The Dynamic Skellam Model with Applications*

Siem Jan Koopman[(a,b,c)], *Rutger Lit*[(a,b)] *and* *André Lucas*[(a,b)]

[(a)]VU University Amsterdam, The Netherlands

[(b)]Tinbergen Institute Amsterdam

[(c)]CREATES, Aarhus University, Denmark

March 6, 2014

Abstract

We introduce a dynamic statistical model for Skellam distributed random variables. The Skellam distribution can be obtained by taking differences between two Poisson distributed random variables. We treat cases where observations are measured over time and where possible serial correlation is modeled via stochastically time-varying intensities of the underlying Poisson counts. The likelihood function for our model is analytically intractable and we evaluate it via a multivariate extension of numerically accelerated importance sampling techniques. We illustrate the new model by two empirical studies and verify whether our framework can adequately handle large data sets. First, we analyze long univariate high-frequency time series of U.S. stock price changes, which evolve as discrete multiples of a fixed tick size of one dollar cent. In a second illustration, we analyze the score differences between rival soccer teams using a large, unbalanced panel of seven seasons of weekly matches in the German Bundesliga. In both empirical studies, the new model provides interesting and non-trivial dynamics with a clear interpretation.

Key words: dynamic count data models; non-Gaussian multivariate time series models; importance sampling; numerical integration; volatility models; sports data.

*We thank Asger Lunde and Albert J. Menkveld for their comments on an earlier draft. Lit and Lucas acknowledge the financial support of the Dutch Science Foundation (NWO, grant VICI453-09-005).

1 Introduction

Various recent contributions have raised renewed interest in the Skellam distribution to model integer outcomes; see for example Karlis and Ntzoufras (2006), Karlis and Ntzoufras (2009) and Barndorff-Nielsen et al. (2012). The Skellam distribution can be viewed as a distribution on positive and negative integers, but can also be constructed from differences in pairs of Poisson counts; see Skellam (1946). Paired count observations and their differences appear in many situations and research fields. For example, in low-scoring sports such as ice-hockey and soccer, the score difference between the teams can be viewed as the difference between two Poisson counts and thus be modeled by a Skellam distributed random variable. In medical research, experiments for measuring the effect of treatments and drug intake lead to paired counts. A famous example is the decayed, missing and filled teeth (DMFT) index for a region that measures the effect of preventive methods in dental care; see Bohning et al. (1999) and Karlis and Ntzoufras (2006). The change of the DFMT index over time or between regions can be modeled by the Skellam distribution. Finally, in financial markets, price changes from one trade to the next can be modeled by the Skellam distribution, as prices move on a grid of integer multiples of a fixed 'tick size'.

The Skellam distribution is characterized by two 'intensity' parameters. In earlier studies, the Skellam distribution is used in a static perspective. When we analyze time series of differences in counts, we often obtain significant improvements in model fit and forecasting performance if the parameters of the Skellam distribution are allowed to vary over time. Time variation in the parameters of the Skellam distribution may capture the developments of relative team strengths over longer periods of time in sports applications, trends in health and demography in medical applications, or market circumstances and risk attitudes in economic and finance applications.

We present a novel dynamic Skellam model with stochastically time-varying intensities. We formulate the model in terms of a nonlinear non-Gaussian state space process. It is well documented that the likelihood function for this class of models is analytically intractable. Therefore we rely on numerical (possibly) simulation based methods. Many non-Gaussian state space models can be analyzed efficiently by the methods developed in Shephard and Pitt (1997) and Durbin and Koopman (1997), hereafter referred to as SPDK. A major

advantage of the SPDK method is the combination of speed and the ability to analyze high dimensional state space models using an approximating linear Gaussian state space model for the evaluation of the likelihood integral of the true likelihood via simulation.

An important alternative to SPDK is the efficient importance sampling (EIS) method of Liesenfeld and Richard (2003) and Richard and Zhang (2007). The EIS method uses a more accurate (global) approximation of the likelihood integral than the local approximation of SPDK. Koopman et al. (2013) further extend the global approximating method of Liesenfeld and Richard (2003) and Richard and Zhang (2007) by solving for the parameters of the importance sampling distribution using Gauss-Hermite quadrature rather than simulation. This results in further efficiency improvements in estimating the likelihood. They call their method numerically accelerated importance sampling (NAIS) and show that it is applicable for a high dimensional state vector and a univariate signal. Numerical integration techniques in time series have been proposed earlier by Kitagawa (1987) and Fridman and Harris (1998). However, their methods are only applicable for models with a low-dimensional state vector. In this paper we extend the NAIS methodology to a high-dimensional state with a bivariate signal by adopting bivariate Gauss-Hermite quadrature.

To study the performance of our new model and the resiliency of the associated estimation methodology based on the bivariate NAIS, we present the results of two empirical studies based on two large scale applications. Each empirical study reveals its own challenges for the developed methodology. In our first application, we consider long time series of stock price changes. These changes are recorded as integer multiples of a fixed tick size of one dollar cent, suggesting the use of the dynamic Skellam model. Long time series are known to pose particular efficiency problems for importance sampling estimation methods; see Robert and Casella (2004, §3.3) and Cappé et al. (2005, §6.1 and 9.1). We find, however, that the dynamic Skellam model can be efficiently estimated using bivariate NAIS for time series of more than 23,000 trades. In our second application, we consider score differences of soccer matches of 29 teams observed over 7 seasons of the German Bundesliga. The resulting panel data set has many missing values and is clearly high dimensional. In addition, we model the score difference for each match by a dynamic Skellam distribution with intensity parameters that vary with the attack and defense strengths of the home and away teams. Given the large number of teams in the Bundesliga, the state vector in the state space representation of the

model is also high dimensional. The combination of missing values and high dimensions poses well-known challenges to the computational feasibility of the estimation methodology. We show, however, that the dynamic Skellam model for this complex data set can be estimated successfully using NAIS in a feasible way. Several interesting extensions of the basic model are also considered.

The remainder of the paper is organized as follows. We present the new dynamic Skellam model in Section 2 and explain how it can be cast in nonlinear non-Gaussian state space form. In Section 3 we discuss parameter estimation and state smoothing using a bivariate version of numerically accelerated importance sampling. Section 4 applies the dynamic Skellam model to high-frequency U.S. stock price changes to show how our methodology can handle long time series. Section 5 treats a large unbalanced panel data set of German Bundesliga soccer matches to show how the method performs for high dimensional data sets, missing values and high dimensional state vectors. Section 6 concludes.

2 The dynamic Skellam model

2.1 Skellam distribution

The probability mass function (pmf) of a Skellam distributed random variable $Y \in \mathbb{Z}$ with parameters $\lambda_1, \lambda_2 \in \mathbb{R}^+$ is given by

$$\mathbb{P}(Y = y; \lambda_1, \lambda_2) = \exp\left(-\lambda_1 - \lambda_2\right) \left(\lambda_1/\lambda_2\right)^{y/2} I_{|y|}(2\sqrt{\lambda_1\lambda_2}), \tag{1}$$

where $I_{|y|}(\,\cdot\,)$ is the modified Bessel function of order $|y|$, see Abramowitz and Stegun (1972) for more details. Following Skellam (1946), we can derive the Skellam distribution by defining Y as the difference $C_1 - C_2$ of a bivariate Poisson count pair (C_1, C_2), see also Mardia (1970). If C_1 and C_2 are independent Poisson, λ_1 and λ_2 can be directly interpreted as the Poisson intensities for C_1 and C_2, respectively. More background information on the Skellam distribution and further references are provided by Johnson et al. (1992).

The mean and variance of Y are given by

$$\mathbb{E}(Y) = \lambda_1 - \lambda_2, \qquad \mathbb{V}\mathrm{ar}(Y) = \lambda_1 + \lambda_2. \tag{2}$$

Moreover,

$$p(Y = y; \lambda_1, \lambda_2) = p(Y = -y; \lambda_2, \lambda_1),$$

such that the Skellam distribution is symmetric for $\lambda_1 = \lambda_2$, right-skewed for $\lambda_1 > \lambda_2$, and left-skewed for $\lambda_1 < \lambda_2$. Just as for the Poisson distribution, we can also construct a zero-inflated version of the Skellam distribution; see for example Bohning et al. (1999) and Karlis and Ntzoufras (2009). This transfers probability mass from $Y \neq 0$ towards $Y = 0$ if the latter is over-represented. The zero-inflated Skellam distribution is defined by its pmf

$$p_z(Y = y; \lambda_1, \lambda_2, \gamma) = \begin{cases} (1-\gamma)\, p(Y = y; \lambda_1, \lambda_2), & \text{for } y \neq 0, \\ \gamma + (1-\gamma)\, p(Y = 0; \lambda_1, \lambda_2), & \text{for } y = 0, \end{cases} \tag{3}$$

with $\gamma \in [0, 1)$ an additional unknown and fixed parameter, and $p(y; \lambda_1, \lambda_2)$ as defined in (1). For $\gamma = 0$, we recover the original Skellam distribution. The mean and variance of the zero-inflated Skellam distribution are

$$\mathbb{E}(Y) = (1-\gamma)(\lambda_1 - \lambda_2), \qquad \mathbb{V}\text{ar}(Y) = (1-\gamma)(\lambda_1 + \lambda_2) + \gamma(1-\gamma)(\lambda_1 - \lambda_2)^2. \tag{4}$$

The inflation of probability mass to non-zero values of Y can be achieved in a similar way. In Figure 1 we present a few examples of Skellam and zero-inflated Skellam distributions. The figure shows that the distribution is highly peaked at the center for low values of λ_1 or λ_2. The effects of $\lambda_1 \neq \lambda_2$ and $\gamma \neq 0$ are also clearly visible.

2.2 Dynamic specification of intensities

In the dynamic Skellam model, we replace Y, y, λ_1, and λ_2 in (1) by their time-varying counterparts Y_t, y_t, λ_{1t}, and λ_{2t}, respectively. We denote the dynamic model as

$$Y_t \sim \text{Skellam}(\lambda_{1t}, \lambda_{2t}), \qquad t = 1, \ldots, n, \tag{5}$$

where n is the length of the time series. We assume that the serial correlation in Y_t is accounted for by the time-variation in the intensities λ_{1t} and λ_{2t}. We first treat the case of univariate observations and extend our modeling framework to the multivariate case later.

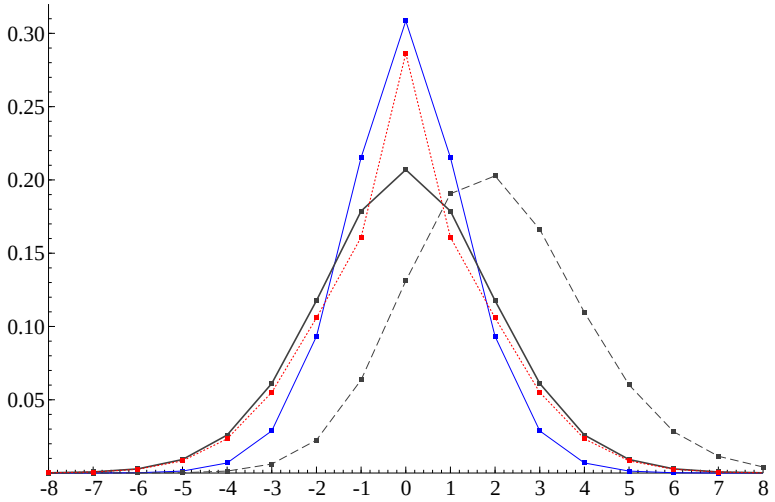

Figure 1: Skellam and zero-inflated Skellam distributions with density functions (1) and (3), respectively, for different λ_1, λ_2 and γ coefficients. These are discrete distributions, the connecting lines are drawn for clarity and do not indicate continuity. (———) $\lambda_1 = \lambda_2 = 1$; (———) $\lambda_1 = \lambda_2 = 2$; (− − −) $\lambda_1 = 3, \lambda_2 = 1$; (· · · · · ·) $\lambda_1 = \lambda_2 = 2, \gamma = 0.1$.

We model the dynamics in λ_{1t} and λ_{2t} by a nonlinear transformation of autoregressive processes,

$$\lambda_{it} = s_i(\boldsymbol{\theta}_t), \tag{6}$$

$$\boldsymbol{\theta}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_t, \tag{7}$$

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

for $i = 1, 2$ and $t = 1, \ldots, n$, where $s_i(\,\cdot\,) : \mathbb{R}^{p\times 1} \to \mathbb{R}^+$ is a nonlinear scalar function referred to as a link function, $\boldsymbol{\theta}_t \in \mathbb{R}^{p\times 1}$ is called the signal vector, $\boldsymbol{\alpha}_t \in \mathbb{R}^{m\times 1}$ is referred to as the state vector, $\boldsymbol{c}_t \in \mathbb{R}^{p\times 1}$ and $\boldsymbol{d}_t \in \mathbb{R}^{m\times 1}$ are vectors of intercepts, $\boldsymbol{Z}_t \in \mathbb{R}^{p\times m}$ is a matrix of zeroes, ones and possibly some unknown coefficients, $\boldsymbol{T}_t \in \mathbb{R}^{m\times m}$ is a transition matrix, and the disturbances $\boldsymbol{\eta}_t$ are normally and independently distributed with mean zero and variance matrix $\boldsymbol{Q}_t \in \mathbb{R}^{m\times m}$. The vectors $\boldsymbol{c}_t$, $\boldsymbol{d}_t$ and the matrices $\boldsymbol{Z}_t$, $\boldsymbol{T}_t$ and $\boldsymbol{Q}_t$ are possibly time-varying but in a deterministic manner. Typical examples of link functions $s_i(\,\cdot\,)$ are the exponential function (to ensure positive support for the intensities) and the scaled logistic function (to preserve a lower and upper bound for the intensities). When the link function $s_i(\,\cdot\,)$ directly requires the state vector $\boldsymbol{\alpha}_t$ as argument, we simply set $p = m$, $\boldsymbol{c}_t = \boldsymbol{0}$, and

$\boldsymbol{Z}_t = \mathrm{I}_{\boldsymbol{m}}$. The initial conditions for the elements of the state vector $\boldsymbol{\alpha}_1$ depend on their dynamic properties. The variance matrix $\boldsymbol{Q}_t$ is possibly positive semi-definite and hence the vector $\boldsymbol{\eta}_t$ may contain zeroes.

The model specified in equations (6)–(8) allows for a wide variety of dynamic patterns for the intensities λ_{1t} and λ_{2t}, including autoregressive moving average dynamics, time varying seasonal and cyclical patterns, deterministic and stochastic trends, and their combinations. Regression and intervention effects can be incorporated as well; see Durbin and Koopman (2012, Chapter 3) for more details and more illustrations.

The dynamic Skellam model as specified above falls within the class of non-Gaussian nonlinear state space models

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

with $\boldsymbol{\alpha}_1 \sim p_g(\boldsymbol{\alpha}_1;\boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is an unknown and fixed parameter vector gathering all the parameters in $\boldsymbol{c}_t$, $\boldsymbol{Z}_t$, $\boldsymbol{d}_t$, $\boldsymbol{T}_t$, and $\boldsymbol{Q}_t$, and possibly in the signal functions $s_i(\,\cdot\,)$ for $i = 1, 2$. The observation density $p(y_t|\boldsymbol{\theta}_t;\boldsymbol{\psi})$ refers to the dynamic (possibly zero inflated) Skellam distribution from Section 2.1 with parameters λ_{1t} and λ_{2t}, the updating Gaussian state density $p_g(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t;\boldsymbol{\psi})$ refers to the linear Markov process (8), and $p_g(\boldsymbol{\alpha}_1;\boldsymbol{\psi})$ represents the initial condition for $\boldsymbol{\alpha}_1$. We assume that for given realisations of the signal $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_n')$ the observations $\boldsymbol{y} = (y_1, \ldots, y_n)'$ are conditionally independent, and also write $\boldsymbol{\theta} = \boldsymbol{c} + \boldsymbol{Z}\boldsymbol{\alpha}$ with $\boldsymbol{c}' = (\boldsymbol{c}_1', \ldots, \boldsymbol{c}_n')$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_n')'$, and $\boldsymbol{Z}$ a block-diagonal matrix with blocks $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ on the leading diagonal. The joint conditional density for all observations and the marginal density for all states can now be written as

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

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

The state space perspective in equations (9)–(10) for the dynamic Skellam model allows us to build on a well developed framework for parameter estimation and inference and state filtering and smoothing in such models; see for example Durbin and Koopman (2012).

As for other non-Gaussian nonlinear state space models, the main complication for the dynamic Skellam model is that the likelihood function $\int p(\boldsymbol{y}|\boldsymbol{\theta};\boldsymbol{\psi})p_g(\boldsymbol{\alpha};\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\alpha}$ is analytically intractable. We therefore adopt the method of simulated maximum likelihood estimation for parameter estimation. In particular, we develop a bivariate extension of the numerically accelerated importance sampling (NAIS) method of Koopman et al. (2013) and show that it can efficiently handle long univariate time series (large n) as well as high dimensional unbalanced panel time series models for Skellam distributed random variables. We first present the details of our estimation methods in Section 3. Then we proceed with our empirical applications in Sections 4 and 5.

3 Monte Carlo likelihood and parameter estimation

Before presenting our bivariate extension of the numerically accelerated importance sampling techniques (NAIS) of Koopman et al. (2013), we first present a short review of Monte Carlo likelihood evaluation for state space models and the efficient importance sampling (EIS) technique of Liesenfeld and Richard (2003) and Richard and Zhang (2007) to establish the necessary notation.

3.1 Likelihood evaluation and importance sampling

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

$$L(\boldsymbol{y};\boldsymbol{\psi}) = \int p(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\theta} = \int p(\boldsymbol{y}|\boldsymbol{\theta};\boldsymbol{\psi})p_g(\boldsymbol{\theta};\boldsymbol{\psi})\,\mathrm{d}\boldsymbol{\theta}. \tag{11}$$

An analytical expression is not available for this high dimensional integral. In cases where the model is linear and Gaussian, the Kalman filter can be used for likelihood evaluation, signal extraction and forecasting. Here we rely on numerical integration techniques that need to be both practical and feasible.

It is well established that we can use Monte Carlo simulation methods for the evaluation of (11); see Ripley (1987) for a general introduction. In particular, in various contributions in statistics and econometrics it is argued that (11) can be evaluated efficiently using the method of importance sampling; see, for example, Shephard and Pitt (1997), Durbin and

Koopman (1997), Liesenfeld and Richard (2003) and Richard and Zhang (2007). For a feasible implementation of this method we require a Gaussian importance density $g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)$ from which the $\boldsymbol{\theta}$s are sampled conditional on the observation vector $\boldsymbol{y}$, where $\boldsymbol{\psi}^*$ denotes a fixed parameter vector, containing $\boldsymbol{\psi}$ as well as parameters $\tilde{\boldsymbol{\psi}}$ particular to the importance density $g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})$, i.e., $\boldsymbol{\psi}^* = (\boldsymbol{\psi}', \tilde{\boldsymbol{\psi}}')'$. Under the assumption that a numerically efficient device can be developed for sampling $\boldsymbol{\theta}$ from $g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)$, we can express the likelihood function (11) in terms of the importance density as

$$L(\boldsymbol{y};\boldsymbol{\psi}) = \int \frac{p(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)} g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)\,\mathrm{d}\boldsymbol{\theta}, \tag{12}$$

with the importance sampling estimate given by

$$\frac{1}{S}\sum_{k=1}^{S} \omega(\boldsymbol{y},\boldsymbol{\theta}^{(k)};\boldsymbol{\psi}^*), \qquad \omega(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi}^*) = \frac{p(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)}, \qquad \boldsymbol{\theta}^{(k)} \sim g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*), \tag{13}$$

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

$$\omega(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi}^*) = \frac{p(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta};\boldsymbol{\psi})p_g(\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})g(\boldsymbol{\theta};\boldsymbol{\psi})/g(\boldsymbol{y};\boldsymbol{\psi}^*)} = g(\boldsymbol{y};\boldsymbol{\psi}^*)\,\frac{p(\boldsymbol{y}|\boldsymbol{\theta};\boldsymbol{\psi})}{g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})}, \tag{14}$$

see for example Durbin and Koopman (2012). The density $g(\boldsymbol{y};\boldsymbol{\psi}^*)$ can be taken as a scaling function since it does not depend on $\boldsymbol{\theta}$. The function $\omega(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi}^*)$ is usually referred to as the importance sampling weight function. If the variance of $\omega(\boldsymbol{y},\boldsymbol{\theta};\boldsymbol{\psi}^*)$ exists, the estimate (13) is consistent for any $g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})$ and a central limit theorem applies; see Geweke (1989) and Koopman et al. (2009). We may expect that a well-behaved weight function leads to an efficient importance sampling estimate of the likelihood function.

3.2 Construction of the importance density

The key choice in selecting an importance density $g(\boldsymbol{\theta}|\boldsymbol{y};\boldsymbol{\psi}^*)$ is numerical efficiency. We follow the predominant choice in the literature and opt for the Gaussian density; we construct

$g(\cdot)$ efficiently using standard techniques such as regression analysis and the Kalman filter.

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

$$\arg\min_{\tilde{\boldsymbol{\psi}}_t} \int \tilde{\lambda}^2(y_t,\boldsymbol{\theta}_t;\boldsymbol{\psi}^*)\,\omega_t(y_t,\boldsymbol{\theta}_t;\boldsymbol{\psi}^*)\,g(\boldsymbol{\theta}_t|\boldsymbol{y};\boldsymbol{\psi}^*)\,\mathrm{d}\boldsymbol{\theta}_t, \tag{15}$$

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

$$\tilde{\lambda}(y_t,\boldsymbol{\theta}_t;\boldsymbol{\psi}) := \log\,\omega_t(y_t,\boldsymbol{\theta}_t;\boldsymbol{\psi}^*) := \log\,p(y_t|\boldsymbol{\theta}_t;\boldsymbol{\psi}) - \log\,g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t). \tag{16}$$

The importance density is effectively determined by the minimization of the variance of the log weight ω_t, for each t. Richard and Zhang (2007) evaluate the integral in (15) using importance sampling and perform its minimization via weighted least squares regression.

In a further development of EIS, Koopman et al. (2013) replace the evaluation of the integral in (15) by standard Gauss-Hermite quadrature methods. This results in a highly numerically efficient importance sampling technique, that can be augmented with easy-to-compute control variates to increase efficiency even further. They label their method numerically accelerated importance sampling (NAIS). The key to NAIS is the availability of analytic expressions for the marginal densities $g(\boldsymbol{\theta}_t|\boldsymbol{y};\boldsymbol{\psi}^*)$ given the Gaussian importance densities $g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})$ and a Gaussian marginal density $g(\boldsymbol{\theta};\boldsymbol{\psi}) = p_g(\boldsymbol{\theta};\boldsymbol{\psi})$. Although NAIS was originally developed for a univariate signal $\theta_t \in \mathbb{R}$, the method can easily be extended to multiple dimensions; see Scharth (2012, Chapter 5) and the discussions in Koopman et al. (2013). Scharth (2012) proposes Halton sequences and quasi-Monte Carlo integration for the evaluation of high dimensional integrals. In the case of our dynamic Skellam model, the signal is only two-dimensional and hence we can still rely on Gauss-Hermite quadrature methods efficiently. We explore this further below.

3.3 Bivariate numerically accelerated importance sampling

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

$$g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}}) = \prod_{t=1}^{n} g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t), \qquad g(y_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t) = \exp\left(a_t + \boldsymbol{b}_t'\boldsymbol{\theta}_t - \frac{1}{2}\boldsymbol{\theta}_t'\boldsymbol{C}_t\boldsymbol{\theta}_t\right), \tag{17}$$

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

$$\arg\min_{\tilde{\boldsymbol{\psi}}_t^{(k+1)}} \int\int \tilde{\lambda}^2(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^{*(k+1)})\omega_t(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^{*(k)})g(\boldsymbol{\theta}_t|\boldsymbol{y}; \boldsymbol{\psi}^{*(k)})\, d\theta_{1t}\, d\theta_{2t}, \tag{18}$$

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

$$g(\boldsymbol{\theta}_t|\boldsymbol{y};\boldsymbol{\psi}^{*(k)}) = \mathrm{N}(\widehat{\boldsymbol{\theta}}_t^{(k)}, \boldsymbol{V}_t^{(k)}) = \frac{1}{2\pi|\boldsymbol{V}_t^{(k)}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_t^{(k)})'(\boldsymbol{V}_t^{(k)})^{-1}(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_t^{(k)})\right), \tag{19}$$

where $\widehat{\boldsymbol{\theta}}_t^{(k)}$ and $\boldsymbol{V}_t^{(k)}$ are obtained from the Kalman filter and smoother, for given $\boldsymbol{\psi}^* = \boldsymbol{\psi}^{*(k)}$, applied to the linear Gaussian model $\boldsymbol{x}_t = \boldsymbol{\theta}_t + \boldsymbol{u}_t$ with disturbance $\boldsymbol{u}_t \sim \mathrm{N}(0, \boldsymbol{C}_t^{-1})$ and pseudo-observation $\boldsymbol{x}_t = \boldsymbol{C}_t^{-1}\boldsymbol{b}_t$, for $t = 1, \ldots, n$. It is straightforward to verify that the observation density $\prod_{t=1}^n g(\boldsymbol{x}_t|\boldsymbol{\theta}_t;\tilde{\boldsymbol{\psi}}_t)$ is equivalent to $g(\boldsymbol{y}|\boldsymbol{\theta};\tilde{\boldsymbol{\psi}})$ in (17).

We numerically implement the minimization in (18) by the Gauss-Hermite quadrature method; see, for example, Monahan (2001). For this purpose we define

$$\varphi(y_t, \boldsymbol{\theta}_t; \tilde{\boldsymbol{\psi}}_t^{(k+1)}, \boldsymbol{\psi}^{*(k)}) = \tilde{\lambda}^2(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^{*(k+1)})\omega_t(y_t, \boldsymbol{\theta}_t; \boldsymbol{\psi}^{*(k)}), \tag{20}$$

and we select a set of abscissae $\{z_i\}_{i=1}^M$ with associated Gauss-Hermite weights $h(z_i)$, for

$i = 1, \ldots M$. The numerical implementation of the minimization (18) becomes

$$\arg \min_{\tilde{\boldsymbol{\psi}}_t^{(k+1)}} \sum_{i=1}^{M} \sum_{j=1}^{M} w_{ij} \cdot \varphi(y_t, \tilde{\boldsymbol{z}}_{ij,t}^{(k)}; \tilde{\boldsymbol{\psi}}_t^{(k+1)}, \boldsymbol{\psi}^{*(k)}), \tag{21}$$

with weight $w_{ij} = h(z_i)h(z_j)\exp(\frac{1}{2}z_i^2)\exp(\frac{1}{2}z_j^2)$ and $\tilde{\boldsymbol{z}}_{ij,t}^{(k)} = \widehat{\boldsymbol{\theta}}_t + \boldsymbol{F}_t^{(k)} \boldsymbol{z}_{ij}$, where the 2×2 square root matrix $\boldsymbol{F}_t^{(k)}$ is the result of the decomposition $\boldsymbol{V}_t^{(k)} = \boldsymbol{F}_t^{(k)} \boldsymbol{F}_t^{(k)\prime}$ and $\boldsymbol{z}_{ij} = (z_i\,,\, z_j)'$ for $i, j = 1, \ldots, M$. In this implementation we have used the fact that $g(\tilde{\boldsymbol{z}}_{ij,t}^{(k)}|\boldsymbol{y}; \boldsymbol{\psi}^{*(k)}) \propto \exp(-\frac{1}{2}\boldsymbol{z}_{ij}'\boldsymbol{z}_{ij})$; see Koopman et al. (2013) and Scharth (2012, Chapter 5). The decomposition of $\boldsymbol{V}_t^{(k)}$ is needed because the joint set of M^2 abscissae $\boldsymbol{z}_{ij}$, for $i, j = 1, \ldots, M$, is associated with the bivariate standard normal distribution.

We can express the minimization problem (21) as a standard weighted least squares computation applied to M^2 observations for the regression equation

$$\log p(y_t|\tilde{\boldsymbol{z}}_{ij,t}^{(k)}) \quad = \quad \text{constant} + \boldsymbol{\kappa}'\tilde{\boldsymbol{z}}_{ij,t}^{(k)} - \frac{1}{2}\boldsymbol{\xi}'\text{vech}(\tilde{\boldsymbol{z}}_{ij,t}^{(k)}\tilde{\boldsymbol{z}}_{ij,t}^{(k)\prime}) + \text{error}, \tag{22}$$

where $\boldsymbol{\kappa}$ and $\boldsymbol{\xi}$ are regression coefficient vectors and the regression weights are given by $w_{ij} \cdot \omega_t(y_t, \tilde{\boldsymbol{z}}_{ij,t}^{(k)}; \boldsymbol{\psi}^{*(k)}) \cdot g(\tilde{\boldsymbol{z}}_{ij,t}^{(k)}|\boldsymbol{y}; \boldsymbol{\psi}^{*(k)})$, and where vech$(\cdot)$ stacks elements of the upper triangular part of a symmetric matrix into a vector. The resulting weighted least squares estimates for $\boldsymbol{\kappa}$ and $\boldsymbol{\xi}$ yield the new values for $\boldsymbol{b}_t^{(k+1)}$ and vech$(\boldsymbol{C}_t^{(k+1)})$, respectively. Hence, new values for $\tilde{\boldsymbol{\psi}}_t^{(k+1)}$ are obtained for each $t = 1, \ldots, n$. Using these new estimates, we can determine a new $g(\boldsymbol{\theta}_t|\boldsymbol{y}; \boldsymbol{\psi}^{*(k+1)})$ in (19) by constructing a new time series $\boldsymbol{x}_t$ and applying the Kalman filter and smoother to the linear Gaussian model given below (19). In this last step we obtain new values for $\widehat{\boldsymbol{\theta}}_t^{(k+1)}$ and $\boldsymbol{V}_t^{(k+1)}$, which we require in (19).

This procedure is iterated until convergence. Typically, we only need a small (< 10) number of iterations for the applications in this paper. For the case of a univariate signal, Richard and Zhang (2007) and Koopman et al. (2013) argue that an unweighted least squares regression can be performed without significant loss of efficiency. In our bivariate case, however, we have found that the omission of the weights resulted in an efficiency loss of about 10%. Therefore, we include the weights for the computation of $\tilde{\boldsymbol{\psi}}^{(k+1)}$. Finally, we emphasize that the regression computations can be carried out in parallel over t, leading to a highly efficient implementation.

4 Analysis of high-frequency Skellam returns

High-frequency (intra-day) analysis of prices changes in financial markets has received much attention in the recent literature; see, amongst others, Andersen (2000) and Tsay (2005). Analyzing the dynamics of prices at high frequencies is important for providing insights into market (in)stabilities in view of today's algorithmic, computer-based trading strategies and for developing appropriate regulations and fee structures for algorithmic traders, if needed; see, for example, Hendershott et al. (2011), Kervel (2012) and Chaboud et al. (2013).

We study the high-frequency dynamics of U.S. stock price changes listed at the New York Stock Exchange (NYSE) using our new dynamic Skellam model. The Skellam distribution is appropriate because at very high frequencies stock prices evolve as integer multiples of a fixed tick size. The tick size is defined as the smallest permitted price increment for a traded financial instrument. At the NYSE, the tick size is 1 dollar cent ($0.01), but the tick size may vary across exchanges. Alzaid and Omair (2010) analyze data from the Saudi Stock Exchange at a one-minute frequency using a static (rather than dynamic) Skellam model. Barndorff-Nielsen et al. (2012) study Lévy processes with static Skellam marks as the basis of price processes for high-frequency data.

The dimension of high-frequency data sets can easily become very large; it is no exception that more than $20,000$ trades occur per day for a liquid stock. We therefore have a setting which is extremely suited for testing the resiliency of our dynamic Skellam model and the corresponding NAIS based estimation methodology for these long samples.

Rather than aggregating the data to one-minute or five-minute intervals, we analyze stock price changes on a trade-by-trade basis within a single working day. Let τ_t denote the calendar time stamp for the tth trade (in hundredth of a second), and define y_t as the stock price change between τ_{t-1} and τ_t. We define the dynamic Skellam model as

$$
\begin{aligned}
&y_t|\boldsymbol{\theta}_t \sim \text{Skellam}(\lambda_{1t}, \lambda_{2t}), \qquad \lambda_{it} = \exp(\theta_{it}), \qquad i = 1, 2, \\
&\boldsymbol{\theta}_{t+1} = \bar{\boldsymbol{\theta}} + \boldsymbol{\phi}\boldsymbol{\theta}_t + \boldsymbol{\epsilon_t}, \qquad \boldsymbol{\epsilon_t} \sim \text{NID}\left(\mathbf{0}, \boldsymbol{Q}_\epsilon\right), \qquad \boldsymbol{\theta}_t = (\theta_{1,t}, \theta_{2,t})',
\end{aligned}
$$

where $\bar{\boldsymbol{\theta}}$ is a 2×1 vector of constants, $\boldsymbol{\phi}$ a 2×2 is a diagonal matrix with elements ϕ_1 and ϕ_2 and $\boldsymbol{\epsilon_t}$ is normally and independently distributed with mean zero and variance matrix $\boldsymbol{Q}_\epsilon$.

Table 1: Descriptive statistics, sample May 18, 2010. The table reports data characteristics of tick changes for eight selected stocks traded on NYSE during May 18, 2010, between 9:35am and 4:00pm. We report the "opening price" at 9:35 am (OP), the number of trades (#Trades), the percentage of zero price changes in the time series (%0), the sample mean (Mean) and variance (Var.), and the largest up tick (Max) and down tick (Min).

Code	Company	Sector	OP	#Trades	%0	Mean	Var.	Max	Min
JPM	JPMorgan Chase	Financials	40.21	23,359	55.26	-0.005	1.336	7	-8
IBM	IBM	IT	131.41	15,505	29.57	-0.007	6.896	16	-18
CAT	Caterpillar Inc.	Industrials	65.09	11,516	31.17	-0.015	4.282	12	-12
VZ	Verizon Com.	Telecom.	28.81	8,278	74.05	-0.004	0.336	3	-3
KO	Coca-Cola Company	Consumer	53.51	7,190	53.74	-0.004	1.001	6	-6
PFE	Pfizer Inc.	Health Care	16.10	7,141	79.96	-0.003	0.229	3	-3
LMT	Lockheed Martin	Aerospace	82.20	4,929	30.13	-0.024	4.386	10	-15
RDSa	Royal Dutch Shell A	Oil & Gas	54.35	3,727	33.97	-0.038	3.532	9	-9

4.1 Data

We obtain the data from the Thomson Reuters Tick History Database. We download the prices for eight different stocks traded on the New York Stock Exchange (NYSE) on May 18, 2010. The stocks come from a range of different industries and have different trading intensity levels. We provide descriptive statistics of the data in Table 1. It is clear that even though we restrict our attention to one day, the time series dimension of these eight series is considerable, with the number of trades ranging from a minimum of 3,727 to a maximum of 23,359. We notice that the mean price change is negative for all stocks; it implies that all stocks lost value on this particular day.

A number of data features is worth noting. First, most of our stocks have many trades without a price change. The percentage ranges from 29% for IBM to 80% for Pfizer Inc. We therefore consider the zero inflated dynamic Skellam model for these series.

Second, given data availability, we analyze tick-by-tick data without the "odd-lot". The odd-lot are trades with volumes less than 100 and are not recorded on the consolidated tape; see the discussion in O'Hara et al. (2012).

Third, tick data require pre-processing. A review of high frequency data cleaning procedures is, for example, provided by Falkenberry (2002). We apply the data cleaning algorithm proposed by Brownlees and Gallo (2006) with some adjustments as described in Barndorff-Nielsen et al. (2008). Before we start the data cleaning algorithm, we apply a rudimentary filter with the following steps:

(1) The NYSE is open from 9:30 am to 4:00 pm. We follow the standard practice of discarding the trades in the first 5 minutes of trading. Hence we only consider the trades from 9:35 am to 4:00 pm.

(2) All entries with transaction prices which are missing or equal to zero are removed.

(3) If multiple transactions have the same time stamp, we use the volume weighted average of the transaction prices and round it to the nearest tick size.

Table 1 further shows that the variances and price ranges (Max − Min) vary across the different stocks. For example, Verizon and Pfizer Inc. show very little volatility with a price change range of 6 ticks, while IBM is much more volatile with a range as high as 34 ticks.

4.2 Estimation results

We estimate the autoregressive constant $\bar{\boldsymbol{\theta}} = (\bar{\theta}, \bar{\theta})'$ which we assume to be equal for both signals, the autoregressive parameters on the diagonal of $\boldsymbol{\phi}$, the variances on the diagonal of $\boldsymbol{Q}_\epsilon$ and γ as introduced in equation (3). This gives us the following parameter vector,

$$\boldsymbol{\psi} = \left(\bar{\theta}, \phi_1, \phi_2, \sigma^2_{1\epsilon}, \sigma^2_{2\epsilon}, \gamma\right). \tag{23}$$

We maximize the likelihood function (11) in logs and as computed via the NAIS algorithm numerically using a quasi-Newton optimization scheme based on the numerically evaluated score function. The Gauss-Hermite polynomial in NAIS is based on $M^2 = 100$ abscissae points. The actual computation of the likelihood integral is based on $S = 100$ simulations from the importance sampling density. To ensure a smooth likelihood function, its evaluation for each possible value of the parameter vector is started with the same seed of the random number generator.

The maximum likelihood estimates of parameter vector $\boldsymbol{\psi}$ in (23) are presented in Table 2, together with some related statistics, for each selected stock. We discuss a couple of noteworthy features from Table 2.

First, the variance of the importance sampling weights is very small for such long time series. The longest series consists of 23,359 observations which implies that we have 46,718 latent factors (given that the state dimension is two). We are confident that our proposed

Table 2: Estimates of parameter vector $\boldsymbol{\psi}$: high-frequency Skellam returns. The table presents the maximum likelihood estimates of the parameter vector $\boldsymbol{\psi}$ and some related statistics. The parameter estimates are based on likelihood evaluations with $M = 10$ and $S = 100$. All estimates are significant at the 99% confidence level. The sample mean (Mean ω) and sample variance (Var ω) of the $S = 100$ generated importance weights used for the evaluation of the maximized likelihood estimate are reported.

Company	$\bar{\theta} \times 10^3$	ϕ_1	ϕ_2	$\sigma_{1\epsilon}$	$\sigma_{2\epsilon}$	γ	Mean ω	Var ω
JPMorgan Chase	-0.2166	0.9980	0.9980	0.0193	0.0199	0.3255	1.1502	0.4008
IBM	1.0471	0.9992	0.9992	0.0086	0.0086	0.1734	1.0525	0.1204
Caterpillar Inc.	1.5970	0.9980	0.9980	0.0123	0.0130	0.1464	1.0777	0.1827
Verizon Com.	-3.5298	0.9981	0.9980	0.0172	0.0153	0.0480	1.0240	0.0511
Coca-Cola Company	-0.7299	0.9988	0.9987	0.0195	0.0206	0.1641	1.0389	0.0836
Pfizer Inc.	-3.1338	0.9986	0.9985	0.0042	0.0207	0.0012	1.0150	0.0311
Lockheed Martin	0.8542	0.9991	0.9991	0.0137	0.0141	0.1357	1.0340	0.0753
Royal Dutch Shell A	0.9452	0.9985	0.9985	0.0121	0.0120	0.1647	1.0135	0.0285

methodology can handle even longer time series given the low variances of the considered series.

Second, although the maximum likelihood estimates of ϕ_1 and ϕ_2 as well as $\sigma^2_{1\epsilon}$ and $\sigma^2_{2\epsilon}$ are often very close to each other, they cannot be set equal ex-ante. For example, for the case of Pfizer Inc the estimates of $\sigma^2_{1\epsilon}$ and $\sigma^2_{2\epsilon}$ are clearly different.

Third, the estimates of γ reveal that the Skellam returns for our selected stocks have an overrepresentation of zeroes with respect to the Skellam distribution. Amongst our selected stocks, the density mass for the JP Morgan Chase stock has been moved to zero most severely. All estimates of γ are highly significant, at the 99% confidence level, except for the Verizon Com and Pfizer Inc stock. We notice that although these stocks have the largest percentage of zero trades, their return series varied between 3 ticks up and 3 ticks down during this day. This small price range resulted in relatively low intensities λ_1 and λ_2 and therefore no transfer of probability mass towards zero was needed.

4.3 Performance of NAIS

The estimation and signal extraction results partly rely on the bivariate importance sampling method NAIS. We first scrutinize the sample mean and variance of the $S = 100$ generated weights used for computing the maximized log likelihood value; they are reported in Table 2. To verify formally whether the Monte Carlo error is sufficiently small, we investigate whether a central limit theorem (CLT) applies to the importance sampling estimate of the likelihood

Table 3: Log likelihood values for a range of choices of M and S. The table presents the absolute differences in the log likelihood values, which are evaluated with different values for the number of abscissae M and the number of importance sampling simulations S, with respect to $\text{LL}^* = -45,215.935$ which is the maximized log likelihood value evaluated with $M = 30$ and $S = 10,000$. The log likelihoods are computed for the Caterpillar Inc. series and for the model with the parameter estimates reported in Table 2, that is $\boldsymbol{\psi} = (0.001597, 0.9980, 0.9980, 0.0123, 0.0130, 0.1464)$.

$S\backslash$M	6	8	10	12	14	16	18	20	30
20	0.007	0.011	0.012	0.012	0.012	0.012	0.012	0.011	0.012
50	0.010	0.007	0.006	0.006	0.006	0.006	0.006	0.006	0.006
100	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
200	0.004	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
500	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
1000	0.006	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
10000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	LL^*

function. Geweke (1989) argues that when the variance of the importance density exists, the CLT applies. We cannot determine the variance of the importance density analytically. Therefore, we use the t-test procedure from Koopman et al. (2009) by generating 100,000 realizations from the importance density and estimating the shape parameter of a generalized Pareto distribution by maximum likelihood based on the largest 100 realizations. For all stocks, the t-test statistics are negative and hence they provide strong evidence that a CLT can be applied safely. We conclude that all estimation results are reliable.

To verify the reliability of our results further, we consider the Caterpillar stock in more detail. The benchmark estimation results reported in Table 3 use $M = 10$ and $S = 100$. To verify how much the actual maximized log likelihood changes for different values of M and S, we report the absolute numerical differences with respect to the log likelihood value LL^* that is computed with $M = 30$ and $S = 10,000$ and which we regard as the "true" maximized log likelihood value. We report the value $|\text{LL}_{m,s} - \text{LL}^*|$ where $\text{LL}_{m,s}$ is the log likelihood value computed for $M = m$ and $S = s$. All reported values are small, except if particularly S becomes very low. We therefore regard $M = 10$ and $S = 100$ as a reasonable choice. For the other stocks, we obtain similar results. Also, all parameter estimates remain stable if we use different values of M and S.

4.4 Signal extraction

Figure 2 presents the estimates of the volatility $\sqrt{\lambda_{1,t} + \lambda_{2,t}}$ together with its two standard error confidence interval. The signal extraction methods of Durbin and Koopman (2000) are implemented using the same importance density that is used for the evaluation of the maximized log likelihood values in Table 2 and that is based on $M = 10$ and $S = 100$.

The estimates of the intra-day volatility from the Skellam returns differ substantially between stocks. Verizon, Pfizer, Lockheed Martin, and Royal Dutch Shell have rather smooth volatility processes within the day. For Caterpillar and IBM, the intra-day volatility process is less persistent and varies more throughout the day. The valatility of JPMorgan Chase shows most more noise around a slowly moving trend. In all cases we can observe that changes in the volatility pattern are reflected accurately by the ranges in which the returns vary locally.

5 Analyzing soccer scores

As our second application, we consider score differences for soccer matches in the German Bundesliga. The number of goals per match in a soccer game is typically low, such that the score balance can easily be viewed as a difference of two Poisson count variables, see Section 2. Let $C_{1,ijt}$ and $C_{2,ijt}$ denote the number of goals scored by the home team i and the visiting team j in week t, respectively, in a match of team i versus j. Our dependent variable is the score balance $y_{ij,t} = C_{1,ijt} - C_{2,ijt}$, which determines whether the match is won or lost, or ends in a tie. We assume $y_{ij,t}$ has a Skellam distribution.

Our data consists of weekly match results for 7 seasons of the German Bundesliga for the period from 2006/07 to 2012/13. The number of teams active in the Bundesliga during one season is 18. Each week, 9 matches are played and the total season consists of 34 weeks. Due to team promotions and relegations, we have $J = 29$ teams in total that have played in the Bundesliga for at least one season during the sample period. The total sample thus consists of an unbalanced panel over 238 weeks for 29 teams and 2142 team pairs (i, j).

Since we model the match outcomes in the Bundesliga over a prolonged period, team performance and the ability to score goals may vary over the sample, possibly due to changes

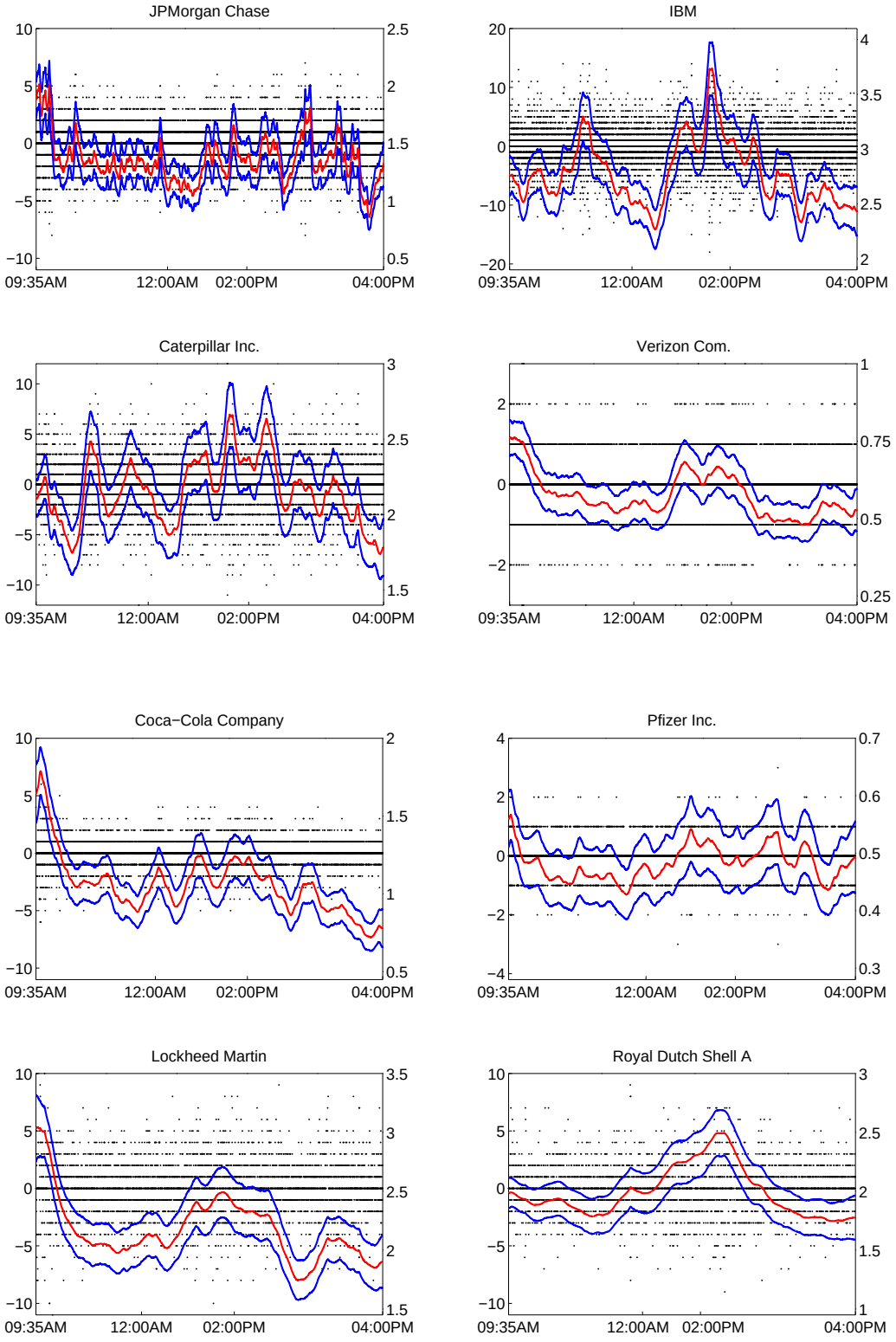

Figure 2: Extracted volatility with confidence intervals. We present the extracted volatility $\sqrt{\lambda_{1,t} + \lambda_{2,t}}$ with 95% confidence intervals of the tick returns of the investigated stocks. The dots are the tick returns with the magnitude of the tick indicated by the y-axis on the left. The magnitude of the volatility is denoted by the right y-axis. The number of trades are on the x-axis.

in the composition and management of the teams. We can handle this directly using our dynamic Skellam model. The current data set allows us to investigate the performance of our model and the associated NAIS estimation methodology for large unbalanced panels with many missing observations. Our state space modeling framework turns out to be well suited for the analysis of such data.

We extend our dynamic Skellam model to a panel setting and specify the model as

$$y_{ij,t} \sim \text{Skellam}(\lambda_{1,ijt}, \lambda_{2,ijt}), \qquad i \neq j, \quad i, j = 1, \ldots, J, \quad t = 1, \ldots, n,$$

where $\lambda_{1,ij,t}$ and $\lambda_{2,ij,t}$ are the intensities of scoring goals for the home and away teams, respectively, during a match played in week t. Team i is likely to win on its home ground from team j if $\lambda_{1,ij,t} > \lambda_{2,ij,t}$. We assume that these intensities depend on the attack (μ_{it} and μ_{jt}) and defense (β_{it} and β_{jt}) strengths of both teams in week t. We assume a fixed time-invariant home ground advantage ζ for all teams and model the scoring intensities as

$$\lambda_{1,ij,t} = \exp(\theta_{1,ij,t}) = \exp\big(\zeta + \mu_{it} - \beta_{jt}\big), \qquad \lambda_{2,ij,t} = \exp(\theta_{2,ij,t}) = \exp\big(\mu_{jt} - \beta_{it}\big). \tag{24}$$

This parsimonious modeling framework for league match results using attack and defense strengths is based on Maher (1982) and provides our benchmark model. The generalization towards a dynamic bivariate Poisson model is developed by Koopman and Lit (2014), building on the work of Dixon and Coles (1997) and Rue and Salvesen (2000). The home ground advantage assumes that a team scores, on average, more goals in a home game than in an away game; see Pollard (2008) for a review.

We collect the home ground advantage coefficient and the time-varying attack and defense strengths for each team in the state vector $\boldsymbol{\alpha}_t$ in (8), that is

$$\boldsymbol{\alpha}_t = (\zeta, \mu_{1t}, \ldots, \mu_{Jt}, \beta_{1t}, \ldots, \beta_{Jt})'. \tag{25}$$

We note that the state vector $\boldsymbol{\alpha}_t$ has 59 elements in our analysis for the Bundesliga. For each week t with K_t scheduled matches, we collect the log intensity pairs in the signal vector

$$\boldsymbol{\theta}_t = (\theta_{1,i_1 j_1,t}, \theta_{2,i_1 j_1,t}, \ldots, \theta_{1,i_{K_t} j_{K_t},t}, \theta_{2,i_{K_t} j_{K_t},t})'.$$

When all teams play in week t, $K_t = 9$ and $\boldsymbol{\theta}_t$ has 18 entries with a result out of a total of 29 (total number of teams) and has 11 missing entries. The vector $\boldsymbol{\theta}_t$ can be constructed from (7) using the state vector $\boldsymbol{\alpha}_t$ in (25) with $\boldsymbol{c}_t = \mathbf{0}$ and $\boldsymbol{Z}_t$ an appropriate selection matrix as implied by (24).

The attack and defense strengths for each team evolve separately over time as

$$\begin{aligned} \mu_{i,t} &= (1-\phi_\mu)\bar{\mu}_i + \phi_\mu \mu_{i,t-1} + \varepsilon_{\mu,it}, \\ \beta_{i,t} &= (1-\phi_\beta)\bar{\beta}_i + \phi_\beta \beta_{i,t-1} + \varepsilon_{\beta,it}, \end{aligned} \qquad \begin{pmatrix} \varepsilon_{\mu,it} \\ \varepsilon_{\beta,it} \end{pmatrix} \sim \text{NID}\left(0, \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}\right),$$

for $i = 1, \ldots, J$, and where the ε_{it} are mutually independent over i and t. Although each team has its own unique attack and defense strength, the persistence coefficients ϕ_μ and ϕ_β and the innovation variances σ_μ^2 and σ_β^2 are common to all teams. This again results in a highly parsimonious model. The fixed means $\bar{\mu}_i$ and $\bar{\beta}_i$ are part of the initial state vector α_1 and can be estimated as part of the signal extraction of the state vectors using importance sampling methods; see Durbin and Koopman (2012, Chapter 11). We retain 4 parameters $\boldsymbol{\psi} = (\phi_\mu, \phi_\beta, \sigma_\mu^2, \sigma_\beta^2)'$, which we estimate using the techniques outlined in Section 3.

5.1 Estimation results

To verify the effect of different values of S and M on the estimation results, we present the estimation results for $\boldsymbol{\psi}$ in Table 4 using $S = 50, 200, 1000$ and $M = 10, 20$. The values of the maximized log likelihood function are also presented. All reported estimates differ only slightly for different values of S and M. In particular, the differences are negligible compared to the estimated standard errors. We conclude that reasonable choices for S and M have no major impact on the results.

Table 4 shows that the attack and defense strengths on a weekly basis are highly persistent and smooth. This is to be expected: the attack or defense strength of a particular team does not change dramatically from one week to another. When we consider the persistence year-by-year (34 weeks), the corresponding persistence coefficients are $0.9958^{34} = 0.87$ and $0.9912^{34} = 0.74$ for the defense and attack strength, respectively. These values clearly point to stationary dynamics. Interestingly, the attack strength evolves more persistently over time than the defense strength.

Table 4: Estimates of parameter vector $\boldsymbol{\psi}$: Bundesliga team strengths. The table presents the Monte Carlo estimates of the five model coefficients, where ζ is estimated as part of the state vector. The remaining parameters are estimated using non-linear numerical optimization. The estimates are given for different values of M and S (in columns). The standard errors of the estimates are presented in italics below. The last row contains the maximized estimated log likelihood values.

	$S = 50$		$S = 200$		$S = 1,000$	
	$M = 10$	$M = 20$	$M = 10$	$M = 20$	$M = 10$	$M = 20$
ϕ_μ	0.9958	0.9958	0.9958	0.9958	0.9958	0.9958
	0.0019	*0.0019*	*0.0018*	*0.0018*	*0.0019*	*0.0019*
ϕ_β	0.9911	0.9911	0.9911	0.9911	0.9912	0.9912
	0.0048	*0.0050*	*0.0050*	*0.0046*	*0.0050*	*0.0050*
$\sigma_\mu^2 \times 10^4$	8.649	8.653	8.607	8.615	8.658	8.654
	3.950	*3.833*	*3.807*	*3.711*	*3.892*	*3.859*
$\sigma_\beta^2 \times 10^4$	5.738	5.744	5.698	5.704	5.685	5.672
	3.706	*3.851*	*3.849*	*3.542*	*3.865*	*3.814*
ζ	0.2617	0.2617	0.2617	0.2617	0.2617	0.2617
	0.0262	*0.0262*	*0.0262*	*0.0262*	*0.0262*	*0.0262*
LL	-8137.50	-8137.50	-8137.54	-8137.54	-8137.51	-8137.51

5.2 Model extensions

The benchmark Skellam dynamic panel model for match results as described above can be extended in many different ways. In this section, we explore a number of such extensions and provide further results. We introduce the extensions briefly and then comment on the empirical findings for the German Bundesliga data. A compilation of the various extensions is presented in Table 5. All results are computed for $S = 100$ and $M = 10$.

Away ground disadvantage

Apart from the fixed effect ζ for home ground advantage in the scoring intensity of the home team, we can also introduce a fixed effect ζ_a for the disadvantage of scoring by the away team: $\lambda_{2,ij,t} = \zeta_a + \mu_{jt} - \beta_{it}$. We find $\zeta_a = 0.088$ with a standard error of 0.075, such that this effect is not statistically significant in our study. Also, the positive sign of the estimate suggests an advantage rather than a disadvantage of scoring by the away team.

Zero inflated model

If we compare the observed number of draws $y_{ij,t} = 0$ in our data set with the expected number of draws based on the empirical estimates in Table 4, we find that the benchmark model underestimates the number of draws. We therefore consider a zero-inflated version of the model using (3) and add an extra parameter γ that accounts for the possible transfer of probability mass towards zero. Although the expected number of draws based on this model is closer to the observed number of draws, the estimate for γ is not statistically significant.

Heterogeneous panel

The benchmark model assumes that the dynamic properties of the attack and defense processes are the same for all teams, i.e., the coefficients ϕ_μ, ϕ_β, σ_μ^2 and σ_β^2 do not depend on i. It is possible that these processes behave differently over time for the different teams. A first attempt to relax this constraint is to have different properties for the more constant performing teams. For the definition of a constant performing team we look at the final tables of the four most recent years before the start of our dataset, the period from 2002/03 to 2005/06. We make two groups. In the first group labeled group I we have the teams that finished four years in a row in the top half, i.e. best 9 out of 18 teams. Group I consists of {Dortmund, Hamburg, Bayern Munich, Werder Bremen, Schalke 04, Stuttgart}. The second group labeled group II holds all other teams. The four additional parameters are placed in $\boldsymbol{\psi}$. This heterogeneous panel specification leads to a much better in-sample fit with a likelihood ratio test value of 14.74 for 4 additional parameters. The estimates for this model are

$$\begin{aligned} &\phi_{\mu,I} = 0.9986, \quad \phi_{\beta,I} = 0.9958, \quad \sigma^2_{\mu,I} \times 10^4 = 6.35, \quad \sigma^2_{\beta,I} \times 10^4 = 5.22, \quad \zeta = 0.263, \\ &\phi_{\mu,II} = 0.9737, \quad \phi_{\beta,II} = 0.9852, \quad \sigma^2_{\mu,II} \times 10^4 = 30.9, \quad \sigma^2_{\beta,II} \times 10^4 = 8.00. \end{aligned}$$

We conclude that the traditionally better performing teams have more persistent attack and defense strength processes.

Home stadium capacity

Home ground advantage may depend on the stadium capacity of the home team. A larger stadium that can contain a larger crowd may have a larger impact on the performance of the two teams and perhaps the referee; see the discussion in Pollard (2008). The team specific home ground advantage is added as a regression effect to the home team scoring intensity, that is

$$\lambda_{1,ij,t} = \exp(\zeta + \zeta_x x_i + \mu_{it} - \beta_{jt}),$$

where x_i is the stadium capacity of team i (measured in multiples of 10,000) and ζ_x is the regression coefficient that is placed in the parameter vector $\boldsymbol{\psi}$. This model specification does not lead to a significant improvement of the log likelihood value.

Correlated attack and defense strengths

The attack and defense strengths of teams are typically related; both should be good for a successful team. We therefore consider the innovations of the autoregressive processes for attack and defense strength to be correlated, that is

$$\begin{pmatrix} \varepsilon_{\mu,it} \\ \varepsilon_{\beta,it} \end{pmatrix} \sim \text{NID}\left(0, \begin{bmatrix} \sigma_\mu^2 & \rho\sigma_\mu\sigma_\beta \\ \rho\sigma_\mu\sigma_\beta & \sigma_\beta^2 \end{bmatrix}\right), \qquad -1 < \rho < 1, \qquad i = 1, \ldots, J,$$

where ρ is the correlation coefficient which is common to all teams. The estimate of ρ is 0.97 while the other estimates are given by

$$\begin{aligned} &\phi_{\mu,I} = 0.9987, \quad \phi_{\beta,I} = 0.9942, \quad \sigma^2_{\mu,I} \times 10^4 = 5.12, \quad \sigma^2_{\beta,I} \times 10^4 = 4.40, \quad \zeta = 0.267, \\ &\phi_{\mu,II} = 0.9263, \quad \phi_{\beta,II} = 0.9894, \quad \sigma^2_{\mu,II} \times 10^4 = 92.6, \quad \sigma^2_{\beta,II} \times 10^4 = 3.84. \end{aligned}$$

The likelihood ratio test for one extra coefficient is 5.96, which is statistically significant at the 5% level.

Synthesis

Table 5 reviews our empirical findings for the Bundesliga. For each model extension the estimated log likelihood function is maximized and the Akaike information criterion (AIC)

Table 5: Model comparison. We present the fit improvements of different model specifications discussed in Section 5.2. Each row represents an extension of the model. The sign ✓ is used to indicate whether the model extension is adopted in the final model. The number of parameters ($\#\boldsymbol{\psi}$), the log likelihood (LL) value, the likelihood ratio (LR) test and the Akaike information criterion (AIC) are reported for each extension. In case the model extension only concerns a single parameter, the null hypothesis and the t-test are reported as well. The LR test with * and ** indicates significance at the 5% and 1% significance level respectively.

Model specification		$\#\boldsymbol{\psi}$	LL	LR test	H_0	t-test	AIC
Basic model		4	-8137.51				7.6027
Away ground disadvantage		5	-8136.87	1.28	$\zeta_a = 0$	1.17	7.6030
Zero inflated model		5	-8136.89	1.24	$\gamma = 0$	1.17	7.6030
Heterogeneous panel	✓	8	-8130.14	14.74^{**}			7.5996
Home stadium capacity		9	-8130.92	-1.56	$\zeta_x = 0$	1.23	7.6012
Correlated attack and defense	✓	9	-8127.16	5.96^{*}			7.5977

is computed to facilitate model comparison. We conclude that team heterogeneity in the dynamics, and correlation between attack and defense strengths are two important extensions for soccer match outcomes in our Bundesliga data set.

5.3 Signal extraction

We present the estimated attack and defense strengths in Figure 3 using the bivariate NAIS methodology as described in Section 4.4. The smooth and persistent processes for the attack and defense strengths of the overall stronger teams in group I are clearly visible. The teams in group II have not played in the Bundesliga for all seasons in our sample. A number of those teams have only played during one season. It emphasizes that our estimation procedure can handle such a large and unbalanced panel of teams without much problems, despite the intricacies of the Skellam distribution and the high dimensional state and signal vector.

The correlation between the attack and defense strengths for each team is also clearly visible. The teams of Dortmund and Bayern Munich have increased their attack strength consistently from season to season which is also true for their defense strength. On the other hand, the attack strength of Werder Bremen has deteriorated over time, which only partly applies to its defense strength.

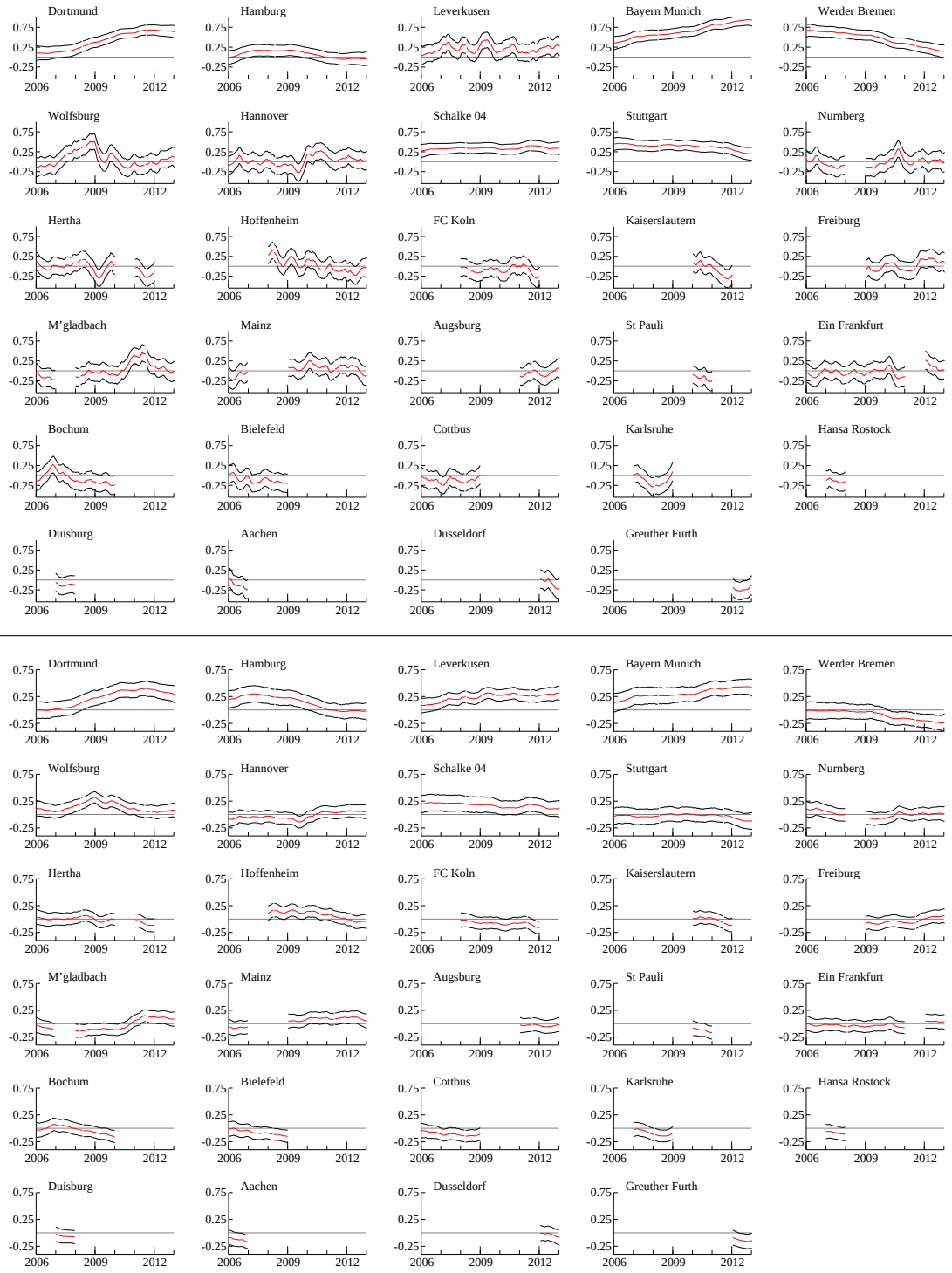

Figure 3: Attack and defense strengths of teams in the Bundesliga. The two panels of graphs present respectively the extracted attack and defense strengths for all teams in the Bundesliga from the season 2006/07 towards 2012/13, together with confidence intervals based on one standard error. The more persistent attack and defense strength processes of the group I teams are clearly visible. Group I = {Dortmund, Hamburg, Bayern Munich, Werder Bremen, Schalke 04, Stuttgart}. Group II is formed by the remaining teams.

6 Conclusions

In this paper we introduced a general dynamic model for Skellam distributed counts. We made the two intensity parameters of the Skellam distribution time-varying and showed how to formulate the resulting model as a non-Gaussian state space model. We then performed a likelihood-based analysis of the model using importance sampling methods. In particular, we showed how to estimate the parameters and states of the dynamic Skellam model using a bivariate extension of the numerically accelerated importance sampling (NAIS) method of Koopman et al. (2013). In contrast to the higher dimensional generalization of NAIS based on Halton sequences in Scharth (2012, Chapter 5), we were still able to use bivariate Gauss-Hermite numerical integration techniques to compute the appropriate integrals.

Based on two different illustrations, we demonstrated the versatility of our dynamic Skellam analysis. In our first application, we studied price changes (denoted as the number of discrete ticks) for 8 stocks listed on the New York Stock Exchange. The number of trades considered for these stocks ranged from almost 4,000 to more than 23,000. The dynamic Skellam distribution turned out to be extremely useful to model the dynamic volatility of stocks based on discrete price changes based on high-frequency tick data. In our second application, we showed that our analysis can handle a large panel of differences of scores by matches played in the German Bundesliga. We showed that the dynamic Skellam model in its state space formulation can handle large sections of missing data. We also showed how the model can be extended to include regression effects, heterogeneous dynamics in the panel, and extensions of the Skellam distribution that assign different probability mass to a small number of discrete outcomes. A key example of the latter is the dynamic zero inflated Skellam model.

We conclude that the new dynamic Skellam model is robust and computationally feasible for both long univariate time series and large unbalanced panels. The models may even rely on high dimensional state vectors. Our flexible modeling framework for time series may provide a useful benchmark for empirical applications based on integer outcomes that can take both positive and negative values.

References

Abramowitz, M. and I. A. Stegun (1972). *Handbook of mathematical functions*. New York: Dover publications.

Alzaid, A. and M. A. Omair (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society 33(1)*, 17–45.

Andersen, T. G. (2000). Some reflections on analysis of high-frequency data. *J. Business and Economic Statist. 18(2)*, 146–153.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Realised kernels in practice: Trades and quotes. *Econometrics Journal 4*, 1–33.

Barndorff-Nielsen, O. E., D. G. Pollard, and N. Shephard (2012). Integer-valued levy processes and low latency financial econometrics. *Quantitative Finance 12(4)*, 587–605.

Bohning, D., E. Dietz, P. Schlattmann, L. Mendonca, and U. Kirchner (1999). The zero inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *J. Royal Statistical Society A 162(2)*, 195–209.

Brownlees, C. T. and G. M. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis 51*, 2232–2245.

Cappé, O., E. Moulines, and T. Ryden (2005). *Inference in Hidden Markov Models*. New York: Springer.

Chaboud, A., B. Chiquoine, E. Hjalmarsson, and C. Vega (2013). Rise of the machines: Algorithmic trading in the foreign exchange market. Discussion paper, Federal Reserve Board.

Dixon, M. J. and S. G. Coles (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics 46(2)*, 265–280.

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

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Royal Statistical Society B 62(1)*, 3–56.

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

Falkenberry, T. N. (2002). High frequency data filtering. Technical report, Tick Data.

Fridman, M. and L. Harris (1998). A maximum likelihood approach for non-gaussian stochastic volatility models. *J. Business and Economic Statist. 16(3)*, 284–291.

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

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *Journal of Finance 66*, 1–33.

Johnson, N., S. Kotz, and A. W. Kemp (1992). *Univariate discrete distributions*. New York: Wiley.

Karlis, D. and I. Ntzoufras (2006). Bayesian analysis of the differences of count data. *Statistics in medicine 25(11)*, 1885–1905.

Karlis, D. and I. Ntzoufras (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics 20*, 133–145.

Kervel, V. L. (2012). Liquidity, what you see is what you get? Discussion paper, VU University Amsterdam and Tilburg Law and Economics Center.

Kitagawa, G. (1987). Non-gaussian state-space modeling of nonstationary time series. *J. American Statistical Association 82(400)*, 1032–1041.

Koopman, S. J. and R. Lit (2014). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *J. Royal Statistical Society A 177(4)*, forthcoming.

Koopman, S. J., A. Lucas, and M. Scharth (2013). Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. Discussion paper, Tinbergen Institute.

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

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

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica 36(3)*, 109–118.

Mardia, K. V. (1970). *Families of Bivariate Distributions*. London: Griffin.

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

O'Hara, M., C. Yao, and M. Ye (2012). Whats not there: The odd-lot bias in taq data. Research paper series 16-2012, Johnson School.

Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The open sports sciences journal 1*, 12–14.

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

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

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer.

Rue, H. and O. Salvesen (2000). Prediction and retrospective analysis of soccer matches in a league. *The Statistician 49(3)*, 399–418.

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

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

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.

Tsay, R. S. (2005). *Analysis of financial time series* (2nd ed.). New Jersey: Wiley-Interscience.