin{aligned}
p(r_t^*|\gamma, \nu, h_t, y_t) &= p(r_t^*|\gamma, \nu, h_t, \lambda_t, y_t, \text{zero})p(\text{zero}|\gamma, h_t, \lambda_t, y_t) \\
&+ p(r_t^*|\gamma, h_t, \lambda_t, y_t, \text{non-zero})p(\text{non-zero}|\gamma, h_t, \lambda_t, y_t)
\end{aligned} \quad (\text{A49})$$

Where $p(r_t^*|\gamma, h_t, \lambda_t, y_t, \text{zero})$ is a normal density with zero mean and variance $\lambda_t \exp(h_t)$ truncated to the interval $[y_t - 0.5, y_t + 0.5]$. If $y_t = 0$ then

$$\begin{aligned}
p(\text{zero}|\gamma, h_t, y_t = 0) &= \frac{p(\text{zero}, \gamma, h_t, y_t = 0)}{p(\gamma, h_t, y_t = 0)} = \frac{p(y_t = 0|\text{zero}, \gamma, h_t)p(\text{zero}|\gamma, h_t)}{p(y_t = 0|\gamma, h_t)} \\
&= \frac{1 \times \gamma}{\gamma + (1-\gamma) \left[\Phi\left(\frac{0.5}{\sqrt{\lambda_t} \exp(h_t/2)}\right) - \Phi\left(\frac{-0.5}{\sqrt{\lambda_t} \exp(h_t/2)}\right) \right]}
\end{aligned} \quad (\text{A50})$$

If $y_t = k \neq 0$ then

$$\begin{aligned}
p(\text{zero}|\gamma, h_t, y_t = k) &= \frac{p(\text{zero}, \gamma, h_t, y_t = k)}{p(\gamma, h_t, y_t = k)} \\
&= \frac{p(y_t = k|\text{zero}, \gamma, h_t)p(\text{zero}|\gamma, h_t)}{p(y_t = k|\gamma, h_t)} = 0
\end{aligned} \quad (\text{A51})$$

Moreover $p(\text{non-zero}|\gamma, h_t, y_t) = 1 - p(\text{zero}|\gamma, h_t, y_t)$.

C.4 Generating ν and λ

To sample ν and λ we use the method by Stroud and Johannes (2014). We can decompose the posterior density as

$$p(\nu, \lambda | \gamma, \varphi, \sigma_\eta^2, h, C, y, r^*) = p(\nu, \lambda | h, r^*) = p(\lambda | \nu, h, r^*) p(\nu | h, r^*) \quad (\text{A52})$$

Note that we have to following mixture representation

$$r_t^* = \exp(h_t/2) \sqrt{\lambda_t} \varepsilon_t \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad \lambda_t \sim \text{IG}(\nu/2, \nu/2) \quad (\text{A53})$$

which implies

$$p(\nu | h, r^*) \propto \prod_{t=1}^T p\left(\frac{r_t^*}{\exp(h_t/2)} \middle| h_t, \nu\right) p(\nu) \quad (\text{A54})$$

where

$$p\left(\frac{r_t^*}{\exp(h_t/2)} \middle| h_t, \nu\right) \sim t_\nu(0, 1) \quad (\text{A55})$$

and the prior $\nu \sim \mathcal{DU}(2, 128)$ which leads to the posterior

$$p(\nu | h, r^*) \propto \prod_{t=1}^T p\left(\frac{r_t^*}{\exp(h_t/2)} \middle| h_t, \nu\right) = \prod_{t=1}^T g_{\nu^*}\left(\frac{r_t^*}{\exp(h_t/2)}\right) = \prod_{t=1}^T g_{\nu^*}(w_t) \quad (\text{A56})$$

where $w_t = r_t^* / \exp(h_t/2)$.

To avoid the computationally intense evaluation of these probabilities we can use a Metropolis-Hastings update. We can draw the proposal ν^* from the neighbourhood of the current value $\nu^{(i)}$ using a discrete uniform distribution $\nu^* \sim \mathcal{DU}(\nu^{(i)} - \delta, \nu^{(i)} + \delta)$ and accept with probability

$$\min\left\{1, \frac{\prod_{t=1}^T g_{\nu^*}(y_t)}{\prod_{t=1}^T g_{\nu^{(i)}}(y_t)}\right\} \quad (\text{A57})$$

δ is chosen such that the acceptance rate is reasonable.

$$p(\lambda | \nu, h, r^*) = \prod_{t=1}^T p(\lambda_t | \nu, h_t, r_t^*) \propto \prod_{t=1}^T p(r_t^* | \lambda_t, \nu, h_t) p(\lambda_t | \nu) \quad (\text{A58})$$

where

$$p\left(\frac{r_t^*}{\exp(h_t/2)} \middle| \lambda_t, \nu, h_t\right) \sim \mathcal{N}(0, \lambda_t) \quad (\text{A59})$$

$$p(\lambda_t | \nu) \sim \text{IG}(\nu/2, \nu/2) \quad (\text{A60})$$

$$p(\lambda_t | \nu, h_t, r_t^*) \sim \text{IG} \left(\frac{\nu + 1}{2}, \frac{\nu + \left(\frac{r_t^*}{\exp(h_t/2)} \right)^2}{2} \right) \quad (\text{A61})$$

D MCMC estimation of the dynamic ΔNB model

In this section, the t element vectors (v_1, \dots, v_t) containing time dependent variables for all time time periods, are denoted by v , the variable without a subscript.

D.1 Generating the parameters $x, \mu_h, \varphi, \sigma_\eta^2$ (Step 2)

Notice that conditional on $C = \{c_{tj}, t = 1, \dots, T, j = 1, \dots, \min(N_t + 1, 2)\}$, τ, N, γ and s we have

$$-\log \tau_{t1} = \log(z_{t1} + z_{t2}) + \mu_h + s_t + x_t + m_{c_{t1}}(1) + \varepsilon_{t1}, \quad \varepsilon_{t1} \sim \mathcal{N}(0, v_{c_{t1}}^2(1)) \quad (\text{A62})$$

and

$$-\log \tau_{t2} = \log(z_{t1} + z_{t2}) + \mu_h + s_t + x_t + m_{c_{t2}}(N_t) + \varepsilon_{t2}, \quad \varepsilon_{t2} \sim \mathcal{N}(0, v_{c_{t2}}^2(N_t)) \quad (\text{A63})$$

which implies the following state space form

$$\underbrace{\tilde{y}_t}_{\min(N_t+1,2) \times 1} = \underbrace{\begin{bmatrix} 1 & w_t & 1 \\ 1 & w_t & 1 \end{bmatrix}}_{\min(N_t+1,2) \times (K+2)} \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_t \end{bmatrix}}_{(K+2) \times 1} + \underbrace{\varepsilon_t}_{\min(N_t+1,2) \times 1}, \quad \varepsilon_t \sim \mathcal{N}(0, H_t) \quad (\text{A64})$$

$$\alpha_{t+1} = \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_{t+1} \end{bmatrix}}_{(K+2) \times 1} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & I_K & 0 \\ 0 & 0 & \varphi \end{bmatrix}}_{(K+2) \times (K+2)} \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_t \end{bmatrix}}_{(K+2) \times 1} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \eta_{t+1} \end{bmatrix}}_{(K+2) \times 1}, \quad (\text{A65})$$

where $\eta_{t+1} \sim \mathcal{N}(0, \sigma_\eta^2)$ and

$$\underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_1 \end{bmatrix}}_{(K+2) \times 1} \sim \mathcal{N} \left(\underbrace{\begin{bmatrix} \mu_0 \\ \beta_0 \\ 0 \end{bmatrix}}_{(K+2) \times 1}, \underbrace{\begin{bmatrix} \sigma_\mu^2 & 0 & 0 \\ 0 & \sigma_\beta^2 I_K & 0 \\ 0 & 0 & \sigma_{eta}^2 / (1 - \varphi^2) \end{bmatrix}}_{(K+2) \times (K+2)} \right) \quad (\text{A66})$$

$H_t = \text{diag}(v_{c_{t1}}^2(1), v_{c_{t2}}^2(N_t))$ and

$$\underbrace{\tilde{y}_t}_{\min(N_t+1,2) \times 1} = \begin{pmatrix} -\log \tau_{t1} - m_{r_{t1}}(1) - \log(z_{t1} + z_{t2}) \\ -\log \tau_{t2} - m_{r_{t2}}(N_t) - \log(z_{t1} + z_{t2}) \end{pmatrix} \quad (\text{A67})$$

First we draw φ, σ_η^2 from $p(\varphi, \sigma_\eta^2 | \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$. Notice that

$$p(\varphi, \sigma_\eta^2 | \gamma, \nu, C, \tau, N, z_1, z_2, s, y) = p(\varphi, \sigma_\eta^2 | \tilde{y}_t, C, N) \propto p(\tilde{y}_t | \varphi, \sigma_\eta^2, C, N) p(\varphi) p(\sigma_\eta^2), \quad (\text{A68})$$

where \tilde{y}_t is defined above in equation (A67). The likelihood can be evaluated using standard Kalman filtering and prediction error decomposition (see e.g, Durbin and Koopman (2012)) taking advantage of fact that conditional on the auxiliary variables we have a linear Gaussian state space form given by equation (A64),(A65), (A66) and (A67). We draw from the posterior using an adaptive random walk Metropolis-Hastings step proposed by Roberts and Rosenthal (2009). Conditional on φ, σ_η^2 we draw μ_h, s and x from $p(\mu_h, s, x | \varphi, \sigma_\eta^2, \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$, which is done simulating from the smoothed state density of the linear Gaussian state space model given by (A64),(A65), (A66) and (A67). We use the simulation smoother proposed by Durbin and Koopman (2002).

D.2 Generating γ (Step 3)

$$p(\gamma | \nu, \mu_h, \varphi, \sigma_\eta^2, x, C, s, \tau, N, z_1, z_2, y) = p(\gamma | \nu, \mu_h, s, x, y) \quad (\text{A69})$$

because given ν, λ and y , the variables C, τ, N, z_1, z_2 are redundant.

$$p(\gamma | \nu, \mu_h, s, x, y) \propto p(y | \gamma, \nu, \mu_h, s, x) p(\gamma | \nu, \mu_h, s, x) = p(y | \gamma, \nu, \mu_h, s, x) p(\gamma) \quad (\text{A70})$$

as γ is independent from ν and $\lambda_t = \exp(\mu_h + s_t + x_t)$.

$$\begin{aligned} p(y | \gamma, \nu, \mu_h, s, x) p(\gamma) &= \prod_{t=1}^T \left[\gamma \mathbf{1}_{\{y_t=0\}} + (1 - \gamma) \left(\frac{\nu}{\lambda_t + \nu} \right)^{2\nu} \left(\frac{\lambda_t}{\lambda_t + \nu} \right)^{|y_t|} \frac{\Gamma(\nu + |y_t|)}{\Gamma(\nu)\Gamma(|y_t|)} \right. \\ &\quad \times \left. F \left(\nu + y_t, \nu, y_t + 1; \left(\frac{\lambda_t}{\lambda_t + \nu} \right)^2 \right) \right] \frac{\gamma^{a-1} (1 - \gamma)^{b-1}}{B(a, b)} \\ &\propto \prod_{t=1}^T \left[\gamma^a (1 - \gamma)^{b-1} \mathbf{1}_{\{y_t=0\}} + \gamma^{a-1} (1 - \gamma)^b \left(\frac{\nu}{\lambda_t + \nu} \right)^{2\nu} \left(\frac{\lambda_t}{\lambda_t + \nu} \right)^{|y_t|} \right. \\ &\quad \times \left. \frac{\Gamma(\nu + |y_t|)}{\Gamma(\nu)\Gamma(|y_t|)} F \left(\nu + y_t, \nu, y_t + 1; \left(\frac{\lambda_t}{\lambda_t + \nu} \right)^2 \right) \right] \end{aligned}$$

We sample from this posterior using an adaptive random walk Metropolis-Hastings sampler.

D.3 Generating $C, \tau, N, z_1, z_2, \nu$ (Step 4)

We can decompose the joint posterior of $C, \tau, N, z_1, z_2, \nu$ into

$$\begin{aligned}
p(C, \tau, N, z_1, z_2, \nu | \gamma, \mu_h, \varphi, \sigma_\eta^2, s, x, y) &= p(C | \tau, N, z_1, z_2, \gamma, \mu_h, \varphi, \sigma_\eta^2, s, x, y) \\
&\times p(\tau | N, z_1, z_2, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, y) \\
&\times p(N | z_1, z_2, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, y) \\
&\times p(z_1, z_2 | \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, y) \\
&\times p(\nu | \gamma, \mu_h, \varphi, \sigma_\eta^2, s, x, y)
\end{aligned} \tag{A71}$$

Generating ν (Step 4a)

Note that

$$\begin{aligned}
p(\nu | \gamma, \mu_h, \varphi, \sigma_\eta^2, s, x, y) &= p(\nu | \gamma, \lambda, y) \\
&\propto p(\nu, \gamma, \lambda, y) \\
&= p(y | \gamma, \lambda, \nu) p(\lambda | \gamma, \nu) p(\gamma | \nu) p(\nu) \\
&= p(y | \gamma, \lambda, \nu) p(\lambda) p(\gamma) p(\nu) \\
&\propto p(y | \gamma, \lambda, \nu) p(\nu)
\end{aligned} \tag{A72}$$

where $p(y | \gamma, \lambda, \nu)$ is a product of zero inflated Δ NB probability mass functions.

We draw ν using a discrete uniform prior $\nu \sim DU(2, 128)$ and a random walk proposal in the following fashion as suggested by Stroud and Johannes (2014) for degree of freedom parameter of a t density. We can write the posterior as a multinomial distribution $p(\nu | \mu_h, x, z_1, z_2) \sim M(\pi_2^*, \dots, \pi_{128}^*)$ with probabilities

$$\pi_\nu^* \propto \prod_{t=1}^T [\gamma \mathbb{I}_{\{y_t=0\}} + (1 - \gamma) f_{\Delta\text{NB}}(y_t; \lambda_t, \nu)] = \prod_{t=1}^T g_\nu(y_t) \tag{A73}$$

To avoid the computationally intense evaluation of these probabilities we can use a Metropolis-Hastings update. We can draw the proposal ν^* from the neighbourhood of the current value $\nu^{(i)}$ using a discrete uniform distribution $\nu^* \sim DU(\nu^{(i)} - \delta, \nu^{(i)} + \delta)$ and accept with probability

$$\min \left\{ 1, \frac{\prod_{t=1}^T g_{\nu^*}(y_t)}{\prod_{t=1}^T g_{\nu^{(i)}}(y_t)} \right\} \tag{A74}$$

δ is chosen such that the acceptance rate is reasonable.

Generating z_1, z_2 (Step 4b)

Notice that z_1, z_2 are independent given γ, μ_h, s, x, y .

$$p(z_1, z_2 | \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, y) = \prod_{t=1}^T p(z_{t1}, z_{t2} | \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t, y_t) \quad (\text{A75})$$

$$\begin{aligned} p(z_{t1}, z_{t2} | \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t, y_t) &\propto p(z_{t1}, z_{t2}, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t, y_t) \\ &= p(y_t | z_{t1}, z_{t2}, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t) \\ &\times p(z_{t1}, z_{t2} | \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t) \end{aligned} \quad (\text{A76})$$

$$p(z_{t1}, z_{t2} | \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t, y_t) \propto g(z_{t1}, z_{t2}) \frac{\nu^\nu z_{t1}^\nu e^{-\nu z_{t1}}}{\Gamma(\nu)} \frac{\nu^\nu z_{t2}^\nu e^{-\nu z_{t2}}}{\Gamma(\nu)} \quad (\text{A77})$$

where

$$g(z_{t1}, z_{t2}) = \left[\gamma \mathbf{1}_{\{y_t=0\}} + (1 - \gamma) \exp[-\lambda_t(z_{t1} + z_{t2})] \left(\frac{z_{t1}}{z_{t2}} \right)^{\frac{y_t}{2}} I_{|y_t|}(2\lambda_t \sqrt{z_{t1} z_{t2}}) \right] \quad (\text{A78})$$

with $\lambda_t = \exp(\mu_h + s_t + x_t)$. We can carry out an independent MH step by sampling z_{1t}^*, z_{2t}^* from $\text{Ga}(\lambda_t, \nu)$ and accept it with probability

$$\min \left\{ \frac{g(z_{1t}^*, z_{2t}^*)}{g(z_{t1}, z_{t2})}, 1 \right\} \quad (\text{A79})$$

Generating N (Step 4c)

Note that condition on on z_{t1}, z_{t2} and the intensity λ_t the N_t are independent over time, hence

$$p(N | \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, z_1, z_2, y) = \prod_{t=1}^T p(N_t | \gamma, \lambda_t, z_{t1}, z_{t2}, y_t). \quad (\text{A80})$$

For a given t we can draw N_t from a discrete distribution with

$$\begin{aligned} p(N_t | \gamma, \lambda_t, z_{t1}, z_{t2}, y_t) &= \frac{p(N_t, y_t | \gamma, \lambda_t, z_{t1}, z_{t2})}{p(y_t | \gamma, \lambda_t, z_{t1}, z_{t2})} \\ &= \frac{p(y_t | N_t, \gamma, \lambda_t, z_{t1}, z_{t2}) p(N_t | \gamma, \lambda_t, z_{t1}, z_{t2})}{p(y_t | \gamma, \lambda_t, z_{t1}, z_{t2})} \\ &= [\gamma \mathbf{1}_{\{y_t=0\}} + (1 - \gamma) p(y_t | N_t, \lambda_t, z_{t1}, z_{t2})] \\ &\times \frac{p(N_t | \gamma, \lambda_t, z_{t1}, z_{t2})}{p(y_t | \gamma, \lambda_t, z_{t1}, z_{t2})} \end{aligned} \quad (\text{A81})$$

The denominator in equation (A81) is a Skellam distribution with intensity $\lambda_t z_{t1}$ and $\lambda_t z_{t2}$. We can calculate probability

$$p(y_t | N_t, \lambda_t, z_{t1}, z_{t2}) \quad (\text{A82})$$

using the results from equation (12) condition on λ_t , z_{t1} and z_{t2} , y_t is distributed as a marked Poisson process with marks given by

$$M_i = \begin{cases} 1, & \text{with } P(M_i = 1) = \frac{z_{t1}}{z_{t1} + z_{t2}} \\ -1, & \text{with } P(M_i = -1) = \frac{z_{t2}}{z_{t1} + z_{t2}} \end{cases}, \quad (\text{A83})$$

which implies that we can represent y_t as $\sum_{i=0}^{N_t} M_i$.

$$p(y_t | N_t, \lambda_t, z_{t1}, z_{t2}) = \begin{cases} 0, & \text{if } y_t > N_t \text{ or } |y_t| \bmod 2 \neq |N_t| \bmod 2 \\ \binom{N_t}{\frac{N_t + y_t}{2}} \left(\frac{z_{t1}}{z_{t1} + z_{t2}} \right)^{\frac{N_t + y_t}{2}} \left(\frac{z_{t2}}{z_{t1} + z_{t2}} \right)^{\frac{N_t - y_t}{2}}, & \text{otherwise} \end{cases}, \quad (\text{A84})$$

Conditional on z_{t1} , z_{t2} and λ_t , N_t is a realization of a Poisson process on $[0, 1]$ with intensity $(z_{t1} + z_{t2})\lambda_t$, hence the probability $p(N_t | \gamma, \lambda_t, z_{t1}, z_{t2})$ is a Poisson random variable with intensity equal to $\lambda_t(z_{t1} + z_{t2})$. We can draw N_t parallel over $t = 1, \dots, T$ by drawing a uniform random variable $u_t \sim U[0, 1]$ and

$$N_t = \min \left\{ n : u_t \leq \sum_{i=0}^n p(i | \gamma, \lambda_t, z_{t1}, z_{t2}, y_t) \right\} \quad (\text{A85})$$

Generating τ (Step 4d)

Notice that $p(\tau | N, z_1, z_2, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, x, y) = p(\tau | N, \mu_h, z_1, z_2, s, x)$. Moreover

$$\begin{aligned} p(\tau | \mu_h, z_1, z_2, s, x) &= \prod_{t=1}^T p(\tau_{1t}, \tau_{2t} | N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) \\ &= \prod_{t=1}^T p(\tau_{1t} | \tau_{2t}, N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) p(\tau_{2t} | N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) \end{aligned}$$

where we can sample from $p(\tau_{2t} | N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t)$ using the fact that conditionally on N_t the arrival time τ_{2t} of the N_t th jump is the maximum of N_t uniform random variables and it has a $Beta(N_t, 1)$ distribution. The arrival time of the $(N_t + 1)$ th jump after 1 is exponentially distributed with intensity $\lambda_t(z_{t1} + z_{t2})$, hence

$$\tau_{1t} = 1 + \xi_t - \tau_{2t} \quad \xi_t \sim \text{Exp}(\lambda_t(z_{t1} + z_{t2})) \quad (\text{A86})$$

Generating C (Step 4e)

Notice that

$$p(C|\tau, N, z_1, z_2, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, y) = p(C|\tau, N, z_1, z_2, \nu, s, x) \quad (\text{A87})$$

Moreover

$$p(C|\tau, N, z_1, z_2, \nu, s, x) = \prod_{t=1}^T \prod_{j=1}^{\min(N_t+1, 2)} p(r_{tj}|\tau_t, N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) \quad (\text{A88})$$

Sample c_{t1} from the following discrete distribution

$$p(c_{t1}|\tau_t, N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) \propto w_k(1)\varphi(-\log \tau_{1t} - \log[\lambda_t(z_{t1} + z_{t2})], m_k(1), v_k^2(1)) \quad (\text{A89})$$

where $k = 1, \dots, C(1)$ If $N_t > 0$ then draw r_{t2} from the discrete distribution

$$p(c_{t2}|\tau_t, N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) \propto w_k(N_t)\varphi(-\log \tau_{1t} - \log[\lambda_t(z_{t1} + z_{t2})], m_k(N_t), v_k^2(N_t))$$

for $k = 1, \dots, C(N_t)$

These algorithmic details also apply to the dynamic Skellam model. Illustrations of the resulting posterior distributions of the parameters are given in Figure 1 for our ΔNB model and in Figure 9 for the dynamic Skellam model.

E Log returns versus price changes

Stock prices can be quoted as a multiple of the tick size. As a consequence prices are defined on a discrete grid, where the grid points are a tick size distance away from each other. We can write the prices at time t_j as

$$p(t_j) = n(t_j)g \quad (\text{A90})$$

where g is the tick size which can be the function of the price on some exchanges and $n(t_j)$ is a natural number, denoting the location of the price on the grid. Modelling trade by trade returns can pose difficulty as the effect of price discreteness on a few seconds frequency is pronounced compared to lower frequencies such as one hour or one day. As described in Münnix et al. (2010), the problem is that the return distribution is a mixture of return distributions r^i , which correspond to fix price changes ig

$$r^i = \left\{ \frac{p(t_j) - p(t_{j-1})}{p(t_{j-1})} \mid p(t_j) - p(t_{j-1}) = [n(t_j) - n(t_{j-1})]g = i(t_j)g = ig \right\} \quad (\text{A91})$$

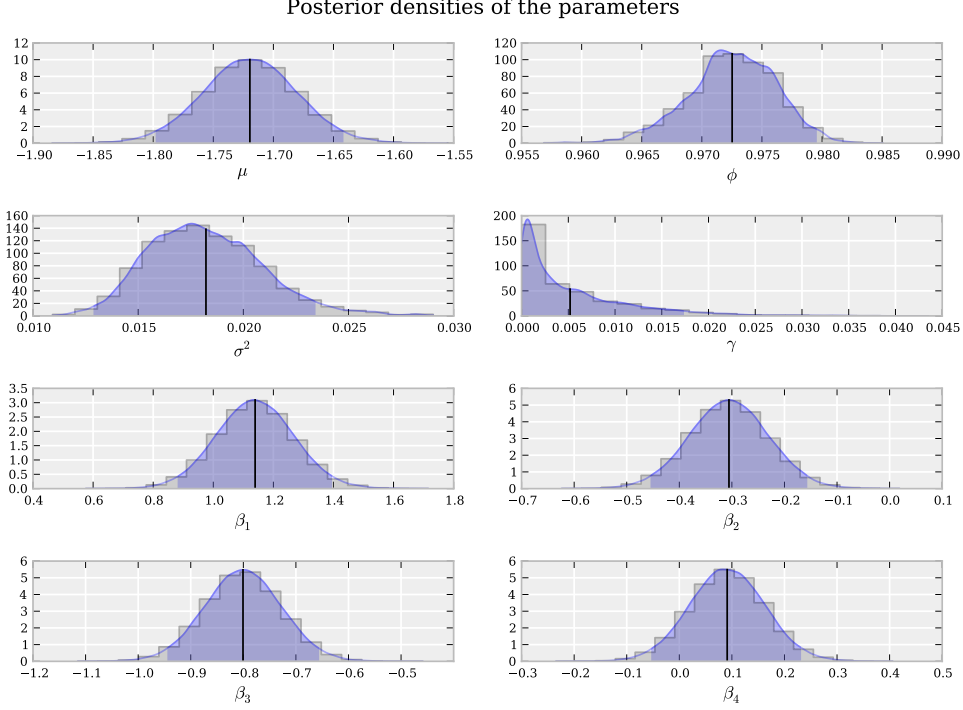


Figure 9: The posterior distribution of the parameters from a dynamic Skellam model based on 20000 observations and 100000 iterations from which 20000 used as a burn in sample. Each picture shows the histogram of the posterior draws the kernel density estimate of the posterior distribution, the HPD region and the posterior mean. The true parameters are $\mu = -1.7$, $\phi = 0.97$, $\sigma = 0.02$, $\gamma = 0.001$

where $i(t_j)$ is an integer, which express the price change in terms of ticks. The r^i distributions are on the intervals (for positive i)

$$\left[\frac{ig}{\max p^i}, \frac{ig}{\min p^i} \right], \quad (\text{A92})$$

where

$$p^i = \left\{ p(t_{j-1}) \mid p(t_j) - p(t_{j-1}) = [n(t_j) - n(t_{j-1})]g = i(t_j)g = ig \right\} \quad (\text{A93})$$

These interval and the center of the intervals c_i can be approximated by

$$\left[\frac{ig}{\bar{p}}, \underbrace{\frac{ig}{p}} \right], \quad (\text{A94})$$

and

$$c_i \approx \frac{ig}{2} \left(\underbrace{\frac{1}{p}} - \frac{1}{\bar{p}} \right) \quad (\text{A95})$$

as $\max p^i \approx \bar{p}$ and $\min p^i \approx \underbrace{p}$ for i close to 0.

First, note that the intervals corresponding to zero price change and one tick changes are always non-overlapping. Secondly, the center of the intervals are approximately

equally spaced, however the intervals for higher absolute value changes are wider, which means that the intervals are getting more and more overlapping as $|i|$ is increasing. Thirdly, the intervals are less overlapping when the price is lower, the volatility is higher or the tick size is bigger. Figure 10 shows the empirical trade by trade return distribution of several stocks from the New York Stock Exchange (NYSE).

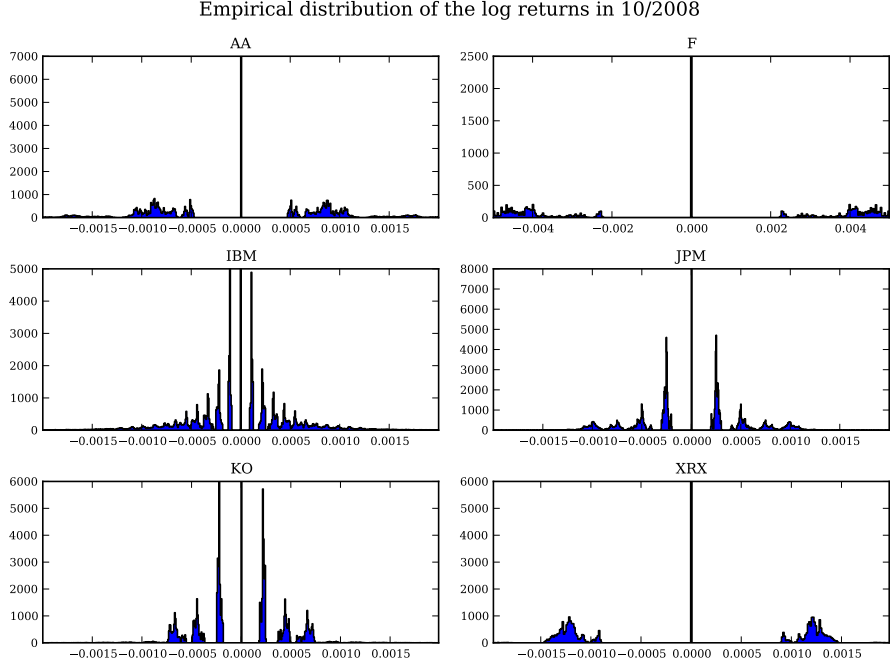


Figure 10: Empirical distribution of the tick by tick log returns during October 2008 for Alcoa (AA), Ford (F), International Business Machines (IBM),JP Morgan (JPM), Coca-Cola (KO) and Xerox (X)

Empirical distribution of the tick returns in 10/2008

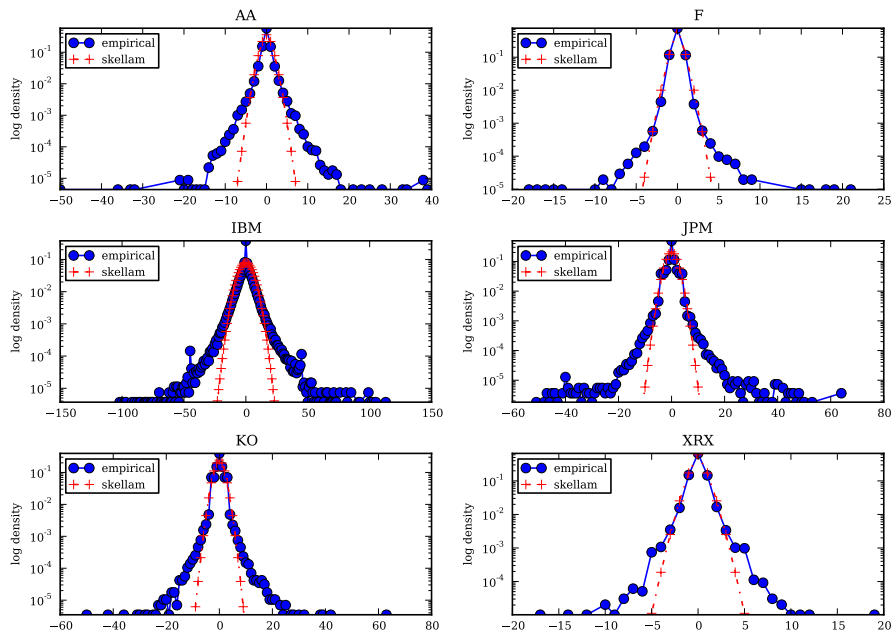


Figure 11: Empirical distribution of the tick returns along with fitted Skellam density during October 2008 for Alcoa (AA), Ford (F), International Business Machines (IBM), JP Morgan (JPM), Coca-Cola (KO) and Xerox (X)

Data cleaning

Table 10: Summary of the cleaning and aggregation procedure on the data from 3rd to 10th October 2008 for Alcoa (AA), Coca-Cola (KO) International Business Machines (IBM), J.P. Morgan (JPM), Ford (F), Xerox (XRX from the NYSE).

	AA		F		IBM		JPM		KO		XRX	
	#	% dropped	#	% dropped	#	% dropped	#	% dropped	#	% dropped	#	% dropped
Raw quotes and trades	511 185		311 914		688 805		984 526		541 616		371 065	
Trades	107 448	78.98	59 749	80.84	128 589	81.33	298 773	69.65	126 509	76.64	40 846	88.99
Non missing price and volume	107 434	0.01	59 737	0.02	128 575	0.01	298 761	0	126 497	0.01	40 834	0.03
Trades between 9:30 and 16:00	107 421	0.01	59 724	0.02	128 561	0.01	298 744	0.01	126 484	0.01	40 820	0.03
Aggrageted trades	79 623	25.88	47 146	21.06	89 517	30.37	188 469	36.91	96 482	23.72	34 722	14.94
Without outliers	79 198	0.53	47 075	0.15	88 808	0.79	186 103	1.26	95 398	1.12	34 649	0.21
Without opening trades	79 192	0.01	47 069	0.01	88 802	0.01	186 097	0	95 392	0.01	34 643	0.02

Table 11: Summary of the cleaning and aggregation procedure on the data from 23rd to 30th April 2010 for Alcoa (AA), Coca-Cola (KO) International Business Machines (IBM), J.P. Morgan (JPM), Ford (F), Xerox (XRX from the NYSE).

	AA		F		IBM		JPM		KO		XRX	
	#	% dropped	#	% dropped	#	% dropped	#	% dropped	#	% dropped	#	% dropped
Raw quotes and trades	1 487 382		2 737 300		803 648		2 109 770		692 657		1 038 502	
Trades	33 684	97.74	77 778	97.16	53 346	93.36	126 153	94.02	41 184	94.05	43 170	95.84
Non missing price and volume	33 675	0.03	77 765	0.02	53 332	0.03	126 142	0.01	41 173	0.03	43 155	0.03
Trades between 9:30 and 16:00	33 666	0.03	77 757	0.01	53 324	0.02	126 136	0	41 164	0.02	43 149	0.01
Aggrageted trades	32 446	3.62	73 160	5.91	52 406	1.72	122 579	2.82	40 573	1.44	40 673	5.74
Without outliers	32 439	0.02	73 141	0.03	52 199	0.39	122 494	0.07	40 548	0.06	40 664	0.02
Without opening trades	32 433	0.02	73 135	0.01	52 193	0.01	122 488	0	40 542	0.01	40 658	0.01