TIME-VARYING PARAMETER MODELS FOR DISCRETE VALUED TIME SERIES

ISBN 978 90 5170 755 7

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul
Image: Wiki Commons (template) and Marcin Zamojski (design)

This book is no. **642** of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

TIME-VARYING PARAMETER MODELS FOR DISCRETE VALUED TIME SERIES

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Economische Wetenschappen en Bedrijfskunde
op dinsdag 9 februari 2016 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Rutger Lit
geboren te Alkmaar

promotoren: prof.dr. A. Lucas
prof.dr. S.J. Koopman

Acknowledgements

The last two and a half years went by at a fast pace. It was a period of time that I greatly enjoyed and the end product, this dissertation, is something that I am proud of. I could not have done this research alone. I am especially grateful to my supervisors Siem Jan Koopman and André Lucas who were essential for the realisation of this dissertation. Their expert advice on numerous topics in combination with frequent meetings, despite their busy schedules, were very important and much appreciated. I look forward to continuing our collaboration.

From all my colleagues at the finance and econometrics department at the VU Amsterdam, a special thanks goes out to István Barra. The originality and quality of our research discussions were very helpful to me and very much appreciated.

Another thanks goes out to Marcin Zamojski for his advice on several econometric and finance topics as well as making my digital life easier and more secure. I would also like to thank Marco Bazzi, Falk Bräuning, Artem Duplinskiy, Wenqian Huang (黄文倩), Anne Opschoor, Kristian Støre, Andries van Vlodrop, and Marius Zoican for the discussions and laughs we shared.

I am grateful to the Dutch National Science Foundation (NWO) and the Tinbergen Institute for the financial support. Last but not least I thank my parents for never doubting the research path that I have chosen.

Amsterdam, November 2015

Contents

Acknowledgements **v**

1 Introduction **1**
1.1 Introduction for a general audience . . . 1
1.2 Econometric methodologies . . . 2
1.2.1 Non-Gaussian state space models . . . 2
1.2.2 Score-driven models . . . 4
1.3 Contributions . . . 4
1.3.1 Chapter 2 . . . 4
1.3.2 Chapter 3 . . . 5
1.3.3 Chapter 4 . . . 6
1.3.4 Chapter 5 . . . 7

2 A Dynamic Bivariate Poisson Model for Analysing Football Match Results **9**
2.1 Introduction . . . 9
2.2 The statistical modelling framework . . . 12
2.2.1 Bivariate Poisson model . . . 12
2.2.2 Dynamic specification for goal scoring intensities . . . 13
2.2.3 Some extensions of the basic model . . . 14
2.2.4 General state space representation of the model . . . 16
2.2.5 Evaluation of likelihood function and estimation . . . 17
2.3 Empirical application . . . 19
2.3.1 Data description . . . 19
2.3.2 Details of the basic model . . . 20
2.3.3 Parameter estimates . . . 21
2.3.4 Signal estimates of strengths of attack and defence . . . 22
2.3.5 Model evaluations: in-sample and out-of-sample . . . 24
2.4 Out-of-sample performance in a betting strategy . . . 28
2.5 Conclusions . . . 32
Appendices . . . 33

A Likelihood evaluation . . . 33
B Construction of approximating model . . . 34
C The derivatives for the model observation density . . . 36
D Computational details . . . 37
E Tables and figures . . . 40

3 Intraday Stochastic Volatility in Discrete Price Changes 49
3.1 Introduction . . . 49
3.2 The dynamic Skellam model . . . 52
3.2.1 The Skellam distribution . . . 52
3.2.2 The modified Skellam distribution . . . 52
3.2.3 The Skellam model with dynamic mean and variance . . . 55
3.3 Analysis of high-frequency Skellam price changes . . . 57
3.3.1 Data . . . 57
3.3.2 Dynamic Skellam with Intraday Stochastic Volatility . . . 58
3.3.3 Parameter estimation results . . . 60
3.3.4 Signal extraction . . . 62
3.3.5 Goodness-of-fit . . . 62
3.3.6 Diagnostic checking . . . 64
3.3.7 Forecasting study . . . 66
3.4 Conclusions . . . 69
Appendices . . . 70
A Modified Skellam distribution of type I . . . 70
B Moments of the MSKII(i, j, k) distribution . . . 70
C Simulation study . . . 71
D Numerically accelerated importance sampling . . . 73
E Intradaily time series of price changes in 2012 . . . 77

4 A Skellam Model for Analysing the Differences in Count Data 79
4.1 Introduction . . . 79
4.2 The dynamic Skellam model . . . 80
4.2.1 Skellam distribution . . . 80
4.2.2 Dynamic specification of intensities . . . 81
4.3 Analysing football scores . . . 82
4.3.1 Estimation results . . . 84
4.3.2 Model extensions . . . 85
4.3.3 Signal extraction . . . 88
4.4 Conclusions . . . 90

5 Dynamic Discrete Copula Models for High Frequency Stock Price Changes 93
5.1 Introduction . . . 93
5.2 Score-driven dynamic discrete copula model . . . 96
5.3 Simulation study . . . 100
5.3.1 Estimating parameters when model is correctly specified . . . 100
5.3.2 Estimating time-varying paths when model is misspecified . . . 101
5.4 Dependence between discrete price changes . . . 103
5.4.1 Data description . . . 104
5.4.2 Missing values . . . 105
5.4.3 Copula selection . . . 106
5.4.4 Full year results . . . 107
5.4.5 Comparison with intraday spline . . . 110
5.5 Conclusions . . . 111
Appendices . . . 114
A Derivation of the score vector . . . 114
B Tables and figures . . . 115

Bibliography 119

Summary 127

Chapter 1

Introduction

The introduction of this dissertation is organised as follows. Section 1.1 explains the title of this dissertation to a general audience and uses a football example as a lively illustration of the use of time-varying parameter models. Here, the use of mathematical notation is avoided as much as possible. Starting from Section 1.2, the focus is on the more experienced reader in the field of econometrics and statistics.

1.1 Introduction for a general audience

Suppose in the very unlikely event that a football aficionado with knowledge of econometrics, statistics, and finance is interested in predicting the outcome of the next football match. This dissertation, with the title *'time-varying parameter models for discrete valued time series'*, can assist with these predictions.

Let us start by getting a good understanding of what a *time series* is. Here, I consider a time series as a set of observations taken at (possibly unequally spaced) intervals over time. For example, we can observe the number of goals scored by a football team over a period of time (for example each week for a period of five years). It is crucial that the order of the observations is preserved since without this we cannot do any meaningful time series analysis. If we denote the number of goals scored by the football team at time index t by y_t for $t = 1, \ldots, n$, then $y = (y_1, \ldots, y_n)'$ is a time series of length n.

The part *discrete valued* refers to the type of data the time series consists of. In this dissertation I focus on integers (whole numbers) and do not consider categorical data which could also be regarded as discrete. An example of discrete data are the number of goals scored by a football team. This number is in this case restricted to a positive integer since a football team cannot score a negative number of goals. Negative integers, however, play a prominent role in this dissertation in Chapter 3–5.

Time-varying parameters form the core of this dissertation. To make a forecast about the next football match we could be interested in a measure of strength of the football

teams. It is an unrealistic assumption that the strength of a team is constant over time since the composition of football teams change over time as well. Also, recent match results probably tell us more about the current strength of a team than match results from the more distant past. Therefore, it can be expected that time-varying parameters generally allow us to obtain (much) more accurate forecasts than their static counterparts. The evolution of the team's strength over time can be used in, for example, the analysis of the performance of the teams in a competition or they can form the basis of a betting strategy.

The dynamics of the parameters are determined by the econometric *model* that is developed and/or applied. There are many econometric models, varying largely in complication. The next section discusses the two classes of time-varying parameter models that form the econometric modelling framework of this dissertation.

The football illustration is a nice example of the wide applicability of time-varying parameter models for discrete valued time series. There are, however, many other fields where these models play a role in analysing data and forecasting future observations. One can think about the number of hospital admissions in the field of medical research or the number of earthquakes of magnitude > 5 along the San Andreas fault line in the field of geology. A prominent example in this dissertation is the extraction of volatility from discrete stock price changes in the field of econometrics and finance.

1.2 Econometric methodologies

1.2.1 Non-Gaussian state space models

Consider a parametric model for an observed time series $y = (y_1', \ldots, y_n')'$ that is formulated conditionally on a latent $m \times 1$ time-varying parameter vector α_t, for time index $t = 1, \ldots, n$. We are interested in the statistical behavior of the state vector, α_t, given a subset of the data, i.e. the data up to time $t-1$ (forecasting), the data up to time t (filtering) or the whole data set (smoothing). One possible framework for such an analysis is the state space model, the general form of which is given by

$$y_t|\alpha_t \sim p(y_t|\alpha_t;\psi), \qquad \alpha_{t+1} \sim p(\alpha_{t+1}|\alpha_t;\psi), \qquad \alpha_1 \sim p(\alpha_1;\psi), \tag{1.1}$$

where $p(y_t|\alpha_t;\psi)$ is the observation density, $p(\alpha_{t+1}|\alpha_t;\psi)$ is the state transition density with initial density $p(\alpha_1;\psi)$ and ψ is a static parameter vector.

Minimum mean square error (MMSE) estimates of α_t and MMSE forecasts for y_t can be obtained by the Kalman filter and related smoother methods if the following three conditions are met: (i) the state transition density $p(\alpha_{t+1}|\alpha_t;\psi)$ for α_t is linear and Gaussian, (ii) the relation between y_t and α_t in $p(y_t|\alpha_t;\psi)$ is linear and (iii) the

observation y_t is, conditional on α_t, normally distributed. In other words, $p(y_t|\alpha_t;\psi)$, $p(\alpha_{t+1}|\alpha_t;\psi)$ and $p(\alpha_1;\psi)$ are Gaussian and the observation and transition relations are linear. If all three conditions are satisfied, the state space model of (1.1) reduces to the linear Gaussian state space model, see for example Durbin and Koopman (2012, Part I). The violation of at least one of the three properties means that the state space model becomes nonlinear and/or non-Gaussian for which we have to rely on other methods to obtain optimal estimates.

Several methods to analyse nonlinear non-Gaussian state space models are available in the literature. For example, particle filters, Markov Chain Monte Carlo methods and Gaussian approximations. In this dissertation we opt for a classical analysis and work with the principle of maximum likelihood estimation (MLE). More specifically, we rely on Monte Carlo simulation methods based on importance sampling as proposed by Shephard and Pitt (1997), Durbin and Koopman (1997) and the improvements on these methods by Koopman, Lucas, and Scharth (2014). The main motivation to use MLE are the well established and well documented properties of MLE.

We define $\alpha = (\alpha_1', \ldots, \alpha_n')'$ and assume that, given the unobserved state, the observations are conditionally independent which implies,

$$p(y|\alpha;\psi) = \prod_{t=1}^{n} p(y_t|\alpha_t;\psi). \tag{1.2}$$

Importance sampling techniques are employed because the likelihood function for y, based on the observation density $p(y_t|\alpha_t;\psi)$ and given by

$$\ell(\psi) = p(y;\psi) = \int p(y,\alpha;\psi)\mathrm{d}\alpha = \int p(y|\alpha;\psi)p(\alpha;\psi)\mathrm{d}\alpha, \tag{1.3}$$

does not have an analytical solution and requires simulation methods since numerical integration of a multi-dimensional integral quickly becomes infeasible. A naive Monte Carlo estimate of the likelihood function is given by

$$\hat{\ell}(\psi) = \frac{1}{M}\sum_{k=1}^{M} p(y|\alpha^{(k)};\psi), \qquad \alpha^{(k)} \sim p(\alpha;\psi), \tag{1.4}$$

where M is the number of Monte Carlo replications and draws $\alpha^{(1)}, \ldots, \alpha^{(M)}$ are generated independently from one another. This Monte Carlo estimate is numerically not efficient since the simulated paths have no support from the observed data y. A more effective approach for the evaluation of the likelihood function is to adopt Monte Carlo simulation methods based on importance sampling for which details are given in Chapter 2.

1.2.2 Score-driven models

In the class of score-driven models, the latent time-varying parameter vector α_t is updated over time using an autoregressive updating function based on the score of the conditional observation probability density function, see Creal, Koopman, and Lucas (2013) and Harvey (2013). The updating function for α_t is given by

$$\alpha_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j \alpha_{t-j+1},$$

where ω is a vector of constants, A and B are fixed coefficient matrices and s_t is the scaled score function which is the driving force behind the updating equation. The unknown coefficients ω, A and B depend on the static parameter vector ψ. The definition of s_t is

$$s_t = S_t \cdot \nabla_t, \qquad \nabla_t = \frac{\partial \log p(y_t|\alpha_t, \mathcal{F}_{t-1}; \psi)}{\partial \alpha_t}, \qquad t = 1, \ldots, n,$$

where ∇_t is the score vector of the (predictive) density $p(y_t|\alpha_t, \mathcal{F}_{t-1}; \psi)$ of the observed time series $y = (y_1', \ldots, y_n')'$. The information set $\mathcal{F}_{t-1}$ usually consists of lagged variables of α_t and y_t but can contain exogenous variables as well. To introduce further flexibility in the model, the score vector ∇_t can be scaled by a matrix S_t. Common choices for S_t are unit scaling, the inverse of the Fisher information matrix, or the square root of the Fisher inverse information matrix. The latter has the advantage of giving s_t a unit variance since the Fisher information matrix is the variance matrix of the score vector. In this framework and given past information, the time-varying parameter vector α_t is perfectly predictable one-step-ahead.

The score-driven model has three main advantages: (i) the 'filtered' estimates of the time-varying parameter are optimal in a Kullback-Leibler sense, see Blasques, Koopman, and Lucas (2015); (ii) since the score-driven models are observation driven, their likelihood is known in closed-form; and (iii) the forecasting performance of these models is comparable to their parameter-driven counterparts, see Koopman, Lucas, and Scharth (2015). The second point emphasizes that static parameters can be estimated in a straightforward way using maximum likelihood methods.

1.3 Contributions

1.3.1 Chapter 2

In the non-Gaussian state space models that we consider in this dissertation, we can reformulate the observation density $p(y_t|\alpha_t; \psi)$ by a density that can be written in terms

of the signal vector θ_t, i.e. $p(y_t|\theta_t;\psi)$. The $r \times 1$ signal vector is defined as $\theta_t = Z_t\alpha_t$ where Z_t is a selection matrix often consisting of zeros and ones, and possibly some unknown coefficients that are collected in ψ. In contrast to the state vector, the dimension of the signal vector θ_t is often low (typically $r = 1$) since increasing the dimension of the signal vector has the drawback of reducing the efficiency of the importance sampler. We show that the methodology of Shephard and Pitt (1997) and Durbin and Koopman (1997) can be extended to a large dimensional signal that consists of bivariate signal building blocks. We apply this methodology to a large panel of football match results which assumes a bivariate Poisson distribution with intensity coefficients that change stochastically over time. The importance sampling methods are computationally efficient despite the high dimensional signal and state vector ($r = 36$ and $m = 72$ respectively). This can be regarded as an achievement as in no other contributions in the field such high dimensions are ever used.

1.3.2 Chapter 3

Stochastic volatility is typically associated with the time-varying variance in time series of daily continuously compounded rates of financial returns. Münnix, Schäfer, and Guhr (2010) show that returns concentrate around the tick-size, are severely multi-modal and, consequently, highly non-Gaussian. We propose to model stochastic volatility in discrete price changes of a stock which are measured on a grid of one dollar cent and hence we face the challenge to model positive, zero, or negative tick-by-tick price changes. One possible option is to consider such data as Skellam distributed random variables that take values in $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. We develop a new statistical model that builds on a dynamic modified Skellam distribution to make the model congruent with the realised data and we analyse the model with the state space methodology discussed in Section 1.2.1. Our modified Skellam distribution features a dynamic variance parameter that is allowed to be different over the course of a trading day due to intraday seasonal patterns, which we capture by including a spline function over the time of day. On top of this, we also allow for autoregressive intraday stochastic volatility dynamics to capture any remaining volatility dynamics over the course of the trading day that cannot be attributed to seasonal patterns. Finally, our data requires a careful treatment when the observed price changes are equal to 0, 1, or -1 dollar cents. For this purpose, we modify the dynamic Skellam distribution by allowing for a probability mass transfer between these different price change realisations. The probability mass transfer needs to vary over time as well because it turns out that trades with a zero price-change are not spread evenly across the trading day. We analyse stock price changes on a second-by-second basis within a single trading day on the New York Stock Exchange. As a consequence, all series have the same length of $n = 23{,}400$ (6.5 hours $\times$ 3600 seconds) with many missing values. Our

state space framework for the dynamic modified Skellam model is able to account for the possibly many missing values efficiently. Long time series (large n) are known to reduce the efficiency of importance sampling, see the discussion in Cappé, Moulines, and Ryden (2005). We show that time series with a length of $n = 23{,}400$ can efficiently be evaluated by the Numerically Accelerated Importance Sampling (NAIS) methodology of Koopman et al. (2014). The resulting new model with the new features embedded performs well in terms of fit, diagnostics, and forecasting power compared to a range of alternative models. Hence we may conclude that a satisfactory modelling solution is developed.

1.3.3 Chapter 4

We show in Chapter 3 that the univariate NAIS methodology of Koopman et al. (2014) is able to efficiently analyse long univariate time series. In Chapter 4, we extend the univariate NAIS methodology to a bivariate signal to accommodate two 'intensity' parameters that are part of the Skellam distribution as originally derived by Skellam (1946). The bivariate NAIS methodology adopts a bivariate Gauss-Hermite quadrature to solve the integral that describes the variance of the importance sampling log weights and determines the parameters of the importance density by minimizing this integral. To avoid any unnecessary repetition of notation and equations, the bivariate NAIS methodology is already introduced in Appendix D of Chapter 3.

We test the bivariate NAIS methodology in a large scale application by studying the score differences of football matches of 29 teams observed over seven seasons of the German Bundesliga. We allow the intensities of the Skellam distribution to vary stochastically over time in the state space framework that was discussed in Section 1.2.1. The two intensities of the Skellam distribution correspond with the scoring intensities of the two football teams that face each other. After estimating and presenting the results of the basic model, which serves as a benchmark, several extensions are proposed. We introduce an 'away ground disadvantage' to test for a disadvantage of scoring by the away team and we allow the panel to be heterogeneous by assigning individual dynamic properties to groups of teams. Moreover, we test whether home ground advantage may depend on the stadium capacity of the home team. A larger stadium may have a larger impact on the performance of the two teams and perhaps the referee. Furthermore, we apply a zero inflated Skellam model since the observed number of draws in the data set is higher than the expected number of draws based on the benchmark model. Finally, we let the strengths of attack and defence to be correlated since they are typically related due to, for example, overlap in training or investments in a team.

1.3.4 Chapter 5

In many studies into intraday tick data of continuously compounded rates of financial returns, it is found that stock return volatility is higher during opening hours than during the rest of the day; see, for example, Andersen and Bollerslev (1997) and Tsay (2005). It is confiremed in Chapter 3 that this intraday pattern is present for discrete stock price changes as well. We continue to study discrete stock price changes and focus on a much less known topic, namely the intraday pattern of the dependence structure between stock price changes. We study the pattern of intraday dependence dynamics (beyond correlation structures) by adopting a flexible dynamic copula framework for the modelling of the dependence structure. We analyse intraday dependence structures for each trading day in 2012 and allow for marginal distributions and the dependence structure to vary over time. The stock price changes are assumed to be Skellam distributed and this distribution is adopted for the marginal distributions of the copula distribution. We provide a novel and parsimonious framework that is congruent with the empirical data. In particular, the dynamic parameters in our model, including stock return volatilities and dependence parameters, are updated using an observation-driven, autoregressive updating function based on the score of the conditional observation probability mass function, see Section 1.2.2.

Chapter 2

A Dynamic Bivariate Poisson Model for Analysing Football Match Results

Almost all results in this chapter previously appeared in Koopman and Lit (2015).

2.1 Introduction

Predicting the outcome of a football match is a challenging task. The pundit usually has strong beliefs about the outcomes of games. Bets can be placed on a win, a loss, a draw or on the match score itself. The collection of the predictions is reflected by the bookmaker's odds. In this chapter we study a history of nine years of football match results from the English Premier League. The number of goals scored by a team may depend on the strength of attack of the team, the strength of defence of the opposing team, the home ground advantage (when applicable) and the development of the match itself. We analyse the match results on the basis of a dynamic statistical modelling framework in which the strengths of attack and defence of the teams can vary over time. We show that the forecasts from this model are sufficiently accurate to gain a positive return over the bookmaker's odds.

Many statistical analyses of match results are based on the product of two independent Poisson distributions, which is also known as the double-Poisson distribution. The means of the two distributions can be interpreted as the goal scoring intensities of the two competing teams. In our modelling framework, the bivariate Poisson distribution is used which includes a dependence parameter that allows for correlation between home and away match scores. It represents the phenomenon that the ability or the effort of a team for a particular game is influenced by the other team or by the way that the match progresses. The performances of the teams due to the interactions between teams are captured by the dependence parameter. Furthermore, we let the goal scoring intensities

of the two teams depend on the strengths of attack and defence of the two teams. These strengths for each team are allowed to change stochastically over time. This time-varying feature becomes more important when we jointly analyse the match results for a series of consecutive football seasons. For example, when an excellent scorer leaves a team, the strength of attack weakens. Overall we expect that strengths of attack and defence change slowly over time.

The basis of our modelling approach was proposed by Maher (1982). In this study, the double-Poisson distribution, with the means expressed as team-specific strengths of attack and defence, is adopted as the underlying distribution for goal scoring. Maher (1982) explored the existence of a small correlation between home and away scores; he found a considerable improvement in model fit by trying a range of values for the dependence parameter. He did not provide parameter estimates of the correlation or dependence parameter. Furthermore, Maher's basic model is static; the team's strengths of attack and defence do not vary over time. Dixon and Coles (1997) consider the double-Poisson model with a dependence parameter that is estimated together with the other parameters. They suggested that the assumption of independence between goal scoring is reasonable except for the match results 0-0, 1-0, 0-1 and 1-1. They also introduced a weighting function to downweight likelihood contributions of observations from the more distant past. Karlis and Ntzoufras (2003) also used a bivariate Poisson distribution for match results; they showed that even a small value for the dependence parameter leads to a more accurate prediction of the number of draws. However, strengths of attack and defence are kept static over time in their analysis. Rue and Salvesen (2000) incorporated the framework of Dixon and Coles (1997) within a dynamic generalized linear model and adopted Markov Chain Monte Carlo methods to study the time-varying properties of the football teams in continuous time. In their analysis of match results, they truncated the number of goals to a maximum of five because they argued that the number of goals beyond five provides no further information about the strengths of attack and defence of a team. Owen (2011) adopted a similar dynamic generalized linear model and also uses Markov Chain Monte Carlo methods for estimation. However, he argued strongly for a model in discrete time. He found that the evolution of parameters over time, the role of strengths of attack and defence and the effect of home and away match scores are more effectively analysed in discrete time. We also formulate the model in discrete time but our model is based on the bivariate Poisson distribution and we estimate the parameters by simulated maximum likelihood methods rather than Markov Chain Monte Carlo methods. Owen (2011) empirically verified the model for a low dimensional data set whereas we consider all matches in the English Premier League for nine years.

The following contributions in the literature involve multivariate time series models and sports but are less relevant to our study. Ord, Fernandes, and Harvey (1993) consid-

ered a moderate multivariate extension of a Bayesian dynamic count data model for the analysis and forecasting of the number of goals scored by a small number of teams over a period of time. Furthermore their modelling framework is not based on Maher (1982). In Crowder, Dixon, Ledford, and Robinson (2002), the dynamic generalized linear model of Dixon and Coles (1997) is formulated as a non-Gaussian state space model with time-varying strengths of attack and defence as well. However, they used approximate methods for parameter estimation as they stated that an exact analysis is computationally too expensive. Given the rapid development of simulation methods for time series models, we shall show that exact maximum likelihood methods for an extensive analysis of match results can be carried out as a matter of routine. Our empirical study aims to analyse match results from the English Premier League. Earlier and leading studies have analysed match results from other sport leagues. In particular, Glickman and Stern (1998) and Glickman (2001) have considered match results from the American Football League, Fahrmeir and Tutz (1994) from the German Bundesliga and Knorr-Held (2000) from the American National Basketball Association.

We show that football match results from a high dimensional data set can be analysed effectively within a non-Gaussian state space model where the observed pairs of counts are assumed to come from a bivariate Poisson distribution. We have strengths of attack and defence that are stochastically evolving over time. The statistical analysis is based on exact maximum likelihood and signal extraction methods which rely on efficient Monte Carlo simulation techniques such as importance sampling. Several extensions can be considered within our modelling framework. For example, we introduce a parameter that accounts for the transition of summer and winter breaks. We also introduce the diagonal inflation method of Dixon and Coles (1997) for the bivariate Poisson distribution to account for the overrepresentation of draws. Finally we emphasize that we do not need to truncate the observed match outcomes to some maximum value in our analysis.

The remainder of this chapter is organised as follows. Our dynamic statistical modelling framework for the bivariate Poisson distribution is introduced and discussed in detail in Section 2.2. It is shown how we can represent the dynamic model in a non-Gaussian state space form. The statistical analysis relies on advanced simulation-based time series methods which are developed elsewhere. We provide the implementation details and some necessary modifications of the methods. The analysis includes maximum likelihood estimation, signal extraction of the strengths of attack and defence of a team and the forecasting of match results. In Section 2.3 we illustrate the methodology for a high dimensional data set of football match results from the English Premier League during the seasons 2003–2004 to 2011–2012. The first seven seasons are used for parameter estimation and in-sample diagnostic checking of the empirical results whereas the last two seasons are used for the out-of-sample forecast evaluation of the model. A forecast-

ing study is presented in Section 2.4 where we give evidence that our model can turn a positive return over the bookmakers' odds by applying a simple betting strategy during the seasons of 2010–2011 and 2011–2012. Concluding remarks are given in Section 2.5.

2.2 The statistical modelling framework

We analyse football match results in a competition for a number of seasons as a time series panel of pairs of counts. We assume that an even number of J teams play in a competition and hence each week $J/2$ matches are played. It also follows that a season consists of $2(J-1)$ weeks in which each team plays against another team twice, as a home team and as a visiting team. The specific details of our data set for the empirical study is discussed in Section 2.3.

2.2.1 Bivariate Poisson model

The result or outcome of a match between the home football team i and the visiting football team j in week t is taken as the pair of counts $(X, Y) = (X_{it}, Y_{jt})$, for $i \neq j = 1, \ldots, J$ and $t = 1, \ldots, n$ where n is the number of weeks available in our data set. The first count X_{it} is the non-negative number of goals scored by the home team i and the second count Y_{jt} is the number of goals scored by the visiting team j, in week t. Each pair of counts (X, Y) is assumed to be generated or sampled from the bivariate Poisson distribution with probability density function

$$p_{BP}(X, Y; \lambda_x, \lambda_y, \gamma) = \exp\left(-\lambda_x - \lambda_y - \gamma\right) \frac{\lambda_x^X}{X!} \frac{\lambda_y^Y}{Y!} \sum_{k=0}^{min(X,Y)} \binom{X}{k} \binom{Y}{k} k! \left(\frac{\gamma}{\lambda_x \lambda_y}\right)^k, \quad (2.1)$$

for $X = X_{it}$ and $Y = Y_{jt}$, with λ_x and λ_y being the intensities for X and Y respectively, and γ being a coefficient for the dependence between the two counts in the pair, X and Y. In short notation, we write

$$(X, Y) \sim BP(\lambda_x, \lambda_y, \gamma).$$

The means, variances and covariance for the home team score X and the away team score Y are

$$\mathbb{E}(X) = \mathbb{V}\text{ar}(X) = \lambda_x + \gamma, \qquad \mathbb{E}(Y) = \mathbb{V}\text{ar}(Y) = \lambda_y + \gamma, \qquad \text{Cov}(X, Y) = \gamma, \quad (2.2)$$

and hence the correlation coefficient between X and Y is given by

$$\rho = \frac{\gamma}{\sqrt{(\lambda_x + \gamma)(\lambda_y + \gamma)}}.$$

This definition of the bivariate Poisson distribution is not unique: other formulations have also been considered; see, for example, Kocherlakota and Kocherlakota (1992) and Johnson, Kotz, and Balakrishnan (1997). A different formulation for the bivariate Poisson distribution also implies different means, variances and covariances in expression (2.2).

The difference between the counts X and Y determines whether the match is a win, a loss or a draw for the home team. The variable $X - Y$ has a discrete probability distribution known as the Skellam distribution and it is invariant to γ when $(X, Y) \sim BP(\lambda_x, \lambda_y, \gamma)$ for $\gamma > 0$; see Skellam (1946). Karlis and Ntzoufras (2006, 2009) have used the Skellam distribution to analyse differences in scores in football matches.

2.2.2 Dynamic specification for goal scoring intensities

The scoring intensities of two teams playing against each other are determined by λ_x, λ_y and γ. In our modelling framework, we let λ_x and λ_y to vary with the pairs of teams that play against each other. Furthermore, we allow these intensities to change slowly over time since the composition and the performance of the teams will change over time. The intensity of scoring for team i, when playing against team j, is assumed to depend on the strength of attack of team i and the strength of defence of team j. We also acknowledge the home ground advantage in scoring by having the coefficient δ; this relative advantage is considered to be the same for all teams and constant over time. In section 2.2.3, we introduce a model extension in which δ is not the same for all teams. The strength of attack of the home team i in week t is denoted by ξ_{it} and its strength of defence is denoted by β_{it} for $i = 1, \ldots, J$. The goal scoring intensities for home team i and away team j in week t are then specified as

$$\lambda_{x,ijt} = \exp(\delta + \xi_{it} - \beta_{jt}), \qquad \lambda_{y,ijt} = \exp(\xi_{jt} - \beta_{it}). \tag{2.3}$$

In a football season with $J(J-1)$ matches, $2J(J-1)$ goal counts and for some time index t, we can identify the unknown signals for attack ξ_{it}'s and defence β_{it}'s together with coefficient δ, i.e. $2J + 1$ unknowns, when the number of teams is $J > 2$. The time variation of the strengths of attack and defence can be identified when we analyse match results from a series of football seasons.

All teams in the competition are assumed to have unique strengths of attack and defence which we do not relate to each other. In effect we assume that each team can compose their teams independently of each other. The strengths of attack and defence of

the team can change over time since the composition of the team will not be constant over time. Also the performances of the teams are expected to change over time. We therefore specify the strengths of attack and defence as auto-regressive processes. We have

$$\xi_{it} = \mu_{\xi,i} + \phi_{\xi,i}\xi_{i,t-1} + \eta_{\xi,it}, \qquad \beta_{it} = \mu_{\beta,i} + \phi_{\beta,i}\beta_{i,t-1} + \eta_{\beta,it}, \tag{2.4}$$

where $\mu_{\xi,i}$ and $\mu_{\beta,i}$ are unknown constants, $\phi_{\xi,i}$ and $\phi_{\beta,i}$ are auto-regressive coefficients and the disturbances $\eta_{\xi,it}$ and $\eta_{\beta,it}$ are normally distributed error terms which are independent of each other for all $i = 1, \ldots, J$ and all $t = 1, \ldots, n$. We assume that the dynamic processes are independent of each other and that they are stationary. It requires that $|\phi_{\kappa,i}| < 1$ for $\kappa = \xi, \beta$ and $i = 1, \ldots, J$. The independent disturbance sequences are stochastically generated by

$$\eta_{\kappa,it} \sim \text{NID}(0, \sigma^2_{\kappa,i}), \qquad \kappa = \xi, \beta, \tag{2.5}$$

where $\text{NID}(c, d)$ refers to normally independently distributed with mean c and variance d, for $i = 1, \ldots, J$ and $t = 1, \ldots, n$.

The initial conditions for the auto-regressive processes ξ_{it} and β_{it} can be based on means and variances of their unconditional distributions, which are given by

$$\mathbb{E}(\kappa_{it}) = \mu_{\kappa,i} \,/\, (1 - \phi_{\kappa,i}), \qquad \mathbb{V}\text{ar}(\kappa_{it}) = \sigma^2_{\kappa,i} \,/\, (1 - \phi^2_{\kappa,i}), \qquad \kappa = \xi, \beta.$$

Other, and possibly more complicated, dynamic structures for ξ_{it} and β_{it} can be considered as well but in our current study we shall consider only the first-order auto-regressive processes as given in expression (2.4).

2.2.3 Some extensions of the basic model

Our basic modelling framework for match results can be extended in several directions. First, we address the tendency of the bivariate Poisson distribution (2.1) to underestimate draws in match results, in particular when $\gamma = 0$, that is the double-Poisson model. For example, Dixon and Coles (1997) find that the scores 1-0 and 0-1 are underrepresented in their extended data set in favour of 0-0 and 1-1. They proposed to adjust their double-Poisson model by introducing an adjustment term that shifts probability mass from 1-0 and 0-1 towards 0-0 and 1-1. The adjustment is referred to as diagonal inflation and we apply it to the bivariate Poisson density function (2.1). The resulting density function is

obtained by multiplying the term $\pi_{\lambda_x,\lambda_y}(X,Y)$ with density function (2.1) where

$$\pi_{\lambda_x,\lambda_y}(X,Y) = \begin{cases} 1+\lambda_x\lambda_y\omega, & \text{if } (X,Y)=(0,0), \\ 1-\lambda_x\omega, & \text{if } (X,Y)=(0,1), \\ 1-\lambda_y\omega, & \text{if } (X,Y)=(1,0), \\ 1+\omega/\left\{1+(\gamma/\lambda_x\lambda_y)\right\}, & \text{if } (X,Y)=(1,1), \\ 1, & \text{otherwise,} \end{cases} \tag{2.6}$$

where coefficient ω determines how much probability mass is shifted. The multiplication leads to a proper density with moments that are the same as those of the bivariate Poisson distribution. A different but related adjustment was considered by Karlis and Ntzoufras (2003).

Another extension of our basic model is to allow for summer and winter breaks in league matches. In most football leagues, players can be bought or hired only during the summer and winter breaks. A change in the composition of a team can lead to changes in their strengths of attack and defence. We allow for such changes in the paths of ξ_{it} and β_{it} by letting the random shocks $\eta_{\xi,it}$ and $\eta_{\beta,it}$ respectively have different scalings for the first time period after the summer and winter breaks. When the processes for ξ_{it} and β_{it} are sufficiently persistent, large random shocks will lead to breaks in these processes. Hence we replace the distributions for the disturbance sequences $\eta_{\kappa,it}$ in expression (2.5) by

$$\eta_{\kappa,it} \sim \text{NID}\left\{0, \sigma^2_{\kappa,i} + \sigma^2_{\kappa,S}\tau_S(t) + \sigma^2_{\kappa,W}\tau_W(t)\right\}, \qquad \kappa = \xi, \beta, \tag{2.7}$$

for team $i = 1, \ldots, J$, where the indicator variables $\tau_S(t)$ and $\tau_W(t)$ are set equal to 1 at the end of the summer and winter breaks respectively, and to 0 otherwise, with $\sigma^2_{\kappa,S} > 0$ and $\sigma^2_{\kappa,W} > 0$. As a result, all disturbance variances are time-varying. The two additional variances for the strengths of both attack and defence are estimated jointly with the other parameters in ψ of equation (2.18) in Section 2.3.2.

Our final extension concerns the home ground advantage δ which is the same for all teams. It is realistic to expect that the home ground advantage has different effects on different teams. By introducing a team-specific home ground advantage in the model, the number of parameters increases and it will slow down the estimation process. A more feasible option is to limit this extension by pooling home ground advantage coefficients for groups of teams. For example, in Section 2.3.5 we consider a different home ground coefficient for the traditionally well-performing teams in the English Premier League: Arsenal, Chelsea, Liverpool, Manchester City and Manchester United. We may expect that for these teams the effect of home ground advantage on match results is higher. An interesting discussion of what home ground advantage represents is given by Pollard (2008).

2.2.4 General state space representation of the model

For our model-based analysis, it is convenient to present the model in the general state space form. The pair (X_{it}, Y_{jt}) is the observed outcome of the match of home team i against the visiting team j which is played at time t. The statistical dynamic model for the match result (X_{it}, Y_{jt}) of home team i against team j is given by

$$(X_{it}, Y_{jt}) \sim BP(\lambda_{x,ijt}, \lambda_{y,ijt}, \gamma), \tag{2.8}$$

where BP refers to the bivariate Poisson distribution with density function (2.1) and with the goal scoring intensities $\lambda_{x,ijt}$ and $\lambda_{y,ijt}$ specified via the link functions

$$\lambda_{x,ijt} = s_{x,ij}(\alpha_t), \qquad \lambda_{y,ijt} = s_{y,ij}(\alpha_t), \qquad i \neq j = 1, \ldots, J.$$

Here the so-called state vector α_t contains the strengths of attack and defence of all J teams at time t, i.e.

$$\alpha_t = (\xi_{1t}, \ldots, \xi_{Jt}, \beta_{1t}, \ldots, \beta_{Jt})', \qquad t = 1, \ldots, n. \tag{2.9}$$

Hence the dimension of the state vector is $2J \times 1$. We can represent the goal scoring intensity specifications in expression (2.3) by having the link functions as

$$s_{x,ij}(\alpha_t) = \exp(\delta + w_{ij}\,\alpha_t), \qquad s_{y,ij}(\alpha_t) = \exp(w_{ji}\,\alpha_t), \qquad i \neq j = 1, \ldots, J, \tag{2.10}$$

where w_{ij} selects the appropriate ξ_{it} and β_{jt} elements from α_t in expression (2.9). The transformation of the state vector into goal scoring intensities is illustrated in the Appendix. The bivariate Poisson distribution used in expression (2.8) relies further on dependence coefficient γ and $s_{x,ij}(\alpha_t)$ relies also on the home ground advantage coefficient δ. We collect such unknown coefficients in the parameter vector ψ for which more details are given below.

The linear dynamic process for the $2J \times 1$ state vector is given generally by

$$\alpha_t = \mu + T\alpha_{t-1} + \eta_t, \qquad \eta_t \sim \text{NID}(0, Q), \tag{2.11}$$

for $t = 1, \ldots, n$, where μ is the constant vector of dimension $2J \times 1$, T is the auto-regressive coefficient matrix of dimension $2J \times 2J$ and the disturbance vector η_t of dimension $2J \times 1$ is normally independently distributed with mean zero and variance matrix Q. The vector μ and matrices T and Q may rely partly on unknown coefficients which we also collect in the parameter vector ψ. The initial condition for the state vector α_1 can be obtained from the unconditional properties of α_t.

The dynamic specifications of the strengths of attack and defence in expression (2.4)

can be represented in the general form of expression (2.11) as follows. We collect the disturbances of expression (2.4) in $\eta_t = (\eta_{\xi,1t}, \ldots, \eta_{\xi,Jt}, \eta_{\beta,1t}, \ldots, \eta_{\beta,Jt})'$. Next we need to define only the matrices μ, T and Q as

$$\begin{aligned} \mu &= (\mu_{\xi,1}, \ldots \mu_{\xi,J}, \mu_{\beta,1}, \ldots, \mu_{\beta,J})', \\ T &= \text{diag}(\phi_{\xi,1}, \ldots \phi_{\xi,J}, \phi_{\beta,1}, \ldots, \phi_{\beta,J}), \\ Q &= \text{diag}(\sigma^2_{\xi,1}, \ldots \sigma^2_{\xi,J}, \sigma^2_{\beta,1}, \ldots, \sigma^2_{\beta,J}), \end{aligned}$$

where $\text{diag}(v)$ refers to a diagonal matrix with the elements of v on the leading diagonal. The constant vector μ is captured in the unknown initial state vector α_1. The remaining unknown coefficients are then placed in the parameter vector ψ. In this case we have

$$\psi = (\phi', q', \delta, \gamma)',$$

where the column vectors ϕ and q contain the diagonal elements of T and Q respectively. It can imply that the number of unknown coefficients is large and the burden of parameter estimation is high. In practice, we shall pool many unknown coefficients into a smaller set of parameters. This is illustrated in our empirical study of Section 2.3.

2.2.5 Evaluation of likelihood function and estimation

We opt for the method of maximum likelihood to obtain parameter estimates with optimal properties in large samples. Hence we develop an expression for the likelihood function of our model. For the evaluation of the likelihood function we require simulation methods because the multivariate model is non-Gaussian and nonlinear and hence we cannot rely on linear estimation methods for dynamic models such as the Kalman filter.

We have $J/2$ match results for each week t. A specific match result is denoted by (X_{it}, Y_{jt}) with $i \neq j$ and $i, j \in \{1, \ldots, J\}$. The number of goals scored by all teams in week t are collected in the $J \times 1$ observation vector y_t. The observation density of y_t for a given realization of the state vector α_t is then given by

$$p(y_t|\alpha_t; \psi) = \prod_{k=1}^{J/2} p_{BP}(\lambda_{x,ijt}, \lambda_{y,ijt}, \gamma), \tag{2.12}$$

where p_{BP} is the probability density function (2.1) of the bivariate Poisson distribution and where index k represents the kth match between home team i against visiting team j. We note that $\lambda_{x,ijt} = s_{x,ij}(\alpha_t)$ and $\lambda_{y,ijt} = s_{y,ij}(\alpha_t)$ where the link functions can, for example, be based on expression (2.3). In this case we can express the signal vector that

is associated with the density $p(y_t|\alpha_t;\psi)$ as

$$\mathbb{E}(y_t|\alpha_t;\psi) = \exp(a_t\delta + W_t\alpha_t), \tag{2.13}$$

where vector a_t consists of elements equal to 1 when the scores of the corresponding elements in y_t are from the home team and 0 otherwise, whereas matrix W_t is composed of the appropriate row vectors w_{ij} as introduced in expression (2.10). The home ground advantage coefficient δ is part of the parameter vector ψ.

We define $y = (y_1', \ldots, y_n')'$ and $\alpha = (\alpha_1', \ldots, \alpha_n')'$ for which it follows that

$$p(y|\alpha;\psi) = \prod_{t=1}^{n} p(y_t|\alpha_t;\psi). \tag{2.14}$$

It implies that, given the strengths of attack and defence in $\alpha_1, \ldots, \alpha_n$ and given the home ground advantage δ and the dependence coefficient γ, the scores from week to week are conditionally independent. Finally, we can express the joint density as $p(y, \alpha;\psi) = p(y|\alpha;\psi)p(\alpha;\psi)$ where

$$p(\alpha;\psi) = p(\alpha_1;\psi)\prod_{t=2}^{n} p(\alpha_t|\alpha_1, \ldots, \alpha_{t-1};\psi). \tag{2.15}$$

Given the linear Gaussian auto-regressive process for the state vector α_t in expression (2.11), the evaluation of $p(\alpha_t|\alpha_1, \ldots, \alpha_{t-1};\psi)$ is straightforward. The parameter vector ψ includes the coefficients $\phi_{\kappa,i}$ and $\sigma^2_{\kappa,i}$ for $\kappa = \xi, \beta$ and $i = 1, \ldots, J$. The evaluation of the initial density $p(\alpha_1;\psi)$ can be based on the unconditional properties of α_t. The constants $\mu_{\kappa,i}$, for $\kappa = \xi, \beta$ and $i = 1, \ldots, J$, are incorporated in the initial condition for α_1.

The likelihood function for y is based on the observation density (2.1) and is given by

$$\ell(\psi) = p(y;\psi) = \int p(y, \alpha;\psi)\mathrm{d}\alpha = \int p(y|\alpha;\psi)p(\alpha;\psi)\mathrm{d}\alpha, \tag{2.16}$$

which we want to evaluate for different values of the parameter vector ψ. An analytical solution to evaluate this integral is not available and therefore we rely on numerical evaluation methods. It is well established that numerical integration of a multi-dimensional integral becomes quickly infeasible when the dimension increases. We therefore adopt Monte Carlo simulation methods. We can use such methods since explicit expressions for the densities $p(y|\alpha;\psi)$ and $p(\alpha;\psi)$ are available. A naive Monte Carlo estimate of the likelihood function is given by

$$\hat{\ell}(\psi) = \frac{1}{M}\sum_{k=1}^{M} p(y|\alpha^{(k)};\psi), \qquad \alpha^{(k)} \sim p(\alpha;\psi), \tag{2.17}$$

where M is the number of Monte Carlo replications. Since the state vector density $p(\alpha;\psi)$ is associated with the auto-regressive process (2.11), we obtain $\alpha^{(k)}$ simply via the simulation of auto-regressive processes for a given parameter vector ψ. The draws $\alpha^{(1)},\ldots,\alpha^{(M)}$ are generated independently from each other. This Monte Carlo estimate is numerically not efficient (nor feasible) since the simulated paths are having no support from the observed data y. A more effective approach for the evaluation of the likelihood function is to adopt Monte Carlo simulation methods based on importance sampling as proposed by Shephard and Pitt (1997) and Durbin and Koopman (1997). The details of this estimation methodology for likelihood evaluation and for the signal extraction of strengths of attack and defence are discussed in the Appendix.

Parameter estimation is carried out via the maximisation of the likelihood function with respect to ψ by using standard numerical maximisation procedures. To obtain a smooth multi-dimensional likelihood surface in ψ for its maximisation, each likelihood evaluation is based on the same random numbers for the generation of M simulated paths of α. The method of maximum likelihood produces parameter estimates with optimal properties in large samples. These optimal properties remain when using Monte Carlo simulation methods whereas the estimates are subject to simulation error.

2.3 Empirical application

2.3.1 Data description

We analyse a panel time series of nine years of football match results from the English Premier League for which 20 football clubs are active in each season. The 20 football clubs that participate in a season vary because the three lowest placed teams at the end of the season are relegated. In the new season they are replaced by three other teams. The number of different teams in the panel is 36. Only 11 teams have played in all nine seasons of our sample and 10 teams have only played in one season. In the time dimension, we span a period from the season 2003–2004 to the season 2011–2012. The seasons run from August to May. Each team plays 38 matches in a season (19 home and 19 away games) so that in total we have 380 matches in the season. Most games are played in the afternoons of Saturdays and Sundays; the other games are played during weekday evenings (often on Mondays). The total number of matches played in our data set is $9 \times 380 = 3420$. The first seven years are used for parameter estimation and the last two years are used to explore the out-of-sample performance of the model. The data used in our study can be found at `http://www.football-data.co.uk`. Our data set of football match results can be treated as a time series panel of low counts. In approximately 85% of all matches in our sample, a team has scored only 0, 1 or 2 goals. The distribution of home and away

goals scored during the nine seasons is presented in the Appendix. Although working with low counts, a significant difference can be identified in the number of goals scored and conceded between the competing teams. A low ranking team rarely scores more than two goals in an away match while the top ranking teams sometimes reach scores of five or higher.

2.3.2 Details of the basic model

Our analysis of the Premier League football match results is based on the modelling framework presented in Section 2.2. The panel data set has $J = 36$ teams and we therefore need to estimate 36 attack strengths and 36 defence strengths over time; the dimension of the state vector α_t is 72. In comparison with other empirical studies where also state space time series analyses are carried out, the state vector is high dimensional. Since only 20 teams are active during a season, we need to treat large sections of the observations in the time series panel as missing. The state space methodology can treat missing observations in a routine manner; see the discussion in the Appendix. The time index t in our analysis does not refer to calendar weeks. Only weeks in a football season for which at least one match is played officially for the Premier League are indexed. The last week of football matches in one season and the first week in the next football season have then consecutive time indices. In our basic model of Section 2.2, the summer and winter breaks are not taken into account. In Section 2.2.3 we discuss a modification of our model that accounts for summer and winter breaks. If all teams play their matches weekly, each season consists of 38 weeks. However, owing to unforeseen circumstances, specific matches are postponed and extra time periods need to be added in the data set. The resulting calendar is adopted for the time index t in our analysis.

The dynamic processes of the strengths of attack and defence are given by expression (2.4) or collectively for the state vector by expression (2.11). Given the high number of teams, we restrict the auto-regressive coefficients and the disturbance variances to be the same among the teams:

$$\phi_{\xi,i} = \phi_\xi, \qquad \phi_{\beta,i} = \phi_\beta, \qquad \sigma^2_{\xi,i} = \sigma^2_\xi, \qquad \sigma^2_{\beta,i} = \sigma^2_\beta,$$

for $i = 1, \ldots, J$. These restrictions are not strong since we expect the persistence and the variation of the time-varying strengths of attack and defence to be small and similar between the teams. In other words, we expect the strengths of attack and defence for all teams to be evolving slowly over time. However, the strengths of attack and defence of the different teams can still evolve over time by following very different time paths. For the basic model, the home ground advantage δ and the dependence γ are assumed to be

the same for all teams and matches. The parameter vector is then given by

$$\psi = (\phi_\xi, \phi_\beta, \sigma_\xi^2, \sigma_\beta^2, \delta, \gamma)', \tag{2.18}$$

and is estimated by the method of Monte Carlo maximum likelihood of Section 2.2.5. The parameters in ψ are transformed during the estimation process so that the parameter values are within their restrictive ranges, which are

$$0 < \phi_\kappa < 1, \qquad \sigma_\kappa^2 > 0, \qquad \delta > 0, \qquad 0 < \gamma < c,$$

for $\kappa = \xi, \beta$ and where c represents the upper bound that is implied by the model and derived in the Appendix. The transformations for the elements in expression (2.18) are given by

$$\psi_j = \begin{cases} \frac{\exp \psi_j^*}{1+\exp \psi_j^*}, & j = 1, 2, \\ \exp(\psi_j^*), & j = 3, 4, 5, \\ \psi_j^* & j = 6, \end{cases} \tag{2.19}$$

where ψ_j is the jth element of ψ and ψ_j^* is the transformed coefficient that is actually estimated, for $j = 1, \ldots, 6$. We note that $\psi_6 = \gamma$ is not restricted because the upperbound c is implied by the model.

The signal extraction of the time-varying strengths of attack and defence has been carried out by the Monte Carlo methods described in Section 2.2.5. We have used a common set of random numbers to generate M simulated paths for α. The choice of M can be relatively low because we use efficient importance sampling methods; the details are provided in the Appendix. The computations have been implemented using the numerical routines developed and presented in Koopman, Shephard, and Doornik (2008); they are carried out on a standard computer. We have not encountered numerical problems while the computing times have been relatively low despite the high-dimensional state vector.

2.3.3 Parameter estimates

We present in Table 2.1 the parameter estimates for our time series panel of number of goals scored by teams in the English Premier League during the seven seasons from 2003–2004 to 2009–2010. To show the robustness of our Monte Carlo maximum likelihood methods, we present the estimates for various importance sampling replications M. The parameter estimates are robust to different choices of M. We may conclude that the choice of $M = 200$ is sufficient in our analysis but that we can also take $M = 50$ for repeated analyses of the model. Further evidence of the reliability of our results is presented in the Appendix.

The estimates of the auto-regressive coefficients of the latent dynamic processes for

the signals related to the strengths of attack and defence are close to one. They imply that the strengths of attack and defence are highly persistent and behave as random-walk processes. However, the auto-regressive coefficients reflect the persistence from week to week during the football seasons for which we do not expect many changes. More changes are expected from season to season in which a season consists of 38 weeks. When we consider the persistence of the signals from season to season, the implied estimates of the auto-regressive coefficients are equal to $(0.9985)^{38} = 0.94$ and $(0.9992)^{38} = 0.97$ which still imply persistent processes for the signals but they are stationary.

The estimated disturbance variances for the signals are relatively small, which illustrate that the attack and defence signals do vary over time in a smooth way. We emphasize that the estimated variances determine the scale of the fluctuations from week to week which we expect to be very small. We do not expect that a top team turns into a relegation candidate during one season. Furthermore, the number of goals in a match scored by one team is typically low. The main changes in the signals for strengths of attack and defence take place in the data over longer time periods.

Table 2.1: Monte Carlo estimates for the parameter vector ψ together with the value of the maximised log-likelihood value for different numbers of simulated paths $M = 50, 200, 1000$. The Monte Carlo estimates of the standard errors are given below the estimates and between parentheses. The dataset that was used for estimation covers seven seasons of the English Premier League (from 2003-2004 to 2009-2010)

ψ	$M = 50$	$M = 200$	$M = 1000$
ϕ_ξ	0.9985	0.9985	0.9985
	(0.00044)	(0.00044)	(0.00044)
ϕ_β	0.9992	0.9992	0.9992
	(0.00027)	(0.00027)	(0.00027)
σ_ξ^2	0.000205	0.000206	0.000206
	(2.20e-05)	(2.27e-05)	(2.28e-05)
σ_β^2	0.000141	0.000143	0.000143
	(2.05e-05)	(2.02e-05)	(2.02e-05)
δ	0.3662	0.3643	0.3641
	(0.0196)	(0.0269)	(0.0252)
γ	0.0966	0.0966	0.0966
	(0.0232)	(0.0232)	(0.0232)
$\hat{\ell}(\psi)$	-9608.56	-9608.38	-9608.38

2.3.4 Signal estimates of strengths of attack and defence

By replacing the parameter vector ψ with its estimate as given in Table 2.1, we can apply the Monte Carlo simulation methods of Section 2.2.5 to obtain the estimates for the attack and defence signals. The state vector α contains the strengths of attack and defence for all time periods and for all football teams. Once we have computed $\hat{\alpha}$, the importance

sampling estimate of the state vector, we can graphically present the estimated attack and defence signals over time together with their standard errors. We note that the standard errors are also computed using the importance sampling method; see the Appendix for details.

The estimation results of the previous section have indicated that the strengths of attack and defence do not fluctuate strongly from week to week but from season to season they can be more substantial. We present in Figure 2.1 the signal estimates for the time-varying strengths of attack and defence of the well-known football teams of Manchester United and Manchester City. The strength of attack of United has remained relatively constant from 2006 onwards whereas in the earlier years we observe an upwards trend in their strength of attack. The strength of attack of City has increased much more dramatically since 2007 and stabilized somewhat in the most recent season of 2011–2012. Manchester City has been able to invest more in high quality players in the previous five years owing to the new owners of the club. It is interesting to observe that the investments by City have been more directed towards forward players since the upward trend of the strength of attack is stronger than the trend of the strength of defence. An assessment of the strengths of attack and defence for all teams is presented in the Appendix.

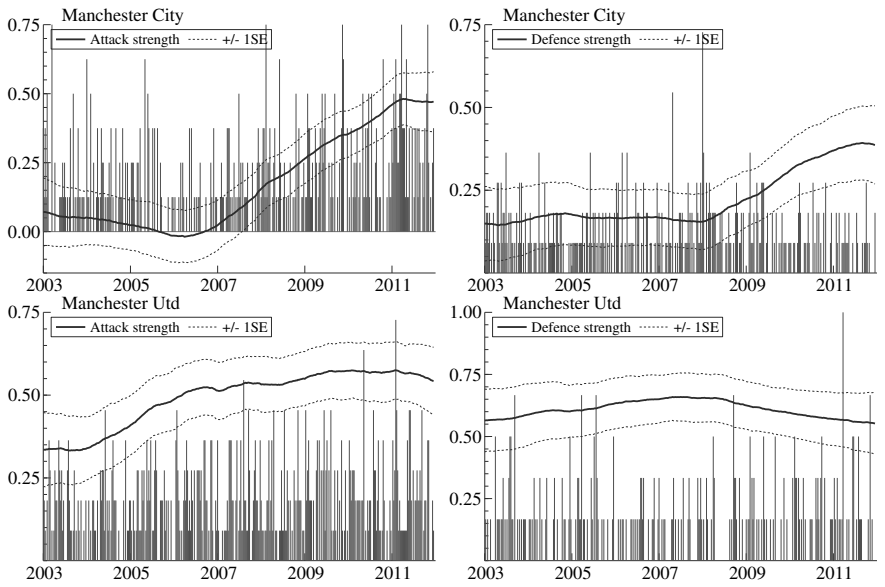

Figure 2.1: Strengths of attack and defence of the two highest ranking teams at the end of the 2011-2012 season of the English Premier League. The solid lines are the estimated strengths of attack and defence. The dotted lines provide the symmetric confidence intervals based on one standard error. The bars represent the number of goals scored and conceded from the 2003-2004 towards the 2011-2012 season which accounts for 404 time periods.

2.3.5 Model evaluations: in-sample and out-of-sample

To validate in-sample estimation and out-of-sample prediction results for the basic model, we present a selection of estimation and testing results for a set of extended, restricted and related model specifications. This study considers the following seven model specifications.

(a) the basic model with parameter estimates presented in Table 2.1.

(b) the basic model with time invariant strengths of attack and defence. (the auto-regressive processes (2.4) for ξ_{it} and β_{it} are replaced by fixed coefficients; the state vector (2.9) reduces to $\alpha_t = \mu$ in expression (2.11); hence we can adopt the same state space time series analysis but with system matrices $T = 0$ and $Q = 0$ in expression (2.11); the parameter vector consists only of the dependence parameter γ and home ground advantage δ);

(c) the basic model with dependence parameter set equal to zero, i.e. $\gamma = 0$ in expression (2.1) (the observation model reduces to a double-Poisson distribution for match outcomes);

(d) the basic model with a time-varying, team-specific dependence parameter given by

$$\gamma_{ijt} = \gamma^* \sqrt{\lambda_{x,ijt}\lambda_{y,ijt}}, \qquad \gamma^* \geq 0, \tag{2.20}$$

where γ^* is a scaling coefficient that replaces γ in the parameter vector given in expression (2.18) (the dependence coefficient is time-varying owing to its dependence on the strengths of attack and defence; this specification was proposed by Goddard (2005) but the time-varying feature of the dependence in expression (2.20) has not been considered before);

(e) the diagonal inflation model for which the density function (2.1) is multiplied by expression (2.6) (the coefficient ω in expression (2.6) is added to the parameter vector (2.18));

(f) the basic model with time-varying strengths of attack and defence that account for the summer and winter breaks (the disturbance variances for the state vector are time-varying as specified in expression (2.7) and the additional variance parameters are added to the parameter vector (2.18); here we concentrate on only the long summer break);

(g) the basic model with two home ground advantage parameters, δ_1 for the group {Arsenal, Chelsea, Liverpool, Manchester City, Manchester United} and δ_2 for the group with all other teams; see the discussion in Section 2.2.3.

In-sample evaluation

For all the model specifications $(a) - (g)$ reported above, we have estimated the parameter vector by the method of Monte Carlo maximum likelihood using the match results in seven seasons of the English Premier League, from 2003–2004 to 2009–2010. Importance sampling methods are used for likelihood evaluation by using a simulation sample size of $M = 50$. The same random draws are used for each model specification, for each parameter vector and for each likelihood evaluation. The usual t-test (for a single restriction) and likelihood ratio statistics are used for the in-sample validation of the restricted and extended model specifications $(a) - (g)$. The test statistics are computed on the basis of maximum likelihood estimates of the parameter vector ψ. Under standard regularity conditions and for sufficiently large sample sizes, the reported t-test and likelihood ratio statistics converge in distribution to a standard normal and a χ^2 distribution with k degrees of freedom, where k is the number of restrictions, respectively. The test statistics are reported in Table 2.2.

A major aspect of our basic model (a) is the inclusion of time-varying strengths of attack and defence. Model (b) reduces the strengths of attack and defence to fixed coefficients. By comparing models (a) and (b) using the likelihood ratio statistic, we conclude that model (b) is not supported by our data set. Another key aspect of our basic model is the use of the bivariate Poisson distribution rather than the double-Poisson distribution as adopted in model (c). The test statistic for model (c) provides clear evidence that our data set favours the model with dependence between the match results. With respect to model (d), we find that the estimated dependence coefficient γ^* in expression (2.20) is significant. However, the dependence as specified by Goddard (2005) is not strongly favoured in our data set since the maximised likelihood value for basic model (a) is somewhat higher than the maximised likelihood value for model (d).

To account for the overrepresentation of draws in the data set, we consider the diagonal inflation model (e). The maximum likelihood estimate of ω in expression (2.6) is not significant although the t-test statistic is positive and close to the critical value of 1.96. Hence the number of draws 0-0 and 1-1 implied by our basic model is somewhat too small for our data set.

Model (f) allows for breaks in the strengths of attack and defence after the winter and summer holidays in the football calendar. It requires the estimation of four additional variances in the parameter vector. The estimated variances for the winter breaks are not significant and close to zero. Hence we have re-estimated the model with two additional variances for the summer break only. The two estimated variances have almost equal values. In our final specification we therefore restrict the two summer break variances to be equal to each other. The restricted variance estimate is highly significant as indicated by the reported t-test statistic in Table 2.2. It also affects the estimates of the other

variances in the model. In particular, the dynamic coefficients for attack, ϕ_ξ and σ_ξ^2 are estimated to be close to one and zero respectively. It implies that the strength of attack is close to a constant within each season and its evolution over time behaves as a step function with breaks at the beginning of each football season. The dynamic coefficients for the strength of defence are not affected in the same way. The strength of defence continues to vary within the season at a slow pace. We present the estimated patterns of attack and defence of Manchester United and Manchester City from model (f) in Figure 2.2. We can compare these patterns with those presented in Figure 2.1 for our basic model (a). The patterns for a selection of other teams in the English Premier League are presented in the Appendix.

Finally, we verify whether the home ground advantage is different for the larger teams in the English Premier League. The home ground advantage parameters δ_1 and δ_2 in model (g) are estimated as an extension of our basic model. The null hypothesis $H_0 : \delta_1 = \delta_2$ cannot be rejected given the low value of the reported t-test in Table 2.2. Hence the home ground advantage is not significantly different for the larger teams in our data set.

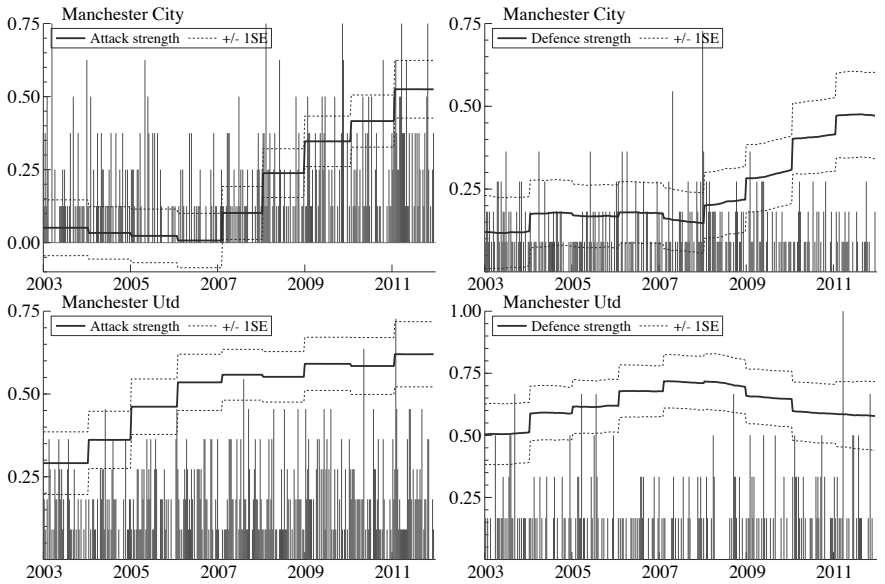

Figure 2.2: Strengths of attack and defence of the two highest ranking teams at the end of the 2011-2012 season of the English Premier League. The stepwise evolution of the patterns is due to an additional variance for the strengths of attack and defence at the start of the new football season after the summer break. The solid lines are the estimated strengths of attack and defence. The dotted lines provide the symmetric confidence intervals based on one standard error. The bars represent the number of goals scored and conceded from the 2003-2004 towards the 2011-2012 season which accounts for 404 time periods.

Out-of-sample evaluation

For the out-of-sample evaluation of the considered models, we carry out a one-step ahead forecasting study. For each model, we forecast the outcome of the matches in the football seasons 2010–2011 and 2011–2012 using a so-called rolling window strategy. We estimate the parameter vector for the time series of all match results from the seven seasons. At time T, the week before the first week of football season 2010–2011, we forecast the match outcomes for the first week of the season 2010–2011, i.e. time $T+1$, based on a specific model and the estimated parameter vector. We then can compare the forecasts with the actual outcomes. The differences between realisations and forecasts are collected in the 20×1 forecast error vector e_{T+1}. Next we compute the sum of squared errors, which we take as our loss function, that is $L_{T+1} = e'_{T+1} e_{T+1}$. This loss function is computed for each model m, i.e. $L_{t+1}^{(m)}$ for $m = a, \ldots, g$. The difference in accuracy compared to our main model can be measured as $d_{T+1}^{(m)} = L_{T+1}^{(a)} - L_{T+1}^{(m)}$ for $m = b, \ldots, g$. For the next period $T+1$, we re-estimate the parameter vector by including the match results of time $T+1$ in our data but removing the match results in the first week of our sample, seven years ago. Hence the estimation sample length remains constant when re-estimating the parameter vector for producing the next forecasts. This procedure of re-estimation and forecasting is then repeated for each week in the two football seasons that we use for our out-of-sample evaluations.

The difference in the one-step-ahead predictions of the models, $d_j^{(m)}$, for $j = T+1, \ldots, T+N$ with out-of-sample length N, are compared with each other with the Diebold-Mariano (DM) test statistic; see Diebold and Mariano (1995). The test is designed for the null hypothesis of equal out-of-sample predictive accuracy between two competing models. The DM test statistic for model m is computed by (i) taking the average of the out-of-sample computed values $d_j^{(m)}$'s over time, for each $m = b, \ldots, g$; (ii) standardizing this average by a consistent measure of the long-term variance of d_j. We require the long-term variance because the time series of d_{t+1} is serially correlated by construction since at least only one of the two competing models can be correctly specified. In general, the DM test statistic should not be applied when we compare the predictive accuracy between two nested models since the numerator and denominator of the DM test statistic have their limits at zero, when the in-sample and out-of-sample dimensions increase. However, it is argued by Giacomini and White (2006) that the DM test statistic can still be applied as long as the forecasts are generated with a rolling window and for a relatively short out-of-sample horizon. Diebold and Mariano (1995) show that the DM test statistic is asymptotically distributed as a standard normal random variable. Hence, we reject the null hypothesis of equal predictive accuracy at the 5% significance level if the absolute value of the DM test statistic is larger than 1.96. The resulting loss function values and DM test statistics in our out-of-sample forecasting study are reported in Table 2.2.

Table 2.2: We compare the in-sample fit and out-of-sample forecasting accuracy for seven model specifications. The number of parameters #pars (p_1/p_0) is given for each model (p_1) and for the model under the null hypothesis H_0 (p_0), see Section 2.3.5 for further details. The in-sample results are based on seven seasons of the English Premier League (from 2003-2004 to 2009-2010). The t-tests are computed and presented for the hypotheses with a single restriction while the likelihood ratio (LR) test is presented for the multiple restriction in (b). The out-of-sample results are based on the two seasons 2010-2011 and 2011-2012. The squared loss functions and the Diebold-Mariano (DM) tests are based on one-step ahead forecasts from a rolling window sample. The test statistic values with ** indicate significance at the 5% significance level.

	Model specification	#pars	H_0	LR test	t-test	Sqr loss	DM
(a)	Basic model	6				2087.10	
(b)	... time-invariant signals	6/2		123.04**		2190.80	–3.67**
(c)	... no dependence	6/5	$\gamma = 0$		4.16**	2087.90	–0.62
(d)	... time-varying dependence	6/5	$\gamma^* = 0$		3.84**	2088.60	–1.51
(e)	Diagonal inflation model	7/6	$\omega = 0$		1.85	2086.70	0.91
(f)	Summer break for signals	7/6	$\sigma^2_{\kappa,S} = 0$		2.84**	2098.50	–1.75
(g)	Two home ground advantages	7/6	$\delta_1 = \delta_2$		0.35	2089.00	–1.08

The out-of-sample squared loss function values reported in Table 2.2 show that model (e) has the smallest loss compared with all other models. Except for models (b) and (f), the forecast losses of the other models are only small and similar in size. This finding is confirmed by the reported DM test statistics which indicate that we cannot reject the hypothesis any of the models (c), (d), (e) and (g) are equally accurate as model (a) in our out-of-sample forecasting exercise. Although the same conclusion can be drawn for model (f), this model is closest to rejection and appears to provide less accurate forecasts. The stepwise evolution of the strengths of attack and defence from season to season may have a negative impact on its forecasting ability. Given the non-significant DM test statistic for model (c) and despite the in-sample significance of the dependence parameter γ, it appears that the presence of γ does not have much impact on the out-of-sample forecast performance of the basic model. This finding may be due to our choice of a relatively short out-of-sample forecasting window. The only significant DM statistic is reported for model (b) which is consistent with our in-sample rejection of the null hypothesis of time invariant signals. Overall we can conclude that the model extensions of Section 2.2.3 do not lead to significant improvements in their forecast performance except for model (e). However, the extensions may be more beneficial for longer forecast horizons and for other data sets.

2.4 Out-of-sample performance in a betting strategy

Finally, we verify the out-of-sample performance of our basic model (a) in a realtime study into the betting on a win, a loss or a draw of the home team for a weekly selection

of matches during the two seasons of 2010–2011 and 2011–2012. The betting on matches in the English Premier League is immense popular and is a truly world-wide activity. In our betting evaluation study we carry out the same out-of-sample rolling window strategy as used in the previous section. At time T, we estimate the model parameters and forecast the intensities $\lambda_{x,ij,T+1}$ and $\lambda_{y,ij,T+1}$. The resulting full distributional properties of the next ten games implied by the bivariate Poisson model (2.1), with its unknown parameters replaced by their estimates, enables us to compute the probabilities of all possible outcomes of a match. Hence we can compute the probabilities of a win, a loss or a draw, for each match. We can now visit the bookmaker's office and bet on matches accordingly.

Different betting strategies can be pursued and we illustrate our basic and conservative strategy by using an example. Consider the first match of the out-of-sample 2010–2011 season where Aston Villa plays against West Ham. The intensity forecasts are $\lambda_{x,ij,t+1} = 1.7272$ and $\lambda_{y,ij,t+1} = 0.8127$ which correspond to win, loss and draw probabilities for the home team of 0.591, 0.174 and 0.235 respectively. The bookmaker offers the following odds for the home team: 1.96 for a win, 4.03 for a loss and 3.30 for a draw. For each outcome, the expected value (EV) of a unit bet on an event A is given by

$$\mathrm{EV}(A) = P(A)\left\{\mathrm{odds}(A) - 1\right\} - P(A^c) \times 1 = P(A)\mathrm{odds}(A) - 1,$$

where event A represents a win, a loss or a draw of the home team, A^c is the complement of A, $P(A)$ is the probability of event A and $\mathrm{odds}(A)$ is the bookmaker's odds for event A. In our illustration we obtain 0.159, -0.300 and -0.224 as expected values for a unit bet on a win, a loss and a draw for the home team respectively. A basic strategy could be to bet on all events for which the expected value is positive, $\mathrm{EV}(A) > 0$. In this illustration we bet on a win for the home team. However, we shall consider a less risky betting strategy which is based on the following guidelines. First, we bet only on 'quality' events which are defined as bets with EVs that exceed some benchmark τ, i.e. $\mathrm{EV}(A) > \tau$ for some $\tau > 0$. Second, we also consider longshot events which are defined as small probability events with very high odds. The probability of losing the bet on a longshot is of course high. We consider events with odds higher than 7 as longshots. Our basic strategy consists of betting a unit value on each quality event for some value of τ. We also bet on longshots but reduce the bet to a fixed value of 0.3 units. The definition of a longshot and the bet sizes in our basic betting strategy are assumptions and are not based on optimizing payoff or minimizing variance betting strategies. A betting strategy that determines which proportion of the bettors bankroll should be risked in a sequence of positive expected value bets to maximize the growth rate of the bankroll was determined by Kelly (1956). Since this betting exercise is only meant as an illustration of the performance of our model we refrain from more advanced betting strategies like the Kelly betting strategy.

The 2010–2011 and 2011–2012 season of the English Premier League consist of 760 matches and therefore 760 is the maximum number of betting opportunities. It is, however, not guaranteed that a match is a betting opportunity since $\mathrm{EV}(A)$ can be negative for all possible match outcomes (win, draw, loss). The expected and actual profit for all our bets in the 2010–2011 and 2011–2012 seasons can be determined for a range of τ values. Confidence intervals for the mean return are obtained by a bootstrap related method in which football matches are sampled with replacement. We emphasize that we do not take model or parameter uncertainty into account here but we solely focus on the effect of '(un)lucky betting streaks'. Match outcomes, bookmakers' odds and time-varying parameters (and thereby match probabilities) are fixed. The only uncertainty comes from the drawing of 760 matches, with replacement, out of the pool of 760 matches which is repeated a 1000 times.

The odds for betting are offered by many different bookmakers. We consider the average odds taken from 28 to 40 bookmakers (depending on the match) which are collected online at `http://www.football-data.co.uk`. In the example match between Aston Villa and West Ham, the implied probabilities given by the bookmakers' odds were, on average, 1/1.96, 1/4.03 and 1/3.30 for a win, a loss and a draw respectively. The sum of these probabilities is 106.1%. Everything above the 100% is the profit of the bookmaker (or the bookmaker's edge) which is 7% on average. This means that the expected profit under random betting of a unit value is -0.07. Random betting is referred to as having a unit bet on a win, a loss or a draw randomly chosen for each match. Hence our betting strategy must achieve an overall return that overtakes the bookmaker's edge of 7% but also generates a positive overall return.

In Figure 2.3 we present the outcomes of our betting strategy for various values of τ. In the first panel (i) the overall return is presented as the full curve and is compared with the negative overall return of 7%, the bookmaker's edge. The 90% bootstrap confidence interval is represented by the dotted curves. A similar graph was presented by Dixon and Coles (1997). For $\tau = 0$, the majority of betting opportunities is marked by the model as quality bets, i.e. $\mathrm{EV}(A) > 0$. We start to obtain positive mean returns at $\tau > 0.12$ and although the confidence interval often includes 0, we expect to outperform a random betting strategy for higher values of τ. The number of betting opportunities becomes small, less than 40, for $\tau = 0.45$. Hence the generated mean returns for $\tau > 0.45$ are not reliable as reflected by the bootstrap confidence intervals. We therefore do not display mean returns for $\tau > 0.45$ in Figure 2.3.

We observe that, for small values of τ, the forecasts of our model imply a zero return on average and a negative return on average also finds support in the 90% interval. When the benchmark τ for a quality bet increases, the number of actual bets decreases in our strategy as is shown in panel (ii) of Figure 2.3. However, the quality bets from a higher

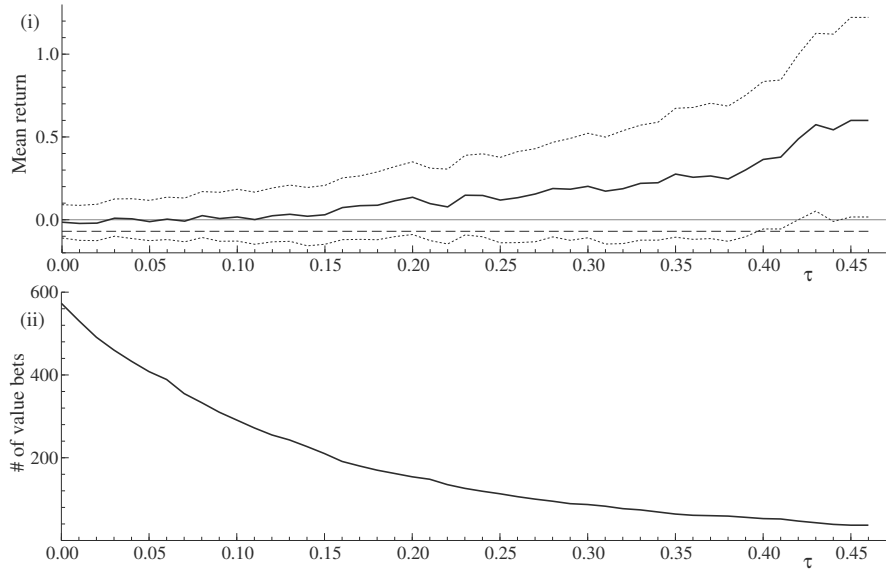

Figure 2.3: Returns of a betting strategy for the 2010-2011 and 2011-2012 seasons of the English Premier League: (i) average return from betting on match outcomes by our strategy for various values of the threshold τ; the dashed line represents the average return under random betting which we have established at -0.07; the dotted curves are 90% bootstrap confidence intervals. (ii) number of quality bets for various values of τ out of the 760 betting opportunities in the two seasons.

benchmark will also provide us with a higher return on average as we learn from panel (i).

The average return curve in Figure 2.3(i) is not smooth in τ. This is partly due to the role of longshots in this exercise. For example, at $\tau = 0.11$, we have 74 longshots from which eight have been correct, resulting in a net profit of 5.07 units. Even when we bet with 0.3 units for longshots, the betting strategy remains highly variable because for another value of τ, another small number of correct longshots is obtained that can lead to a very different net profit. A more advanced betting strategy takes into account the variation of odds. We abstain from such more advanced strategies since we only want to illustrate the performance of our model in a basic and simple betting strategy. The presented results can be used as a benchmark for the more advanced betting strategies based on our model. We regard this validation study as only an example of how our modelling framework can be used in practice. Results are obtained only for the 2010–2011 and 2011–2012 seasons of the English Premier League and results may differ greatly for other seasons and/or football competitions.

2.5 Conclusions

We have presented a non-Gaussian state space model for the analysis and forecasting of football matches. Our basic model takes a match result as a pairwise observation that is assumed to come from a bivariate Poisson distribution with intensity coefficients for the number of goals scored by the two teams and a dependence coefficient for measuring the correlation between the two scores. The intensity coefficients depend on the strengths of attack and defence of the teams and they are allowed to evolve stochastically over time. The intensities are also subject to a fixed coefficient for home ground advantage. The resulting dynamic bivariate Poisson model is a novelty and can be used for the analysis of match results in many different competitions for team sports. Several extensions of the basic model have been considered including amendments for the overrepresentation of draws in data sets, breaks in strengths of attack and defence after winter and summer breaks, and a team-specific home advantage. Our empirical study is for a data set of match results from nine seasons of the English Premier League. The two seasons 2010–2011 and 2011–2012 are used as an out-of-sample evaluation period for the forecasting of football match results. The model-based forecasts are of sufficient accuracy to beat a random betting strategy and can be used as the basis for a more advanced betting strategy. The confidence interval for the return on betting often includes zero return which can be viewed as a sign of market efficiency. The betting market for the English Premier League is also a liquid betting market, especially when compared to betting markets for other football leagues.

Although we have presented promising results for our basic model and some of its extensions, we believe that further improvements can be made in different directions. First, other dynamic model specifications for the strengths of attack and defence can be considered such as random walk or long memory processes. Second, our statistical modelling framework only uses match results as data. The forecasting performance of the model can be further improved by adding more information about the matches. For example, potential explanatory variables for match results are the duration between matches, the traveling distance of the visiting team and shots on target which can be included as covariates or be part of a mixed measurement framework to increase forecasting results further. Third, our statistical analysis is carried out from a classical perspective. Bayesian Markov Chain Monte Carlo methods can be used to obtain predictive densities that account for parameter uncertainty. Fourth, given the popularity of betting on football matches, the odds provided by bookmakers are expected to be highly efficient. In such a liquid market of football betting, one can easily find higher odds than the averages that we have used in our study. More advanced betting strategies, like Kelly (1956) betting, that take account of the variance of a bet can improve (long run) returns further.

Appendices

The following appendices are part of the chapter 'A dynamic bivariate Poisson model for analysing football match results' and are organised as follows. Appendix A, B and C provide the specific details of the estimation methodology that we pursue in the main chapter. Details are given on, likelihood evaluation, construction of the approximating model and the derivatives for the model observation density. Appendix D provides further computational details and Appendix E shows figures of the data used in the main chapter and tables with strengths of attack and defence off all teams at different time points.

A Likelihood evaluation

Given our model specification for the time series of pairs of counts collected in y with its dependence on the states in α, we can express the likelihood function $\ell(\psi)$ as given by (2.16). The individual observations and states at time t are indicated by y_t and α_t, respectively; see the discussion in Section 2.4. We evaluate the integral numerically by the method of importance sampling as developed by Shephard and Pitt (1997) and Durbin and Koopman (1997), hereafter referred to as SPDK. A comprehensive treatment of the method, together with other and related methods, is provided by (Durbin and Koopman, 2012, Part II). The SPDK method is based on an approximating linear Gaussian model $\mathrm{g}(y, \alpha; \psi)$ which allows us to compute the approximate likelihood function $\mathrm{g}(y; \psi)$ by means of the Kalman filter and to simulate random samples for α from $\mathrm{g}(\alpha|y; \psi)$ by means of the simulation smoother; see the discussions in Jungbacker and Koopman (2007). The simulated random samples for α will give a better support to y although they come from an approximating model.

The likelihood function of the Gaussian model $\mathrm{g}(y, \alpha; \psi) = \mathrm{g}(y; \psi)\mathrm{g}(\alpha|y; \psi)$ can be expressed as

$$\ell_{\mathrm{g}}(\psi) = \mathrm{g}(y; \psi) = \frac{\mathrm{g}(y, \alpha; \psi)}{\mathrm{g}(\alpha|y; \psi)} = \frac{\mathrm{g}(y|\alpha; \psi)p(\alpha; \psi)}{\mathrm{g}(\alpha|y; \psi)}, \tag{2.21}$$

since $p(\alpha; \psi) \equiv \mathrm{g}(\alpha; \psi)$. Substituting $p(\alpha; \psi) = \mathrm{g}(y; \psi)\mathrm{g}(\alpha|y; \psi)/\mathrm{g}(y|\alpha; \psi)$ into (2.16), we obtain

$$\ell(\psi) = \mathrm{g}(y; \psi) \int \frac{p(y|\alpha; \psi)}{\mathrm{g}(y|\alpha; \psi)} \mathrm{g}(\alpha|y; \psi)\mathrm{d}\alpha = \ell_{\mathrm{g}}(\psi)\mathbb{E}_{\mathrm{g}} \left\{ \frac{p(y|\alpha; \psi)}{\mathrm{g}(y|\alpha; \psi)} \right\}, \tag{2.22}$$

where $\mathbb{E}_{\mathrm{g}}$ refers to expectation with respect to the Gaussian density $\mathrm{g}(\alpha|y; \psi)$. This method has proved to work effectively for multivariate time series models; see, for example, Koopman and Lucas (2008). In our model specification, the individual observations y_t are independent for given α_t as implied by (2.12) for $t = 1, \ldots, n$. Hence we can also assume that $\mathrm{g}(y|\alpha; \psi) = \prod_{t=1}^{n} \mathrm{g}(y_t|\alpha_t; \psi)$. The construction of an approximating model is discussed in Appendix B.

For a given approximating model, we estimate the likelihood function via Monte Carlo simulation as

$$\hat{\ell}(\psi) = \ell_{\mathrm{g}}(\psi)\frac{1}{M}\sum_{i=1}^{M} w_i, \qquad w_i = \frac{p(y|\alpha^i;\psi)}{\mathrm{g}(y|\alpha^i;\psi)}, \qquad \alpha^i \sim \mathrm{g}(\alpha|y;\psi), \tag{2.23}$$

where w_i is referred to as an importance weight, $\ell_{\mathrm{g}}(\psi)$ is obtained from the Kalman filter and α^i is computed by the simulation smoother for $i = 1, \ldots, M$. We can refer to $\hat{\ell}(\psi)$ as the importance sampling estimate of the likelihood function. For the purpose of likelihood maximisation with respect to ψ, it is preferred to work with the loglikelihood function. Taking the log of $\hat{\ell}(\psi)$ in (2.23) introduces a bias that can be accounted for in the usual way; see Durbin and Koopman (1997).

The effectiveness of the importance sampling method for likelihood evaluation relies on the properties of the importance sampling weight function $w_i = w(y, \alpha; \psi)$; see Geweke (1989) who provides conditions for $w(y, \alpha; \psi)$ under which a central limit theorem is valid for the estimate $\hat{\ell}(\psi)$. An important condition is the existence of a variance for weight function $w(y, \alpha; \psi)$. Based on a sample of importance weights $w_1, \ldots, w_M$, Koopman, Shephard, and Creal (2009) discuss diagnostic test statistics to validate the existence of a variance for the importance sampling weights.

B Construction of approximating model

For the implementation of the SPDK importance sampling method, the approximating linear Gaussian state space model is given by

$$\mathrm{g}(y, \alpha; \psi) = \mathrm{g}(y|\alpha; \psi)\mathrm{g}(\alpha; \psi) = \mathrm{g}(\alpha; \psi)\prod_{t=1}^{n} \mathrm{g}(y_t|\alpha_t; \psi), \tag{2.24}$$

with $\mathrm{g}(\alpha; \psi)$ the density of the dynamic state process (2.11) and we let $\mathrm{g}(y_t|\alpha_t; \psi)$ be represented by the linear Gaussian model equation

$$y_t = a_t\delta + W_t\alpha_t + c_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, V_t), \qquad t = 1, \ldots, n, \tag{2.25}$$

or more explicitly

$$\mathrm{g}(y_t|\alpha_t; \psi) = \mathrm{NID}(a_t\delta + W_t\alpha_t + c_t, V_t), \qquad t = 1, \ldots, n, \tag{2.26}$$

where vector a_t has element 1 if the number of goals in the corresponding element of y_t is from a home team and 0 otherwise, matrix W_t, with elements of 1s, 0s and -1s, selects the attack (+1) and defence (-1) strengths of the relevant teams, and mean correction

c_t and variance V_t are selected such that the first and second derivatives of logdensities $\log p(y_t|\alpha_t;\psi)$ and $\log \mathrm{g}(y_t|\alpha_t;\psi)$ with respect to α_t are equal to each other, for $t = 1,\ldots,n$. We note that $a_t\delta + W_t\alpha_t$ represents the signal as also defined in (2.13). Closed-form solutions of these two sets of n equalities are not available and hence we solve them iteratively with the use of the Kalman filter and smoother; more details and discussions are given by Jungbacker and Koopman (2007). The approximating model $\mathrm{g}(y,\alpha;\psi)$ is effectively a second-order Taylor expansion of the true model and it is also equivalent to computing the mode of $p(\alpha|y;\psi)$ for α; see the discussions in Durbin and Koopman (1997), So (2003) and Jungbacker and Koopman (2007). Our application for the bivariate Poisson model is not straightforward and we require to provide some further clarification. We will briefly discuss these necessary details for a successful implementation next.

To obtain values for c_t and V_t in (2.25), we need to solve the equations

$$\dot{\mathrm{g}}_t(\alpha_t) = \dot{p}_t(\alpha_t), \qquad \ddot{\mathrm{g}}_t(\alpha_t) = \ddot{p}_t(\alpha_t), \qquad t = 1,\ldots,n,$$

where

$$\dot{p}_t(\alpha_t) = \frac{\partial \log p(y_t|\alpha_t;\psi)}{\partial \alpha_t}, \qquad \ddot{p}_t(\alpha_t) = \frac{\partial^2 \log p(y_t|\alpha_t;\psi)}{\partial \alpha_t \partial \alpha_t'},$$

and $\dot{\mathrm{g}}_t(\alpha_t)$ and $\ddot{\mathrm{g}}_t(\alpha_t)$ are defined similarly. It follows straightforwardly that

$$\dot{\mathrm{g}}_t(\alpha_t) \equiv W_t' V_t^{-1}(y_t - c_t - a_t\delta - W_t\alpha_t), \qquad \ddot{\mathrm{g}}_t(\alpha_t) \equiv -W_t' V_t^{-1} W_t, \qquad t = 1,\ldots,n.$$

The derivatives for $\log p(y_t|\alpha_t;\psi)$ are more intricate and we develop expressions for $\dot{p}_t(\alpha_t)$ and $\ddot{p}_t(\alpha_t)$ in the next section. Hence we obtain expressions for c_t and V_t by

$$V_t = -W_t\ddot{p}_t^{-1}(\alpha_t)W_t', \qquad c_t = y_t - a_t\delta - W_t\left[\alpha_t + \ddot{p}_t^{-1}(\alpha_t)\dot{p}_t(\alpha_t)\right], \qquad t = 1,\ldots,n. \quad (2.27)$$

The mean c_t and variance V_t depend on the state vector α_t and hence we solve these equations iteratively. For starting values of c_t and V_t, we construct the linear Gaussian state space model for $\mathrm{g}(y,\alpha;\psi)$ and apply the Kalman filter smoother to obtain $\hat{\alpha} = \mathbb{E}_{\mathrm{g}}(\alpha|y;\psi)$. From the value $\alpha = \hat{\alpha}$, we can obtain new values for c_t and V_t and can construct or update a new approximating model. The Kalman filter smoother produces a new $\hat{\alpha}$ and we iterate this process until convergence. When this process has converged, the linear Gaussian model with the final values for c_t and V_t represents the approximating model $\mathrm{g}(y,\alpha;\psi)$ as given by (2.25). It is well established that the Kalman filter and related methods can treat missing observations straightforwardly; see the discussions in (Durbin and Koopman, 2012, Part I).

C The derivatives for the model observation density

Equation (2.12) implies that the matches played at time t, for a given α_t, are treated as independent events. Hence we can treat each match separately. A match is for home team i and visiting team j. The scoring intensities for both teams are collected in the 2×1 vector $\lambda_{ijt} = (\lambda_{x,ijt}, \lambda_{y,ijt})'$ which are functions of α_t, that is $\lambda_{ijt} = s_{ij}(\alpha_t)$ since $\lambda_{x,ijt} = s_{x,ij}(\alpha_t)$ and $\lambda_{y,ijt} = s_{y,ij}(\alpha_t)$; see the discussion in Section 2.3. The first derivative of the log of the bivariate Poisson density (2.1) with respect to α_t can be obtained via the chain rule as

$$\frac{\partial \log p(X, Y; \lambda_{x,ijt}, \lambda_{y,ijt}; \gamma)}{\partial \alpha_t} = \dot{s}_{ij}(\alpha_t) \times \dot{p}_\lambda(\lambda_{ijt}),$$

where X and Y are specific elements of y_t and represent the numbers of goals scored by teams i and j, respectively, at time t, and where

$$\dot{s}_{ij}(\alpha_t) = \frac{\partial \lambda'_{ijt}}{\partial \alpha_t}, \qquad \dot{p}_\lambda(\lambda_{ijt}) = \frac{\partial \log p(X, Y; \lambda_{x,ijt}, \lambda_{y,ijt}; \gamma)}{\partial \lambda_{ijt}}.$$

The second derivative can be obtained in the same way, that is

$$\frac{\partial^2 \log p(X, Y; \lambda_{x,ijt}, \lambda_{y,ijt}; \gamma)}{\partial \alpha_t \partial \alpha'_t} = \dot{s}_{ij}(\alpha_t) \times \ddot{p}_\lambda(\lambda_{ijt}) \times \dot{s}_{ij}(\alpha_t)',$$

where

$$\ddot{p}_\lambda(\lambda_{ijt}) = \frac{\partial^2 \log p(X, Y; \lambda_{x,ijt}, \lambda_{y,ijt}; \gamma)}{\partial \lambda_{ijt} \partial \lambda'_{ijt}}.$$

An expression for $\dot{s}_{ij}(\alpha_t)$ is obtained easily for link functions $s_{x,ij}(\alpha_t)$ and $s_{y,ij}(\alpha_t)$ as given by (2.3).

The general expressions for $\dot{p}_\lambda(\lambda_{ijt})$ and $\ddot{p}_\lambda(\lambda_{ijt})$ follow from (2.1) and are decomposed as

$$\dot{p}_\lambda(\lambda_{ijt}) = \begin{pmatrix} \dot{p}_{\lambda_x}(\lambda_{ijt}) \\ \dot{p}_{\lambda_y}(\lambda_{ijt}) \end{pmatrix}, \qquad \ddot{p}_\lambda(\lambda_{ijt}) = \begin{bmatrix} \ddot{p}_{\lambda_{xx}}(\lambda_{ijt}) & \ddot{p}_{\lambda_{xy}}(\lambda_{ijt}) \\ \ddot{p}_{\lambda_{xy}}(\lambda_{ijt}) & \ddot{p}_{\lambda_{yy}}(\lambda_{ijt}) \end{bmatrix}. \tag{2.28}$$

The first derivative elements are given by

$$\dot{p}_{\lambda_x}(\lambda_{ijt}) = \lambda_{x,ijt}^{-1}[X - \lambda_{x,ijt} - U(1, \lambda_{ijt})], \quad \dot{p}_{\lambda_y}(\lambda_{ijt}) = \lambda_{y,ijt}^{-1}[Y - \lambda_{y,ijt} - U(1, \lambda_{ijt})],$$

where $U(m, \lambda) = S(m, \lambda)/S(0, \lambda)$ with

$$S(m, \lambda) = \sum_{k=0}^{min(X,Y)} \binom{X}{k} \binom{Y}{k} k!\, k^m \left(\frac{\gamma}{\lambda_x\, \lambda_y}\right)^k,$$

and with $\lambda = (\lambda_x, \lambda_y)'$ for $m = 0, 1, 2$. We note that

$$\frac{\partial S(m,\lambda)}{\partial \lambda_u} = -\lambda_u^{-1} S(m+1,\lambda), \qquad u = x, y, \qquad m = 0, 1,$$

and $S(m, \lambda) = 0$ when $\gamma = 0$, for $m = 1, 2$. We further observe that $S(0, \lambda) = 1$ when $k = 0$ so that function $U(m, \lambda)$ is properly defined for all $\gamma \geq 0$. The second derivative elements are given by

$$\begin{aligned}
\ddot{p}_{\lambda_{xx}}(\lambda_{ijt}) &= -\lambda_{x,ijt}^{-1}\left[1 + \dot{p}_{\lambda_x}(\lambda_{ijt}) - \lambda_{x,ijt}^{-1}\dot{U}(\lambda_{ijt})\right], \\
\ddot{p}_{\lambda_{yy}}(\lambda_{ijt}) &= -\lambda_{y,ijt}^{-1}\left[1 + \dot{p}_{\lambda_y}(\lambda_{ijt}) - \lambda_{y,ijt}^{-1}\dot{U}(\lambda_{ijt})\right], \\
\ddot{p}_{\lambda_{xy}}(\lambda_{ijt}) &= \lambda_{x,ijt}^{-1}\lambda_{y,ijt}^{-1}\dot{U}(\lambda_{ijt}),
\end{aligned}$$

with

$$\dot{U}(\lambda) = U(2,\lambda) - U(1,\lambda)^2, \qquad \frac{\partial U(1,\lambda)}{\partial \lambda_u} = -\lambda_u^{-1}\dot{U}(\lambda), \qquad u = x, y.$$

Finally, it follows that

$$\dot{p}_t(\alpha_t) = \sum_{i,j \in y_t} \dot{s}_{ij}(\alpha_t) \times \dot{p}_\lambda(\lambda_{ijt}), \qquad \ddot{p}_t(\alpha_t) = \sum_{i,j \in y_t} \dot{s}_{ij}(\alpha_t) \times \ddot{p}_\lambda(\lambda_{ijt}) \times \dot{s}_{ij}(\alpha_t)',$$

where the notation $i, j \in y_t$ implies that we consider all matches played at time t with a home team i and a visiting team j, for $t = 1, \ldots, n$. First and second derivatives of the diagonal inflated Bivariate Poisson distribution, discussed in section 2.3 of the main chapter, can be derived straightforwardly from the results in this section.

D Computational details

Selection matrix

Assume a competition with four teams. At time t we have the following match up; team 1 against team 3 and team 2 against team 4. The log scoring intensities (or signal) θ_t, selection matrix W_t and state vector α_t are then given by

$$\begin{pmatrix} \theta_{1t} \\ \theta_{2t} \\ \theta_{3t} \\ \theta_{4t} \end{pmatrix} = \begin{pmatrix} \delta \\ \delta \\ 0 \\ 0 \end{pmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{bmatrix} \alpha_t. \tag{2.29}$$

with

$$\alpha_t = \begin{pmatrix} \xi_{1t} & \xi_{2t} & \xi_{3t} & \xi_{4t} & \beta_{1t} & \beta_{2t} & \beta_{3t} & \beta_{4t} \end{pmatrix}' \tag{2.30}$$

Signal extraction of strengths of attack and defence

We use simulation methods for the signal extraction of the strengths of attack and defence ξ_{it} and β_{it} in a similar fashion as for the Monte Carlo maximum likelihood estimation of the parameters $\phi_{\kappa,i}$, $\sigma^2_{\kappa,i}$, γ and δ, with $i = 1, \ldots, J$, based on the simulated likelihood function (2.23). However, the same drawbacks apply as for likelihood evaluation via (2.23). For a given value of the parameter vector ψ, we estimate the strengths of attack and defence in the state vector α by evaluating the conditional expectation $\hat{\alpha} = \mathbb{E}(\alpha|y;\psi)$ where

$$\mathbb{E}(\alpha|y;\psi) = \int \alpha p(\alpha|y;\psi)\mathrm{d}\alpha = p(y;\psi)^{-1} \int \alpha p(\alpha, y;\psi)\mathrm{d}\alpha = p(y;\psi)^{-1} \int \alpha p(y|\alpha;\psi)p(\alpha;\psi)\mathrm{d}\alpha.$$

Given the Monte Carlo method for computing the observation density $p(y;\psi)$ and given the known expressions for $p(y|\alpha;\psi)$ and $p(\alpha;\psi)$ above, we can estimate $\hat{\alpha}$ by the same Monte Carlo simulation importance sampling method. This argument can be generalized to the estimation of any known (linear and nonlinear) function of the state vector α. It implies that we can evaluate the estimated variance, percentile and distribution of any element of α but also that we can evaluate the estimate of the intensities $\lambda_{x,ijt}$ and $\lambda_{y,ijt}$.

Negative definite variance matrix

The construction of the approximating model and the generation of the importance samples require the application of the Kalman filter smoother applied to the linear Gaussian model (2.25). Since matrix V_t in (2.27) is a variance matrix, we require that V_t is positive definite or that $\ddot{p}_t^{-1}(\alpha_t)$ is negative definite which effectively insists that the 2×2 matrix $\ddot{p}_\lambda(\lambda)$ in (2.28) is negative definite. Jungbacker and Koopman (2007) have argued that even when V_t is not positive definite, the application of the Kalman filter and the corresponding computations are still appropriate for our purposes. However, it is insightful to verify under which conditions $\ddot{p}_\lambda(\lambda)$ in (2.28) is negative. We therefore need to verify the determinant of $\ddot{p}_\lambda(\lambda)$. Without providing the details, we present in Figure 2.4 the values of X and Y for which we obtain a positive definite matrix $\ddot{p}_\lambda(\lambda)$. In case $\gamma = 0$, the variance V_t is well defined since the model reduces to a double Poisson which imposes a proper variance; see (Durbin and Koopman, 2012, Chapter 10.6) for the details. In case $\gamma > 0$, the variance V_t becomes negative when X and/or Y are large in relation to their intensities λ_x and/or λ_y, respectively. The benchmark values can be deduced from Figure 2.4.

Upper bound for correlation coefficient

Assume that X and Y are from the bivariate Poisson distribution with means $\lambda_x + \gamma$ and $\lambda_y + \gamma$, respectively, where $\gamma = \rho\sqrt{m_x m_y}$ with $m_x = \lambda_x + \gamma$ and $m_y = \lambda_y + \gamma$; see the definitions in Section 2.1. Since $\lambda_x, \lambda_y \geq 0$, we have $m_x \geq \gamma$ and hence $\rho \leq \sqrt{m_y/m_x}$. Similarly, we have $m_y \geq \gamma$ and $\rho \leq \sqrt{m_x/m_y}$. The upper bound for ρ is given by

$$\rho \leq \min\left\{\sqrt{\frac{\lambda_x + \gamma}{\lambda_y + \gamma}}, \sqrt{\frac{\lambda_y + \gamma}{\lambda_x + \gamma}}\right\}.$$

E Tables and figures

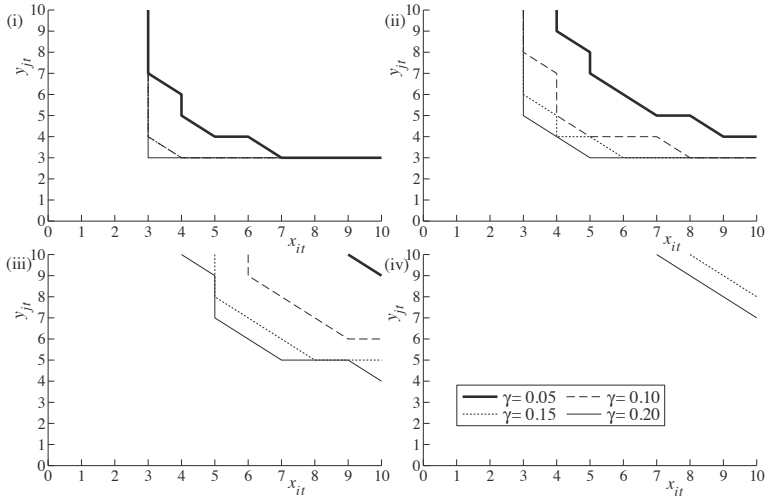

Figure 2.4: The figure illustrates combinations of counts which generate positive, negative and indefinite 'variances' in the approximating model, for various values of λ_x, λ_y and γ. The areas below and left from the lines correspond to counts that generate positive variances. The areas above and right from the lines represent counts that provide negative or indefinite variances. The coefficient γ ranges from 0.05 to 0.20 with 0.05 increments. The panels are for (i) $\lambda_x = \lambda_y = 1.0$; (ii) $\lambda_x = 1.5, \lambda_y = 1.0$; (iii) $\lambda_x = 2.0, \lambda_y = 1.5$; (iv) $\lambda_x = 2.5, \lambda_y = 2.0$.

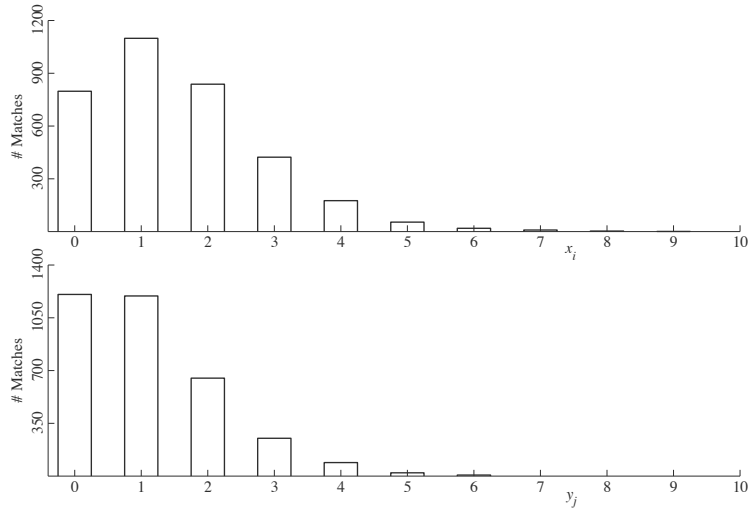

Figure 2.5: Histograms of home and away goals in the English Premier League over nine seasons ranging from 2003-2004 to 2011-2012. The average of home goals and away goals is 1.5287 and 1.0994, respectively. Averages are calculated as the average number of goals scored by the home and visiting teams in official time. No matches are played in overtime or finished with penalties.

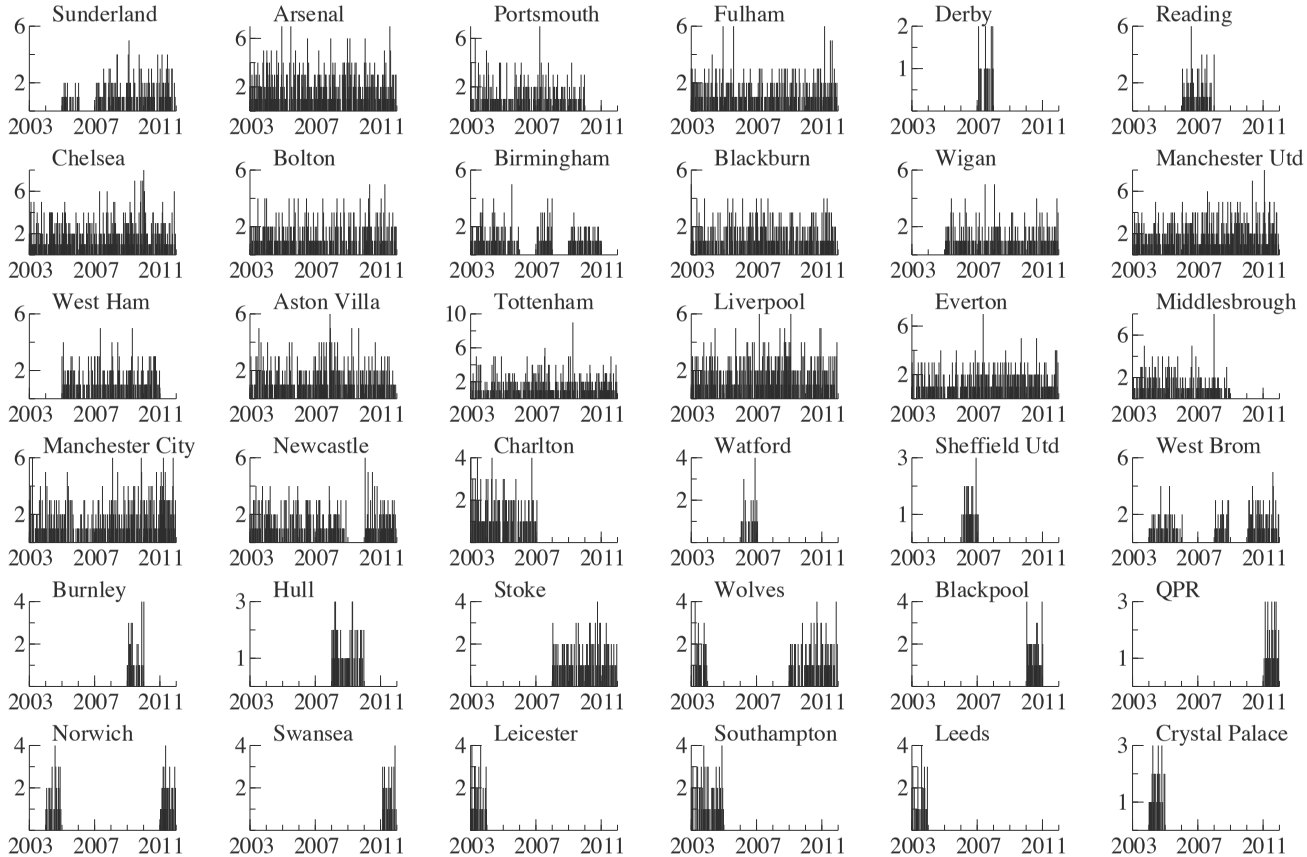

Figure 2.6: The number of goals scored by football teams from the 2003-2004 towards the 2011-2012 season of the English Premier League. Data is given in transaction time meaning that the years on the x-axis are football years and not calendar years. Due to promotion and relegation, many teams did not play in all seasons.

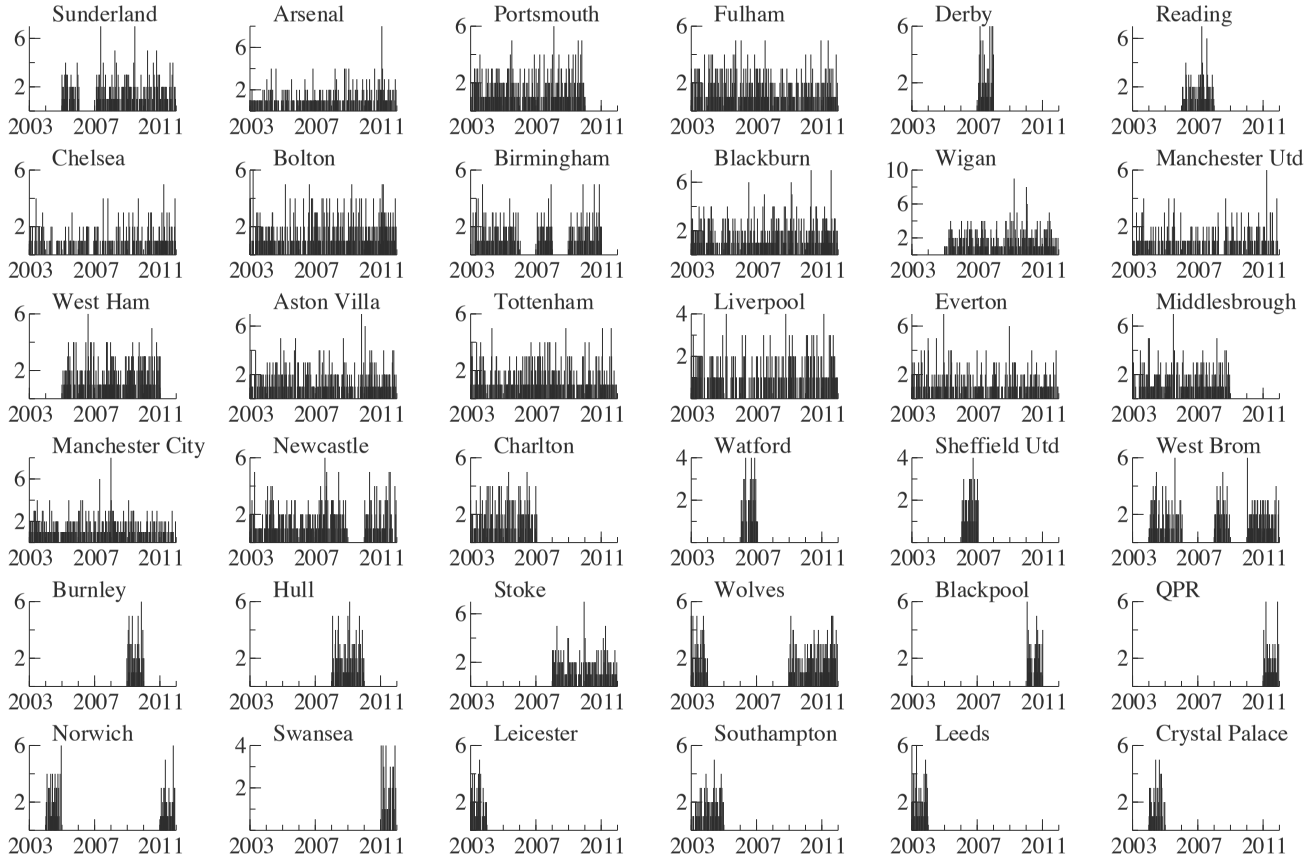

Figure 2.7: The number of goals conceded over time by football teams from the 2003-2004 towards the 2011-2012 season of the English Premier League. Data is given in transaction time meaning that the years on the x-axis are football years and not calendar years. Due to promotion and relegation, many teams did not play in all seasons.

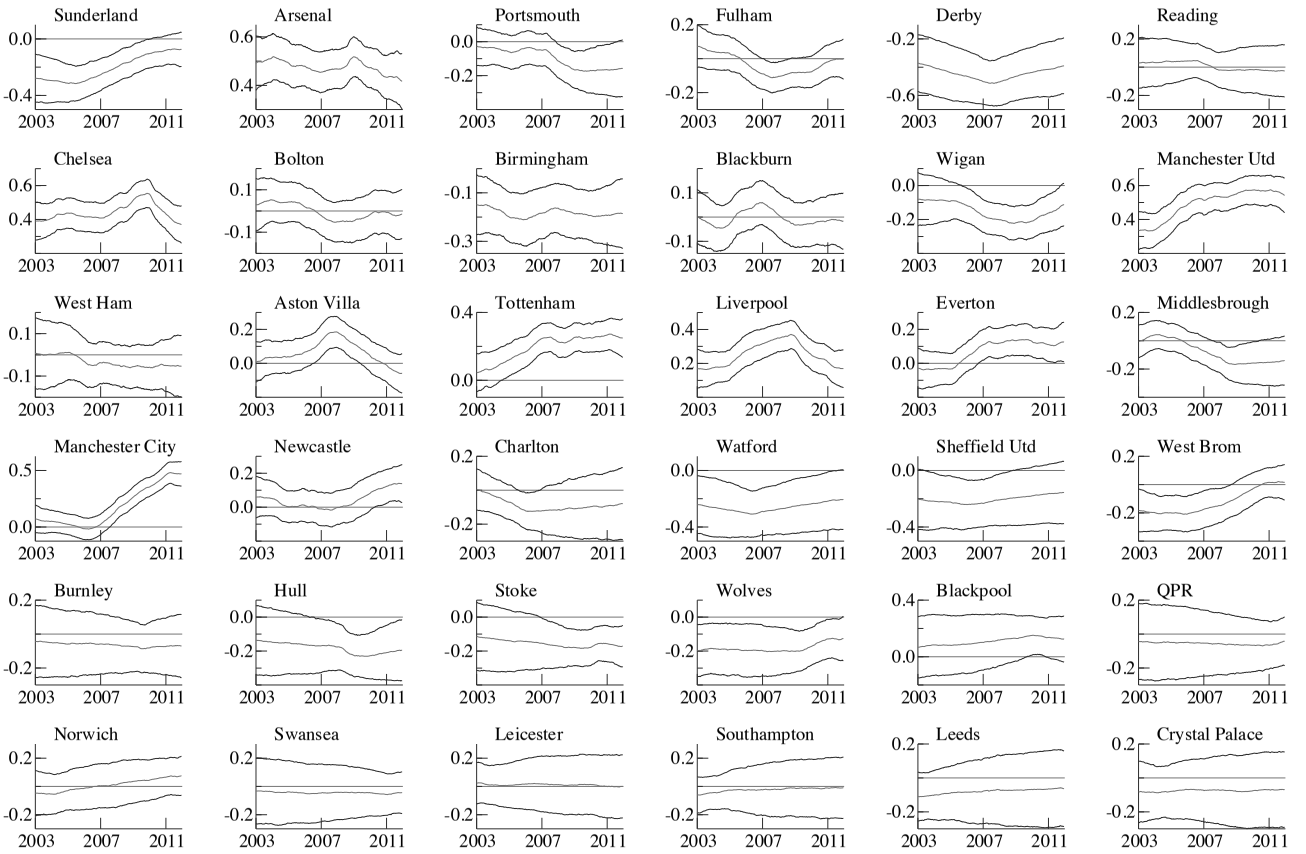

Figure 2.8: Extracted strength of attack of all teams with symmetric confidence intervals based on one standard error.

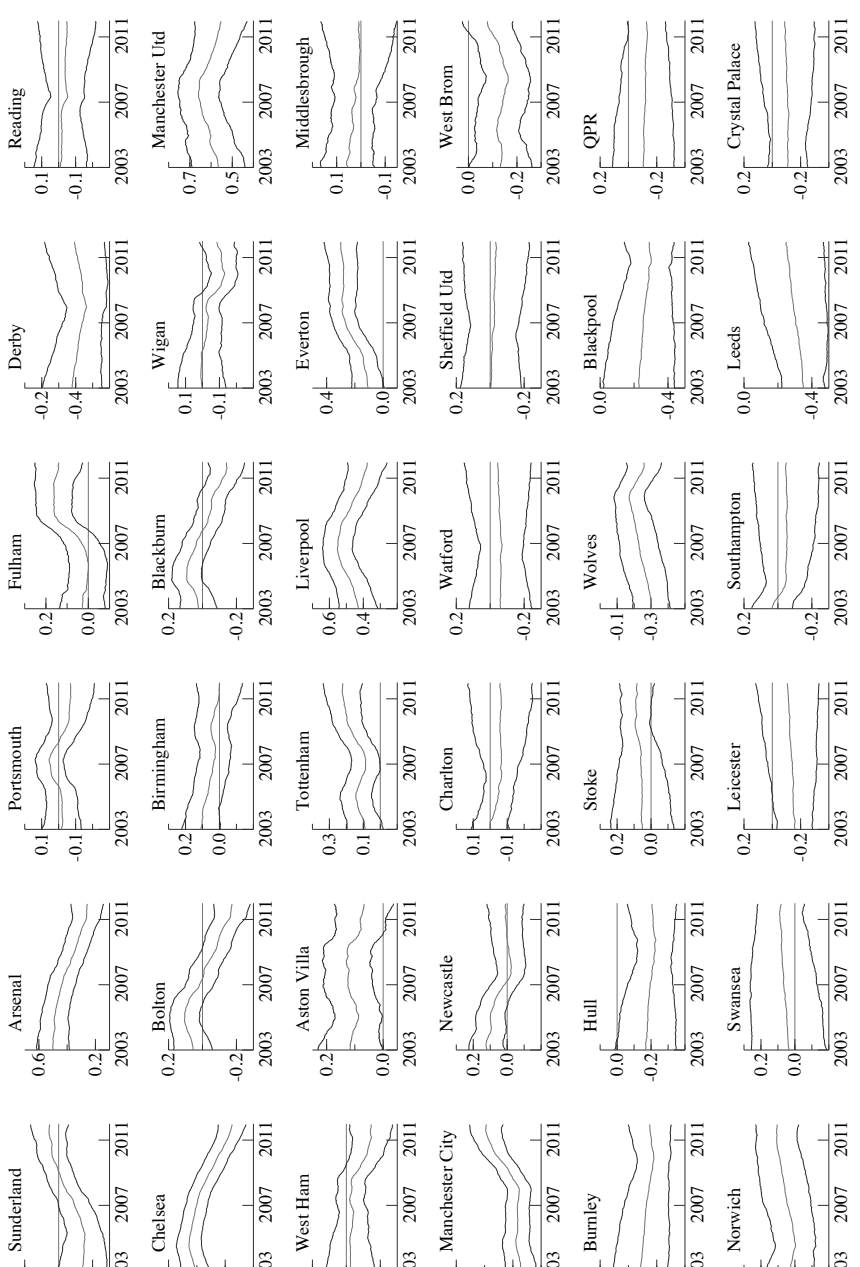

Figure 2.9: Extracted strength of defence of all teams with symmetric confidence intervals based on one standard error.

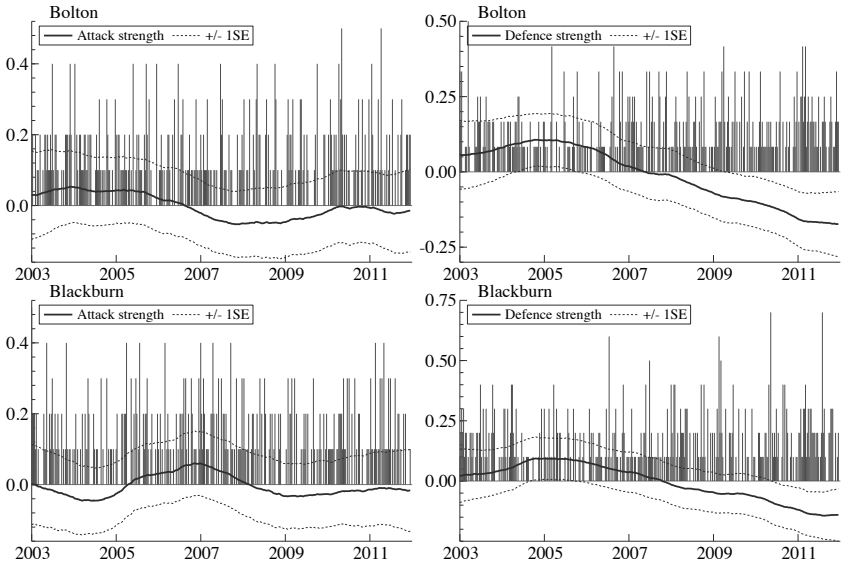

Figure 2.10: The panels show strengths of attack and defence of the two lowest ranking teams at the end of the 2011-2012 season of the English Premier League. The bars represent the number of goals scored and conceded from the 2003-2004 towards the 2011-2012 season which accounts for 404 time periods.

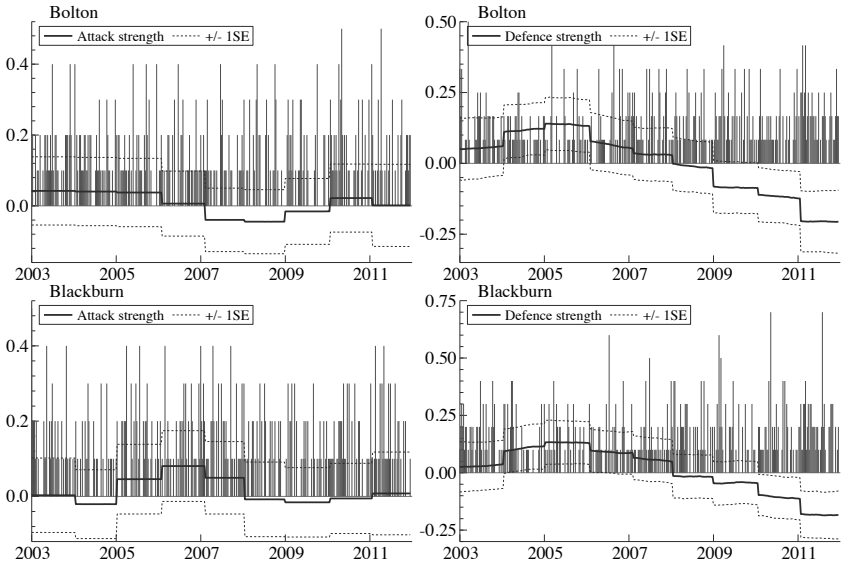

Figure 2.11: The panels show strengths of attack and defence of two low ranking teams at the end of the 2011-2012 season of the English Premier League with summer breaks. The bars represent the number of goals scored and conceded from the 2003-2004 towards the 2011-2012 season which accounts for 404 time periods.

Table 2.3: The table reports average strength of attack $\bar{\xi} = \frac{1}{n}\sum_{t=1}^{n}\xi_{it}$ for team $i, \ldots, J$ of all teams active in the 2003-2004 to 2011-2012 season of the English Premier League. Standard errors are given by $SE(\bar{\xi}) = \frac{1}{n}\sum_{t=1}^{n} SE(\xi_{it})$. The same analogy applies for $\bar{\beta}$. Lowest $\bar{\xi}$ and $\bar{\beta}$ are both from Derby County with average strengths of -0.45 and -0.42 respectively. The one season (2007-2008) Derby County played in our dataset, enters the statistical books as one of the worst in the Premier League ever with 1 win, 8 draws and 29 losses. Derby County was already relegated in March with the last match of the Premier League played on May 11, 2008. The team with the highest $\bar{\xi}$ is Manchester United with 0.49 whereas Chelsea has the highest $\bar{\beta}$ with 0.68. Columns six and eleven show average number of goals scored and conceded which correspond with $\bar{\xi}$ and $\bar{\beta}$.

Teams	$\bar{\xi}$	SE($\bar{\xi}$)	min ξ	max ξ	avg sc	$\bar{\beta}$	SE($\bar{\beta}$)	min β	max β	avg con
Arsenal	0.48	0.09	0.42	0.52	1.94	0.40	0.10	0.26	0.50	0.96
Aston Villa	0.06	0.10	-0.06	0.19	1.29	0.10	0.09	0.07	0.13	1.29
Birmingham	-0.18	0.11	-0.21	-0.15	1.02	0.05	0.10	0.00	0.11	1.36
Blackburn	0.00	0.10	-0.05	0.06	1.20	-0.01	0.09	-0.14	0.09	1.46
Blackpool	0.11	0.18	0.07	0.15	1.45	-0.26	0.17	-0.30	-0.23	2.05
Bolton	0.00	0.10	-0.05	0.05	1.20	-0.01	0.09	-0.17	0.11	1.46
Burnley	-0.06	0.18	-0.09	-0.04	1.11	-0.33	0.15	-0.38	-0.28	2.16
Charlton	-0.09	0.15	-0.13	0.00	1.11	-0.05	0.13	-0.07	0.00	1.47
Chelsea	0.44	0.09	0.37	0.55	1.89	0.68	0.10	0.44	0.82	0.74
Crystal Palace	-0.07	0.19	-0.09	-0.07	1.08	-0.10	0.17	-0.12	-0.09	1.63
Derby County	-0.45	0.18	-0.51	-0.37	0.53	-0.42	0.15	-0.46	-0.38	2.34
Everton	0.07	0.09	-0.04	0.14	1.31	0.23	0.10	0.11	0.30	1.15
Fulham	-0.03	0.10	-0.11	0.07	1.18	0.08	0.09	0.00	0.16	1.34
Hull	-0.18	0.16	-0.23	-0.14	0.96	-0.20	0.14	-0.23	-0.17	1.83
Leeds	-0.08	0.19	-0.11	-0.06	1.05	-0.30	0.18	-0.35	-0.25	2.08
Leicester	0.01	0.19	0.00	0.03	1.26	-0.13	0.17	-0.16	-0.11	1.71
Liverpool	0.25	0.09	0.16	0.37	1.56	0.48	0.10	0.38	0.55	0.89
Man City	0.17	0.09	-0.02	0.48	1.47	0.22	0.09	0.14	0.39	1.16
Man United	0.49	0.09	0.33	0.58	1.98	0.61	0.11	0.55	0.66	0.78
Middlesbrough	-0.08	0.12	-0.17	0.04	1.14	0.03	0.11	0.00	0.06	1.38
Newcastle	0.04	0.10	-0.02	0.14	1.25	0.04	0.09	-0.02	0.13	1.38
Norwich	0.01	0.15	-0.06	0.08	1.24	-0.25	0.13	-0.30	-0.19	1.88
Portsmouth	-0.10	0.11	-0.17	-0.03	1.10	-0.02	0.10	-0.07	0.06	1.43
QPR	-0.06	0.19	-0.07	-0.04	1.13	-0.12	0.18	-0.14	-0.10	1.74
Reading	0.01	0.15	-0.03	0.05	1.22	-0.03	0.13	-0.05	-0.01	1.49
Sheffield Utd	-0.20	0.19	-0.24	-0.16	0.84	-0.02	0.17	-0.04	0.00	1.45
Southampton	-0.02	0.18	-0.06	-0.01	1.17	-0.04	0.16	-0.06	0.04	1.46
Stoke	-0.15	0.14	-0.18	-0.11	1.01	0.07	0.13	0.05	0.09	1.34
Sunderland	-0.21	0.12	-0.32	-0.07	1.03	-0.06	0.10	-0.16	0.06	1.49
Swansea	-0.04	0.19	-0.06	-0.03	1.16	0.07	0.18	0.03	0.09	1.34
Tottenham	0.19	0.09	0.04	0.27	1.47	0.15	0.09	0.09	0.22	1.24
Watford	-0.26	0.19	-0.31	-0.20	0.76	-0.06	0.16	-0.07	-0.04	1.55
West Brom	-0.13	0.12	-0.21	0.02	1.07	-0.13	0.10	-0.16	-0.08	1.63
West Ham	-0.03	0.12	-0.06	0.01	1.14	-0.06	0.10	-0.15	-0.02	1.51
Wigan	-0.15	0.11	-0.22	-0.08	1.01	-0.05	0.10	-0.13	0.01	1.54
Wolves	-0.18	0.14	-0.20	-0.12	1.03	-0.23	0.11	-0.30	-0.17	1.85
Mean	-0.01	0.14	-0.07	0.06	1.20	0.01	0.12	-0.05	0.07	1.49

Table 2.4: The table reports strengths of attack and defence at the start of the denoted seasons. On August 28, 2011 Manchester United defeated Arsenal by 8-2 and reached the highest strength of attack the model identified (0.5748). During the 2004-2005 season Chelsea reached the highest defence strength (0.8308).

Teams	'03 ξ	'05 ξ	'07 ξ	'09 ξ	'11 ξ	'03 β	'05 β	'07 β	'09 β	'11 β
Arsenal	0.49	0.49	0.45	0.52	0.43	0.50	0.49	0.44	0.35	0.27
Aston Villa	0.01	0.04	0.13	0.12	-0.01	0.12	0.09	0.12	0.13	0.08
Birmingham	-0.15	-0.20	-0.18	-0.19	-0.19	0.11	0.08	0.04	0.05	0.01
Blackburn	0.00	-0.03	0.06	-0.03	-0.01	0.02	0.09	0.04	-0.05	-0.12
Blackpool	0.07	0.08	0.10	0.13	0.14	-0.23	-0.24	-0.25	-0.28	-0.30
Bolton	0.03	0.04	-0.02	-0.05	-0.01	0.06	0.10	0.02	-0.07	-0.15
Burnley	-0.04	-0.06	-0.06	-0.07	-0.07	-0.28	-0.31	-0.33	-0.36	-0.36
Charlton	0.00	-0.07	-0.12	-0.11	-0.09	0.00	-0.05	-0.06	-0.06	-0.06
Chelsea	0.39	0.44	0.41	0.51	0.44	0.76	0.82	0.75	0.64	0.49
Crystal Palace	-0.08	-0.08	-0.07	-0.08	-0.07	-0.11	-0.11	-0.11	-0.10	-0.09
Derby County	-0.37	-0.44	-0.50	-0.47	-0.42	-0.38	-0.41	-0.44	-0.44	-0.41
Everton	-0.03	-0.03	0.11	0.13	0.11	0.11	0.14	0.26	0.28	0.29
Fulham	0.07	0.03	-0.09	-0.08	-0.02	0.03	0.00	0.03	0.16	0.16
Hull	-0.14	-0.15	-0.17	-0.22	-0.21	-0.17	-0.18	-0.19	-0.22	-0.21
Leeds	-0.11	-0.09	-0.08	-0.07	-0.06	-0.35	-0.34	-0.30	-0.28	-0.26
Leicester	0.02	0.01	0.02	0.01	0.00	-0.16	-0.15	-0.14	-0.12	-0.11
Liverpool	0.17	0.19	0.32	0.36	0.21	0.43	0.49	0.55	0.49	0.40
Man City	0.07	0.02	0.02	0.26	0.46	0.15	0.18	0.17	0.22	0.37
Man United	0.33	0.41	0.51	0.55	0.57	0.56	0.60	0.65	0.63	0.57
Middlesbrough	0.00	0.02	-0.07	-0.17	-0.15	0.05	0.04	0.02	0.01	0.01
Newcastle	0.06	0.00	-0.01	0.02	0.12	0.12	0.10	0.02	-0.01	0.01
Norwich	-0.05	-0.03	0.00	0.04	0.06	-0.27	-0.30	-0.26	-0.22	-0.19
Portsmouth	-0.03	-0.06	-0.05	-0.16	-0.16	-0.02	-0.02	0.05	-0.03	-0.07
QPR	-0.05	-0.05	-0.05	-0.06	-0.07	-0.11	-0.11	-0.12	-0.12	-0.13
Reading	0.03	0.04	0.03	-0.02	-0.02	-0.01	-0.02	-0.03	-0.04	-0.05
Sheffield Utd	-0.20	-0.22	-0.23	-0.19	-0.16	0.00	-0.01	-0.02	-0.02	-0.03
Southampton	-0.06	-0.03	-0.02	-0.01	-0.01	0.04	-0.05	-0.05	-0.05	-0.05
Stoke	-0.11	-0.14	-0.15	-0.18	-0.16	0.05	0.05	0.06	0.09	0.09
Sunderland	-0.28	-0.31	-0.25	-0.15	-0.08	-0.14	-0.15	-0.09	0.00	0.05
Swansea	-0.03	-0.04	-0.05	-0.04	-0.06	0.03	0.05	0.07	0.08	0.09
Tottenham	0.04	0.12	0.25	0.25	0.27	0.09	0.14	0.09	0.17	0.22
Watford	-0.24	-0.28	-0.29	-0.25	-0.22	-0.06	-0.06	-0.07	-0.05	-0.05
West Brom	-0.18	-0.20	-0.18	-0.10	0.02	-0.13	-0.12	-0.14	-0.15	-0.11
West Ham	0.01	0.01	-0.04	-0.05	-0.05	-0.02	-0.02	-0.03	-0.06	-0.14
Wigan	-0.08	-0.09	-0.18	-0.22	-0.16	0.00	0.01	-0.03	-0.08	-0.11
Wolves	-0.20	-0.19	-0.20	-0.20	-0.13	-0.30	-0.27	-0.23	-0.19	-0.22
Mean	0.02	-0.02	-0.02	-0.01	0.01	0.01	0.02	0.01	0.01	0.00

Chapter 3

Intraday Stochastic Volatility in Discrete Price Changes

3.1 Introduction

Stochastic volatility is typically associated with the time-varying variance in time series of daily continuously compounded rates of financial returns; for a review of the relevant literature, see Shephard (2005). The availability of high-frequency intraday trade information has moved the focus towards the estimation of volatility using realised measures such as realised volatility and realised kernels; see the seminal contributions of Barndorff-Nielsen and Shephard (2001, 2002), Andersen, Bollerslev, Diebold, and Labys (2001) and Hansen and Lunde (2006). Recent research has moved beyond the use of high-frequency data for obtaining daily observations of (realised) variances to the actual modelling of high-frequency price changes themselves at the intraday level. For example, Barndorff-Nielsen, Pollard, and Shephard (2012) and Shephard and Yang (2015) formulate continuous-time stochastic processes and design econometric models based on integer-valued Lévy processes using Skellam distributed random variables. Price changes of a stock are measured on a grid of one dollar cent and hence the tick-by-tick price change can be treated as a Skellam distributed random variable that takes values in $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Also Hansen, Horel, Lunde, and Archakov (2015) study the discrete nature of high-frequency price changes and explore their dynamic properties by formulating a stochastic Markov-chain process.

In our current study we develop a new statistical model that is empirically relevant for the discrete time series of tick-by-tick financial data. Such data enjoy the increasing interest of government regulators as well as industry participants given their potential impact on the stability of financial markets. Our new model has three important features that are needed to capture typical intraday properties of the data. First, the model builds

on a dynamic modified Skellam distribution to make the model congruent with the realised data that consist of discrete-valued tick-size price changes defined on the set of integers $\mathbb{Z}$. Second, our modified Skellam distribution features a doubly dynamic variance parameter. The variance is allowed to be different over the course of a trading day due to intraday seasonal patterns, which we capture by including a spline function over the time of day. On top of this, we also allow for autoregressive intraday stochastic volatility dynamics to capture any remaining volatility dynamics over the course of the trading day that cannot be attributed to seasonal patterns. Third, our data requires a careful treatment of small price changes of the order of 0, 1, or -1 dollar cents. For this purpose, we modify the dynamic Skellam distribution by allowing for a probability mass transfer between these different price change realisations. The probability mass transfer needs to vary over time as well because the data reveal that trades with a zero price-change are not spread evenly across the trading day. The resulting new model with these three features embedded performs well in terms of fit, diagnostics, and forecasting power compared to a range of alternative models.

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

features of intraday financial price changes such as intraday seasonality, long stretches of missing values, and the time-varying modifications for the Skellam distribution. All these features are key for our current analysis of the empirical data. In addition, our new dynamic modified Skellam distribution may also provide a useful flexible modelling framework in other empirical settings.

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

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

The remainder of this chapter is organised as follows. We present the new dynamic modified Skellam model in Section 3.2 and explain how it can be cast into a nonlinear non-Gaussian state space form. Section 3.3 applies the dynamic Skellam model to four stocks, traded on NYSE, for all trading days in the year 2012. This section also contains information on model fit, diagnostic checks and forecasting performance. Section 3.4 concludes.

3.2 The dynamic Skellam model

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

3.2.1 The Skellam distribution

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

$$p(y_t; \mu, \sigma^2) = \exp\left(-\sigma^2\right) \left(\frac{\sigma^2 + \mu}{\sigma^2 - \mu}\right)^{y_t/2} I_{|y_t|}(\sqrt{\sigma^4 - \mu^2}), \tag{3.1}$$

where $I_{|y_t|}(\,\cdot\,)$ is the modified Bessel function of order $|y_t|$; see Abramowitz and Stegun (1972). The Skellam distribution was originally derived from the difference of two Poisson distributions; see Irwin (1937) and Skellam (1946). We then have $\mu = \lambda_1 - \lambda_2$ and $\sigma^2 = \lambda_1 + \lambda_2$, where λ_1 and λ_2 are the intensities of the two underlying Poisson distributions; see also Alzaid and Omair (2010). Karlis and Ntzoufras (2009) show that the underlying Poisson assumption can be dispensed with and that the Skellam distribution can also be considered by itself as an interesting distribution defined on integers.

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

$$p(y_t; 0, \sigma^2) = \exp\left(-\sigma^2\right) I_{|y_t|}(\sigma^2). \tag{3.2}$$

In the upper panels of Figure 3.1 we present examples of Skellam distributions for a range of values for μ and σ^2. The excess kurtosis of the Skellam distribution is $1/\sigma^2$ and the Gaussian distribution is a limiting case of the Skellam distribution; see Johnson, Kotz, and Kemp (1992) and references therein.

3.2.2 The modified Skellam distribution

The upper panels of Figure 3.1 reveal that the Skellam distribution is highly peaked at zero for low values of σ^2. This particular feature does not match the high-frequency tick-by-tick discrete stock price data well in our empirical application. To accommodate some more flexible patterns, we propose a modification of the Skellam distribution to compensate for the over- or under-representation of specific integers. For example, in our empirical application the standard Skellam distribution over-predicts the occurrence of 0s and under-predicts the occurrence of ± 1s.

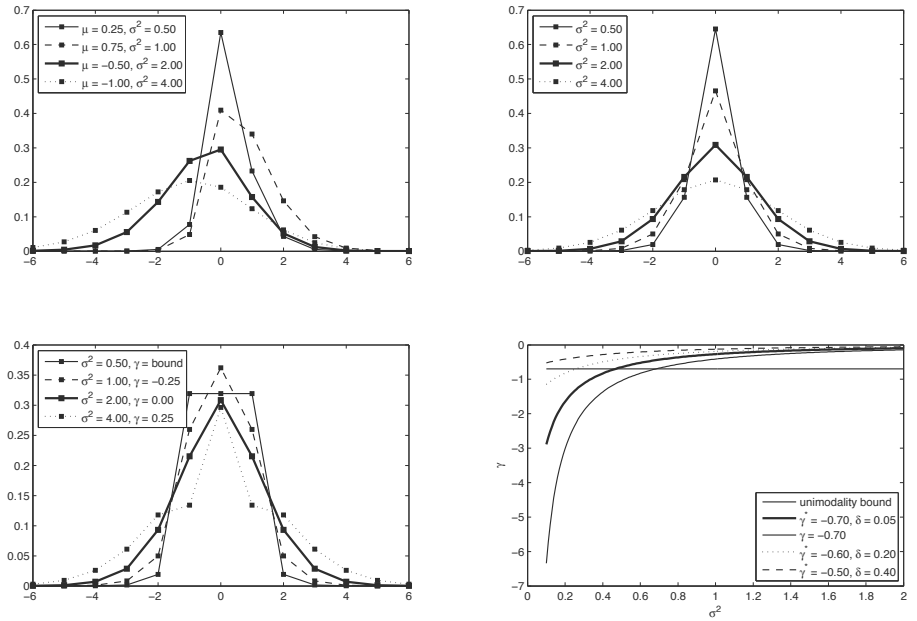

Figure 3.1: Panel 1: Skellam distribution examples with pmf (3.1) for several combinations of μ and σ^2. Panel 2: zero-mean ($\mu = 0$) Skellam distribution (3.2) examples for several combinations of σ^2. Panel 3: MSKII$(-1, 1, 0; \mu, \sigma^2, \gamma)$ distribution examples with pmf (3.3) for $\mu = 0$ and several combinations of σ^2 and γ_t. The distributions provide discrete support: the connecting lines are drawn for clarity and do not indicate continuity. Panel 4: unimodality bound and parameterized zero deflation bounds.

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

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

MSKII$(i, j, k; \mu, \sigma^2, \gamma)$ distribution is defined by its pmf

$$p_{II}(y_t; i, j, k, \mu, \sigma^2, \gamma) = \begin{cases} P_{y_t}, & \text{for } y_t \notin \{i, j, k\}, \\ (1-\gamma)P_i, & \text{for } y_t = i, \\ (1-\gamma)P_j, & \text{for } y_t = j, \\ \gamma P_i + \gamma P_j + P_k, & \text{for } y_t = k, \end{cases} \tag{3.3}$$

where $P_q = p(q; \mu, \sigma^2)$ is defined in equation (3.1) and $q \in \mathbb{Z}$, and with coefficient $\gamma \in \{-P_k/(P_i + P_j), 1\}$. The sign of the coefficient γ determines whether we inflate P_k (positive) or whether we deflate this probability (negative). For $\gamma = 0$, we recover the original Skellam distribution defined in (3.1). The lower bound of γ follows directly from the last equation in (3.3) since $\gamma P_i + \gamma P_j + P_k \geq 0$ implies $\gamma \geq -P_k/(P_i + P_j)$. The mean and variance of the MSKII$(i, j, k; \mu, \sigma^2, \gamma)$ distribution are given by

$$\begin{aligned} \mathbb{E}(Y_t) &= \mu_{II} = \mu - \gamma(i \cdot P_i + j \cdot P_j) + k \cdot \gamma(P_i + P_j), \\ \mathbb{V}\text{ar}(Y_t) &= \sigma^2_{II} = \sigma^2 + \mu^2 + \gamma P_i(k^2 - i^2) + \gamma P_j(k^2 - j^2) - \mu^2_{II}, \end{aligned} \tag{3.4}$$

respectively, see Appendix B for derivations. For $\gamma = 0$, we clearly have $\mu_{II} = \mu$ and $\sigma^2_{II} = \sigma^2$. Given the data in our empirical application below, the MSKII$(-1, 1, 0; 0, \sigma^2, \gamma)$ distribution will prove to be of particular interest.

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

$$\underline{\gamma}(\mu, \sigma^2) = \{\min(P_{-1}, P_1) - P_0\} \,/\, \{\min(P_{-1}, P_1) + P_1 + P_{-1}\}, \tag{3.5}$$

where $P_{q,II} = p_{II}(q; i, j, k, \mu, \sigma^2, \gamma)$ is defined in equation (3.3) and $q \in \mathbb{Z}$. The probability $P_{q,II}$ is a function of μ and σ^2 for all $q \in \mathbb{Z}$. In Panel 3 of Figure 3.1 we present MSKII$(-1, 1, 0; \mu, \sigma^2, \gamma)$ distributions for $\mu = 0$ and different values of σ^2 and γ. The figure reveals the effect of γ on the peakedness of the distribution. Panel 4 of Figure 3.1 presents examples of unimodal bounds $\underline{\gamma}(\mu, \sigma^2)$ for $\mu = 0$ and for different values of σ^2. We can select different model specifications to enforce γ to lie in the unimodality range; see section 3.3.2.

3.2.3 The Skellam model with dynamic mean and variance

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

$$Y_t|\mu_t, \sigma_t^2 \sim \text{MSKII}\big(-1, 1, 0; \mu_t, \sigma_t^2, \gamma_t\big), \qquad t = 1, \ldots, n, \tag{3.6}$$

where $\gamma_t = \gamma(\mu_t, \sigma_t^2)$ is the time-varying coefficient γ in (3.3) and is a function of μ_t and σ_t^2. Hence we assume that the serial dependence in Y_t is accounted for by the time variation in μ_t and σ_t^2 only. In other words, conditional on μ_t and σ_t^2, Y_t is not subject to other dynamic processes. We model the dynamics of μ_t and σ_t^2 by a (possibly) nonlinear transformation of an autoregressive process,

$$\begin{pmatrix} \mu_t \\ \sigma_t^2 \end{pmatrix} = s(\theta_t), \qquad \theta_t = c_t + Z_t\alpha_t, \tag{3.7}$$

$$\alpha_{t+1} = d_t + T_t\alpha_t + \eta_t, \qquad \eta_t \sim \text{NID}(\mathbf{0}, Q_t), \tag{3.8}$$

for $t = 1, \ldots, n$, where vector $s(\,\cdot\,)$ is referred to as the link function, $\theta_t \in \mathbb{R}^{r\times 1}$ is the signal vector, with $r = 2$, $\alpha_t \in \mathbb{R}^{m\times 1}$ is the state vector, $c_t \in \mathbb{R}^{r\times 1}$ is a scalar intercept, $d_t \in \mathbb{R}^{m\times 1}$ is a vector of intercepts, $Z_t \in \mathbb{R}^{r\times m}$ is a matrix of coefficients, $T_t \in \mathbb{R}^{m\times m}$ is a transition matrix, and the disturbances η_t are normally and independently distributed (NID) with mean zero and variance matrix $Q_t \in \mathbb{R}^{m\times m}$. The vectors c_t, d_t and matrices Z_t, T_t, Q_t are typically constant but possibly time-varying in a deterministic manner. Typical examples of link functions $s(\,\cdot\,)$ are the exponential function (to ensure positivity) and the scaled logistic function (to preserve lower and upper bounds). When the link function $s(\,\cdot\,)$ directly requires the state vector α_t as an argument, we simply set $r = m$, $c_t = \mathbf{0}$, and $Z_t = \mathrm{I}_{\boldsymbol{m}}$. For an application with an observation distribution that only requires a time-varying mean or variance, we have a univariate signal and $r = 1$. The initial conditions for the elements of the state vector α_1 depend on their dynamic properties. The variance matrix Q_t is possibly positive semi-definite and hence the vector η_t may contain zeros.

The model specified in equations (3.7)–(3.8) allows for a wide variety of dynamic patterns in μ_t and σ_t^2, including autoregressive moving average dynamics, time-varying seasonal and cyclical patterns, deterministic and stochastic trends, and their combinations. Regression and intervention effects can be added to the signal as well. More details of their formulations in the form of (3.8) are provided in Durbin and Koopman (2012, Ch. 3). The dynamic Skellam model as specified above falls within the class of non-Gaussian

nonlinear state space models which can be represented as

$$y_t \sim p(y_t|\theta_t;\psi), \qquad \theta_t = c_t + Z_t\alpha_t, \qquad \alpha_{t+1} \sim p_g(\alpha_{t+1}|\alpha_t;\psi), \qquad t = 1,\ldots,n, \quad (3.9)$$

with $\alpha_1 \sim p_g(\alpha_1;\psi)$, where ψ is an unknown and fixed parameter vector gathering all the parameters in c_t, Z_t, d_t, T_t, and Q_t, and possibly in the signal function $s(\,\cdot\,)$. The observation density $p(y_t|\theta_t;\psi)$ refers to the dynamic (possibly modified) Skellam distribution from Section 3.2 with signal θ_t representing the dynamic mean μ_t and/or variance σ_t^2. The updating Gaussian state density $p_g(\alpha_{t+1}|\alpha_t;\psi)$ refers to the linear Markov process (3.8), and $p_g(\alpha_1;\psi)$ represents the initial condition for α_1. We assume that for given realisations of the signal $\theta' = (\theta_1',\ldots,\theta_n')$ the observations $y = (y_1,\ldots,y_n)'$ are conditionally independent, and also write $\theta = c + Z\alpha$ with $c' = (c_1',\ldots,c_n')$, $\alpha = (\alpha_1',\ldots,\alpha_n')'$, and Z a block-diagonal matrix with blocks $Z_1,\ldots,Z_n$ on the leading diagonal. The joint conditional density for all observations and the marginal density for all states can now be written as

$$p(y|\theta;\psi) = \prod_{t=1}^{n} p(y_t|\theta_t;\psi), \qquad p_g(\alpha;\psi) = p_g(\alpha_1;\psi)\prod_{t=2}^{n} p_g(\alpha_t|\alpha_{t-1};\psi), \quad (3.10)$$

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

The state space representation implied by equations (3.9) or (3.10) for the dynamic Skellam model allows us to build on a well developed framework for the parameter estimation of ψ, for the signal extraction of θ and the filtering and smoothing of α; see Durbin and Koopman (2012) for a textbook treatment. As for all non-Gaussian nonlinear state space models, the main complication for the dynamic Skellam model is that the likelihood function $\int p(y|\theta;\psi)p_g(\alpha;\psi)\,\mathrm{d}\alpha$ is analytically intractable. We therefore adopt the method of Monte Carlo maximum likelihood for parameter estimation, but also for signal extraction. In particular, we apply the numerically accelerated importance sampling (NAIS) method of Koopman et al. (2014) and show that it can efficiently handle long univariate time series (large n). If we require a time-varying μ_t or σ_t^2, i.e. a univariate signal, $r = 1$, we can apply the NAIS method of Koopman et al. (2014) without extensions. For handling both a time-varying mean μ_t and variance σ_t^2, we have developed a bivariate extension of the NAIS methodology available in Appendix D. In our empirical study below we set $\mu_t = 0$, such that we only consider a stochastic time-varying variance σ_t^2. In the application of Chapter 4, an illustration with a bivariate signal is presented.

3.3 Analysis of high-frequency Skellam price changes

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

Rather than aggregating the data to one-minute or five-minute intervals, we analyse stock price changes on a second-by-second basis within a single trading day. As a consequence, all series have the same length of $n = 23{,}400$ (6.5 hours $\times$ 3600 seconds) with many missing values. By explicitly considering missing values in our analysis we take account of the duration between consecutive trades. Since there is more active trading at the beginning and end of a trading day, the number of missing values also varies throughout the day. We exploit Kalman filter and smoothing methods to handle missing values. Descriptive statistics for the data are reported in Table 3.1 and Table 3.5 and further discussed below.

We analyse the intraday prices using the dynamic Skellam model as developed above. In accordance with other analyses of high-frequency stock returns, the sample mean in price changes for a sufficiently large sample size is typically close to zero; see, for example, Andersen and Bollerslev (1997). Hence we set $\mu_t = 0$ and focus on the modelling of stochastic volatility σ_t^2. This yields a univariate signal ($r = 1$) in our state space representation of the model as discussed in Section 3.2.3.

3.3.1 Data

We use data from the trades and quotes (TAQ) database of the New York Stock Exchange at a one-second frequency. The data consist of the prices of four different stocks traded over the entire year 2012. We select companies from different industries and with different trade intensities. We analyse the tick-by-tick data without the "odd-lots" that represent trades with volumes less than 100 and are not recorded on the consolidated tape; see the discussion in O'Hara, Yao, and Ye (2014). The data require standard pre-processing. For a review of high-frequency data cleaning procedures; see for example Falkenberry (2002). We apply the cleaning algorithm of Brownlees and Gallo (2006) after applying a rudimentary filter corresponding to the cleaning steps P1, P2, P3 and T1, T2, T3 of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008, p. 8).

Table 3.1: Descriptive statistics of the four selected stocks for all trading days in 2012 combined as one sample. The table reports data characteristics of tick changes between 9:30am and 4:00pm. We report the "opening price" at 9:30 am (OP) January 1, 2012, the "closing price" at 16:00 pm (OP) December 31, 2012, the total number of trades in 2012 (#Trades), the percentage of zero price changes (%0), the percentage of $-1, 1$ price changes ($\% \pm 1$), variance (V), skewness (S), kurtosis (K) and the largest up tick (Max) and down tick (Min).

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

The large difference in opening price and closing price for Coca-Cola Company is due to a 2:1 stock split on August 13, 2012. The number of trades ranges from almost 650,000 to more than a million over 2012. At the same time, the column "%0" in Table 3.1 shows that many trades do not result in a price change: the percentage of zeros ranges from 27% for Caterpillar to 58% for Coca-Cola. We can conclude from the "%0" and "$\% \pm 1$" columns that the majority of trades only induce a maximum price change of ± 1. A full breakdown of the empirical distribution of tick-size price changes is provided in the Supplementary Appendix E. The correct handling of zero price change trades is challenging for two reasons. First, zero price changes are not randomly distributed over the trading day. A Wald-Wolfowitz runs test, see Bradley (1968, Ch. 12), strongly rejects the null hypothesis of zeros following a random sequence throughout the trading day. The largest p-value of the runs test is 8.73×10^{-6} out of the 1000 days under consideration (4 stocks $\times$ 250 trading days in 2012). Second, long streaks of zeros and/or missing values occur regularly during slow trading periods of the day. This leads to a low volatility in price changes. Although the majority of observations within a trading day are either missing or are equal to -1, 0 and 1, large price changes (or jumps) do occur as indicated by the "Max" and "Min" columns in Table 3.1. Also the reported yearly sample variance and kurtosis for each stock reflect sufficient variation in the tick-by-tick stock price changes. The challenge for our statistical dynamic model is to address all of these salient features appropriately.

3.3.2 Dynamic Skellam with Intraday Stochastic Volatility

We consider the conditional observation density (3.6) with pmf (3.3). The standard Skellam model is a special case with $\gamma_t = 0$. The model specification for the dynamic variance, or the stochastic volatility, is based on the link function with $r = 1$ given by

$$\sigma_t^2 = s(\theta_t) = \exp(\theta_t), \qquad t = 1, \ldots, n, \tag{3.11}$$

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

$$\theta_t = c + s_t + \alpha_t, \qquad \alpha_{t+1} = \phi\alpha_t + \eta_t, \qquad \eta_t \sim \text{NID}\left(0, \sigma^2_{\eta,t}\right), \tag{3.12}$$

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

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

$$s_t = \beta' \tilde{W}_t, \qquad t = 1, \ldots, n, \qquad \sum_{t=1}^{n} s_t = 0, \tag{3.13}$$

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

The variance of the innovations for the stationary component α_t is time-varying to account for increased volatility due to special news announcements during the trading day. Many of such news announcements are released at pre-set time periods, such as 08:30, 10:00, and other; see Andersen, Bollerslev, Diebold, and Vega (2003). The effect of the news announcement before the opening of the market at 09:30 is captured by the first knot of the spline s_t. The possible effect of, say, a 10:00 news announcement, however, is harder to accommodate by the spline or AR(1) process only. For this purpose we introduce a separate parameter to model a (possible) temporary jump in volatility between 10:00 and 10:01. We do so by defining the indicator variable $\tau_S(t) = 1$ for $t = 1800, \ldots, 1860$

(corresponding to the first minute after 10am), and zero otherwise, thus increasing the variance of η_t from σ_η^2 to $\sigma_\eta^2 + \sigma_{\eta,S}^2$ during this period, where $\sigma_{\eta,S}^2 > 0$. An increase of the variance for η_t allows α_{t+1} to vary more than in other time periods.

We ensure unimodality of the MSKII$(-1, 1, 0; 0, \sigma_t^2, \gamma_t)$ distribution under zero deflation via a parsimonious re-parameterization as follows. We introduce the coefficients $-1 < \gamma^* < 1$ and $\delta > 0$. Then we determine γ_t as $\gamma_t = \gamma^*$, when $\gamma^* \geq 0$, and $\gamma_t = -\gamma^* \times \underline{\gamma}(0, \sigma_t^2 + \delta)$, when $\gamma^* < 0$, since $\underline{\gamma}(0, \cdot) < 0$. The coefficient δ ensures a left-horizontal shift from the lower bound of $\underline{\gamma}(0, \sigma_t^2)$ in order to avoid potential numerical issues for its limit as $\sigma_t^2 \to 0$; see Panel 4 of Figure 3.1. The condition of unimodality stabilizes some numerical issues in likelihood evaluation since the construction of an importance density for bimodal distributions is rather challenging; see the discussion in Durbin and Koopman (2012, p. 253).

3.3.3 Parameter estimation results

The parameter vector for our dynamic Skellam model is given by

$$\psi = (\phi, \sigma_\eta, \sigma_{\eta,S}, c, \delta, \gamma^*, \beta')'.$$

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

The parameter vector is estimated for each stock and each trading day in 2012. Given the large number of estimates, we provide a graphical presentation in Figure 3.2. In particular, we present the parameter estimates of ϕ, σ_η, c, and γ^*. The estimates vary from day to day and characterize the intraday dynamics of price changes for that specific day. We have, for each stock on average, between 2500 and 4000 observations available for the estimation of ψ on daily basis; see Table 3.1. It also allows us to carry out a meaningful forecasting study in Section 3.3.7.

The top row in Figure 3.2 shows the estimates of ϕ. Overall, the estimates indicate a high degree of persistence of the autoregressive process α_t. The average estimate of ϕ

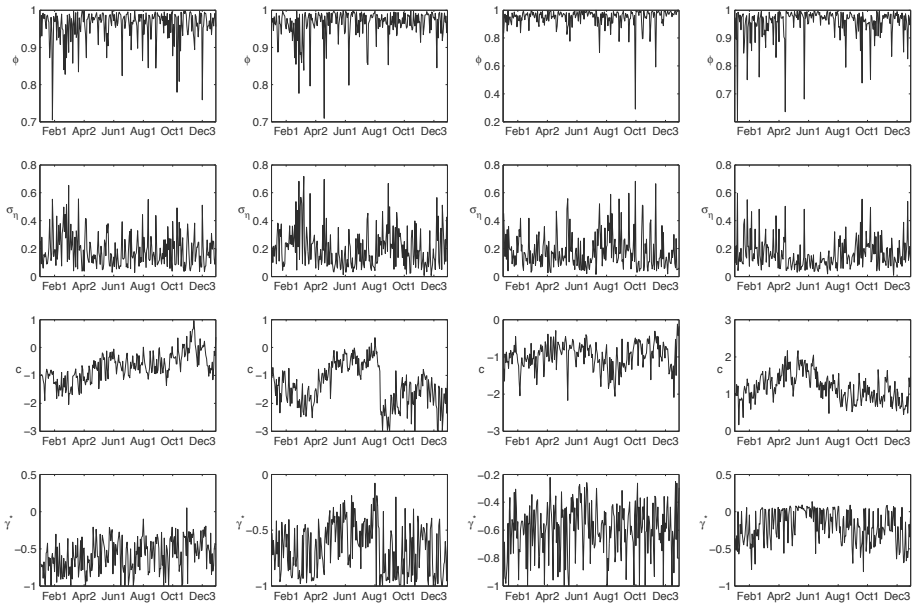

Figure 3.2: The figure shows the maximum likelihood estimates of the first four elements of ψ where each column correspond to one of the four stocks in the order WMT, KO, JPM and CAT and the rows represent the parameter estimates in the order $(\phi, \sigma_\eta, c, \gamma^*)$.

over all trading days of 2012 exceeds 0.94 for each stock. Some individual days exhibit a ϕ estimate that is clearly below the average. It indicates that the cubic spline $c + s_t$ already captures most of the information for that specific day. We investigate the individual contribution of the spline versus the autoregressive component in Section 3.3.6 in more detail.

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

The third row shows the daily estimates of the constant c. For Walmart, the time series of c estimates shows a steady increase of the overall average daily volatility level during the year. For Coca-Cola, the structural break in the daily estimates of c in Augustus 13, 2012, clearly coincides with the 2:1 stock split on that day. The constants c naturally play a dominant role in the overall level of daily log-volatility. As such, they may be compared to alternative estimates of integrated volatility based on high-frequency data. Interestingly, the time series correlations over all trading days in 2012 of our estimates of c with the logged realised volatility (RV) measure as estimated using the algorithm of Aït-Sahalia, Mykland, and Zhang (2011), based on 5-minute intervals, are high. The

correlations are 0.90, 0.88, 0.67, and 0.93 for Walmart, Coca-Cola, JPMorgan, and Caterpillar, respectively.

The bottom panels in Figure 3.2 show the parameter estimates of γ^*. The estimates of γ^* are typically highly statistically significant, which indicates that our modification of the standard Skellam distribution is empirically relevant. For all stocks the 0-deflated model ($\gamma^* < 0$) is clearly preferred. Only for CAT we have that some periods are subject to 0-inflation. CAT has the largest stock price compared to the others stocks, resulting in a larger value of σ_t^2 on average. A larger value of σ_t^2 comes with a lower predicted probability of 0s, such that zero inflation rather than deflation becomes more relevant for CAT compared to the other stocks. Our type II modified Skellam model also outperforms the standard zero-deflation type I modification of the Skellam model of Karlis and Ntzoufras (2006, 2009), which is why we do not report the latter here.

3.3.4 Signal extraction

Figure 3.3 presents the time series average of our estimated zero sum cubic spline s_t, with corresponding 95% confidence bands. Instead of the commonly found volatility U-shape, we only find increased levels of volatility at the start, but not at the end of the average trading day in 2012.

To highlight the possible departures of the the fitted signal from the average spline level across all days, we also present the estimates of the spline plus the autoregressive component $(s_t + \alpha_t)$ for one specific day (August 1, 2012) in Figure 3.3. We find that for each of the four stocks the intraday volatility pattern is close to the overall average spline pattern. At the same time, we observe that particularly the autoregressive component picks up substantial temporary departures from the average level within the day. The size and patterns of the departures vary per stock and per day. For some stocks, departures appear relatively short-lived; see, for example, Caterpillar and JPMorgan. For other stocks, such as Walmart and Coca-Cola, departures are much more persistent. These patterns reveal why the autoregressive component α_t contributes to the model specification and why it is statistically significant. In Section 3.3.7 we also verify whether α_t leads to more precise forecasts of the magnitude of price changes for the next day.

3.3.5 Goodness-of-fit

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

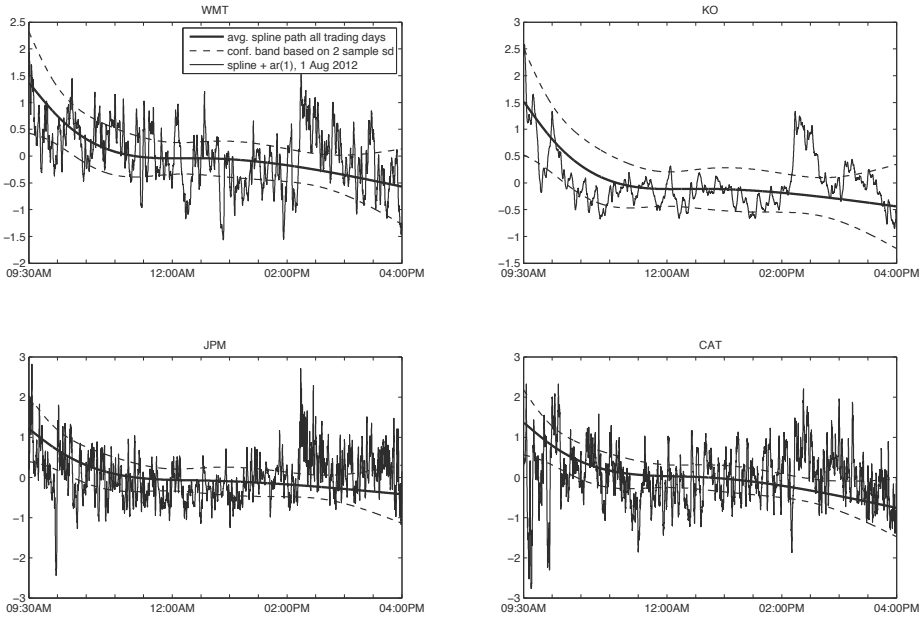

Figure 3.3: The figure shows the time series average of the zero sum spline s_t and a 95% confidence band based on all trading days of 2012. For Aug 1, 2012, it also shows the value of $s_t + \alpha_t$.

1. Model $\mathcal{A}$: the static type II modified Skellam model with $\mu_t = 0$ and static $\sigma_t^2 = \exp(c)$. The parameter vector is given by $\psi = (c, \delta, \gamma^*)'$.

2. Model $\mathcal{B}$: the spline-based model with $\mu_t = 0$ and time-varying $\sigma_t^2 = \exp(c + s_t)$, where s_t is the zero sum cubic spline specified in (3.13). The parameter vector is given by $\psi = (c, \delta, \gamma^*, \beta')'$.

3. Model $\mathcal{C}$: the complete model with $\mu_t = 0$ and $\sigma_t^2 = \exp(c + s_t + \alpha_t)$ as in (3.12). The parameter vector is given in Section 3.3.3.

For each model specification, the parameter vector is estimated by maximum likelihood. Figure 3.4 presents the log-likelihood differences (times 2) between Model $\mathcal{B}$ and Model $\mathcal{C}$ only, because the log-likelihood differences with respect to Model $\mathcal{A}$ are all much larger. For almost all stocks and days, the differences between the maximised log-likelihood values are large and statistically significant. In most cases the differences are so large that also in terms of model selection criteria, such as the Akaike information criterion, model $\mathcal{C}$ is strongly preferred over model $\mathcal{B}$.

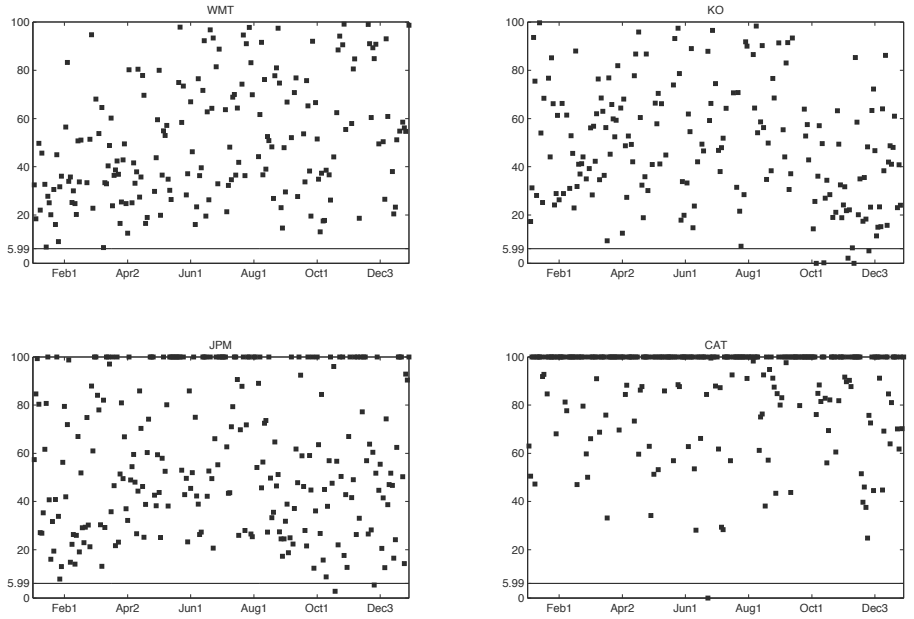

Figure 3.4: Each panel is for a stock and presents the log-likelihood differences (times 2) for all days in 2012. A dot indicates the log-likelihood ratio values for a specific day in 2012 between a model with only a constant and a spline $c+s_t$, Model $\mathcal{B}$, and a model with spline and autoregressive component $c+s_t+\alpha_t$, Model $\mathcal{C}$. The horizontal line indicates the 5% critical value for the $\chi^2(2)$ distribution corresponding to hypothesis H_0: $\phi = 0, \sigma_\eta = 0$. The differences are capped at 100 for visualization purposes.

3.3.6 Diagnostic checking

Variance of importance sampling weights

The estimation results from Section 3.3.3 rely on importance sampling methods. The log importance sampling weights can be used for diagnostic checking purposes. When the sample variance of the importance weights is high, likelihood calculations and signal extraction may change substantially when a different simulation sample is used. Geweke (1989) argues that importance sampling methods should only be used if the variance of the importance weights is known to exist. Robert and Casella (2004) provide examples of importance samplers that do not meet this condition and cases where this leads to biased results.

For our data, we find that sample variances of the importance sampling weights are generally low, typically smaller than 1. To verify more formally whether the variances of the importance weights exist, we follow Koopman et al. (2009). Using maximum likelihood, they estimate the shape parameter ξ and the scale parameter β of a generalized

Pareto distribution for the largest 1% to 50% out of 100,000 importance sampling weights. If the null hypothesis $H_0 : \xi \leq 1/2$ cannot be rejected, they conclude that the variance of the importance sampling weights is finite and that results can be trusted.

Pearson residuals

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

$$e_t = \frac{y_t - \mathbb{E}(y_t|y_{1:t-1})}{\sqrt{\mathbb{V}\mathrm{ar}(y_t|y_{1:t-1})}}, \qquad t = 1, \ldots, n, \tag{3.14}$$

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

Forecast distribution tests

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

Diagnostic testing results

We apply the above diagnostic tests to our MSKII$(-1, 1, 0; 0, \sigma_t^2, \gamma(\sigma_t^2))$ model, Model $\mathcal{C}$. We benchmark the results against the two alternative specifications, Models $\mathcal{A}$ and $\mathcal{B}$. We

select the first trading day of every even month and present the corresponding diagnostic test results for this day in Table 3.2.

Table 3.2 shows that except for the single case of Caterpillar on Dec 03, 2012, the null hypothesis of a finite variance of the importance sampling weights is never rejected. The results also clearly support that allowing for intraday dynamics in σ_t^2 is important. The static model $\mathcal{A}$ is uniformly rejected based on all versions of the Ljung-Box test statistics. Interestingly, the results for the spline-based model $\mathcal{B}$ and the dynamic model $\mathcal{C}$ appear to be more similar. Based on autocorrelations in the levels of e_t or e_t^* the two models perform very similar, with a slight advantage for model $\mathcal{C}$. However, the dynamic model is much more adequate in filtering out the serial dependence in the second order moments, as revealed by the test results for e_t^2 and e_t^{*2}. Whereas model $\mathcal{B}$ has unacceptable diagnostics for most stocks and days, the diagnostic tests for model $\mathcal{C}$ are mostly insignificant. We conclude that the autoregressive intraday component present in our new dynamic modified Skellam model is key to the good performance of the model. It results in a better performance than the commonly used intraday spline-based model.

3.3.7 Forecasting study

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

$$p_{|II|}(X_t = x_t; \sigma_t^2, \gamma_t) = \begin{cases} p_{II}(Y_t = 0; -1, 1, 0, 0, \sigma_t^2, \gamma_t), & \text{for } x_t = 0, \\ 2 \cdot p_{II}(Y_t = x_t; -1, 1, 0, 0, \sigma_t^2, \gamma_t), & \text{for } x_t \geq 1. \end{cases} \tag{3.15}$$

The five considered models have in common that they all derive probabilities according to the type II modified Skellam distribution. They differ in the way the Skellam parameters σ_{t+1}^2 and γ_{t+1} are obtained. Models $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are the parametric models as listed in Section 3.3.5. Models $\mathcal{D}$ and $\mathcal{E}$ are nonparametric benchmarks that are specified as follows.

(iv) Model $\mathcal{D}$: we estimate σ_{t+1}^2 using the sample variance using all observations in a rolling window of the past 900 seconds. We set $\gamma_t = 0$, such that the model collapses to the standard Skellam model.

(v) Model $\mathcal{E}$: both σ_{t+1}^2 and γ_{t+1} are obtained non-parametrically from the data. Define the empirical probability of a zero as $\hat{P}_0$ and $\hat{\sigma}_{t+1}^2$ as obtained under model $\mathcal{D}$. We

Table 3.2: The table reports diagnostic test results of the importance sampling weights, the Pearson residuals, and the quasi-residuals e_t^* as explained in Section 3.3.6. We report results for the first trading day of every even month in 2012. The 1% critical value of the Ljung-Box statistic up to 20 lags is 37.6.

WMT	Model	feb-01	apr-02	jun-01	aug-01	oct-01	dec-03	KO	feb-01	apr-02	jun-01	aug-01	oct-01	dec-03
$H_0 : \xi \leq 1/2$	$\mathcal{C}$	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
$LB_{20}(e_t)$	$\mathcal{A}$	107.84	32.43	52.46	29.47	16.01	62.02		21.86	100.71	31.68	55.44	55.10	45.47
	$\mathcal{B}$	37.60	30.77	54.04	31.98	18.02	45.49		18.99	97.56	27.12	48.92	34.50	45.47
	$\mathcal{C}$	32.43	30.21	53.84	23.99	18.34	41.78		17.27	94.31	24.86	35.21	34.22	45.88
$LB_{20}(e_t^2)$	$\mathcal{A}$	1893.70	317.90	112.74	302.89	143.87	560.00		298.91	308.66	1376.82	1199.65	2653.58	171.70
	$\mathcal{B}$	271.03	38.21	74.90	177.44	28.97	116.42		43.48	44.62	185.47	179.98	124.89	65.59
	$\mathcal{C}$	11.49	33.19	18.70	33.53	14.67	91.42		12.01	27.26	19.53	11.43	27.18	13.63
$LB_{20}(e_t^*)$	$\mathcal{A}$	63.28	21.24	47.26	33.64	15.44	49.68		19.11	130.12	29.76	43.39	37.28	39.62
	$\mathcal{B}$	35.80	22.11	47.85	33.58	18.04	43.99		15.48	95.50	34.47	39.92	32.92	41.84
	$\mathcal{C}$	32.44	21.96	45.67	25.38	17.93	39.47		13.18	93.45	31.68	28.29	30.49	42.48
$LB_{20}(e_t^{*2})$	$\mathcal{A}$	1652.03	182.91	103.38	373.31	170.06	532.21		223.94	188.60	1079.65	1315.75	1636.69	128.31
	$\mathcal{B}$	191.08	22.42	60.92	228.01	27.79	111.69		34.06	27.52	176.30	246.09	104.17	63.28
	$\mathcal{C}$	14.89	16.90	18.97	35.59	15.09	79.97		19.30	22.94	19.20	19.96	19.72	19.92
JPM								CAT						
$H_0 : \xi \leq 1/2$	$\mathcal{C}$	✓	✓	✓	✓	✓	✓		✓	✓	✓	✗	✓	✓
$LB_{20}(e_t)$	$\mathcal{A}$	20.64	23.74	17.75	63.33	28.06	34.22		66.64	54.98	26.58	46.93	41.67	31.76
	$\mathcal{B}$	25.19	27.33	17.66	60.74	25.91	27.33		98.35	37.02	20.75	53.73	29.32	24.97
	$\mathcal{C}$	24.76	26.38	15.72	45.04	27.87	24.12		81.57	36.06	21.60	38.71	30.42	20.24
$LB_{20}(e_t^2)$	$\mathcal{A}$	702.73	262.51	1322.65	176.69	633.21	89.10		1425.48	740.73	190.91	555.87	442.44	670.38
	$\mathcal{B}$	67.68	54.69	39.41	150.90	140.74	29.82		98.19	14.39	39.31	508.94	105.64	82.30
	$\mathcal{C}$	17.63	21.40	24.28	34.57	34.76	9.26		22.18	13.67	18.67	11.64	17.64	18.77
$LB_{20}(e_t^*)$	$\mathcal{A}$	17.30	21.76	13.85	39.26	20.46	36.52		63.80	47.24	28.31	46.75	37.59	29.43
	$\mathcal{B}$	18.10	24.65	12.26	38.25	23.61	31.05		90.74	34.23	23.09	52.55	27.23	23.64
	$\mathcal{C}$	17.69	23.32	11.12	35.31	25.16	29.06		75.58	33.24	23.64	41.04	28.36	20.40
$LB_{20}(e_t^{*2})$	$\mathcal{A}$	421.24	148.03	1314.13	293.28	659.93	112.03		1651.56	804.09	219.01	724.89	523.39	777.93
	$\mathcal{B}$	57.18	41.60	42.59	208.09	144.64	39.43		106.74	20.89	40.18	584.36	104.49	76.31
	$\mathcal{C}$	17.09	18.91	15.93	28.58	33.03	13.42		22.62	13.24	20.97	12.42	22.55	18.65

Table 3.3: The table presents the total log loss (LOGL) of the 21 trading days of June 2012. The losses are based on the forecasting study presented in Section 3.3.7. The DM statistic represents the Diebold and Mariano (1995) statistic which is asymptotically distributed as a standard normal random variable and hence rejects the null hypothesis of equal predictive accuracy at the 5% level of significance in favour of Model $\mathcal{C}$ if the DM test statistic is smaller than -1.65.

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

then solve two equations for two unknowns, namely

$$\hat{\sigma}^2_{t+1} = \sigma^2_{t+1} - 2\gamma_{t+1}P_1, \tag{3.16}$$

$$\hat{P}_0 = P_0 + 2\gamma_{t+1}P_1, \tag{3.17}$$

where equations (3.16) and (3.17) follow from equations (3.4) and (3.3), respectively. By the substitution of (3.17) into (3.16), we obtain $\hat{\sigma}^2_{t+1} = \sigma^2_{t+1} - \hat{P}_0 + P_0$ which we solve numerically for σ^2_{t+1} using a binary search algorithm. The resulting σ^2_{t+1} is substituted into (3.16) to obtain γ_{t+1}.

We emphasize that Models $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ use the subsequent estimated parameter vectors from the day before. Further extensions can be obtained by considering a forecasting model for the daily estimates of ψ; for instance, see Diebold and Li (2006). Even without these modifications, the forecasting experiment already produces some clear advantages of the new dynamic Skellam model, Model $\mathcal{C}$. For all models and all trading days, we start our forecast evaluation after a burn-in period of $\tau = 60$ seconds. Models $\mathcal{D}$ and $\mathcal{E}$ subsequently extend the burn-in window to 900 seconds, after which the forecasts are updated using a rolling window. The results are presented in Table 3.3.

The performance of the models is first assessed in terms of an out-of-sample probabilistic loss function LOGL, which can be classified as a proper scoring rule; see Winkler (1969). LOGL_h sums the log probabilities for Model $h \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}\}$ using the model's predictive pmf and the realised absolute tick-size change x_{t+1}. A loss of zero indicates that the absolute tick-size change x_{t+1} was perfectly predicted by the model. The log loss differences can also be compared between models using the Diebold Mariano (DM) test statistic; see Diebold and Mariano (1995). The DM statistic is asymptotically normally distributed under the null hypothesis of equal predictive accuracy. We take Model $\mathcal{C}$ as our benchmark in the computation of the Diebold Mariano statistics.

Table 3.3 shows that the forecasts based on Model $\mathcal{C}$ have always the lowest log loss. The new fully dynamic type II modification of the Skellam model clearly outperforms its static (Model $\mathcal{A}$) and spline-based (Model $\mathcal{B}$) counterparts, as well as the non-parametric zero-inflation model, Model $\mathcal{E}$. Using a one-sided test, Model $\mathcal{C}$ also significantly outperforms the nonparametric benchmark Model $\mathcal{D}$ for 3 out of the 4 stocks. Only for Caterpillar, the two models cannot be distinguished in a statistically significant manner. However, the excellent forecasting performance of Model $\mathcal{C}$ remains despite its use of the estimate of the constant, spline, and autoregressive parameters of the day before. We emphasize that the parameter estimates are not recursively updated during the day. Models $\mathcal{D}$ and $\mathcal{E}$, by contrast, do not rely on any parameter estimates from the previous day.

3.4 Conclusions

We have modelled tick-by-tick discrete price changes for U.S. stocks listed on the New York Stock Exchange. The analysis of high-frequency data attracts ever more attention from both government regulators and the financial industry. Hence the understanding of the dynamics in high-frequency data has become important. We have shown that the empirical analysis of high-frequency tick-by-tick data can be based on modifications and dynamic extensions of the Skellam distribution. Our type II modified Skellam distribution features a dynamic variance parameter, and a dynamic transfer of probability mass to accommodate the non-standard properties of the data in terms of the occurrence of zero-price-changes. These features of our model are needed to have a stable importance sampling estimation procedure, a good in-sample fit, an adequate diagnostic performance, and an accurate out-of-sample forecasting performance, in comparison to a number of relevant benchmark models. We conclude that the new dynamic modified Skellam model provides a flexible modelling framework that can be effectively employed to capture the dynamics in high-frequency tick-by-tick data with many missing entries. Since the model produces intraday patterns of high-frequency volatility dynamics, it may provide an interesting and complementary perspective to the literature on nonparametric realised volatility measures and realised kernels which are proposed by Barndorff-Nielsen and Shephard (2001, 2002) and Andersen et al. (2001). Further research could be directed towards the comparison of the Skellam stochastic volatility model and the 'standard' stochastic volatility literature that usually applies continuous distributions like the Gaussian or Students t distribution. Interesting results could be obtained by comparing the efficiency of both models and the characteristics of the discovered dynamics.

Appendices

The following appendices are part of the chapter 'Intraday Stochastic Volatility in Discrete Price Changes' and are organised as follows. Appendix A and B provide moments of the modified Skellam distribution, type I and II. Appendix C shows a simulation study and provides evidence of the accuracy of the novel Skellam model. Appendix D discusses the numerically accelerated importance sampling methodology and the extensions to a bivariate signal framework. This part contains more material than strictly necessary, however, Chapter 4 relies on Appendix D as well. Additional tables and figures are provided in Appendix E.

A Modified Skellam distribution of type I

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

$$p_I(y_t; \mu, \sigma^2, \gamma) = \begin{cases} (1-\gamma)p(Y_t = y_t; \mu, \sigma^2), & \text{for } y_t \neq 0, \\ \gamma + (1-\gamma)p(Y_t = 0; \mu, \sigma^2), & \text{for } y_t = 0, \end{cases} \tag{3.18}$$

where $\gamma \in (\frac{P_0}{P_0-1}, 1)$ and $P_q = p(q; \mu, \sigma^2)$ as defined in equation (3.1), $q \in \mathbb{Z}$. For $\gamma = 0$ we recover the Skellam distribution as defined in (3.1) and for $\gamma = \frac{P_0}{P_0-1}$ we have the lower bound $P_{0,I} = 0$ with $P_{q,I} = p_I(q; \mu, \sigma^2, \gamma)$ as defined in equation (3.18). If unimodality is required the zero deflation should be bounded as $\gamma \in (\frac{\min(P_{-1}, P_1) - P_0}{1 + \min(P_{-1}, P_1) - P_0}, 1)$ which ensures $P_{0,I} \geq \min(P_{-1,I}, P_{1,I})$. The mean and variance of the MSKI distribution are $\mathbb{E}(Y_t) = (1-\gamma)\mu$ and $\mathbb{V}\text{ar}(Y_t) = (1-\gamma)\sigma^2 + \gamma(1-\gamma)\mu^2$ which follows from

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

with $\sum_{x=-\infty}^{\infty} x^2 \, p(Y_t = x; \mu, \sigma^2) = \sigma^2 + \mu^2$ being the second moment of the Skellam distribution of (3.1). The inflation/deflation of probability mass to non-zero values of Y_t can be achieved in a similar way.

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

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

$$
\begin{aligned}
\mathbb{E}(Y_t) &= \sum_{x\in\mathbb{Z}} x\, p_{II}(Y_t = x; \mu, \sigma^2, \gamma) \\
&= \left[\sum_{x\in\mathbb{Z}\setminus\{i,j,k\}} x\, p(Y_t = x; \mu, \sigma^2)\right] + i(1-\gamma)P_i + j(1-\gamma)P_j + k(\gamma P_i + \gamma P_j + P_k) \\
&= \left[\sum_{x\in\mathbb{Z}\setminus\{i,j,k\}} x\, p(Y_t = x; \mu, \sigma^2)\right] + iP_i + jP_j + kP_k - i\gamma P_i - j\gamma P_j + k\gamma P_i + k\gamma P_j \\
&= \mu - i\gamma P_i - j\gamma P_j + k\gamma P_i + k\gamma P_j,
\end{aligned}
\tag{3.19}
$$

which is equal to the first equation of (3.4).
The second moment of the MSKII$(i, j, k, \mu, \sigma^2, \gamma)$ distribution is given by

$$
\begin{aligned}
\mathbb{E}(Y_t^2) &= \sum_{x\in\mathbb{Z}} x^2\, p_{II}(Y_t = x; \mu, \sigma^2, \gamma) \\
&= \left[\sum_{x\in\mathbb{Z}\setminus\{i,j,k\}} x^2\, p(Y_t = x; \mu, \sigma^2)\right] + i^2(1-\gamma)P_i + j^2(1-\gamma)P_j + k^2(\gamma P_i + \gamma P_j + P_k) \\
&= \left[\sum_{x\in\mathbb{Z}\setminus\{i,j,k\}} x^2\, p(Y_t = x; \mu, \sigma^2)\right] + i^2P_i + j^2P_j + k^2P_k \\
&\qquad\qquad - i^2\gamma P_i - j^2\gamma P_j + k^2\gamma P_i + k^2\gamma P_j \\
&= \sigma^2 + \mu - i^2\gamma P_i - j^2\gamma P_j + k^2\gamma P_i + k^2\gamma P_j.
\end{aligned}
\tag{3.20}
$$

Combining (3.19) and (3.20) leads to the variance of the MSKII(i, j, k) distribution as presented in the second equation of (3.4).

C Simulation study

We conduct a simulation study to verify the performance of the importance sampling estimation methodology explained in Appendix D in combination with the Skellam model as presented in (3.12). The case of zero inflation, zero deflation and zero neutral is covered in this study. We assume that the Skellam model of (3.12) is the true data generating process and we simulate time series of Skellam variables with length $n = 23{,}400$ which is equal to the length of the tick price change series in the application of this chapter. To incorporate missing values in the simulated data sets we denote $P_{.NaN}$ which is the probability of no trade at time t. We set $P_{.NaN} = 0.85$ at 09:30 and 16:00 and

Table 3.4: This table reports simulation averages of maximum likelihood estimates of the static parameters for the dynamic Skellam model of Section 3.3. The simulation averages are calculated with $R = 100$ replications of time series with length $n = 23{,}400$. The true parameter values are in the table above the simulated values. Standard deviations of the estimates over the Monte Carlo simulations are in parentheses. The column $t(s)$ denotes the average computation time (in seconds) for finding the maximum of the log likelihood function. Computations are carried out on a i7-2600, 3.40 GHz desktop PC using four cores.

	ϕ	σ_η	c	γ^*	β_1	β_2	$t(s)$
true	0.99	0.05	-0.30	0.00	1.00	-0.40	
	0.987	0.055	-0.298	-0.024	1.005	-0.400	356.24
	(0.007)	(0.022)	(0.065)	(0.082)	(0.131)	(0.064)	
true	0.95	0.15	0.10	-0.50	1.00	-0.40	
	0.944	0.154	0.101	-0.498	0.997	-0.395	271.71
	(0.022)	(0.046)	(0.059)	(0.140)	(0.110)	(0.055)	
true	0.95	0.15	0.10	0.25	1.00	-0.40	
	0.945	0.150	0.104	0.252	0.996	-0.396	269.58
	(0.030)	(0.054)	(0.056)	(0.028)	(0.107)	(0.054)	

$P_{.NaN} = 0.95$ at 13:00. Every $P_{.NaN}$ between the time points 09:30$-$13:00 and 13:00$-$16:00 is determined by two triangles with the hypotenuses connecting $P_{.NaN} = 0.95$ in the middle of the day and $P_{.NaN} = 0.85$ at the beginning and end of the day. With the probability of a missing value over the day, missing values are randomly positioned at time points with the idea that the probability of a missing values is highest when trading activity is lowest. We refer to, for example, Koopman, Lit, and Lucas (2015) for graphs of trading patterns. For this simulation study, we obtain an average of 2000-2500 simulated trades out of 23,400 which is just below average.

The simulated data comes from a slightly more parsimonious model specification than (3.12). We set $\delta = 0.30$, $\sigma_{\eta,S} = 0$ and the vector of hyper parameters has dimension 6 and is given by

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

where the elements of the 2×1 vector β correspond to a zero sum spline with spline knots placed at {09:30, 12:30, 16:00}. We present the estimation results in Table 3.4.

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

D Numerically accelerated importance sampling

Likelihood evaluation and importance sampling

We can express the likelihood function for the non-Gaussian nonlinear state space model (3.9) as

$$L(y;\psi) = \int p(y,\theta;\psi)\,\mathrm{d}\theta = \int p(y|\theta;\psi)p_g(\theta;\psi)\,\mathrm{d}\theta. \tag{3.21}$$

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 (3.21); see Ripley (1987) for a general introduction. A naive Monte Carlo estimate of $L(y;\psi)$ is given by

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

where S is the number of Monte Carlo replications and the simulated value of $\theta^{(k)}$ is obtained by simulating the state vectors from the vector autoregressive process (3.8) and with $\theta = c + Z\alpha$ for a given parameter vector ψ. This Monte Carlo estimate is numerically highly inefficient since the simulated paths have no support from y.

In various contributions in statistics and econometrics it is argued that (3.21) 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 $g(\theta|y;\psi^*)$ from which the θs are sampled conditional on the observation vector y, where ψ^* denotes a fixed parameter vector, containing ψ as well as parameters $\tilde{\psi}$ particular to the importance density $g(y|\theta;\tilde{\psi})$, i.e., $\psi^* = (\psi', \tilde{\psi}')'$. Under the assumption that a numerically efficient device can be developed for sampling θ from $g(\theta|y;\psi^*)$, we can express the likelihood function (3.21) in terms of the importance density as

$$L(y;\psi) = \int \frac{p(y,\theta;\psi)}{g(\theta|y;\psi^*)} g(\theta|y;\psi^*)\,\mathrm{d}\theta, \tag{3.22}$$

with the importance sampling estimate given by

$$\frac{1}{S}\sum_{k=1}^{S} \omega(y,\theta^{(k)};\psi^*), \qquad \omega(y,\theta;\psi^*) = \frac{p(y,\theta;\psi)}{g(\theta|y;\psi^*)}, \qquad \theta^{(k)} \sim g(\theta|y;\psi^*), \tag{3.23}$$

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

$$\omega(y, \theta; \psi^*) = \frac{p(y, \theta; \psi)}{g(\theta|y; \psi^*)} = \frac{p(y|\theta; \psi) p_g(\theta; \psi)}{g(y|\theta; \tilde{\psi}) g(\theta; \psi) / g(y; \psi^*)} = g(y; \psi^*) \frac{p(y|\theta; \psi)}{g(y|\theta; \tilde{\psi})}, \tag{3.24}$$

see, for example, Durbin and Koopman (2012). The density $g(y; \psi^*)$ can be taken as a scaling function since it does not depend on θ. The function $\omega(y, \theta; \psi^*)$ is usually referred to as the importance sampling weight function. If the variance of $\omega(y, \theta; \psi^*)$ exists, the estimate (3.23) is consistent for any $g(y|\theta; \tilde{\psi})$ 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.

Construction of the importance density

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

$$\underset{\tilde{\psi}_t}{\operatorname{argmin}} \int \lambda^2(y_t, \theta_t; \psi^*)\, \omega_t(y_t, \theta_t; \psi^*)\, g(\theta_t|y; \psi^*)\, \mathrm{d}\theta_t, \tag{3.25}$$

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

$$\lambda(y_t, \theta_t; \psi^*) := \log\, \omega_t(y_t, \theta_t; \psi^*) := \log\, p(y_t|\theta_t; \psi) - \log\, g(y_t|\theta_t; \tilde{\psi}_t). \tag{3.26}$$

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 (3.25) using importance sampling and perform its minimization via weighted least squares regression. Koopman, Lit, and Nguyen (2012) show that the EIS method can also fully rely on computationally efficient Kalman filter and smoothing methods. Their modification leads

to a faster and efficient importance sampling method, especially for large state dimensions.

In a further development of EIS, Koopman et al. (2014) replace the evaluation of the integral in (3.25) 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 $g(\theta_t|y;\psi^*)$ given the Gaussian importance densities $g(y|\theta;\tilde{\psi})$ and a Gaussian marginal density $g(\theta;\psi) = p_g(\theta;\psi)$. Although NAIS was originally developed for a univariate signal $\theta_t \in \mathbb{R}$, the method can easily be extended to multiple dimensions; see Scharth (2012, Ch. 5) and the discussions in Koopman et al. (2014). Scharth (2012) proposes Halton sequences and quasi-Monte Carlo integration for the evaluation of high dimensional integrals. In the case of our dynamic Skellam model, the signal is only two-dimensional and hence we can still rely on Gauss-Hermite quadrature methods efficiently.

Bivariate numerically accelerated importance sampling

To facilitate the exposition, we express the Gaussian density as a kernel function in θ_t,

$$g(y|\theta;\tilde{\psi}) = \prod_{t=1}^{n} g(y_t|\theta_t;\tilde{\psi}_t), \qquad g(y_t|\theta_t;\tilde{\psi}_t) = \exp\left(a_t + b_t'\theta_t - \frac{1}{2}\theta_t' C_t \theta_t\right), \tag{3.27}$$

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

$$\underset{\tilde{\psi}_t^{(k+1)}}{\operatorname{argmin}} \int\int \lambda^2(y_t, \theta_t; \psi^{*(k+1)}) \omega_t(y_t, \theta_t; \psi^{*(k)}) g(\theta_t|y;\psi^{*(k)})\, d\theta_{1t}\, d\theta_{2t}, \tag{3.28}$$

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

$$g(\theta_t|y;\psi^{*(k)}) = \mathrm{N}(\widehat{\theta}_t^{(k)}, V_t^{(k)}) = \frac{1}{2\pi|V_t^{(k)}|^{1/2}} \exp\left(-\frac{1}{2}(\theta_t - \widehat{\theta}_t^{(k)})'(V_t^{(k)})^{-1}(\theta_t - \widehat{\theta}_t^{(k)})\right), \tag{3.29}$$

where $\widehat{\theta}_t^{(k)}$ and $V_t^{(k)}$ are obtained from the Kalman filter and smoother, for given $\psi^* = \psi^{*(k)}$, applied to the linear Gaussian model $x_t = \theta_t + u_t$ with disturbance $u_t \sim \mathrm{N}(0, C_t^{-1})$

and pseudo-observation $x_t = C_t^{-1} b_t$, for $t = 1, \ldots, n$. It is straightforward to verify that the observation density $\prod_{t=1}^{n} g(x_t|\theta_t; \tilde{\psi}_t)$ is equivalent to $g(y|\theta; \tilde{\psi})$ in (3.27).

We numerically implement the minimization in (3.28) by the Gauss-Hermite quadrature method; see, for example, Monahan (2001). For this purpose we define

$$\varphi(y_t, \theta_t; \tilde{\psi}_t^{(k+1)}, \psi^{*(k)}) = \lambda^2(y_t, \theta_t; \psi^{*(k+1)}) \omega_t(y_t, \theta_t; \psi^{*(k)}), \tag{3.30}$$

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 (3.28) becomes

$$\underset{\tilde{\psi}_t^{(k+1)}}{\operatorname{argmin}} \sum_{i=1}^{M} \sum_{j=1}^{M} w_{ij} \cdot \varphi(y_t, \tilde{\boldsymbol{z}}_{ij,t}^{(k)}; \tilde{\psi}_t^{(k+1)}, \psi^{*(k)}), \tag{3.31}$$

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 $\tilde{\boldsymbol{z}}_{ij,t}^{(k)} = \widehat{\theta}_t + F_t^{(k)} \boldsymbol{z}_{ij}$, where the 2×2 square root matrix $F_t^{(k)}$ is the result of the decomposition $V_t^{(k)} = F_t^{(k)} F_t^{(k)\prime}$ and $\boldsymbol{z}_{ij} = (z_i\,,\, z_j)'$ for $i, j = 1, \ldots, M$. In this implementation we have used the fact that $g(\tilde{\boldsymbol{z}}_{ij,t}^{(k)}|y; \psi^{*(k)}) \propto \exp(-\frac{1}{2}\boldsymbol{z}_{ij}'\boldsymbol{z}_{ij})$; see Koopman et al. (2014) and Scharth (2012, Ch. 5). The decomposition of $V_t^{(k)}$ 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 (3.31) as a standard weighted least squares computation applied to M^2 observations for the regression equation

$$\log p(y_t|\tilde{\boldsymbol{z}}_{ij,t}^{(k)}) \;\; = \;\; \text{constant} + \kappa' \tilde{\boldsymbol{z}}_{ij,t}^{(k)} - \frac{1}{2}\xi' \text{vech}(\tilde{\boldsymbol{z}}_{ij,t}^{(k)} \tilde{\boldsymbol{z}}_{ij,t}^{(k)\prime}) + \text{error}, \tag{3.32}$$

where κ and ξ are regression coefficient vectors and the regression weights are given by $w_{ij} \cdot \omega_t(y_t, \tilde{\boldsymbol{z}}_{ij,t}^{(k)}; \psi^{*(k)}) \cdot g(\tilde{\boldsymbol{z}}_{ij,t}^{(k)}|y; \psi^{*(k)})$, and where $\text{vech}(\cdot)$ stacks elements of the upper triangular part of a symmetric matrix into a vector. The resulting weighted least squares estimates for κ and ξ yield the new values for $b_t^{(k+1)}$ and $\text{vech}(C_t^{(k+1)})$, respectively. Hence, new values for $\tilde{\psi}_t^{(k+1)}$ are obtained for each $t = 1, \ldots, n$. Using these new estimates, we can determine a new $g(\theta_t|y; \psi^{*(k+1)})$ in (3.29) by constructing a new time series x_t and applying the Kalman filter and smoother to the linear Gaussian model given below (3.29). In this last step we obtain new values for $\widehat{\theta}_t^{(k+1)}$ and $V_t^{(k+1)}$, which we require in (3.29).

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

NAIS: the algorithm

The minimum of (3.31) is obtained when $\log p(y_t|\theta_t;\psi) = \log g(y_t|\theta_t;\tilde{\psi}_t)$. Therefore we regress the log Gaussian density $\log g(y_t|\theta_t;\tilde{\psi}_t)$ as given by (3.27) on the log observation density $\log p(y_t|\theta_t;\psi)$ by use of weighted least squares. The regression coefficient vector at time t, Ψ_t, consists of the intercept a_t, the individual components of the 2×1 vector b_t and the 2×2 matrix C_t at time t, i.e. $\Psi_t = (a_t, \kappa', \xi')'$. The optimum values $\widehat{\Psi}_t$ are obtained by applying the following iterative algorithm

(i) Find appropriate starting values for κ and ξ with $t = 1, ..., n$ and set $s = 1$ and $\Psi_t^{(s)} = (a_t, \kappa', \xi')'$. In most cases the algorithm is not very sensitive to starting values so κ consisting of ones and $C_t(\xi)$ set to I_2 suffices.

(ii) Construct the linear Gaussian state space model with observation equation $x_t = \theta_t + u_t$ with disturbance $u_t \sim \mathrm{N}(0, C_t^{-1})$ and pseudo-observation $x_t = C_t^{-1} b_t$, for $t = 1, \ldots, n$ and apply the Kalman filter and smoother to obtain $\widehat{\theta}_t^{(k)}$ and $V_t^{(k)}$ and use these to calculate $\tilde{\boldsymbol{z}}_{ij,t}^{(k)}$ as described below equation (3.31).

(iii) Minimize equation (3.31) by weighted least squares and obtain $\Psi_t^{(s+1)}$.

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

Once the iterative algorithm has converged in step (iv), $\Psi_t^{(s+1)}, t = 1, \ldots, n$ represents the new importance density. The number of times the algorithm needs to be called before convergence depends on the model and the size of the dataset. Starting from init values the algorithm converges most of the time in 10 steps or less. The minimization of (3.31) can be carried out independently for all time points t and can therefore be done in parallel over t.

E Intradaily time series of price changes in 2012

Table 3.5: The table reports the empirical distribution (in percentage points) of tick price changes for the four stocks Walmart (WMT), Coca-Cola (KO), JPMorgan (JPM), and Caterpillar (CAT), in 2012. The majority of the observations are -1, 1 and 0, the distribution is close to symmetric and it centers around zero which validates the use of the MSKII(-1,1,0) distribution presented in (3.3).

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

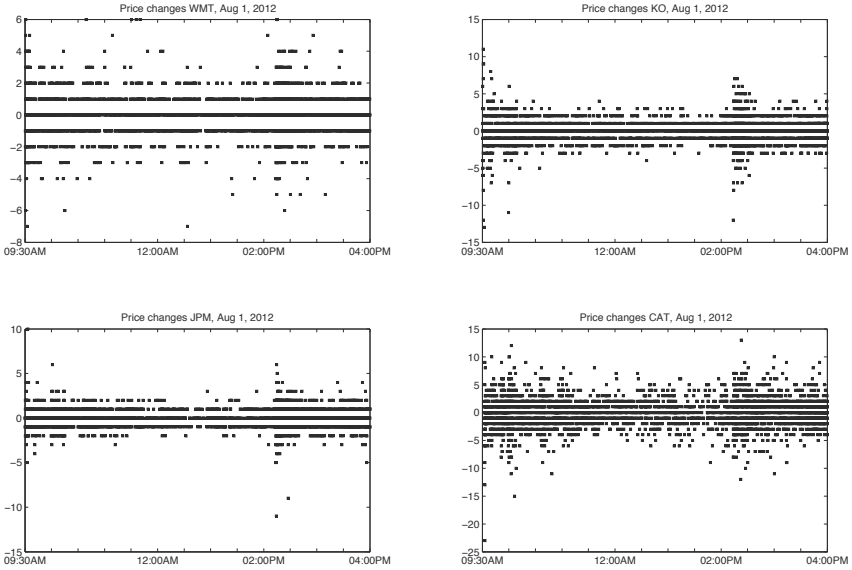

Figure 3.5: The panels show the observed price changes for August 1, 2012 for the four stocks {WMT,KO,JPM,CAT}.

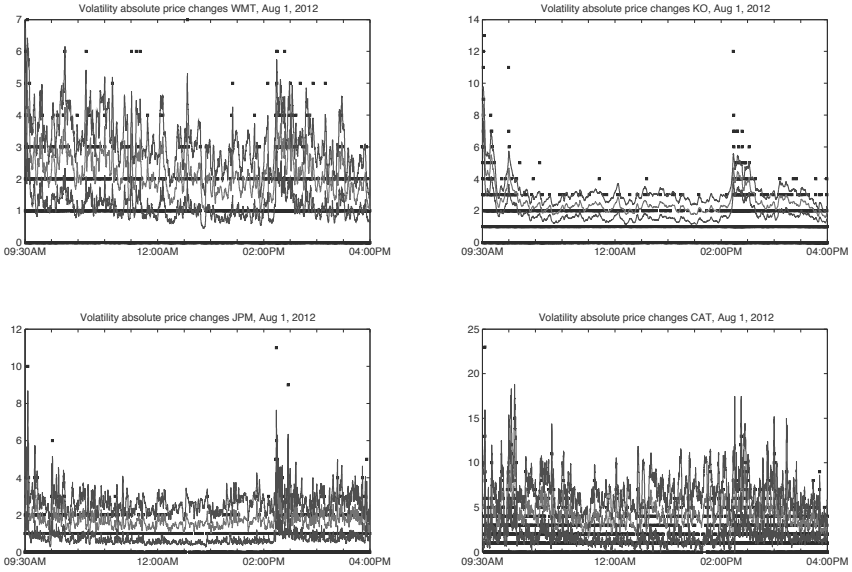

Figure 3.6: The panels show the absolute values of observed price changes for August 1, 2012 for the four stocks {WMT,KO,JPM,CAT}. Furthermore, in each panel the estimate of $2 \times \sigma_t$ is presented together with its estimated 95% confidence interval.

Chapter 4

A Skellam Model for Analysing the Differences in Count Data

4.1 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), Barndorff-Nielsen et al. (2012) and Chapter 3 and 5 of this dissertation. 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 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, Dietz, Schlattmann, Mendonca, and Kirchner (1999) and Karlis and Ntzoufras (2006). The change of the DFMT index over time or between regions can be modelled by the Skellam distribution. Another example is low-scoring sports such as ice-hockey and football where the score difference between the teams can be viewed as the difference between two Poisson counts and thus be modelled by a Skellam distributed random variable.

The Skellam distribution that we apply in this chapter is originally derived by Skellam (1946) and is characterized by two 'intensity' parameters. The Skellam distribution of Chapter 3 and 5 are a re-parameterization of the one we apply in this chapter. In earlier studies, the Skellam distribution is used from a perspective of static parameters. When we analyse 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 for which we rely on the numerical and simulation based methods (NAIS methodology) as described in Chapter 3. The difference between the methodology of this chapter and Chapter 3 is the use of the extended bivariate NAIS methodology by adopting bivariate Gauss-Hermite quadrature which we presented in Appendix D of Chapter 3.

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 a large scale application. We consider score differences of football 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 strengths of attack and defence 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 this chapter is organised as follows. We present the new dynamic Skellam model in Section 4.2 and explain how it can be cast in nonlinear non-Gaussian state space form. Section 4.3 treats a large unbalanced panel data set of German Bundesliga football matches to show how the method performs for high dimensional data sets, missing values and high dimensional state vectors. Section 4.4 concludes.

4.2 The dynamic Skellam model

4.2.1 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

$$\mathbb{P}(Y = y; \lambda_1, \lambda_2) = \exp\left(-\lambda_1 - \lambda_2\right) \left(\lambda_1/\lambda_2\right)^{y/2} I_{|y|}(2\sqrt{\lambda_1\lambda_2}), \tag{4.1}$$

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

$$\mathbb{E}(Y) = \lambda_1 - \lambda_2, \qquad \mathbb{V}\text{ar}(Y) = \lambda_1 + \lambda_2. \tag{4.2}$$

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 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

$$p_z(Y = y; \lambda_1, \lambda_2, \gamma) = \begin{cases} (1-\gamma)\, p(Y = y; \lambda_1, \lambda_2), & \text{for } y \neq 0, \\ \gamma + (1-\gamma)\, p(Y = 0; \lambda_1, \lambda_2), & \text{for } y = 0, \end{cases} \tag{4.3}$$

with $\gamma \in [0, 1)$ an additional unknown and fixed parameter, and $p(y; \lambda_1, \lambda_2)$ as defined in (4.1). For $\gamma = 0$, we recover the original Skellam distribution. The mean and variance of the zero-inflated Skellam distribution are

$$\mathbb{E}(Y) = (1-\gamma)(\lambda_1 - \lambda_2), \qquad \mathbb{V}\text{ar}(Y) = (1-\gamma)(\lambda_1 + \lambda_2) + \gamma(1-\gamma)(\lambda_1 - \lambda_2)^2. \tag{4.4}$$

The inflation of probability mass to non-zero values of Y can be achieved in a similar way. In Figure 4.1 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.

4.2.2 Dynamic specification of intensities

In the dynamic Skellam model, we replace Y, y, λ_1, and λ_2 in (4.1) by their time-varying counterparts Y_t, y_t, λ_{1t}, and λ_{2t}, respectively. We denote the dynamic model as

$$Y_t \sim \text{Skellam}(\lambda_{1t}, \lambda_{2t}), \qquad t = 1, \ldots, n, \tag{4.5}$$

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} which means that, conditional on λ_{1t} and λ_{2t}, Y_t is not subject to other dynamic processes. The dynamics

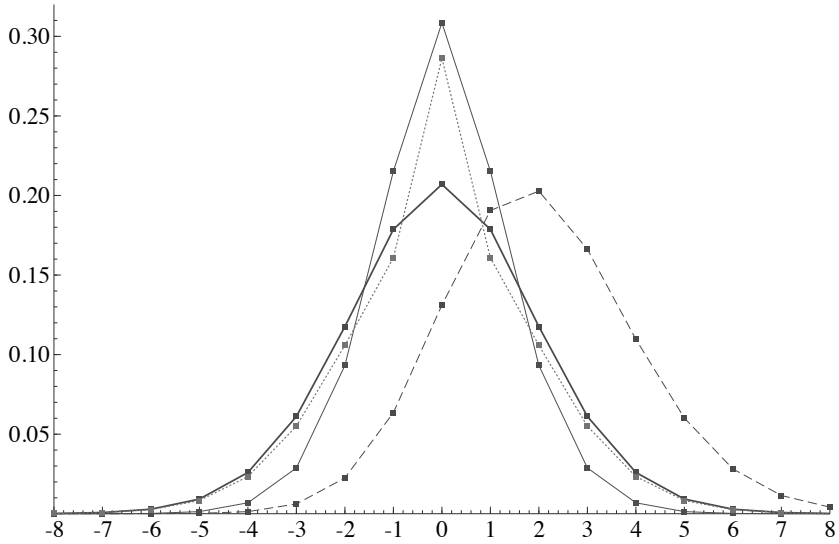

Figure 4.1: Skellam and zero-inflated Skellam distributions with density functions (4.1) and (4.3), respectively, for different λ_1, λ_2 and γ coefficients. These are discrete distributions, the connecting lines are drawn for clarity and do not indicate continuity. (———) $\lambda_1 = \lambda_2 = 1$; (———) $\lambda_1 = \lambda_2 = 2$; (− − −) $\lambda_1 = 3, \lambda_2 = 1$; (· · · · · ·) $\lambda_1 = \lambda_2 = 2, \gamma = 0.1$.

of λ_{1t} and λ_{2t} are modelled by a nonlinear transformation of an autoregressive process,

$$\lambda_{it} = s_i(\theta_t), \tag{4.6}$$

$$\theta_t = c_t + Z_t\alpha_t, \tag{4.7}$$

$$\alpha_{t+1} = d_t + T_t\alpha_t + \eta_t, \qquad \eta_t \sim \text{NID}(\mathbf{0}, Q_t), \tag{4.8}$$

for $i = 1, 2$ and $t = 1, \ldots, n$. The link functions $s_i(\theta_t)$, $i = 1, 2$ are exponential functions to ensure positivity of the intensities λ_{1t} and λ_{2t}. We refer to Section 3.2.3 for the specifications of equations (4.7) and (4.8) and we note that, except for the link functions in (4.6), the dynamic Skellam model specified above is identical to that of Section 3.2.3. We therefore refer to Section 3.2.3 for details of the non-Gaussian nonlinear state space model specified above and to the bivariate NAIS methodology in Appendix D of Chapter 3 for likelihood evaluation, importance sampling and construction of the importance density.

4.3 Analysing football scores

We consider score differences for football matches in the German Bundesliga. The number of goals per match in a football game is typically low, such that the score balance can

easily be viewed as a difference of two Poisson count variables, see Section 4.2. 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 that $y_{ij,t}$ is Skellam distributed.

Our data consists of weekly match results for 7 seasons of the German Bundesliga for the period from 2006–2007 to 2012–2013. 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). In each of the seasons in our data set, matches are postponed and extra time periods need to be added in the data set. The resulting calendar is adopted for the time index t in our analysis. This means that on the added time periods several teams do not play and missing observations need to be added to the data set which can be treated by the state space methodology in a routine manner; see also the discussion in the Appendix of Chapter 2.

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 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 bivariate NAIS estimation methodology for large unbalanced panels with many missing observations. Our state space modelling 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 strengths of attack (ξ_{it} and ξ_{jt}) and strengths of defence (β_{it} and β_{jt}) of both teams in week t. We assume a fixed time-invariant home ground advantage δ for all teams and model the scoring intensities as

$$\lambda_{1,ij,t} = \exp(\theta_{1,ij,t}) = \exp\big(\delta + \xi_{it} - \beta_{jt}\big), \qquad \lambda_{2,ij,t} = \exp(\theta_{2,ij,t}) = \exp\big(\xi_{jt} - \beta_{it}\big). \quad (4.9)$$

This parsimonious modelling framework for league match results using strengths of attack and defence is based on Maher (1982) and provides our benchmark model. The gener-

alization towards a dynamic bivariate Poisson model is developed in Chapter 2, 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 strengths of attack and defence for each team in the state vector α_t in (4.8), i.e.

$$\alpha_t = (\delta, \xi_{1t}, \ldots, \xi_{Jt}, \beta_{1t}, \ldots, \beta_{Jt})'. \tag{4.10}$$

We note that the state vector α_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 θ_t. When all teams play in week t, $K_t = 9$ and θ_t has 18 entries with a result out of a total of 29 (total number of teams) and has 11 missing entries. The vector θ_t can be constructed from (4.7) using the state vector α_t in (4.10) with $c_t = \mathbf{0}$ and Z_t an appropriate selection matrix as implied by (4.9). The strengths of attack and defence for each team evolve separately over time as

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

for $i = 1, \ldots, J$, and where the ε_{it} are mutually independent over i and t. Although each team has its own unique strength of attack and defence, the persistence coefficients ϕ_ξ and ϕ_β and the innovation variances σ_ξ^2 and σ_β^2 are common to all teams. This again results in a highly parsimonious model. We retain 4 parameters $\psi = (\phi_\xi, \phi_\beta, \sigma_\xi^2, \sigma_\beta^2)'$, which we estimate using the techniques outlined in Section 3.2.3.

An important difference between Chapter 3 and Chapter 4 is made in the applications. The Skellam distribution of Chapter 3 has one time-varying parameter, i.e. a time-varying variance (and a mean set to zero) whereas in the application of this chapter, λ_{1t} and λ_{2t} are both time-varying. In other words, here we have a non-Gaussian nonlinear state space model with a bivariate signal for which we need the bivariate NAIS methodology as presented in Appendix D of Chapter 3 in contrast to the univariate signal and methodology in the application of Chapter 3.

4.3.1 Estimation results

To verify the effect of different values of the number of importance draws S and the number of Gauss-Hermite quadrature points M on the estimation results, we present the estimation results for ψ in Table 4.1 using $S = 50$, 200, 1000 and $M = 10$, 20. The values of the maximised 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

Table 4.1: Estimates of parameter vector ψ: Bundesliga team strengths. The table presents the Monte Carlo estimates of the five model coefficients, where δ 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 maximised estimated log likelihood values (ℓ).

	$S = 50$		$S = 200$		$S = 1000$	
	$M = 10$	$M = 20$	$M = 10$	$M = 20$	$M = 10$	$M = 20$
ϕ_ξ	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*
$\sigma_\xi^2 \times 10^4$	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*
δ	0.2617	0.2617	0.2617	0.2617	0.2617	0.2617
	0.0262	*0.0262*	*0.0262*	*0.0262*	*0.0262*	*0.0262*
ℓ	-8137.50	-8137.50	-8137.54	-8137.54	-8137.51	-8137.51

compared to the estimated standard errors. We conclude that reasonable choices for S and M have no major impact on the results which shows the robustness of the methodology.

Table 4.1 shows that the strengths of attack and defence on a weekly basis are highly persistent and smooth. This is to be expected: the strengths of attack or defence of a particular team do 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 defence and attack strength, respectively. These values clearly point to stationary dynamics. Interestingly, the strengths of attack evolve more persistently over time than the strength of defence.

4.3.2 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 4.2. All results are computed for $S = 100$ and $M = 10$.

Heterogeneous panel

The benchmark model assumes that the dynamic properties of the attack and defence processes are the same for all teams, i.e., the coefficients ϕ_ξ, ϕ_β, σ_ξ^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 data set, the period from 2002–2003 to 2005–2006. 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 ψ. 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

$$\begin{aligned} &\phi_{\xi,I} = 0.9986, \quad \phi_{\beta,I} = 0.9958, \quad \sigma^2_{\xi,I} \times 10^4 = 6.43, \quad \sigma^2_{\beta,I} \times 10^4 = 5.24, \quad \delta = 0.263, \\ &\phi_{\xi,II} = 0.9737, \quad \phi_{\beta,II} = 0.9851, \quad \sigma^2_{\xi,II} \times 10^4 = 30.8, \quad \sigma^2_{\beta,II} \times 10^4 = 8.00. \end{aligned}$$

We conclude that the traditionally better performing teams have more persistent strengths of attack and defence processes. Although the values $\phi_{\xi,I}$ and $\phi_{\beta,I}$ are estimated close to one, which points towards random walks, we maintain the autoregressive processes since $0.9986^{34} = 0.95$ still indicates to stationary dynamics if we look at the persistence year-by-year (34 weeks).

Correlated strengths of attack and defence

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

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

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

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

The likelihood ratio test for one extra coefficient is 5.96, which is statistically significant at the 5% level. The estimated value of $\rho = 0.97$ is close to one which indicates that the dynamic development of the strengths of attack and defence are almost perfectly co-dependent. A possible explanation for this might be that clubs when building their teams make progressive steps in acquisitions both on the attack and defence side: having a team with a great scoring potential but a poor defence is much less effective for securing a high final position in the competition. The reverse is also true, explaining why the two processes will probably be heavily correlated.

Away ground disadvantage

Apart from the fixed effect δ for home ground advantage in the scoring intensity of the home team, we can also introduce a fixed effect δ_a for the disadvantage of scoring by the away team: $\lambda_{2,ij,t} = \exp(\delta_a + \xi_{jt} - \beta_{it})$. We find $\delta_a = 0.066$ with a standard error of 0.054, such that this effect is not statistically significant in our study.

Zero inflated model

In order to capture a (possible) excess of draws, $y_{ij,t} = 0$, we consider a zero-inflated version of the model using equation (4.3) where an extra parameter γ is added that accounts for the possible transfer of probability mass towards zero. The log likelihood for this model is not significantly higher than the benchmark model and the estimate for γ is not statistically significant. We conclude that an excess of draws is not present in the Bundesliga data set as modelled by the dynamic Skellam model.

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, i.e.

$$\lambda_{1,ij,t} = \exp(\delta + \delta_x x_i + \xi_{it} - \beta_{jt}),$$

where x_i is the stadium capacity of team i (measured in multiples of 10,000) and δ_x is the regression coefficient that is placed in the parameter vector ψ. This model specification does not lead to a significant improvement of the log likelihood value and the estimate for $\delta_x = 0.00$.

Table 4.2: Model comparison. We present the fit improvements of different model specifications discussed in Section 4.3.2. 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 dimension of ψ denoted by k, the log likelihood value (ℓ), the likelihood ratio (LR), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are reported for each extension. The AIC is calculated as $\frac{1}{n^*}(2k - 2\ell)$ and the BIC is calculated as $\frac{1}{n^*}(\log(n^*)k - 2\ell)$where $n^* = 2142$ are the number of matches in the data set. 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		k	ℓ	LR test	H_0	t-test	AIC	BIC
Basic model		4	-8137.51				7.6018	7.6124
Heterogeneous panel	✓	8	-8130.14	14.74^{**}			7.5986	7.6198
Correlated attack and defence	✓	9	-8127.16	5.96^{*}			7.5968	7.6206
Away ground disadvantage		10	-8126.59	1.14	$\delta_a = 0$	1.22	7.5972	7.6237
Zero inflated model		10	-8126.73	0.86	$\gamma = 0$	0.92	7.5973	7.6238
Home stadium capacity		10	-8127.16	0.00	$\delta_x = 0$	0.00	7.5977	7.6242

Synthesis

Table 4.2 reviews our empirical findings for the Bundesliga. For each model extension the estimated log likelihood function is maximised and the Akaike information criterion (AIC) and the Bayesian Information Criterion (BIC) are computed to facilitate model comparison. Based on AIC, we can conclude that team heterogeneity in the dynamics, and correlation between the strengths of attack and defence are two important extensions for football match outcomes in our Bundesliga data set. These findings are backed up by likelihood ratio tests. However, BIC, which penalizes the addition of parameters more than AIC does, favours the basic model.

4.3.3 Signal extraction

We present the estimated strengths of attack and defence in Figure 4.2 using the bivariate NAIS methodology as described in Section 3.2.3. The smooth and persistent processes for the strengths of attack and defence 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 strengths of attack and defence for each team is also clearly visible. The teams of Dortmund and Bayern Munich have increased their strength of attack consistently from season to season which is also true for their strength of defence. On the other hand, the strength of attack of Werder Bremen has deteriorated over time, which only partly applies to its strength of defence.

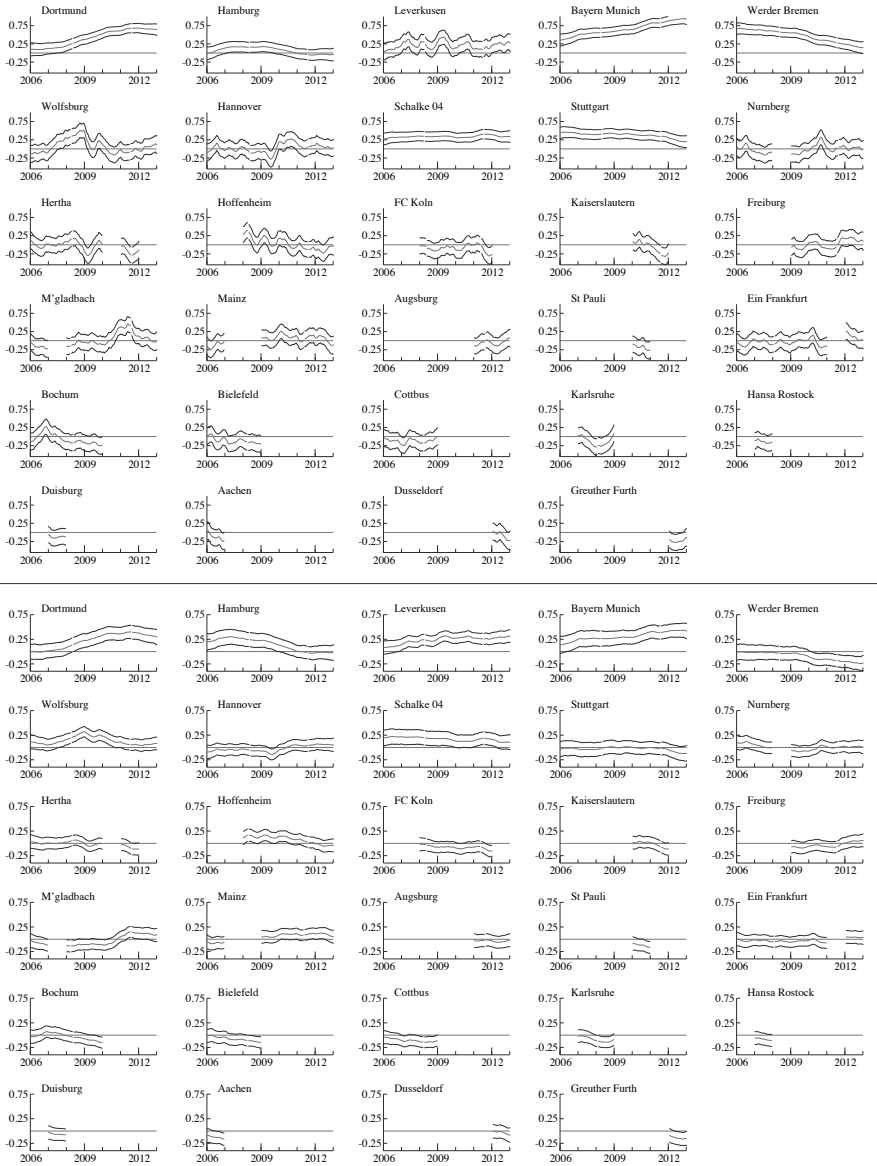

Figure 4.2: Strengths of attack and defence of teams in the Bundesliga. The two panels of graphs present respectively the extracted strengths of attack and defence for all teams in the Bundesliga from the season $2006-2007$ towards $2012-2013$, together with confidence intervals based on one standard error. The more persistent strengths of attack and defence 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.

4.4 Conclusions

In this chapter we introduced a general dynamic model for Skellam distributed difference in 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. (2014). In contrast to the higher dimensional generalization of NAIS based on Halton sequences in Scharth (2012, Chapter 5), we were still able to use the bivariate Gauss-Hermite numerical integration techniques to compute the appropriate integrals. In Table 4.1 we presented maximised log likelihood values and parameter estimates for various values of the number of importance draws S and number of Gauss-Hermite quadrature points M and concluded that reasonable choices for S and M did not have a major impact on the results which showed the robustness of the methodology.

Based on an illustration, we demonstrated the versatility of our dynamic Skellam analysis. In our 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. Results were summarized in Table 4.2 from which we concluded that heterogeneity in the dynamics, and correlation between the strengths of attack and defence significantly improved model fit based on the Akaike information criterion and likelihood ratio tests. We conclude that the new dynamic Skellam model is robust and computationally feasible for large unbalanced panels. The model may even rely on high dimensional state vectors. Our flexible modelling framework for time series may provide a useful benchmark for empirical applications based on integer outcomes that can take both positive and negative values.

Further research should be focused in the direction of including additional explanatory variables, see also the conclusion and further research recommendations of Chapter 2. It is not yet clear if modelling the home and away scores as two observations (Chapter 2) is better than modelling the difference between scores (Chapter 4). One could argue that with the modelling of the difference between scores, information is lost. However, if one is interested in the probabilities of a home win, draw or away win, it can be beneficial to model the differences. The reasoning behind this is the accumulation of modelling error. In Chapter 2, the probability of a home win is the sum of all the individual probabilities

for which $x > y$ where x and y are the number of goals scored by the home and away team respectively. By modelling the difference between x and y, the probability of a home win, draw or away win consists of less individual probability components. It should be further investigated which method yield better results.

Chapter 5

Dynamic Discrete Copula Models for High Frequency Stock Price Changes

5.1 Introduction

A key empirical finding from many analyses of intraday tick data is that stock price volatility is higher during opening hours than during the rest of the day; see, for example, Andersen and Bollerslev (1997) and Tsay (2005). Much less is known about the intraday pattern of the dependence structure between stock price changes. These dependence structures are of direct importance for intraday risk management, for example, when managing a book of multiple stocks that are traded repeatedly over the course of the day. In this chapter we investigate the pattern of intraday dependence dynamics (beyond correlation structures) for a number of U.S. financial stocks observed at the tick-by-tick frequency. Earlier studies typically analysed lower-frequency data (5 minutes) using standard correlation models; see, for example, Allez and Bouchaud (2011). We account for the discreteness of tick-by-tick stock price changes in our analysis by adopting a flexible dynamic copula framework for the modelling of the dependence structure. Bibinger, Hautsch, Malec, and Reiss (2014) developed a realised intraday covariance measure which also relies on high frequency data, but does not rely on copula functions and does not account for the discrete nature of the tick data.

There are at least two possible reasons to expect intraday time-variation in the dependence structure of stock price changes. First, news may accumulate overnight. As many of the firm-specific announcements are scheduled after trading hours whereas most common macro announcement are scheduled during normal trading hours, a relatively higher percentage of idiosyncratic, firm-specific news is impounded in stock prices during the first minutes after the opening. Such increased information flows are known to affect intraday volatilities upwards immediately after the opening of the exchange; see

for example Wood, McInish, and Ord (1985) and Admati and Pfleiderer (1988). Given the relatively higher fraction of idiosyncratic information directly after the opening, price changes are likely to exhibit lower dependence during the first minutes after the opening compared to during the rest of the day. Second, idiosyncratic components may also play an important role towards the end of the day. This may result in lower levels of dependence during the closing minutes of the trading day. In particular, we expect many players to unwind inventory positions that are built up over the course of the day due to idiosyncratic liquidity shocks, in order to limit the (overnight) risk. The positions at the end of the day are therefore likely to contain relatively more idiosyncratic components. Hence the expected dependence between price changes at the end of the trading day is lower.

We study intraday dynamics in price changes using tick-by-tick data observed at the one-second frequency over the year of 2012 for four financial stocks that are heavily traded. As the tick-size for our stocks is 1 dollar cent, prices as well as price changes move on a discrete grid. It is well-established that intraday price changes are subject to time-varying volatility and hence a time-varying marginal distribution. Many econometric challenges arise in the modelling of the dependence structure between discrete variables in case both the marginal distributions and the dependence structure are allowed to vary over time. The main methodological contribution of this chapter is that we provide a novel framework to address these issues in a way that is congruent with the empirical data and parsimonious. In particular, the dynamic parameters in our model, including stock return volatilities and dependence parameters, are updated using an observation-driven, autoregressive updating function based on the score of the conditional observation probability mass function; for an introduction to the score-driven approach, see Creal, Koopman, and Lucas (2011); Creal et al. (2013) and Harvey (2013), and for successful applications see, for example, Lucas, Schwaab, and Zhang (2014), Harvey and Luati (2014), Creal, Schwaab, Koopman, and Lucas (2014), and De Lira Salvatierra and Patton (2015). The score-driven model has several favorable features: (i) the 'filtered' estimates of the time-varying parameter are optimal in a Kullback-Leibler sense, see Blasques et al. (2015); (ii) it is an inherently observation-driven model rather than a parameter-driven model in the classification of Cox (1981), such that its likelihood is known in closed-form; and (iii) its forecasting performance is at least comparable to parameter-driven counterparts, even when the latter constitute the true data generating process, see Koopman et al. (2015). The second point emphasizes that static parameters can be estimated in a straightforward way using maximum likelihood methods.

We adopt a dynamic Skellam distribution to model the tick-size price changes on the grid $\ldots, -2, -1, 0, 1, 2, \ldots$; see Irwin (1937) and Skellam (1946). The Skellam distribution has also been used to model price change series in other recent contributions; see

Barndorff-Nielsen et al. (2012) for Skellam Lévy processes, and Shahtahmassebi (2011) for a Bayesian analysis of a parameter-driven Skellam model. Rather than only having a dynamic generalization of the Skellam distribution for the marginal models, our main focus is on formulating a time-varying specification for the dependence structure in discrete data based on a copula framework. In this regard, we amend the time-varying copula models of Creal et al. (2013) and De Lira Salvatierra and Patton (2015) for continuous data towards discrete data.

Discrete copulas and, in particular, dynamic discrete copulas pose a number of challenges. First, copulas for discrete marginals are not unique over the entire domain of the unit hypercube. Second, the copula density is no longer well-defined for discrete marginals, but is replaced by a copula probability mass function. Third, given the time-varying nature of the marginal distributions, the grid that defines the copula uniquely changes from one time period to the next. We address these issues using a parametric copula specification that parsimoniously describes the copula surface. This function should cover grid points over which the copula at the current time point is uniquely defined but also at grid points that may become relevant at future time points given the time-varying nature of the marginal distributions. We further allow for time-variation in the dependence structure by endowing the copula parameters with autoregressive dynamics that are a function of the score of the copula probability mass function. In a Monte Carlo study, we show that our dynamic copula approach works well in uncovering the true parameter dynamics if the model is correctly specified; we can extract the path of the dynamic parameters with high precision. Moreover, when the model is not correctly specified, we show that our approach still accurately extracts the correct parameter path as well.

In our empirical study, we investigate the dependence in tick-by-tick price changes for a selection of four U.S. financial stocks which are traded on the NYSE. We present key evidence that significant intraday time-variation in the dependence structure of these four stocks is present. The intraday dependence in all trading days of 2012 increases during the first 30 minutes after the opening. The average intraday dependence remains relatively constant after the first 30 minutes until, say, 15 minutes before the close when a sharp decrease in the dependence takes place for the six stock pairs considered in our analysis. An alternative approach is to specify the intraday pattern of the dependence as a fixed intraday seasonality pattern based on a flexible spline function. However, we show that in almost all cases our score-driven time-varying copula methodology significantly outperforms the alternative spline-based approach. This indicates that time-variation in the intraday dependence is captured accurately by the score-driven model and varies substantially between days. Furthermore, it suggests that substantial day-to-day deviations from the average intraday pattern occur regularly.

The remainder of this chapter is organised as follows. We introduce the model in

Section 5.2. Section 5.3 presents simulation results on the model's adequacy. Our empirical analysis is presented in Section 5.4, while Section 5.5 contains the conclusions. The Appendix gathers a number of the more technical background expressions for the score-dynamics of the different marginals and copulas used in this chapter.

5.2 Score-driven dynamic discrete copula model

Consider a d-dimensional integer-valued vector $y_t = (y_{1,t}, \ldots, y_{d,t})' \in \mathbb{Z}^d$ with time-varying conditional marginal distributions $F_i(y_{i,t} \mid \mathcal{F}_{t-1}; \theta^m_{i,t})$ for $i = 1, \ldots, d$ and $t = 1, \ldots, T$, where $\theta^m_{i,t}$ is a time-varying parameter vector for the ith marginal distribution, and $\mathcal{F}_t = \{y_t, y_{t-1}, \ldots\}$. The elements of y_t may for instance consist of counts, such as Poisson or binomial counts, or alternatively of changes in counts, such as the Skellam distributed discrete (tick-size) price changes in our empirical application in Section 5.4. The mean and variance of the Skellam distribution for stock i are then part of $\theta^m_{i,t}$. We characterize the dependence structure by a parametric conditional d-dimensional copula function

$$C\left[F_1(y_{1,t} \mid \mathcal{F}_{t-1}; \theta^m_{1,t}), \ldots, F_d(y_{d,t} \mid \mathcal{F}_{t-1}; \theta^m_{d,t}) \mid \mathcal{F}_{t-1}; \theta^c_t\right], \tag{5.1}$$

where θ^c_t is the parameter vector defining the copula function C; see Sklar (1959). The time-varying nature of θ^c_t allows us to study settings where the dependence structure changes over time. For example, in Section 5.4 we study the intraday dependence between discrete stock price changes. For notational simplicity, we suppress the dependence on the conditioning set $\mathcal{F}_{t-1}$ and write the marginal distributions as

$$F_i := F_i(y_{i,t}; \theta^m_{i,t}) \equiv F_i(y_{i,t} \mid \mathcal{F}_{t-1}; \theta^m_{i,t}), \qquad i = 1, \ldots, d,$$

and the copula function as $C\left[F_1(y_{1,t}; \theta^m_{1,t}), \ldots, F_d(y_{d,t}; \theta^m_{d,t}); \theta^c_t\right]$. The dynamic specifications of the parameter vectors $\theta^m_{i,t}$ and θ^c_t are provided below. The dynamic conditional copula formulation presented in equation (5.1) is obtained from Patton (2002, 2006).

A discrete data analysis based on dynamic copulas faces several challenges; see also the review of Genest and Nešlehová (2007) on the use of static copulas for discrete marginals. For example, standard summary dependence measures such as Kendall's τ or Spearman's ρ are no longer guaranteed to lie in the $[-1, 1]$ interval and need to be used with caution in a discrete setting. In addition, we can no longer guarantee the uniqueness of the copula function in the standard Sklar (1959) representation of a distribution in terms of its marginal distributions and a copula function. The copula is only uniquely determined on the set $\operatorname{Ran} F_1 \times \ldots \times \operatorname{Ran} F_d$, where $\operatorname{Ran} F_i$ denotes the range of the cumulative distribution function (cdf) F_i, $i = 1, \ldots, d$. This stands in sharp contrast with the case of continuous marginal distributions, where the copula function is unique over the entire

unit hypercube $[0,1]^d$.

Despite its non-uniqueness, discrete copulas can still be usefully applied in an empirical setting; see, for example, Zimmer and Trivedi (2006). At one extreme, we can model the value of the copula function at each point of its domain separately. This method can work in simple settings where the discrete data only takes a small number of different values; for example, in case of Bernoulli variables. This approach becomes infeasible, however, when the copula is defined over many different points as is the case in the empirical setting of Section 5.4. First, the price changes in our empirical example take values on $\mathbb{Z}$, and are therefore defined on (countably) infinitely many points. Second, and most importantly, the marginal distributions are time-varying. As a result, also the ranges $\operatorname{Ran} F_i(\cdot;\theta^m_{i,t})$ and therefore the domain over which C is uniquely identified are time-varying. Consequently, it is no longer feasible to estimate the value of the copula function over all points in the domain across all time periods, as there will be infinitely many of them. A possible solution is to model the copula in a parsimonious way. For example, we can use a parametric copula function defined over the entire $[0,1]^d$ space even though uniqueness is only guaranteed over a set of discrete points. This is the approach we will adopt in our analyses below.

The dynamic specifications for $\theta^m_{i,t}$ and θ^c_t in (5.1) are based on the score-driven approach of Creal et al. (2011, 2013) and Harvey (2013). We collect the time-varying parameters in $\theta'_t = (\theta^{m\prime}_{1,t}, \ldots, \theta^{m\prime}_{d,t}, \theta^{c\prime}_t)$. The score-driven model represents a class of models in which the update of θ_t over time is formulated as a function of past data $y_{t-1}, y_{t-2}, \ldots$ and past realised parameter values $\theta_{t-1}, \theta_{t-2}, \ldots$. At time t we can write the update function as

$$\theta_{t+1} = \theta_{t+1}\left(y_t, y_{t-1}, \ldots, \theta_t, \theta_{t-1}, \ldots; \psi\right),$$

where ψ is an unknown parameter vector that contains the update coefficients and the remaining static parameters of the marginal distributions and the copula function. It follows that θ_t is $\mathcal{F}_{t-1}$-measurable and the approach is observation-driven in the classification of Cox (1981). The estimation of the static parameter vector ψ is typically carried out by the method of maximum likelihood in a straightforward manner. A score-driven model updates θ_t in the direction of the steepest increase of the log conditional probability mass function (pmf) at time t given the past information set $\mathcal{F}_{t-1}$. Updating θ_t in this way possesses information theoretic optimality properties as shown by Blasques et al. (2015).

Let $p(y_t|\mathcal{F}_{t-1};\theta_t)$ denote the pmf of y_t, which we again write in short-hand notation as $p(y_t;\theta_t)$, suppressing its dependence on the parameter vector ψ. Using the so called 'inclusion-exclusion' formula, we obtain from equation (5.1) that

$$p(y_t;\theta_t) = \sum_{\phi_1=0,1} \cdots \sum_{\phi_d=0,1} (-1)^{\phi_1+\ldots+\phi_d} \times C\left[F_1(y_{1t}-\phi_1;\theta^m_{1,t}), \ldots, F_d(y_{dt}-\phi_d;\theta^m_{d,t}); \theta^c_t\right]. \tag{5.2}$$

For instance, for the bivariate case ($d = 2$), the pmf becomes

$$p(y_t;\theta_t) = C\left[F_1(y_{1,t};\theta^m_{1,t}), F_2(y_{2,t};\theta^m_{2,t});\theta^c_t\right] - C\left[F_1(y_{1,t}-1;\theta^m_{1,t}), F_2(y_{2,t};\theta^m_{2,t});\theta^c_t\right] - C\left[F_1(y_{1,t};\theta^m_{1,t}), F_2(y_{2,t}-1;\theta^m_{2,t});\theta^c_t\right] + C\left[F_1(y_{1,t}-1;\theta^m_{1,t}), F_2(y_{2,t}-1;\theta^m_{2,t});\theta^c_t\right], \tag{5.3}$$

where the evaluation of equation (5.2) requires 2^d evaluations of the copula function, for any t, and is feasible for low values of d as in (5.3). The evaluation of (5.2) clearly becomes more challenging for larger values of d; see, for example, Panagiotelis, Czado, and Joe (2012). The score-based update function for θ_t takes the form

$$\theta_{t+1} = \omega + A\nabla_t + B\theta_t, \qquad \nabla_t = \frac{\partial \log p(y_t;\theta_t)}{\partial \theta_t}, \tag{5.4}$$

where ∇_t is the score vector of the (predictive) mass function $p(y_t;\theta_t)$ in (5.2), ω is a vector of constants, and A and B are fixed coefficient matrices. These coefficients are functions of the parameter vector ψ that also includes the unknown parameters of the marginal distributions F_i and the copula function C in (5.2). Since $p(y_t;\theta_t)$ relies on ψ, it follows that ∇_t is also a function of ψ. The derivative ∇_t in (5.4) is straightforward to obtain because the pmf is typically differentiable in the time-varying parameters θ_t. The updating equation (5.4) corresponds to the unit scaling option of Creal et al. (2013) and can be generalized in different ways; for example, by adding more lagged values of θ_t and ∇_t.

The time-varying parameter vector θ_t is initialized at θ_1, which we include in the parameter vector ψ. In the case of a bivariate copula, the individual components of θ_t consist of two marginal parameter vectors and one copula dependence parameter. To introduce further parsimony, we assume diagonal matrices for A and B, such that each element of θ_t is updated by its own score function only. The static parameter vector ψ becomes $\psi = \{\theta_1', \omega', \mathrm{diag}(A)', \mathrm{diag}(B)'\}$, where $\mathrm{diag}(M)$ denotes the vector of the diagonal elements of any matrix M. A more parsimonious specification is obtained by having θ_1 as to the unconditional mean of θ_t, i.e. $\theta_1 = \omega \oslash (1 - \mathrm{diag}(B))$ where $\oslash$ denotes the Hadamard division (pointwise division).

The score function for $\nabla_t' = (\nabla^{m\prime}_{1,t}, \ldots, \nabla^{m\prime}_{d,t}, \nabla^{c\prime}_t)$ has an analytical solution that is given by the elements $\nabla^m_{k,t}$, $k = 1, \ldots, d$, and ∇^c_t specified as

$$\nabla^m_{k,t} = \frac{\partial \log p(y_t;\theta_t)}{\partial \theta^m_{k,t}} = \frac{\sum_{\phi_1=0,1}\cdots\sum_{\phi_d=0,1}(-1)^{\phi_1+\ldots+\phi_d}\,\dfrac{\partial C(u_{1,t},\ldots,u_{d,t};\theta^c_t)}{\partial u_{k,t}} \cdot \dfrac{u_{k,t}}{\partial \theta^m_{k,t}}}{\sum_{\phi_1=0,1}\cdots\sum_{\phi_d=0,1}(-1)^{\phi_1+\ldots+\phi_d} C(u_{1,t},\ldots,u_{d,t};\theta^c_t)}, \tag{5.5}$$

$$\nabla_t^c = \frac{\partial \log p(y_t;\theta_t)}{\partial \theta_t^c} = \frac{\sum_{\phi_1=0,1}\cdots\sum_{\phi_d=0,1}(-1)^{\phi_1+\ldots s+\phi_d}\partial C(u_{1,t},\ldots,u_{d,t};\theta_t^c)/\partial\theta_t^c}{\sum_{\phi_1=0,1}\cdots\sum_{\phi_d=0,1}(-1)^{\phi_1+\ldots+\phi_d}C(u_{1,t},\ldots,u_{d,t};\theta_t^c)}, \quad (5.6)$$

with $u_{i,t} = F_i(y_{i,t} - \phi_i; \theta_{i,t}^m)$, for $\phi_i \in \{0,1\}$, $i = 1,\ldots,d$, and $t = 1,\ldots,T$. The denominators in (5.5) and (5.6) are equal to the pmf as given in (5.2). In case of the Gaussian copula as well as the commonly encountered copulas from the Archimedean class, analytical expressions for ∇_t are available. We refer to Appendix A for further details and to Schepsmeier and Stöber (2014) for expressions for a range of derivatives of bivariate copulas.

Given that θ_t is $\mathcal{F}_{t-1}$-measurable and our model specification is observation-driven in the classification of Cox (1981), we obtain the likelihood function in closed form by a standard prediction error decomposition,

$$L(y;\psi) = \sum_{t=1}^{T} \log\, p(y_t;\theta_t), \quad (5.7)$$

with $y = (y_1,\ldots,y_T)$. We define the corresponding maximum likelihood estimator (MLE) of ψ as $\hat{\psi} = \arg\max_\psi L(y;\psi)$. In practice we obtain the MLE of ψ via the direct numerical maximisation of $L(y;\psi)$ with respect to ψ.

Example: Frank copula with Skellam marginals

As a concrete example, consider the bivariate Frank copula with Skellam marginals. This combination of copula and marginals is used to perform the simulation study in Section 5.3. The Frank copula is a symmetric copula given by

$$C_{\mathrm{Fr}}(u_{1,t},u_{2,t};\theta_t^c) = \frac{1}{\theta_t^c}\log\left[1 + \frac{(\exp(-\theta_t^c u_{1,t})-1)(\exp(-\theta_t^c u_{2,t})-1)}{\exp(-\theta_t^c)-1}\right], \quad (5.8)$$

with $\theta_t^c \in \mathbb{R}\backslash\{0\}$; see Frank (1979) and Nelsen (2006). When $\theta_t^c \to 0$, the Frank copula converges to the independence copula $C_{\mathrm{Fr}}(u_{1,t},u_{2,t};0) = u_{1,t}u_{2,t}$.

A Skellam pmf with location parameter μ_t and scale parameter σ_t^2 is given by

$$\Pr(Y_t = y_t;\mu_t,\sigma_t^2) = \exp\left(-\sigma_t^2\right)\left(\frac{\mu_t+\sigma_t^2}{\sigma_t^2-\mu_t}\right)^{y_t/2} I_{|y_t|}\left(\sqrt{\sigma_t^4-\mu_t^2}\right), \quad (5.9)$$

with $y_t \in \mathbb{Z}$ and where $I_{|y_t|}(\cdot)$ is the modified Bessel function of order $|y_t|$. The shape of the Skellam distribution depends on μ_t and σ_t^2 and is symmetric for $\mu_t = 0$, skewed right when $\mu_t > 0$, and left-skewed for $\mu_t < 0$. The excess kurtosis of the Skellam pmf is $1/\sigma_t^2$, and it has the Gaussian distribution as a limiting case. The Skellam distribution was originally derived by Irwin (1937) and Skellam (1946) as a distribution for the difference between two Poisson variables. Our parameterization in equation (5.9) is a reparameterization

of the original version and can be transformed back by substituting $\mu_t = \lambda_{1,t} - \lambda_{2,t}$ and $\sigma_t^2 = \lambda_{1,t} + \lambda_{2,t}$ in (5.9), where $\lambda_{1,t}$ and $\lambda_{2,t}$ are the means of the underlying Poisson distributions; see also Alzaid and Omair (2010). The mean μ_t and variance σ_t^2 in the full model in equation (5.9) are time-varying. In our application of Section 5.4, however, the mean turns out to be insignificantly different from zero and not time-varying, whereas the variance remains time-varying. In this case, equation (5.9) simplifies to

$$\Pr(Y_t = y_t; \sigma_t^2) = \exp\left(-\sigma_t^2\right) I_{|y_t|}(\sigma_t^2). \tag{5.10}$$

We then obtain a Frank copula function $C(u_{1t}, u_{2t}; \theta_t^c)$ with $\theta_{i,t}^m = \sigma_{i,t}^2$ and

$$u_{i,t} = \Pr(Y_{i,t} \leq k; \sigma_{i,t}^2) = \exp\left(-\sigma_{i,t}^2\right) \sum_{\phi=-\infty}^{k} I_{|\phi|}(\sigma_{i,t}^2), \qquad i = 1, 2. \tag{5.11}$$

5.3 Simulation study

To investigate the properties of our model in a controlled setting, we carry out two simulation studies. In our first study, we assume that the score-driven model of equations (5.2) and (5.4) is the true data generating process and verify the finite sample behaviour of the maximum likelihood estimates for the parameter vector ψ. In the second study, we consider a misspecified model setting. We assume that the marginal parameters and the dependence parameter come from some exogenous dynamic patterns that do not rely on the score function. We then verify to what extent the score-driven framework is able to recover the true underlying dynamics of the time-varying parameter vector θ_t. In both simulation studies, we focus on a positive dependence between two series, i.e. $\theta_t^c \in \mathbb{R}^+$. We specify $\bar{\theta}_t^c = \log(\theta_t^c)$ as the time-varying parameter rather than θ_t^c itself. We adopt the same specification for the variance of the Skellam distribution, that is $\bar{\theta}_{i,t}^m = \log(\theta_{i,t}^m) = \log(\sigma_{i,t}^2)$ and $\bar{\theta}_{i,t}^m$ varies over time. The score function ∇_t in (5.5) and (5.6) adapt to this reparameterization into $\bar{\theta}_t$ by pre-multiplying ∇_t by $\partial\theta_t'/\partial\bar{\theta}_t$. This reparameterization yields an estimation procedure that is numerically more stable. In both simulation studies, the observation series are simulated from a bivariate Frank copula with Skellam marginals as discussed in Section 5.2.

5.3.1 Estimating parameters when model is correctly specified

We simulate $S = 500$ series of correlated Skellam observations. The length of the sample is set to $T \in \{250, 1000, 3000\}$. To generate the data, we apply the algorithm of Nelsen (2006, p.41) using a numerical inverse cdf of the Skellam distribution. For the log-transforms of the dynamic parameters θ_t, we consider equation (5.4). The estimates of

Table 5.1: Simulation results under correct model specification. This table reports simulation averages of maximum likelihood estimates of the static parameters for the Skellam-Frank score-driven model of Section 5.2. The results use $S = 500$ replications of time series of length $T \in \{250, 1000, 3000\}$. The intercepts ω in (5.4) are set to $(\mathrm{I} - B)\bar{\theta}_1 = \omega$, such that $\bar{\theta}_1$ is the unconditional mean of $\bar{\theta}_t$, where $\bar{\theta}_t$ contains the logs of the elements of θ_t. The matrices A and B are diagonal with elements (a_1, a_2, a_3) and (b_1, b_2, b_3), respectively. Standard deviations of the estimates over the Monte Carlo simulations are in parentheses. The column $t(s)$ denotes the average computation time (in seconds) for finding the maximum of the log likelihood function. Computations are carried out on a i7-2600, 3.40 GHz desktop PC using four cores.

T	$\bar{\theta}_{1,1}$	$\bar{\theta}_{2,1}$	$\bar{\theta}_{3,1}$	a_1	a_2	a_3	b_1	b_2	b_3	$t(s)$
true	1.00	1.00	2.00	0.10	0.05	0.10	0.90	0.95	0.98	-
250	1.00	1.01	2.03	0.11	0.06	0.11	0.74	0.77	0.87	17.76
	(0.14)	(0.14)	(0.28)	(0.08)	(0.06)	(0.08)	(0.28)	(0.27)	(0.21)	-
1000	1.00	1.00	2.00	0.10	0.05	0.10	0.87	0.91	0.97	59.98
	(0.07)	(0.07)	(0.15)	(0.03)	(0.03)	(0.03)	(0.09)	(0.09)	(0.03)	-
3000	1.00	1.00	2.00	0.10	0.05	0.10	0.89	0.94	0.98	108.30
	(0.04)	(0.04)	(0.06)	(0.02)	(0.02)	(0.02)	(0.04)	(0.08)	(0.01)	-

the parameter vector ψ are obtained via the numerical maximisation of the loglikelihood function (5.7) using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.

Table 5.1 presents the results. The method of maximum likelihood is able to estimate the parameters in ψ accurately, even for the small sample size $T = 250$. For $T \in \{1000, 3000\}$, the maximum likelihood estimates for the unconditional mean $\bar{\theta}_1$ and the score loadings $(a_1, a_2, a_3)'$ in the updating equations are virtually equal to the corresponding true parameters. In the case of $T = 250$, the persistence parameters b_1, b_2, b_3 are underestimated, which matches small sample biases encountered in similar studies for standard linear time series models. The biases disappear for larger sample sizes. In the case of $T = 3000$, the b_1, b_2, b_3 parameters are estimated close to their true values. Finally, we can conclude from the average computing times $t(s)$ for estimation, also reported in Table 5.1, that the score-driven methodology applied to the bivariate Frank copula is quite fast. The computing times for parameter estimation ranges from less than 18 seconds, on average, for $T = 250$, to approximately 108 seconds, for $T = 3000$. The computations are carried out by an i7-2600, 3.40 GHz desktop PC using four cores.

5.3.2 Estimating time-varying paths when model is misspecified

Next we deviate from the assumption that the score-driven model (5.2) and (5.4) is the data generation process. In our second Monte Carlo study, the time-varying Skellam variances and the time-varying dependence parameter are generated as sinusoidal patterns with different periods and amplitudes. We investigate to what extent our misspecified

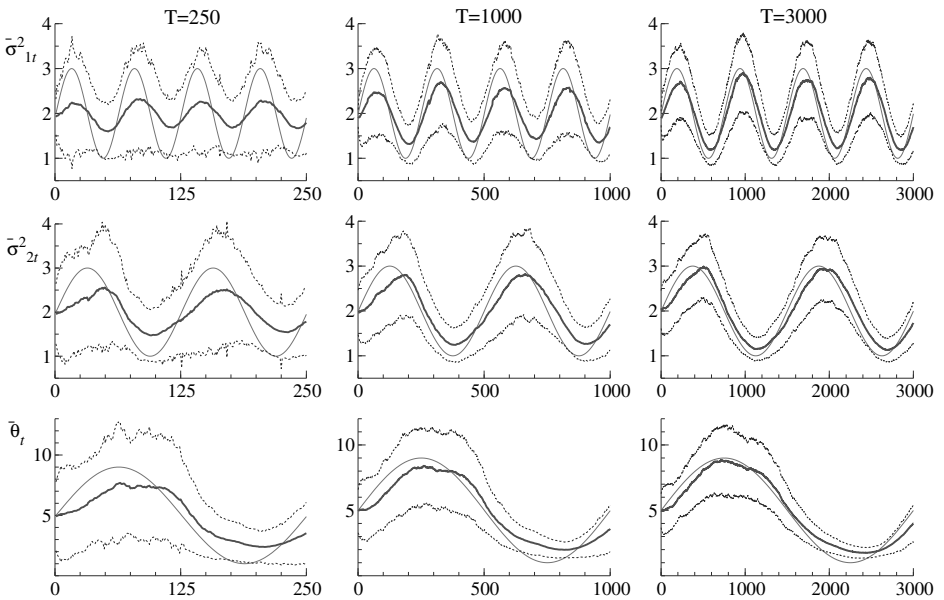

Figure 5.1: Simulation results under misspecification. The figure presents the point wise Monte Carlo averages (solid fat) over 500 replications of the Skellam variances $\sigma^2_{1,t}$ and $\sigma^2_{2,t}$, and of the Frank copula parameter θ^c_t. All three parameters are parametrized in log form in the score-driven specification. Each panel also contains the true time-varying parameter (solid thin) and a band of two times the point wise standard deviations (dotted). From left to right the panels show time series length of $T = 250, 1000, 3000$, respectively.

score-driven framework is able to identify these time-varying patterns. We generate $S = 500$ time series of length $T \in \{250, 1000, 3000\}$ and estimate the parameters in vector ψ by the method of maximum likelihood.

In Figure 5.1, we present the true time-varying parameters σ^2_{1t}, σ^2_{2t} and θ^c_t together with their estimated counterparts $\hat{\sigma}^2_{1t}$, $\hat{\sigma}^2_{2t}$ and $\hat{\theta}^c_t$, respectively. The results for $T = 1000$ and $T = 3000$ show that the score-driven model is able to capture the true paths of the time-varying parameters accurately, despite its misspecification. Only in the case of the small sample size $T = 250$ and the rapidly changing parameter paths for $\sigma^2_{1,t}$, the estimates are less accurate. The fact that the filtered paths lag behind the true path is typical for misspecified observation-driven models, see for example Creal, Koopman, Lucas, and Zamojski (2015). In our empirical study in Section 5.4, we have more than 40,000 observations per month. Hence we expect the score-driven model to perform sufficiently accurate in our empirical study. Admittedly, we only explore one type of misspecification, however, several types of misspecification of score-driven models were studied in Creal et al. (2015) with the same convincing results.

5.4 Dependence between discrete price changes

The dependence measures between price changes of individual stocks or assets are the key ingredients in, for instance, portfolio risk management. In our empirical study, we establish the intraday dependence structure in high-frequency price changes. Whereas most studies concentrate on the intraday dynamics of volatility, our study is, to the best of our knowledge, the first to concentrate on the intraday dynamics of the dependence structure using a copula approach in a tick-by-tick data analysis.

The data sets consist of price changes of stocks traded at the New York Stock Exchange (NYSE). The resulting series consist of discrete, integer multiples of the tick-size of one dollar cent. The observations take values in $\mathbb{Z}$. We model the discrete tick-size price changes instead of the returns. Münnix et al. (2010) argue that the discrete nature of the price grid affects the empirical distribution of returns severely. This distribution concentrates around the actual tick-sizes, is severely multi-modal and, consequently, highly non-Gaussian.

Several models for data in $\mathbb{Z}$ are available in the literature. For example, the model of Rydberg and Shephard (2003) decomposes stock price movements into activity, direction of moves, and size of the moves. Freeland (2010), Alzaid and Omair (2014) and Andersson and Karlis (2014) extend the integer autoregressive (INAR) model for $\mathbb{N}$ variables to the case of $\mathbb{Z}$ variables. They propose the Skellam distribution and use static Skellam parameters. Barndorff-Nielsen et al. (2012) analyse Skellam Lévy processes for intraday price changes. Shahtahmassebi (2011) present a Bayesian analyses based on a Skellam model for $\mathbb{Z}$ variables. The dynamic Skellam model for time series observations in $\mathbb{Z}$ is developed by Koopman et al. (2014) based a non-Gaussian state space analysis. In our current framework, we adopt the Skellam distribution for the marginals and allow the corresponding parameters to vary over time using the score-driven model of Section 5.2.

Although the Skellam distribution is an important ingredient of our analysis, our main focus is on the dependence structure as this feature has received much less attention in other related studies so far. Our analysis proceeds in two steps. First, we study the dependence characteristics between price changes of four major NYSE listed financials over a period of one trading month for a variety of copulas. We consider Bank of America Corporation (BAC), Citigroup Inc. (C), JPMorgan Chase & Co. (JPM) and Wells Fargo & Company (WFC). Based on our initial findings for these four stocks, we select the best copula for the second part of our analysis: an analysis of the intraday dependence dynamics over the long time span of an entire calendar year.

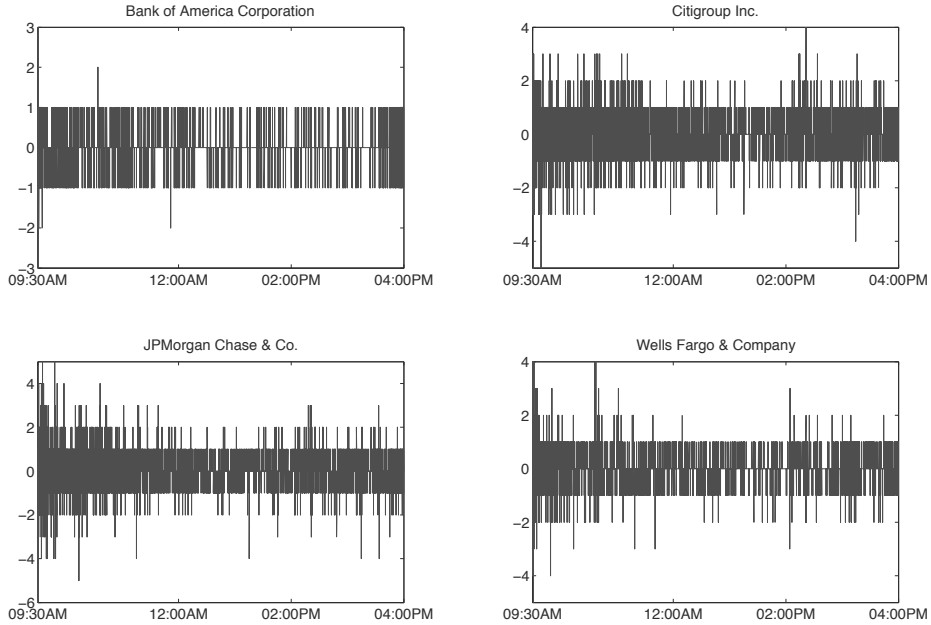

Figure 5.2: Tick price changes of Bank of America Corporation (BAC), Citigroup Inc. (C), JPMorgan Chase & Co. (JPM) and Wells Fargo & Company (WFC) on April 23, 2012.

5.4.1 Data description

We first analyse intraday stock prices obtained from the TAQ database for April 2012. We clean the high-frequency data by following the standard procedures described in Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2008) for TAQ data. This database has a time stamp precision of 1 second so that for many seconds we obtain a number of transactions with the same time stamp. It is common practice to merge these transactions and to replace them by the median price rounded to the nearest tick.

Figure 5.2 presents the intraday tick price changes for our four selected stocks. We present the results for a typical trading day, April 23, 2012. We find that more trades with relatively large price changes occur at the beginning of the day and a quiet period with small or no price changes takes place during lunch-time. Appendix B contains additional descriptive plots of the data.

Table 5.2 presents descriptive statistics of the tick-size price changes. We find that Citigroup and JPMorgan are the most liquid stocks in terms of the number of trades, followed by Wells Fargo and Bank of America. The absolute price level has a clear impact on the tick-size volatility: the minimum and maximum tick-size changes as well as the tick-size variance are substantially lower for Bank of America than for the other three

Table 5.2: The table reports company name, ticker symbol (Code), the number of trades (#Trades), the opening price at 9:30 AM of the first trading day in the sample (P_{open}), the closing price at 16:00 PM of the last trading day in the sample (P_{close}), the largest up-tick ($\uparrow$) measured in multiples of the tick-size, the largest down-tick ($\downarrow$), the variance (Var) and mean (Mean) of the tick-size price changes, and the percentage of 0-trades (%0).

Company	Code	#Trades	P_{open}	P_{close}	$\uparrow$	$\downarrow$	Var	Mean	%0
Apr 2012									
Bank of America Corp.	BAC	41,640	9.53	8.09	7	-6	0.242	-0.004	76.84
Citigroup Inc.	C	93,872	36.34	33.03	8	-11	0.753	-0.004	55.93
JPMorgan Chase & Co.	JPM	90,936	45.79	42.95	8	-8	0.747	-0.001	54.12
Wells Fargo & Company	WFC	64,529	33.85	33.40	8	-9	0.575	0.000	60.77
Jan 2012 - Dec 2012									
Bank of America Corp.	BAC	560,102	5.76	11.62	7	-6	0.232	0.001	77.30
Citigroup Inc.	C	1,084,943	27.20	39.59	11	-15	0.663	0.001	57.87
JPMorgan Chase & Co.	JPM	1,029,844	34.10	44.00	20	-16	0.725	0.001	55.30
Wells Fargo & Company	WFC	766,712	28.00	34.22	13	-14	0.510	0.001	63.30

institutions. We account for this effect by using different parameters in the marginal models for each stock.

5.4.2 Missing values

Our observation-driven model is formulated for a time frequency in seconds. Since we do not observe a trade for every second during the trading day, we encounter many missing observations. We distinguish four situations that can occur at second t during a day.

Situation 1: At time t, stock 1 trades while stock 2 does not trade so that the price change for series 2 is missing at time t. The copula dependence parameter cannot be updated as we require two observations to update the parameter related to instantaneous dependence. Furthermore, the marginal variance $\sigma^2_{1,t}$ cannot be updated by taking derivatives from the copula mass function in (5.6) since both observations from series 1 and 2 are needed as input. In this case, variance $\sigma^2_{1,t}$ is updated by only using the score of the marginal Skellam log pmf in (5.10). No score updating takes place for $\sigma^2_{2,t}$ and θ^c_t and hence these parameters mean revert by setting $\nabla^m_{2,t}$ and ∇^c_t to zero in (5.5) and (5.6). The contribution to the likelihood at time t is given by the logarithm of the pmf in (5.10) with $\sigma^2_{1,t}$ and $y_{1,t}$ as input.
Situation 2: At time t, stock 1 does not trade while stock 2 is traded. This is the converse of Situation 1 and has an analogous solution.
Situation 3: At time t, both stocks trade. The whole time-varying parameter vector θ_t is updated according to (5.4), where the score is obtained by taking derivatives from the

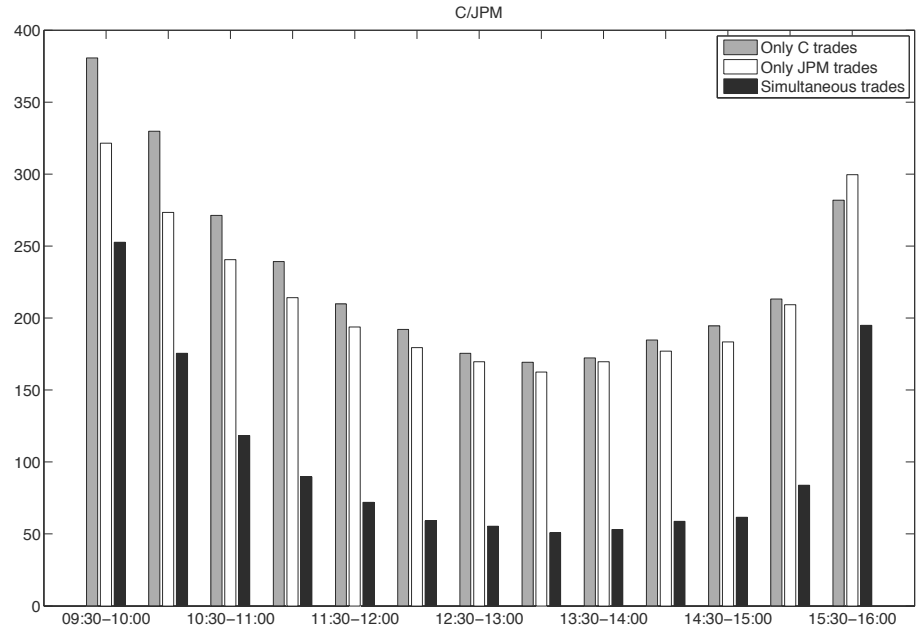

Figure 5.3: The figure displays the number of simultaneous trades per half hour of the trading day as well as the the number of trades if only CitiGroup or JPMorgan trade. The numbers are averaged over all 250 trading days of the year 2012.

copula mass function in equation (5.6). The contribution to the likelihood at time t is made by the logarithm of the copula mass function in (5.5).
Situation 4: At time t neither stock 1 nor stock 2 trades. In this case, none of the parameters is updated and there is no contribution to the likelihood.

For the purpose of estimating a dependence parameter, situation 3 has clearly the most impact. We therefore present in Figure 5.3 the number of simultaneous trades per half hour of the trading day. The numbers are averaged over all 250 trading days of the year 2012. Figure 5.3 reveals more joint trades at the beginning and the end of the day compared to the middle of the day. We may therefore expect more information in the data on the dependence parameter θ_t^c at the start and at the end of the day. Figure 5.7 in the Appendix reveals that the same increased trading intensity at the start and end of the trading day occurs for other stock combinations as well.

5.4.3 Copula selection

We take the independence copula as a benchmark and verify for a range of copulas whether they improve the model fit. The model fits are compared by means of the Bayesian In-

formation Criterion (BIC) for both static dependence θ^c and time-varying dependence θ_t^c. For all models considered, the marginal parameters of the Skellam distribution, σ_{1t}^2 and σ_{2t}^2 are allowed to vary over time. Our selection of copula functions includes the independence copula (Indep), the symmetric Ali-Mikhail-Haq (AMH), Frank, and Gaussian copulas, and the asymmetric Clayton (lower tail dependence), Gumbel (upper tail dependence), Joe (upper tail dependence), and Symmetrized Joe Clayton (SJC) copula (upper and lower tail dependence); see, for instance, Nelsen (2006) and Patton (2006) for the functional specifications of the these copulas.

For each day, the vector of time-varying parameters θ_t is initialized at θ_1 which is estimated as part of the vector of static parameters ψ. Table 5.3 presents the model selection results for all trading days in April 2012. Entries indicate the number of points by which the corresponding copula outperforms the BIC of the independence copula. Higher entries are thus preferred.

From Table 5.3 we learn that dynamic dependence is preferred over static dependence for five out of the six pairwise data sets based on the BIC. The symmetric copulas, Gaussian, AMH, and Frank, are generally preferred over the asymmetric ones. It confirms the somewhat symmetric patterns in the pairwise up and down tick movements encountered in the scatter plots of the data; see the Appendix for more evidence. The main conclusion of our first analysis is clear: both for static dependence as well as for dynamic dependence, the Gaussian copula fits the data best for all stock pairs. The Gaussian copula exhibits zero tail dependence. Given that copula functions with upper and/or lower tail dependence, such as Clayton, Gumbel, Joe, and Symmetrized Joe Clayton copulas, fit the data less well, we infer that tail dependence is not a dominant feature in tick-size price change series.

5.4.4 Full year results

In this section we extend our analysis over the entire year 2012. Descriptive statistics for this larger time span were given in Table 5.2. The characteristics of the data for all trading days in 2012 are broadly similar to those for the trading days in April 2012 only. Therefore, we use the Gaussian copula as our best fitting specification based on our preliminary analysis in Section 5.4.3. For the Gaussian copula correlation parameter ρ_t, we use the time-varying parameter θ_t^c, with

$$\rho_t = \theta_t^c \Big/ \sqrt{1 + (\theta_t^c)^2} \, . \tag{5.12}$$

This parameterization of ρ_t via θ_t^c ensures that the copula dependence parameter is always within the appropriate interval, i.e. $\rho_t \in (-1, 1)$. The likelihood for a full year of tick price changes is maximised in approximately 4 to 15 hours (depending on starting values and

Table 5.3: BIC Improvements compared to the Independence Copula over April 2012. The table reports the difference in Bayesian information criterion for the independence copula vis-à-vis the Gaussian, Ali-Mikhail-Haq (AMH), Frank, Clayton, Gumbel, Joe, and Symmetrized-Joe-Clayton (SJC) copulas: $\mathrm{DB}_s = \mathrm{BIC}_s^{Indep} - \mathrm{BIC}_s^{\tau}$, with $s \in \{st, dy\}$ and where τ denotes the copula under consideration. The data are tick price change series for Bank of America (BAC), Citigroup (C), JPMorgan (JPM), and Wells Fargo (WFC), observed during April 2012. $\#_{st}$ and $\#_{dy}$ denote the number of parameters in the case of a static and dynamic dependence model, respectively. The marginal Skellam distributions are always dynamic. The largest difference in BIC compared to the independence copula is boxed for static dependence and highlighted in gray for dynamic dependence.

			BAC/C		**BAC/JPM**		**BAC/WFC**	
Copula	$\#_{st}$	$\#_{dy}$	DB_{st}	DB_{dy}	DB_{st}	DB_{dy}	DB_{st}	DB_{dy}
Gaussian	9	12	367.72	492.44	454.74	430.19	309.00	320.44
AMH	9	12	348.47	456.91	428.41	407.31	288.07	294.79
Frank	9	12	338.51	465.47	416.51	398.64	283.25	283.45
Clayton	9	12	284.95	398.11	368.14	337.74	257.40	255.70
Gumbel	9	12	253.44	369.16	322.37	301.15	222.86	216.10
Joe	9	12	151.58	260.74	190.49	168.69	136.11	119.80
SJC	10	16	268.74	407.11	350.56	353.04	262.81	233.91
			C/JPM		**C/WFC**		**JPM/WFC**	
Copula	$\#_{st}$	$\#_{dy}$	DB_{st}	DB_{dy}	DB_{st}	DB_{dy}	DB_{st}	DB_{dy}
Gaussian	9	12	4545.80	4793.97	3593.87	3771.53	3929.64	4108.64
AMH	9	12	4264.01	4421.15	3336.53	3441.86	3660.14	3770.96
Frank	9	12	4469.20	4694.16	3593.59	3751.12	3895.69	4029.04
Clayton	9	12	3447.26	3680.29	2653.37	2836.15	3027.43	3200.86
Gumbel	9	12	3868.24	4083.63	3174.97	3311.71	3468.41	3612.73
Joe	9	12	2693.09	2889.16	2294.63	2410.84	2474.92	2621.57
SJC	10	16	4227.30	4413.54	3411.32	3500.54	3733.42	3835.32

data sets) on a i7-2600, 3.40 GHz desktop PC using four cores. The parameter estimates are presented in the Appendix.

We are mainly interested in the intraday pattern of the copula dependence parameter. Therefore, we first compute the point-wise sample mean of the intraday path of the copula dependence parameter over all 250 trading days of 2012. Figure 5.4 presents these sample means together with the confidence bands based on the corresponding sample variances. We compare our estimates of the intraday Gaussian dependence with an adjusted version of Spearman's rank correlation coefficient. This non-parametric rank correlation measure is computed for a rolling window of 600 seconds using only the observations with simultaneous trades. The observations are ordered while ties in ranks are corrected in the usual way by averaging the ranks. The resulting ranks are divided by 1 plus the number of observations. Finally we transform the ranks through the inverse normal cdf. The Pearson correlation between these transformed ranks are presented in Figure 5.4.

We find that the dependence between tick-size price changes exhibits a clear daily

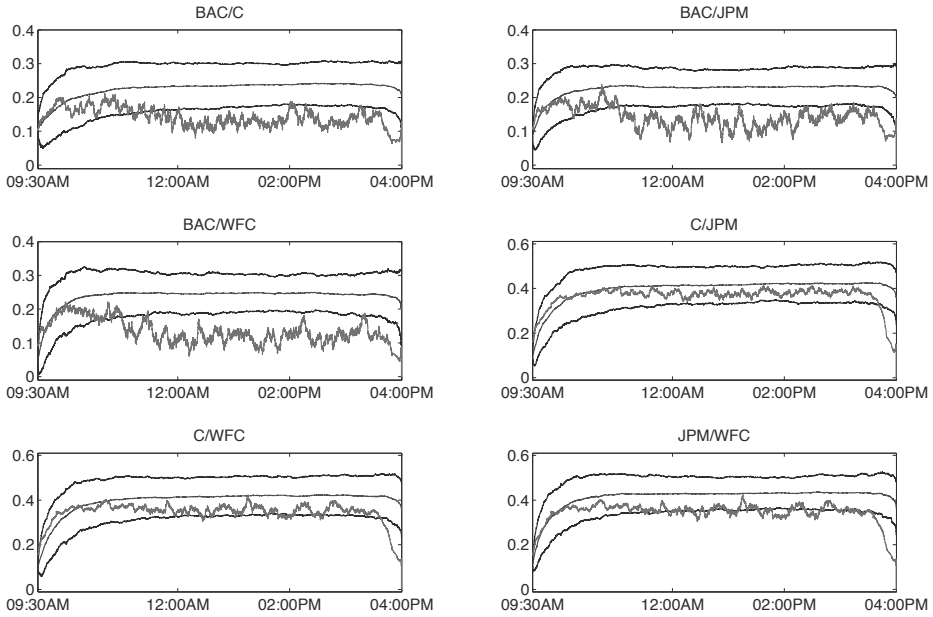

Figure 5.4: Point wise mean copula dependence intraday patterns over the 250 trading days in 2012 based on the Gaussian copula with Skellam marginals (smooth line). The smooth bounds are based on two sample standard deviations. The noisy series is the adjusted version of Spearman's non-parametric rank based estimator.

pattern across all stocks. We see that the trading day starts with a relatively small positive dependence level. But within the first hour of trading, the average dependence increases to a higher level where it remains throughout the trading day. Only during the last 15 minutes of trading, the dependence drops abruptly to a somewhat lower value. This pattern is found across all stock pairs. The point wise sample mean of the non-parametric rank-based dependence measure is much less smooth than our model-based measure. We also observe that the rank-based measure is significantly lower than the score-driven dependence implied by the Gaussian copula, which is partly due to the problems with rank-based statistics such as Spearman's rho for discrete data. We may conclude that our copula framework uses the data more efficiently. We emphasize that the estimated dependence patterns are not due to a lack of observation pairs at the end of the day. By contrast, Figure 5.3 shows that the number of joint observations is relatively higher at the start and at the end of the day.

The empirical intraday pattern for the dependence parameter can be expected given the flow of information over the 24 hour cycle. Throughout the trading day, information becomes available and can immediately be processed and impounded into stock prices due

to active trading. The accumulated overnight information can only be impounded at the opening of the trading day. While most of the common macro announcements are made during the trading day, most major firm-specific information is revealed after the active trading hours. The information available at the opening may therefore have a relatively larger idiosyncratic component. This causes the lower dependence level at the start of the trading day. Interestingly, the lower level in dependence at the opening mirrors the typically higher levels of intraday volatility during the opening.

It is likely that the short, sudden drop in dependence at the end of the day is related to the unwinding of open positions by market participants built up over the trading day. Such unwinding may be spurred by the need to satisfy overnight risk constraints. Hence it comprises a relatively larger idiosyncratic component and therefore also results in a decrease in the dependence parameter.

5.4.5 Comparison with intraday spline

The smooth patterns for the estimated intraday dependence across all stock pairs may prompt the question whether we can alternatively consider a smooth function to capture intraday dependence. We therefore compare our score-driven updating function for the copula dependence parameter ρ_t with a basic cubic spline function to account for the intraday seasonal pattern. The width of the confidence bands around the sample averages of the intraday dependence estimates presented in Figure 5.4 indicate that there exists considerable variation in the dependence parameter across the 250 trading days of 2012. For example, according to the 95% confidence bands the dependence parameter can vary between 0.1 and 0.3 at Noon.

To investigate whether a spline suffices to model the dependence parameter, we keep our score-driven approach for the marginal Skellam distributions, but model the copula dependence path by a cubic spline regression function as proposed by Poirier (1973). For the cubic spline regression, we specify the copula parameter by $\theta_t^c = \kappa' W_t$ where κ is a $q \times 1$ vector of parameters associated with the location of the q spline knots, and W_t is the t-th column of the weight matrix W as constructed in Poirier (1973). We have considered different numbers of knots and different locations for the knots in order to control for the possible sensitivity of the approach. The elements of κ become part of the parameter vector ψ and are jointly estimated by the method of maximum likelihood.

Table 5.4 presents the results for a range of different models. We report the loglikelihood gains and BIC reductions (in parentheses) for the considered spline model compared to the dynamic score-driven Skellam-Gaussian copula model. For almost all combinations, the loglikelihood gains are reported to be negative, indicating that the score-driven model outperforms the spline-based dynamic copula model in terms of fit. Although the models are not nested, the loglikelihood reductions are considerable. It comes as no surprise

Table 5.4: Model comparison: intraday dependence spline versus score-driven dynamics. The entries reflect the gain in log likelihood points (and improvements in BIC in parentheses) of the spline model compared to the dynamic score-driven Skellam-Gaussian copula model. The time points between braces are the positions of the spline knots. $\#_\psi$ denotes the number of estimated parameters, i.e., the dimension of ψ. Stocks are Bank of America (BAC), Citi (C), JPMorgan (JPM), and Wells Fargo (WFC).

Model description	$\#_\psi$	**BAC/C**	**BAC/JPM**	**BAC/WFC**
Spline {09:30, 12:00, 16:00}	11	-60.13	-54.61	-31.27
		(106.06)	(95.05)	(48.53)
Spline {09:30, 10:00, 16:00}	11	-54.97	-45.84	-19.97
		(95.72)	(77.50)	(25.94)
Spline {09:30, 10:00, 13:00, 16:00}	12	-54.76	-44.23	-18.61
		(109.52)	(88.45)	(37.23)
Spline {09:30, 10:00, 12:00, 14:00, 16:00}	13	-13.89	4.88	25.89
		(41.99)	(4.41)	(-37.77)
		C/JPM	**C/WFC**	**JPM/WFC**
Spline {09:30, 12:00, 16:00}	11	-845.33	-604.49	-571.95
		(1676.28)	(1194.70)	(1129.65)
Spline {09:30, 10:00, 16:00}	11	-768.95	-538.29	-504.24
		(1523.50)	(1062.30)	(994.22)
Spline {09:30, 10:00, 13:00, 16:00}	12	-736.52	-513.78	-470.66
		(1473.03)	(1027.57)	(941.32)
Spline {09:30, 10:00, 12:00, 14:00, 16:00}	13	-525.16	-343.89	-254.51
		(1064.71)	(702.06)	(523.27)

therefore that when we compare the models in terms of BIC reductions, we conclude that a fixed intraday spline does not capture the intraday dependence dynamics between discrete price changes as accurately as a model with a time-varying score-driven dependence parameter. The score-driven approach provides a better statistical description of our high-frequency data. To provide further evidence, we graphically display the dynamic copula parameter in Figure 5.5, for three randomly chosen trading days in 2012. These graphs also reveal that the daily pattern of θ_t^c may deviate substantially from the average intraday seasonal pattern.

5.5 Conclusions

Many empirical studies have concentrated on extracting high-frequency intraday volatility measures using tick-by-tick data. Here we have extended this literature to capture the intraday dynamic features of *dependence* using an observation-driven model-based copula approach with discrete marginals. We have developed a new model to capture the intraday seasonal pattern of dependence between discrete tick-size price changes of different stocks. The complete dependence model is composed of dynamic Skellam marginal dis-

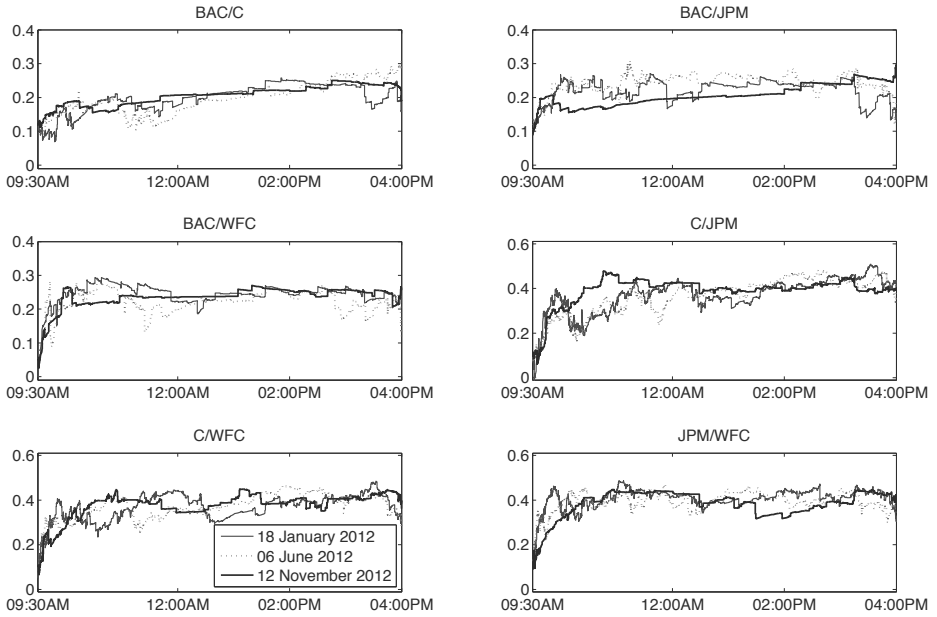

Figure 5.5: Copula dependence intraday patterns for a random selection of three days in 2012 based on the Gaussian copula with Skellam marginals. The selected days are 18 January 2012, 6 June 2012 and 12 November 2012. The panels show that the dependence pattern of a single day can be substantially less smooth than the point wise mean copula dependence path as presented in Figure 5.4.

tributions for the discrete price changes combined with a time-varying copula structure. The dynamic specifications rely on the score of the predictive loglikelihood with respect to the relevant dynamic parameters. The model performs well both in a controlled Monte Carlo setting and in an empirical study using high-frequency data. For four liquid U.S. financial stocks we found that the pairwise dependence varies over time during the trading day. There is a steep increase in dependence within the first hour of trading, and a decrease within the last 15 minutes of trading. We attribute these changes in dependence to the existence of more idiosyncratic risk components in the discrete price changes during the opening and closing hours of trading, in particular overnight firm-specific information accumulation when the market opens and the unwinding of inventory positions when the market closes. The time-varying dependence structures are of direct importance for intraday risk management. When managing a book of multiple stocks that are traded repeatedly over the course of the day, one should take into account the time-varying dependence between the stocks during the day.

Further research should be directed towards extending the bivariate copula model to a multivariate copula model. Pair copula constructions are a flexible way of decomposing

multivariate distributions into a distribution consisting of bivariate building blocks. We expect this approach to be an interesting extension to the current model since in this framework the dependence between several stocks can be investigated.

Appendices

The following appendices are part of the chapter 'Dynamic Discrete Copula Models' and are organised as follows. Appendix A provides derivations of the score vector and Appendix B presents additional tables and figures.

A Derivation of the score vector

The derivations presented here focus on bivariate copulas but can easily be extended to higher dimensions. We assume a time-varying factor θ_t that consist of three elements, where the first two elements correspond to the marginal parameters and the third element corresponds to the copula dependence parameter. We have $\theta_t = (\theta^m_{1,t}, \theta^m_{2,t}, \theta^c_t)'$. The derivative of a bivariate copula with respect to $\theta^m_{1,t}$ is given by

$$\frac{\partial C(u_{1,t}, u_{2,t}; \theta^c_t)}{\partial \theta^m_{1,t}} = \frac{\partial C(u_{1,t}, u_{2,t}; \theta^c_t)}{\partial u_{1,t}} \cdot \frac{\partial u_{1,t}}{\partial \theta^m_{1,t}}. \tag{5.13}$$

We observe that for the continuous parametric copula functions used in this chapter, the first component on the right hand side of (5.13) can be written as a conditional copula

$$P(U_{2,t} \leq u_{2,t} | U_{1,t} = u_{1,t}) = \frac{\partial C(u_{1,t}, u_{2,t}; \theta^c_t)}{\partial u_{1,t}}. \tag{5.14}$$

The second component on the right hand side of (5.13) is the derivative of the first marginal cdf, $u_{1,t} = F_1(y_{1,t}; \theta^m_{t,1})$, with respect to $\theta^m_{1,t}$. The derivative of a bivariate copula with respect to θ^c_t is denoted by $\frac{\partial C(u_{1,t}, u_{2,t}; \theta^c_t)}{\partial \theta^c_t}$.

As a concrete example, consider a bivariate Gaussian copula with Skellam marginals, where $\theta^m_{i,t} = \log(\sigma^2_{i,t})$, and $\rho_t = \theta^c_t / \sqrt{1 + (\theta^c_t)^2}$. This combination of copula, marginals, and parameterization is used in the application of Section 5.4. The Skellam distribution is discussed is Section 5.2. The bivariate Gaussian copula is given by

$$C_{\text{Ga}}(u_{1,t}, u_{2,t}; \rho_t) = \Phi_2 \left(\Phi^{-1}(u_{1,t}), \Phi^{-1}(u_{2,t}); \rho_t \right), \tag{5.15}$$

where Φ_2 is a bivariate standard normal cdf, Φ^{-1} a univariate inverse standard normal cdf, and $\rho_t \in (-1, 1)$ is a correlation parameter. The first expression on the right hand side of (5.13) follows directly from a bivariate normal cdf, we have

$$\frac{\partial C_{\text{Ga}}(u_{1,t}, u_{2,t}, \rho_t)}{\partial u_{1,t}} = \Phi \left(\frac{\Phi^{-1}(u_{2,t}) - \rho_t \Phi^{-1}(u_{1,t})}{\sqrt{1 - \rho^2_t}} \right). \tag{5.16}$$

A probably less well-known, but very useful result is given by Plackett (1954). It states

that for a bivariate standard Gaussian cdf, we have

$$\frac{\partial \Phi_2(x,y;\rho)}{\partial \rho} = (2\pi)^{-1}(1-\rho^2)^{-1/2}\exp\left(-\frac{(x^2-2\rho xy+y^2}{2(1-\rho^2)}\right), \tag{5.17}$$

where we can substitute $x = \Phi^{-1}(u_{1,t})$, $y = \Phi^{-1}(u_{2,t})$ and $\rho = \rho_t$ to obtain the appropriate expression for

$$\frac{\partial C(u_{1,t},u_{2,t};\theta_t^c)}{\partial \theta_t^c} = \frac{\partial C_{\text{Ga}}(u_{1,t},u_{2,t};\rho_t)}{\partial \rho_t}\cdot\frac{\partial \rho_t}{\partial \theta_t^c} = (1+\theta_t^c)^{-3/2}\cdot\frac{\partial C_{\text{Ga}}(u_{1,t},u_{2,t};\rho_t)}{\partial \rho_t}. \tag{5.18}$$

The first derivatives of the marginal Skellam cdfs in (5.13) are given by

$$\frac{\partial u_{i,t}}{\partial \sigma_{i,t}^2} = \exp\left(-\sigma_{i,t}^2\right)\sum_{\nu=-\infty}^{k}\left[\left(\frac{\nu}{\sigma_{i,t}^2}-1\right)I_{|\nu|}(\sigma_{i,t}^2)+I_{|\nu+1|}(\sigma_{i,t}^2)\right], \tag{5.19}$$

with $\frac{\partial u_{i,t}}{\partial \theta_{i,t}^m} = \frac{\partial u_{i,t}}{\partial \sigma_{i,t}^2}\,\frac{\partial \sigma_{i,t}^2}{\partial \theta_{i,t}^m} = \sigma_{i,t}^2\cdot\frac{\partial u_{i,t}}{\partial \sigma_{i,t}^2}$, for $i = 1, 2$.

B Tables and figures

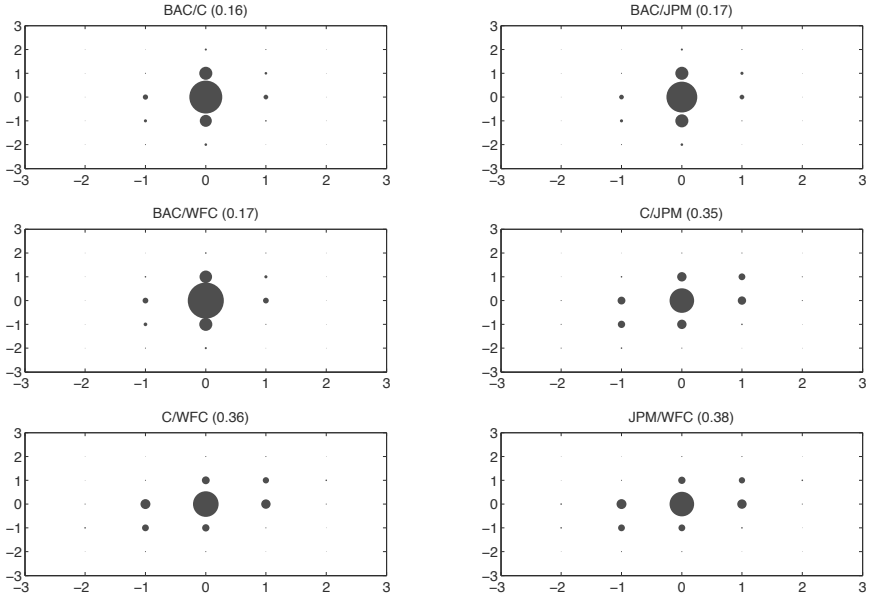

Figure 5.6: The figure shows discrete scatter plots of the bivariate tick price change series in April 2012. The diameter of the circle represents the bivariate observation frequency in the data. We emphasize that the panels only show the situation where both price change series have a trade at time t. The reported value between parenthesis in the panel header is Pearson's linear correlation between the series.

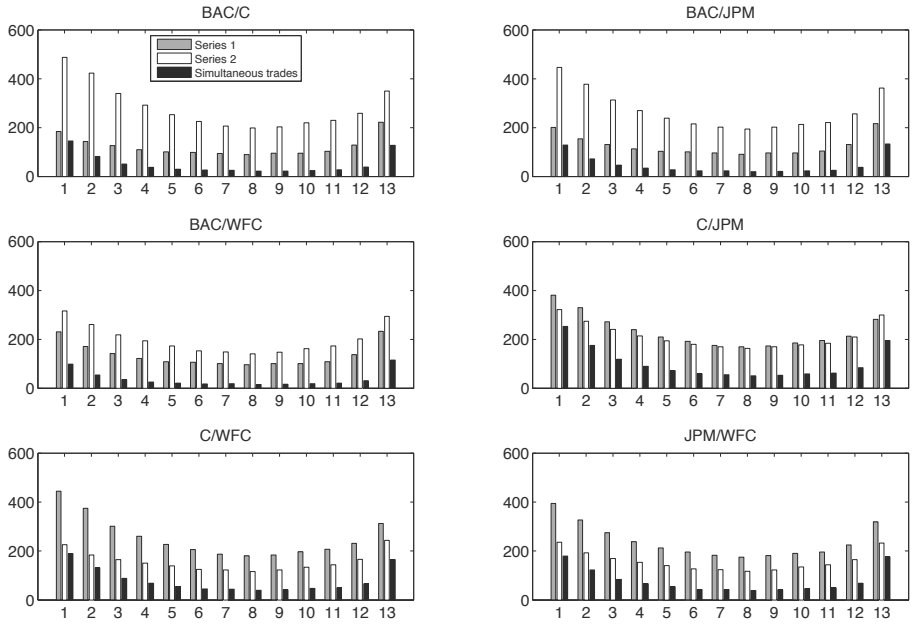

Figure 5.7: The figure displays the number of simultaneous trades per half hour of the trading day as well as the the number of trades if only series 1 or series 2 trade. The numbers are averaged over all 250 trading days of the year 2012. The panels show the six combinations of stocks under consideration. The numbers on the x-axis represent the number of half hours in a trading day (13 in total).

Table 5.5: The table reports maximum likelihood estimates of the Gaussian copula with Skellam marginals fitted to a data set of tick price changes for the year 2012. Standard errors are in parenthesis and are calculated using a numerically obtained Hessian matrix. The parameters are obtained in approximately 4 to 15 hours (depending on starting values and data sets) on a i7-2600, 3.40 GHz desktop PC using four cores.

Series	$\hat{\tilde{\theta}}_{1,1}$	$\hat{\tilde{\theta}}_{2,1}$	$\hat{\tilde{\theta}}_{3,1}$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$
BAC/C	-0.263	0.844	0.116	-0.115	-0.002	2.54E-04	0.709	0.060	0.030	0.910	0.997	0.999
	(0.032)	(0.034)	(0.032)	(0.002)	(0.001)	(0.001)	(0.011)	(0.001)	(0.003)	(0.003)	(1.24E-04)	(2.01E-04)
BAC/JPM	-0.249	1.167	0.089	-0.116	-0.002	4.97E-04	0.695	0.063	0.036	0.910	0.996	0.998
	(0.032)	(0.034)	(0.032)	(0.002)	(0.001)	(0.001)	(0.010)	(0.001)	(0.006)	(0.002)	(1.32E-04)	(0.001)
BAC/WFC	-0.305	0.916	0.049	-0.132	-0.004	0.001	0.677	0.083	0.050	0.899	0.994	0.997
	(0.030)	(0.031)	(0.030)	(0.002)	(0.001)	(0.001)	(0.009)	(0.002)	(0.004)	(0.003)	(2.03E-04)	(3.59E-04)
C/JPM	0.831	1.141	0.103	-0.002	-0.002	0.001	0.057	0.061	0.030	0.997	0.997	0.999
	(0.008)	(0.008)	(0.008)	(1.88E-05)	(1.88E-05)	(1.88E-05)	(0.001)	(0.001)	(0.001)	(9.61E-05)	(1.07E-04)	(7.87E-05)
C/WFC	0.832	0.898	0.100	-0.002	-0.003	0.001	0.057	0.080	0.033	0.997	0.996	0.999
	(0.008)	(0.008)	(0.008)	(2.15E-05)	(2.15E-05)	(2.15E-05)	(0.001)	(0.001)	(0.002)	(1.10E-04)	(2.40E-04)	(2.63E-04)
JPM/WFC	1.145	0.872	0.116	-0.002	-0.003	0.001	0.062	0.076	0.034	0.996	0.996	0.998
	(0.008)	(0.008)	(0.008)	(2.05E-05)	(2.05E-05)	(2.05E-05)	(0.001)	(0.001)	(0.002)	(1.19E-04)	(2.06E-04)	(2.16E-04)

Bibliography

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

Admati, A. R. and P. Pfleiderer (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies 1(1)*, 3–40.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics 160*(1), 160–175.

Allez, R. and J. P. Bouchaud (2011). Individual and collective stock dynamics: intra-day seasonalities. *New Journal of Physics 13*, 1–12.

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

Alzaid, A. A. and M. A. Omair (2014). Poisson difference integer valued autoregressive model of order one. *Bulletin of the Malaysian Mathematical Sciences Society 37(2)*, 465–485.

Andersen, T. G. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance 4*, 115–158.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *J. American Statistical Association 96*(453), 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review 93(1)*, 38–62.

Andersson, J. and D. Karlis (2014). A parametric time series model with covariates for integers in z. *Statistical Modelling 14(2)*, 135–156.

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

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

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. Royal Statistical Society B 63*(2), 167–241.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. Royal Statistical Society B 64*(2), 253–280.

Barreto-Souza, W. and M. Bourguignon (2013). A skew true inar(1) process with application. Discussion paper, arXiv/1306.0156.

Bibinger, M., N. Hautsch, P. Malec, and M. Reiss (2014). Estimating the spot covariation of asset prices – Statistical theory and empirical evidence. Discussion paper, University of Vienna.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika*, forthcoming.

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.

Bradley, J. V. (1968). *Distribution-Free Statistical Tests*. New Jersey: Prentice-Hall.

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

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

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Creal, D., S. J. Koopman, A. Lucas, and M. Zamojski (2015). Generalized autoregressive method of moments. Technical report, Vrije Universiteit Amsterdam.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation driven mixed-measurement dynamic factor models. *Review of Economics and Statistics 96*(5), 898–915.

Creal, D. D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics 29*, 552–563.

Creal, D. D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*, 777–795.

Crowder, M., M. J. Dixon, A. Ledford, and M. Robinson (2002). Dynamic modelling and prediction of english football league matches for betting. *The Statistician 51(2)*, 157–168.

De Lira Salvatierra, I. and A. J. Patton (2015). Dynamic Copula Models and High Frequency Data. *Journal of Empirical Finance 30*, 120–135.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*, 337–364.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253–265.

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 (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford: Oxford University Press.

Fahrmeir, L. and G. Tutz (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *J. American Statistical Association 89(428)*, 1438–1449.

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

Frank, M. J. (1979). On the simultaneous associativity of f(x, y) and x + y - f(x, y). *Aequationes Mathematicae 19*, 194–226.

Freeland, R. K. (2010). True integer value time series. *AStA-Advances in Statistical Analysis 94*, 217–229.

Genest, C. and J. Nešlehová (2007). A primer on copulas for count data. *Astin Bulletin 37(2)*, 475–515.

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

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74(6)*, 1545–1578.

Glickman, M. E. (2001). Dynamic paired comparison models with stochastic variances. *Journal of Applied Statistics 28(6)*, 673–689.

Glickman, M. E. and H. S. Stern (1998). A state-space model for national football league scores. *J. American Statistical Association 93(441)*, 25–35.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting 21*, 331–340.

Hansen, P. R., G. Horel, A. Lunde, and I. Archakov (2015). A markov chain estimator of multivariate volatility from high frequency data. Discussion paper, CREATES, Aarhus University.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise (with discussion). *J. Business and Economic Statist. 24*, 127–161.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Series Monographs. Cambridge: Cambridge University Press.

Harvey, A. C. and S. J. Koopman (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association 88(424)*, 1228–1236.

Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*(507), 1112–1122.

Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same poisson distribution. *J. Royal Statistical Society A 100(3)*, 415–416.

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

Johnson, N. L., S. Kotz, and N. Balakrishnan (1997). *Discrete Multivariate Distributions*. New York: John Wiley & Sons.

Jorgensen, B., S. Lundbye-Christensen, P. Song, and L. Sun (1999). A state space model for multivariate longitudinal count data. *Biometrika 86(1)*, 169–181.

Jung, R. C., M. Kukuk, and R. Liesenfeld (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics & Data Analysis 51(4)*, 2350–2364.

Jungbacker, B. and S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika 94(4)*, 827–839.

Kachour, M. and L. Truquet (2010). A p-order signed integer-valued autoregressive (sinar(p)) model. *Journal of Time Series Analysis 32*, 223–236.

Karlis, D. and I. Ntzoufras (2003). Analysis of sports data by using bivariate Poisson models. *The Statistician 52(3)*, 381–393.

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.

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal 35(4)*, 917–926.

Knorr-Held, L. (2000). Dynamic rating of sports teams. *The Statistician 49(2)*, 261–276.

Kocherlakota, S. and K. Kocherlakota (1992). *Bivariate Discrete Distributions*. New York: Dekker.

Koopman, S. J. and R. Lit (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *J. Royal Statistical Society A 178(1)*, 167–186.

Koopman, S. J., R. Lit, and A. Lucas (2015). Intraday stock price dependence using dynamic discrete copula distributions. Discussion paper, Tinbergen Institute.

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

Koopman, S. J. and A. Lucas (2008). A non-Gaussian panel time series model for estimating and decomposing default risk. *J. Business and Economic Statist. 26*, 510–25.

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

Koopman, S. J., A. Lucas, and M. Scharth (2015). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*, forthcoming.

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

Koopman, S. J., N. Shephard, and J. A. Doornik (2008). *SsfPack 3.0: Statistical algorithms for models in state space form*. London: Timberlake Consultants Press.

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

Lucas, A., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business and Economic Statistics 32*(2), 271–284.

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.

Münnix, M. C., R. Schäfer, and T. Guhr (2010). Impact of the tick-size on financial returns and correlations. *Physica A 389(21)*, 4828–4843.

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer.

O'Hara, M., C. Yao, and M. Ye (2014). What's not there: Odd lots and market data. *Journal of Finance 69(5)*, 2199–2236.

Ord, K., C. Fernandes, and A. C. Harvey (1993). Time series models for multivariate series of count data. In T. S. Rao (Ed.), *Developments in Time Series Analysis*. London: Chapman and Hall.

Owen, A. (2011). Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics 22*, 99–113.

Panagiotelis, A., C. Czado, and H. Joe (2012). Pair copula constructions for multivariate discrete data. *J. American Statistical Association 107*, 1063–1072.

Patton, A. J. (2002). Applications of copula theory in financial econometrics. Ph.D. dissertation, University of California, San Diego.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review 47(2)*, 527–556.

Plackett, R. L. (1954). A reduction formula for Normal multivariate integrals. *Biometrika 41(3,4)*, 351–360.

Poirier, D. J. (1973). Piecewise regression using cubic spline. *Journal of the American Statistical Association 68(343)*, 515–524.

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.

Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics 1(1)*, 2–25.

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

Schepsmeier, U. and J. Stöber (2014). Derivatives and Fisher information of bivariate copulas. *Statistical Papers 55*, 525–542.

Shahtahmassebi, G. (2011). *Bayesian Modelling of Ultra High-Frequency Financial Data*. Doctoral thesis, Research with Plymouth University. University of Plymouth.

Shahtahmassebi, G. and R. Moyeed (2014). Bayesian modelling of integer data using the generalised poisson difference distribution. *International Journal of Statistics and Probability 3*, 24–35.

Shephard, N. (2005). *Stochastic volatility: Selected Readings*. New York: Oxford University Press.

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

Shephard, N. and J. J. Yang (2015). Continuous time analysis of fleeting discrete price moves. Discussion paper, Harvard University.

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society 109(3)*, 296.

Sklar, A. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Publications de l'Institut Statistique de l'Université de Paris, 8, 229-231.

So, M. K. P. (2003). Posterior mode estimation for nonlinear and non-Gaussian state space models. *Statistica Sinica 13*, 255–274.

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

Winkler, R. L. (1969). Scoring rules and the evaluation of probability assessors. *Journal of the American Statistical Association 64(327)*, 1073–1078.

Wood, R. A., T. H. McInish, and J. K. Ord (1985). An investigation of transactions data for nyse stocks. *Journal of Finance 40(3)*, 723–739.

Zhang, H., D. Wang, and F. Zhu (2009). Inference for inar(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference 140*, 667–683.

Zimmer, D. M. and P. K. Trivedi (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *J. Business and Economic Statist. 24(1)*, 63–76.

Summary

This dissertation contains four chapters on time varying parameter models. We summarize the main findings of each chapter.

In Chapter 2 we presented a non-Gaussian state space model for the analysis and forecasting of football matches. The model is a novelty in the sports and statistics literature because it takes a match result as a pairwise observation that is assumed to come from a bivariate Poisson distribution with intensity coefficients for the number of goals scored by the two teams and a dependence coefficient for measuring the correlation between the two scores. The intensity coefficients are functions of the strengths of attack and defence of the teams which evolve stochastically over time. Extensions of the model including amendments for the over representation of draws in data sets, breaks in strengths of attack and defence after summer and winter breaks, and a team-specific home advantage were considered. Since the match results of the teams are analysed for all teams in the competition, and over a period of several seasons, the resulting model is a high dimensional panel time series. Due to promotion and relegation of teams the panel increased with every season and has many missing values. We showed that football match results from a high dimensional panel can be analysed effectively within a non-Gaussian state space model. The statistical analysis is based on exact maximum likelihood and signal extraction methods which rely on efficient Monte Carlo simulation techniques such as importance sampling.

In Chapter 3 we modelled tick-by-tick discrete price changes for U.S. stocks listed on the New York Stock Exchange by a dynamic modified Skellam distribution with a variance parameter that evolves stochastically over time. The price changes were expressed in multiples of the tick size and are therefore in $\mathbb{Z}$. The Skellam distribution is congruent with the discrete price change data and we analysed the model with state space and importance sampling methodology. The new model accounts for a stable importance sampling estimation procedure, a good in-sample fit, an adequate diagnostic performance, and an accurate out-of-sample forecasting performance in comparison to a number of relevant benchmark models. We conclude that the new dynamic modified Skellam model provides a flexible modelling framework that can be effectively employed to capture the dynamics in high-frequency tick-by-tick data with many missing entries. Since the model

produces intraday patterns of high-frequency volatility dynamics, it may provide an interesting comparison with the usual stochastic volatility models that are typically associated with the time varying variance in time series of daily continuously compounded rates of returns.

In Chapter 4 we introduced a general dynamic model for Skellam distributed difference in counts. Our version of the Skellam distribution has two intensity parameters that correspond to the intensities of Poisson distributed counts. We opted for 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 method of Koopman et al. (2014). In the application, we modelled the difference between the number of goals scored by the home and away team in a high dimensional unbalanced panel of football match results. We also extended our benchmark model to a model that included regression effects, heterogeneous dynamics in the panel, and extensions of the Skellam distribution that assign different probability mass to zeros. 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 large unbalanced panels. Our flexible modelling framework for time series may provide a useful benchmark for empirical applications based on integer outcomes that can take both positive and negative values.

In Chapter 5 we continued the modelling of tick-by-tick discrete price changes and extended our research to capture the intraday seasonal pattern of dependence between discrete tick-size price changes of different stocks. We captured the intraday dynamic features of dependence using an observation driven model-based copula approach with discrete marginals. The complete dependence model is composed of dynamic Skellam marginal distributions for the discrete price changes combined with a time varying copula structure. We applied a range of bivariate copulas and the Gaussian copula fitted the data best. For four liquid U.S. financial stocks we found that the dependence structure varies over time during the trading day. There is a steep increase in dependence within the first hour of trading, and a steep decrease within the last 15 minutes of trading. We attribute these changes in dependence to the existence of more idiosyncratic risk components in the discrete price changes during the opening and closing hours of trading, in particular overnight firm-specific information accumulation when the market opens and the unwinding of inventory positions when the market closes.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

592 D.L. IN 'T VELD, *Complex Systems in Financial Economics: Applications to Interbank and Stock Markets*
593 Y. YANG, *Laboratory Tests of Theories of Strategic Interaction*
594 M.P. WOJTOWICZ, *Pricing Credits Derivatives and Credit Securitization*
595 R.S. SAYAG, *Communication and Learning in Decision Making*
596 S.L. BLAUW, *Well-to-do or doing well? Empirical studies of wellbeing and development*
597 T.A. MAKAREWICZ, *Learning to Forecast: Genetic Algorithms and Experiments*
598 P. ROBALO, *Understanding Political Behavior: Essays in Experimental Political Economy*
599 R. ZOUTENBIER, *Work Motivation and Incentives in the Public Sector*
600 M.B.W. KOBUS, *Economic Studies on Public Facility use*
601 R.J.D. POTTER VAN LOON, *Modeling non-standard financial decision making*
602 G. MESTERS, *Essays on Nonlinear Panel Time Series Models*
603 S. GUBINS, *Information Technologies and Travel*
604 D. KOPÁNYI, *Bounded Rationality and Learning in Market Competition*
605 N. MARTYNOVA, *Incentives and Regulation in Banking*
606 D. KARSTANJE, *Unraveling Dimensions: Commodity Futures Curves and Equity Liquidity*
607 T.C.A.P. GOSENS, *The Value of Recreational Areas in Urban Regions*
608 Ł.M. MARĆ, *The Impact of Aid on Total Government Expenditures*
609 C. LI, *Hitchhiking on the Road of Decision Making under Uncertainty*
610 L. ROSENDAHL HUBER, *Entrepreneurship, Teams and Sustainability: a Series of Field Experiments*
611 X. YANG, *Essays on High Frequency Financial Econometrics*
612 A.H. VAN DER WEIJDE, *The Industrial Organization of Transport Markets: Modeling pricing, Investment and Regulation in Rail and Road Networks*
613 H.E. SILVA MONTALVA, *Airport Pricing Policies: Airline Conduct, Price Discrimination, Dynamic Congestion and Network Effects.*
614 C. DIETZ, *Hierarchies, Communication and Restricted Cooperation in Cooperative Games*

615 M.A. ZOICAN, *Financial System Architecture and Intermediation Quality*
616 G. ZHU, *Three Essays in Empirical Corporate Finance*
617 M. PLEUS, *Implementations of Tests on the Exogeneity of Selected Variables and their Performance in Practice*
618 B. VAN LEEUWEN, *Cooperation, Networks and Emotions: Three Essays in Behavioral Economics*
619 A.G. KOPÁNYI-PEUKER, *Endogeneity Matters: Essays on Cooperation and Coordination*
620 X. WANG, *Time Varying Risk Premium and Limited Participation in Financial Markets*
621 L.A. GORNICKA, *Regulating Financial Markets: Costs and Trade-offs*
622 A. KAMM, *Political Actors playing games: Theory and Experiments*
623 S. VAN DEN HAUWE, *Topics in Applied Macroeconometrics*
624 F.U. BRÄUNING, *Interbank Lending Relationships, Financial Crises and Monetary Policy*
625 J.J. DE VRIES, *Estimation of Alonso's Theory of Movements for Commuting*
626 M. POPŁAWSKA, *Essays on Insurance and Health Economics*
627 X. CAI, *Essays in Labor and Product Market Search*
628 L. ZHAO, *Making Real Options Credible: Incomplete Markets, Dynamics, and Model Ambiguity*
629 K. BEL, *Multivariate Extensions to Discrete Choice Modeling*
630 Y. ZENG, *Topics in Trans-boundary River sharing Problems and Economic Theory*
631 M.G. WEBER, *Behavioral Economics and the Public Sector*
632 E. CZIBOR, *Heterogeneity in Response to Incentives: Evidence from Field Data*
633 A. JUODIS, *Essays in Panel Data Modelling*
634 F. ZHOU, *Essays on Mismeasurement and Misallocation on Transition Economies*
635 P. MULLER, *Labor Market Policies and Job Search*
636 N. KETEL, *Empirical Studies in Labor and Education Economics*
637 T.E. YENILMEZ, *Three Essays in International Trade and Development*
638 L.P. DE BRUIJN, *Essays on Forecasting and Latent Values*
639 S. VRIEND, *Profiling, Auditing and Public Policy: Applications in Labor and Health Economics*
640 M.L. ERGUN, *Fat Tails in Financial Markets*
641 T. HOMAR, *Intervention in Systemic Banking Crises*