Financial Models of Interaction Based on Marked Point Processes and Gaussian Fields

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen
Doktorgrades
„Doctor rerum naturalium"
der Georg-August-Universität Göttingen

im Promotionsprogramm Mathematik
der Georg-August University School of Science (GAUSS)

vorgelegt von
Alexander Malinowski
aus Frankenberg/Eder

Göttingen, 2012

Betreuungsausschuss

Prof. Dr. Martin Schlather,
Institut für Mathematik, Universität Mannheim

Prof. Dr. Tatyana Krivobokova, Institut für Mathematische Stochastik,
Courant Forschungszentrum „Armut, Ungleichheit und Wachstum in Entwicklungsländern",
Göttingen

Mitglieder der Prüfungskommission

Referent:
Prof. Dr. Martin Schlather,
Institut für Mathematik, Universität Mannheim

Korreferentin:
Jun.-Prof. Dr. Andrea Krajina,
Institut für Mathematische Stochastik, Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Tatyana Krivobokova, Institut für Mathematische Stochastik,
Courant Forschungszentrum „Armut, Ungleichheit und Wachstum in Entwicklungsländern",
Göttingen

Prof. Dr. Ingo Witt,
Mathematisches Institut, Göttingen

Prof. Dr. Samuel James Patterson,
Mathematisches Institut, Göttingen

Prof. Dr. David Russell Luke,
Institut für Numerische und Angewandte Mathematik, Göttingen

Tag der mündlichen Prüfung: 18. Dezember 2012

Preface

First of all, I would like to thank my supervisor Prof. Martin Schlather for his constant and encouraging support, his friendly way of collaboration, his inspiring ideas, and, above all, for hustling me into doing a PhD at all, against my initial wish. Thank you for your trust in me.

I am deeply grateful for an extremely nice and amicable working group, which created a lovely and relaxed working atmosphere and at the same time brought up ambitious mathematical problems and provided competent support for all of my professional and technical questions. I particularly appreciate the close cooperation with my office mate Sebastian Engelke and our working group member Dr. Marco Oesting in a highly interesting project during summer 2012, the idea for which was brought up by Prof. Zakhar Kabluchko. Pleasant but not crucial is that the project only recently resulted in two submissions (Engelke *et al.*, 2012*a*,*b*).

I also gratefully acknowledge the opportunity to spend three months of my PhD at the University of Wisconsin–Madison with the working group of Prof. Zhengjun Zhang. Working on financial time series analysis and extreme value theory and at the same time being clearly statistically oriented, he provided a lot of valuable input to my work. The two joint manuscripts Malinowski *et al.* (2012*a*,*b*), which have recently been submitted, were initialized during this research stay in Madison. Thanks to all staff members and PhD students who made my stay at the UW as comfortable as possible.

Moreover, I am indebted to all members of the Centre for Statistics and of the DFG Research Training Group 1644 who contributed to my research through useful comments and suggestions, in particular Prof. Walter Zucchini, Prof. Olaf Korn and Dr. Katrin Meyer. I would like to thank Prof. Tatyana Krivobokova for co-supervising and Jun.-Prof. Andrea Krajina for co-refereeing my PhD thesis.

Finally, I thank all my colleagues and fellow PhD students at the IMS and within the RTG 1644 for a great time during my PhD, with lot of fun, many spare time activities and genuine friendship. I warmly thank all my friends for their manifold support, and, above all, my family who clearly provided the basis for what I can look back on now. Special thanks go to my live-in partner Ulrike who still bears with me and supports me more than I deserve. Thank you for your long-standing unconditional love.

Contents

1 Introduction

Classical low-frequency financial data such as daily asset returns, have commonly been analyzed via ARCH and GARCH time-series models during the last decades. These *autoregressive conditional heteroscedastic* processes were introduced in the seminal paper of Engle (1982) and extended into various directions, see Pacurar (2008) for a recent survey. An alternative to the ARCH-based framework is the class of (discrete-time) stochastic volatility (SV) models, in which the variance is assumed to be driven by some latent stochastic process (e.g., Taylor, 1986; Ghysels *et al.*, 1995; Shephard, 1996; Barndorff-Nielsen & Shephard, 2001). The common implicit assumption of most of these time-series models is that measurements are evenly spaced in time.

Besides the large variety in discrete-time models, also continuous-time models have been attracting much attention, most of them being based on stochastic differential equations. The Black-Scholes option pricing model (Black & Scholes, 1973), which describes asset prices by a geometric Brownian motion, is a famous representative of this class of models. In the continuous-time framework, financial data are naturally seen as measurements of the latent process. While for low-frequency data, these measurement locations usually form a regular grid and only the observable or latent values and prices are subject to stochastic behavior, an additional stochastic component arises when financial data at the transaction level are considered. These are often called (ultra-)high-frequency data (e.g., Goodhart & O'Hara, 1997; Ghysels & Jasiak, 1998; Engle, 2000). One of the essential characteristics of this type of data is that the observations are irregularly spaced in time, and it proved indispensable to include the process of transaction times into stochastic models in this context. For this purpose, the concept of marked point processes (MPP) is suitable, as it allows for arbitrary forms of dependence between the point locations and the so-called marks (e.g., Karr, 1991; Møller & Waagepetersen, 2003; Schlather *et al.*, 2004; Daley & Vere-Jones, 2008; Myllymäki & Penttinen, 2009; Ho & Stoyan, 2008; Diggle *et al.*, 2010). In the financial context, the point locations are given by the transaction time stamps, and the marks can be any quantity associated with the transactions; in most cases, price and volume are of practical relevance. The pioneering work of Engle (2000) triggered a considerable amount of research and publications on MPP-based models for high-frequency financial data; a broad survey can be found in Bauwens & Hautsch (2009).

While it is well-known that price volatility of exchange-traded assets increases in times of high trading intensity (e.g., Easley & O'Hara, 1992), which is often modeled by allowing the volatility process to depend on inter-trade durations, not much is known about general influences of the pattern of transaction times on the volatility process at the microscopic scale of minutes and seconds. However, the knowledge of these effects might help to improve prediction of instantaneous volatility compared to prediction solely based on the history of the price process. Important questions arising in this respect are:

- Does the knowledge of past transaction times have an influence on instantaneous characteristics of the price process?

- If so, do the prevalent financial and econometric models account for these effects and, if they do not, can we make them account for these time-stamp-price interactions by suitable modifications?

These questions are considered in Chapter 3 of this thesis. Modeling transaction data as a marked point process and using existing MPP theory and the E- and V-function introduced by Schlather (2001), novel statistics are derived which enable to detect interaction effects between past transaction dates and instantaneous asset price characteristics. In particular, conditional expectations and variances are considered, where "conditional" refers to the existence of other transactions at a fixed temporal distance and thus leads to functions whose natural domain is (temporal) distance. After an application of these statistics to real data, an empirical way of assessing the capability of statistical models to capture possible interactions effects is followed: The model is estimated based on given data, new realizations of that model are simulated, and the statistics are then re-applied to the simulated data. Exemplarily, the famous UHF-GARCH model (Engle, 2000) is examined in Chapter 3. Additionally, an extension of this model is proposed, which is of scientific interest on its own, in order to make the model better account for the above effects of interaction. Statistical inference is based on 7-year transaction data from German stock exchanges.

Besides the simulation-based approach of validating financial models with regard to their capability of representing dependence between marks and point locations, an analytical treatment of the proposed MPP characteristics is certainly preferable, but in many cases unfeasible, e.g., in models that are given by a certain autoregressive specification. Models for which an analytical treatment is at least partially feasible can be found by focusing on a class that is based on Poisson point processes or, as a generalization thereof, on doubly stochastic Poisson processes (DSPP). Then, due to the independence property of a Poisson process, the randomness of the point locations can be integrated out in some sense if the specification of marks depends on the point locations in a sufficiently simple manner. Some examples thereof are presented in Chapter 4. A question that is of rather theoretical interest in this connection is concerned with the range of valid functions for a particular statistic, e.g., for the bi-directional V-function defined in Chapter 3. A contribution to that problem is provided in the second part of Chapter 4.

The stochastic development of asset prices is commonly modeled through continuous-time processes, which is a motivation for considering transaction data as measurements of a latent continuous-time stochastic process, with unevenly spaced measurement locations (e.g., Aït-Sahalia *et al.* 2005; Hansen & Lunde 2006). This perception directly leads to a fruitful combination of two fields of spatial statistics:
On the one hand, the principle goal of *geostatistical methods* is estimation and prediction of a continuous-space process, based on discrete measurements with locations that might be influenced by, but that do in turn *not* impact the underlying continuous-space process. While geostatistical methods (e.g., Goovaerts 1997; Chiles & Delfiner 1999; Wackernagel 2003) were originally developed in the context of environmental applications and thus *space* referred

to (subsets of) $\mathbb{R}^2$ or $\mathbb{R}^3$ only, the concepts can also be applied to temporal processes ($\mathbb{R}^1$), covering financial processes, or to processes on high-dimensional index spaces ($\mathbb{R}^d$, $d \in \mathbb{N}$). On the other hand, MPP approaches put the focus on data, which, in general, cannot be considered as measurements of a continuous-space process. Then, the properties of a *typical point* or a typical pattern of points are of primary interest. Examples of genuine MPPs can easily be found when interacting objects are observed in space or time. Trees in a forest, for instance, compete with each other for resources—a fact that makes trees in a cluster of other trees tend to be smaller. In particular, the variable "height of a tree" is unobservable at locations where there is no tree and if there was a tree, this would possibly have influenced the height of the surrounding trees. Similarly, financial transactions usually have an impact on the future as well as on the instantaneous price process; hence this type of data exemplifies how the distinction between the two perspectives can become blurred. However, this problem has not been addressed in the literature up to now and will be dealt with in Chapter 5.

For a marked point process $\Phi = \{(t_i, y_i) : i \in \mathbb{Z}\}$, where t_i and y_i denote the point locations and marks, respectively, the definition of (higher-order) mean marks is based on measures of the form $\alpha(B) = \mathbb{E} \sum_{(t,y) \in \Phi} f(y) \mathbf{1}_B(t)$ or $\alpha(B) = \mathbb{E} \sum_{(t_1,y_1),(t_2,y_2) \in \Phi} f(y_1,y_2) \mathbf{1}_B((t_1,t_2))$, for a Borel set B in a suitable space. However, when the process is non-stationary or non-ergodic, the definition of averages becomes ambiguous as the process might have a different stochastic behavior in different realizations (non-ergodicity) or in different areas of the observation window (non-stationarity). Then, also depending on which of the above perspectives one takes and depending on the statistical questions at hand, it might be sensible to replace the ordinary expectation functional in the above moment measures α by weighted means. In Chapter 5, different definitions for the moments are investigated, including a new hierarchical definition for non-ergodic MPPs, and embedded into a family of weighted mean marks. Examples of applications are presented, in which different weighted mean marks all have a sensible meaning at the same time, but for different statistical questions. Besides the relation to methods from continuous-time and continuous-space process analysis, also asymptotic properties of the corresponding estimators as well as optimal weighting procedures will be illuminated.

The analysis of the tail behavior of financial returns is of great importance for the assessment of financial risk (e.g., Embrechts *et al.*, 1997; Embrechts, 2000; Bouchaud & Potters, 2003; Gilli & Këllezi, 2006). Thickness of tails and tail dependence in multivariate distributions, measured, for example, in terms of the extreme value index and in terms of extremal coefficients (Smith, 1990), respectively, crucially determine the probability of large portfolio losses, e.g., in times of financial crises. It is therefore of scientific interest to define and analyze analogue quantities for high-frequency financial data by combining extreme value theory (EVT) with MPP methods. In extreme value analysis, mostly only those observations that are extreme in some sense are used for estimation, which can reduce the amount of effectively used data dramatically. This might amplify the impact of structural breaks or, more generally, of non-stationarity and non-ergodicity of financial processes. Based on the MPP theory developed in Chapter 5, Chapter 6 provides new summary statistics for the (multivariate) tail behavior of irregularly spaced data. By applying MPP versions of the classical tail index to transaction data from German stock exchanges in a non-ergodic

framework, new insights into the basic structure of this type of data are obtained, which are of interest for risk management and algorithmic trading purposes, for instance. A simulation study underpins the observed effects and enables assessing the finite sample properties of the respective estimators.

The above conditional tail index is a *univariate* statistic, adapted to the MPP framework as a function of the underlying point pattern. The next natural step from an EVT point of view is the analysis of second- and higher-order characteristics that describe extremal dependence between the marks of an MPP, similarly to the mark correlation function for the central part of the distribution (in case second moments exist). As an example for a second-order characteristic, the *extremal correlation* for a bivariate vector (Z_0, Z_1) in the MDA of some extreme-value distribution and with identically distributed margins is defined as $\lim_{u \to x_0} \mathbb{P}(Z_1 > u \,|\, Z_0 > u)$, where x_0 is the upper endpoint of the distribution of Z_0 and Z_1. These characteristics can in principle be carried over to the MPP setup (cf. Section 6.5 for extremal coefficients of MPPs). The ultimate goal in the sense of full information about the tail behavior and its interdependence with the irregular pattern of points would be (conditional) MPP versions of the so-called exponent measure, which, for a multivariate extreme-value distribution, describes the full dependence structure.
However, even in classical extreme value statistics, i.e. without the additional challenges of MPPs, estimation procedures for max-stable processes and multivariate max-stable distributions are still in development. Due to the lack of analytical tractability of many max-stable distributions, maximum likelihood methods based on blockmaxima are often unfeasible. *Peaks-over-threshold* (POT) methods form the second big class of estimation approaches and make use of the fact that suitably defined exceedances over an increasing critical value converge to a Poisson process (Leadbetter, 1991; Embrechts *et al.*, 1997), from which the max-stable process can be recovered. This is true for a large class of processes. POT methods are particularly attractive from an MPP point of view as they consider all "large" events and can thus be easily applied to the MPP framework, in contrast to the idea of building blockwise maxima. POT methods can also be expected to yield stable results when only small datasets are available.
Chapter 7 provides the basis for new POT approaches of inference for max-stable processes on general spaces that admit a certain incremental representation. In important cases, this representation has a much simpler structure than the max-stable process itself. The resulting POT methods will incorporate all single events, for which a fixed component is extreme, i.e. conditional distributions of the form $[(Z_0, \ldots, Z_k) \,|\, Z_0 > u]$ will be considered. The results of Chapter 7 are a contribution to the foundations of inference for max-stable processes; further development of these concepts, in particular towards the MPP framework, are left for future research work.

Except for the introductory part in Chapter 2 and the entire Chapter 4, all chapters are based on manuscripts that are either submitted to or in revision for peer-reviewed journals. In fact, Chapter 2 starts with some definitions and basic results from (marked) point process theory. The second part of Chapter 2 is based on Appendix A in Malinowski & Schlather (2011*b*) and introduces the conditioning on the existence of other points of the MPP, which is the basis for the measures of interaction considered in this thesis. The manuscript Malinowski

& Schlather (2011*b*), which is currently in revision for the *Journal of Financial Econometrics*, provides the basis for Chapter 3. The manuscript is in turn a condensed, enriched, and completely rewritten version of Chapter 4 of the diploma thesis Malinowski (2009). Chapter 4 contains a collection of models, for which the afore introduced measures of interaction can be treated analytically. Parts of this chapter (the additive marking case in Section 4.2.2 and Section 4.2.3) are already contained in the diploma thesis Malinowski (2009) in a slightly different form. Chapter 5 studies more general summary statistics for possibly non-ergodic MPPs on $\mathbb{R}^d$. This chapter is based on the manuscript Malinowski *et al.* (2012*a*), which is currently under review in *Advances in Applied Probability*. Chapter 6 combines MPP theory with aspects from extreme value analysis and is based on the manuscript Malinowski *et al.* (2012*b*), which has been submitted to *Journal of Business & Economic Statistics*. Chapter 7 is based on the manuscript Engelke *et al.* (2012*b*), which has been submitted to *Advances in Applied Probability*. The results in this chapter apply to a family of max-stable processes that admit a certain incremental representation. The manuscript additionally contains similar results for the class of mixed moving maxima processes, which are not part of this thesis as they were contributed by the co-authors.

In order to allow for a selective reading of the single chapters, they are coherent but not constitutive, except for Chapter 2, which is required for Chapters 3 and 4. Chapter 6 is an application of Chapter 5.

Within the framework of this PhD project, two further sub-projects have been carried out, one of which can be seen as the statistical counterpart of Chapter 7 for the class of Brown-Resnick processes (Brown & Resnick, 1977; Kabluchko *et al.*, 2009). This was joint work with Sebastian Engelke, a fellow PhD student, and resulted in the manuscript Engelke *et al.* (2012*a*), which has been submitted to the *Journal of the Royal Statistical Society, Series B* and in which the first author had the major contribution.

The other sub-project was deeply situated in the area of genetics and dealt with the temporal development of stochastic dependence between different SNP-positions on the DNA, modeled via Markov chains. This was joint work with Ulrike Ober, also a fellow PhD student, and resulted in the manuscript Ober *et al.* (2012), which is at the moment being prepared for submission.

A further publication is Malinowski & Schlather (2011*a*), which reviews the book Gentleman (2008).

2 Marked point processes: Preliminaries and measures of interaction

In the following, some of the basic definitions and results from marked point process (MPP) theory needed for the subsequent chapters are briefly reviewed. A survey of the theory of (marked) point processes can be found in Daley & Vere-Jones (2003, 2008) or Stoyan *et al.* (1995).

2.1 Definition and properties of marked point processes

Definition 2.1.1 (Point processes on general spaces). A *point process* Φ on a complete separable metric space (c.s.m.s.) $\mathcal{X}$ with Borel-σ-algebra $\mathcal{B}(\mathcal{X})$ is a measurable mapping from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into $(\mathbb{M}_0(\mathcal{X}), \mathcal{M}_0(\mathcal{X}))$, where $\mathbb{M}_0(\mathcal{X})$ is the space of all locally finite counting measures on $\mathcal{X}$ and $\mathcal{M}_0(\mathcal{X})$ is the smallest σ-algebra with respect to which all mappings $\varphi \mapsto \varphi(A)$, $A \in \mathcal{B}(\mathcal{X})$, are measurable. The induced probability measure is given by $\mathbb{P}^{\Phi}(Y) = \mathbb{P}(\Phi \in Y)$, $Y \in \mathcal{M}_0(\mathcal{X})$. The process Φ is called *simple*, if almost all realizations φ of Φ satisfy $\varphi(\{x\}) \in \{0, 1\} \; \forall x \in \mathcal{X}$.

Definition 2.1.2 (Moment measure). For a point process Φ on $\mathcal{X}$, we define a measure Λ by $\Lambda(A) = \mathbb{E}\Phi(A)$, $A \in \mathcal{B}(\mathcal{X})$. If Λ is locally finite, we call it *intensity measure* or *first moment measure* of Φ.

If the intensity measure of Φ exists, by the usual argument of algebraic induction, we get the most basic version of the class of *Campbell theorems*: $\mathbb{E}\sum_{x \in \Phi} f(x) = \int_{\mathcal{X}} f(x)\Lambda(dx)$ for $f : \mathcal{X} \to \mathbb{R}$ being measurable and non-negative or Λ-integrable. Summation over "$x \in \Phi$" actually means integration w.r.t. the random measure Φ.

A useful generalization of the first moment measure is the Campbell measure, which we introduce in the following. With a slight abuse of notation we will also write $\mathbb{P}$ for the induced probability measure $\mathbb{P}^{\Phi}$. We define set functions $C_{\mathbb{P}}$ and $C_{\mathbb{P}}^{!}$ through

$$C_{\mathbb{P}}(A \times Y) = \int_{\mathbb{M}_0(\mathcal{X})} \int_{\mathcal{X}} \mathbf{1}_A(x)\mathbf{1}_Y(\varphi)\varphi(dx)\mathbb{P}(d\varphi) = \mathbb{E}(\Phi(A)\mathbf{1}_Y(\Phi)),$$

$$C_{\mathbb{P}}^{!}(A \times Y) = \int_{\mathbb{M}_0(\mathcal{X})} \int_{\mathcal{X}} \mathbf{1}_A(x)\mathbf{1}_Y(\varphi - \delta_x)\varphi(dx)\mathbb{P}(d\varphi),$$

$A \in \mathcal{B}(\mathcal{X})$, $Y \in \mathcal{M}_0(\mathcal{X})$, and where δ_x denotes the Dirac measure at point x. It can be shown that both set functions extend uniquely to σ-finite measures on the product-σ-algebra $\mathcal{B}(\mathcal{X}) \otimes \mathcal{M}_0(\mathcal{X})$.

Definition 2.1.3 (Campbell measures). The unique extensions of $C_{\mathbb{P}}$ and $C_{\mathbb{P}}^!$ to σ-finite measures on $\mathcal{B}(\mathcal{X}) \otimes \mathcal{M}_0(\mathcal{X})$ are called *Campbell measure* and *reduced Campbell measure*, respectively.

Since for each fixed $Y \in \mathcal{M}_0(\mathcal{X})$, the measures defined by $A \mapsto C_{\mathbb{P}}(A \times Y)$ and $A \mapsto C_{\mathbb{P}}^!(A \times Y)$, $A \in \mathcal{B}(\mathcal{X})$, are absolutely continuous w.r.t. the intensity measure Λ of Φ, there exist kernels $P = \{P_x(Y) : x \in \mathcal{X}, Y \in \mathcal{M}_0(\mathcal{X})\}$ and $P^! = \{P_x^!(Y) : x \in \mathcal{X}, Y \in \mathcal{M}_0(\mathcal{X})\}$, respectively, such that

$$C_{\mathbb{P}}(A \times Y) = \int_A P_x(Y) \Lambda(dx), \tag{2.1}$$

$$C_{\mathbb{P}}^!(A \times Y) = \int_A P_x^!(Y) \Lambda(dx) \tag{2.2}$$

for all $A \in \mathcal{B}(\mathcal{X})$, $Y \in \mathcal{M}_0(\mathcal{X})$. The measures P_x and $P_x^!$ are often referred to as *(reduced) Palm distribution at the point x*.

Definition 2.1.4 (Poisson point process). A point process Φ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is called *Poisson process* if there exists a locally finite measure $\Lambda : \mathcal{B}(\mathcal{X}) \to [0, \infty]$ such that for every finite family of disjoint bounded sets $A_i \in \mathcal{B}(\mathcal{X})$, $i = 1 \ldots, k$,

$$\mathbb{P}(\Phi(A_i) = n_i, \, i = 1, \ldots, k) = \prod_{i=1}^{k} \frac{\Lambda(A_i)^{n_i}}{n_i!} \exp(-\Lambda(A_i)).$$

The measure Λ is called the *parameter measure* of Φ.

It is easily seen that the parameter measure of a Poisson process coincides with the intensity measure.

Lemma 2.1.5. *Let Φ be a Poisson point process on $\mathcal{X}$ with intensity measure Λ and reduced Campbell measure $C_{\mathbb{P}}^! : \mathcal{B}(\mathcal{X}) \otimes \mathcal{M}_0(\mathcal{X}) \to [0, \infty]$. Then $C_{\mathbb{P}}^!$ can be decomposed into $C_{\mathbb{P}}^! = \Lambda \times \mathbb{P}^{\Phi}$.*

Proof. By definition of the reduced Palm distribution $P_x^!(\cdot)$, it is $C_{\mathbb{P}}^!(A \times Y) = \int_A P_x^!(Y) \Lambda(dx)$ for all $A \in \mathcal{B}(\mathcal{X}), Y \in \mathcal{M}_0(\mathcal{X})$. Then the assertion is a direct consequence of Slivnyak's theorem (e.g., Daley & Vere-Jones, 2008, Prop. 13.1.VII) stating that P_x is the convolution of $\mathbb{P}^{\Phi}$ and δ_x, which is equivalent to $P_x^! = \mathbb{P}^{\Phi}$. $\qquad\square$

Definition 2.1.6 (Doubly Stochastic Poisson Process (DSPP)). Let Ξ be a random measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ whose realizations are a.s. locally finite and countably additive. Then a point process Φ is a *Doubly Stochastic Poisson Process*, also called *Cox process*, directed by Ξ, when, conditional on Ξ, realizations of Φ are those of a Poisson process with intensity measure Ξ. We write $\Phi = \Phi_{\Xi}$.
If Ξ is induced by a Gaussian random field $(Z(x), x \in \mathcal{X})$ via $\Xi(A) = \int_A \exp(Z(x)) \, dx$, $A \in \mathcal{B}(\mathcal{X})$, we call Φ_{Ξ} a *log Gaussian Cox process* (LGCP) and also write $\Phi_{\Xi} = \Phi_{\exp(Z)}$.

Definition 2.1.7 (Marked point process)**.** A *marked point process (MPP) Φ with locations on $\mathbb{R}^d$ and marks in a c.s.m.s.* $\mathbb{M}^*$ is a point process on $\mathbb{R}^d \times \mathbb{M}^*$ with the additional property that the *ground process Φ_{g}*, defined by $\Phi_{\mathrm{g}}(B) = \Phi(B \times \mathbb{M}^*), B \in \mathcal{B}^d$, itself is a simple point process on $\mathbb{R}^d$. Let $\mathcal{M}^*$ be the Borel-σ-algebra of $\mathbb{M}^*$. The MPP Φ is called *stationary* and *isotropic* if the probability law $\mathbb{P}^\Phi$ is invariant under translations of the location component and under rotations of the location component about the origin of $\mathbb{R}^d$, respectively.

Let Φ be an MPP on $\mathbb{R}^d$ in the following and let $\mathcal{B}^d$ be the Borel-σ-algebra of $\mathbb{R}^d$, $d \in \mathbb{N}$. The mark of a point t of an MPP is denoted by $y(t)$. Besides the measure-theoretic notation $\Phi = \sum_{i \in \mathbb{Z}} \delta_{(t_i, y(t_i))}$, we also use the notation $\Phi = \{(t_i, y(t_i)) : i \in \mathbb{Z}\}$, which rather reflects a set-theoretic perspective. The intensity measure Λ is now a function on $\mathcal{B}^d \otimes \mathcal{M}^*$.

Similarly to the definition of the Palm distribution in (2.1) and (2.2), the measures on $\mathcal{B}^d$, defined by $A \mapsto \Lambda(A \times L), L \in \mathcal{M}^*$, are absolutely continuous w.r.t. the intensity measure Λ_{g} of the ground process. Hence, there exists a kernel $\{M_t : \mathcal{M}^* \to [0, \infty], \ t \in \mathbb{R}^d\}$, such that

$$\Lambda(A \times L) = \int_A M_t(L) \Lambda_{\mathrm{g}}(dt), \quad A \in \mathcal{B}^d, L \in \mathcal{M}^*. \tag{2.3}$$

The measure M_t is called (Palm) mark distribution at location t.
If Φ is stationary, we have that $\Lambda(\cdot \times L) = \lambda_L \nu(\cdot)$ for some $\lambda_L \geq 0$, $L \in \mathcal{M}^*$, and that $M(L) = M_t(L) = \lambda_L/\lambda_{\mathbb{M}^*}$ is independent of t. Then Λ can be decomposed into $\Lambda = \Lambda_{\mathrm{g}} \times M = \lambda_{\mathrm{g}} \nu \times M$, where $\lambda_{\mathrm{g}} = \lambda_{\mathbb{M}^*}$. We call $M(\cdot)$ the *mark distribution*. For the stationary case, Campbell's theorem then reads as

$$\lambda_{\mathrm{g}} \int_{\mathbb{R}^d} \int_{\mathbb{M}^*} f(t, y) \, M(dy) \nu(dt) = \mathbb{E} \sum_{(t, y(t)) \in \Phi} f(t, y(t)) \tag{2.4}$$

for every measurable function $f : \mathbb{R}^d \times \mathbb{M}^* \to \mathbb{R}$ that is either non-negative or Λ-integrable.

2.2 Measures of mark-location interaction

In the following, only MPPs on $\mathbb{R}^d$ are considered.

Definition 2.2.1 (Factorial moment measures, cf. Stoyan & Stoyan (1994); Schlather (2001))**.** Let Φ be an MPP with locations on $\mathbb{R}^d$ and let f be a measurable, non-negative function on $(\mathbb{M}^*)^2$. Then the measure $\alpha_f^{(2)}$ on $\mathbb{R}^d \times \mathbb{R}^d$, defined by

$$\alpha_f^{(2)}(C) = \mathbb{E} \sum_{(t_1, y_1), (t_2, y_2) \in \Phi}^{\neq} f(y_1, y_2) \mathbf{1}_C((t_1, t_2)), \quad C \in \mathcal{B}^{2d}, \tag{2.5}$$

is called *generalized second-order factorial moment measure of Φ*. Here, the "$\neq$" over the summation sign means that the sum runs over all pairs $(t_1, y_1), (t_2, y_2)$ with $t_1 \neq t_2$.

We assume Φ to be stationary from now on. In the following, we generalize the above definition w.r.t. the non-negativity assumption on f. Let $W \in \mathcal{B}^d$ be a bounded window and

f a real-valued function on $(\mathbb{M}^*)^2$ such that at least one of the two functions $f_+ = \max\{0, f\}$ and $f_- = \max\{0, -f\}$ satisfies the condition

$$\mathbb{E} \sum_{\substack{\neq \\ (t_1, y_1), (t_2, y_2) \in \Phi}} f_\pm(y_1, y_2) \mathbf{1}_{[\mathbf{0},\mathbf{1}]}(t_1) \mathbf{1}_W(t_2 - t_1) < \infty. \tag{2.6}$$

Then, using the notation

$$\begin{aligned} C(I) &= \{(t_1, t_2) : t_1 \in [\mathbf{0}, \mathbf{1}], t_2 \in t_1 + I\}, \\ C_W(I) &= C(I \cap W), \end{aligned}$$

for $I, W \in \mathcal{B}^d$,

$$\alpha_f^{(2)}(C_W(\,\cdot\,))$$

is a signed measure on $\mathbb{R}^d$. Note that the intersection with W is introduced for technical reasons: We can easily construct examples of stationary MPPs for which both $-\infty$ and $+\infty$ occur as values of $\alpha_f^{(2)}(C(\,\cdot\,))$, while, if f is suitably behaved, $\alpha_f^{(2)}(C_W(\,\cdot\,))$ has values in either $(-\infty, \infty]$ or $[-\infty, \infty)$.

Obviously, $\alpha_f^{(2)}(C_W(\,\cdot\,))$ is absolutely continuous w.r.t. $\alpha_1^{(2)}(C_W(\,\cdot\,))$. Thus, we may define

$$\mu_f^{(2)}(r) = \frac{\partial \alpha_f^{(2)}(C_W(\,\cdot\,))}{\partial \alpha_1^{(2)}(C_W(\,\cdot\,))}(r), \quad r \in W \backslash \{\mathbf{0}\}.$$

As a Radon-Nikodym derivative, $\mu_f^{(2)}$ is only uniquely determined $\alpha_1^{(2)}(C_W(\,\cdot\,))$-almost everywhere. Since $\{\mathbf{0}\}$ is a $\alpha_1^{(2)}(C_W(\,\cdot\,))$-null set for any $W \in \mathcal{B}^d$, we define

$$\mu_f^{(2)}(\mathbf{0}) = \frac{\mathbb{E} \sum_{(t_1, y_1) \in \Phi} f(y_1, y_1) \mathbf{1}_{[\mathbf{0},\mathbf{1}]}(t_1)}{\mathbb{E} \sum_{(t_1, y_1) \in \Phi} \mathbf{1}_{[\mathbf{0},\mathbf{1}]}(t_1)},$$

which is the expectation of f w.r.t. the ordinary Palm mark distribution M, i.e., $\mu_f^{(2)}(\mathbf{0}) = \int_{\mathbb{M}^*} f(y, y)\, M(dy)$.

Remark 2.2.2. *(a) We can also express $\mu_f^{(2)}(\mathbf{0})$ in terms of the* non-factorial *second-order moment measure $\tilde{\alpha}_f^{(2)}$, which is defined analogously to* (2.5) *but without the $\neq$ over the summation sign. Then, $\mu_f^{(2)}(\mathbf{0}) = \tilde{\alpha}_f^{(2)}(C(\{\mathbf{0}\}))/\tilde{\alpha}_1^{(2)}(C(\{\mathbf{0}\}))$.*

(b) Due to stationarity of Φ, we can also write $\mu_f^{(2)}(r)$ as a two-dimensional derivative, evaluated at an arbitrary tuple of locations with distance r:

$$\mu_f^{(2)}(r) = \frac{\partial \alpha_f^{(2)}(\bullet)}{\partial \alpha_1^{(2)}(\bullet)}\bigg|_{\bullet = (t_1, t_1 + r)}$$

for any $t_1 \in \mathbb{R}^d$. If the densities of $\alpha_f^{(2)}(\cdot)$ w.r.t. the two-dimensional Lebesgue measure exist, we denote them by $\rho_f^{(2)}(\cdot, \cdot)$ and get

$$\mu_f^{(2)}(r) = \frac{\rho_f^{(2)}(t_1, t_1 + r)}{\rho^{(2)}(t_1, t_1 + r)}$$

for any $t_1 \in \mathbb{R}^d$.

In several MPP models, the mark of a point at location t can (conditionally on some random driving mechanisms) be expressed as a deterministic function of t and of the process of point locations (up to time t). To provide a unifying notation, we introduce the following extension of $\alpha_f^{(2)}$: For any $\mathcal{B}^d \otimes \mathcal{M}_0(\mathbb{R}^d)$-measurable function $g : \mathbb{R}^d \times \mathbb{M}_0(\mathbb{R}^d) \to \mathbb{R}$, we define a set function $\alpha_{f,g}^{(2)}$ on $\mathcal{B}^{2d}$ by

$$\alpha_{f,g}^{(2)}(C) = \mathbb{E} \sum_{(t_1, y_1), (t_2, y_2) \in \Phi}^{\neq} f(y_1, y_2) \mathbf{1}_C((t_1, t_2)) g(t_1, \Phi_{\mathrm{g}} - \delta_{t_1}), \quad C \in \mathcal{B}^{2d}.$$

The following lemma provides a representation of $\alpha_{f,g}^{(2)}$ in terms of Palm distributions.

Lemma 2.2.3. *Let Φ be a possibly non-stationary MPP on $\mathbb{R}^d$ with distribution $\mathbb{P}$ and intensity measure Λ. Let $f : (\mathbb{M}^*)^2 \to \mathbb{R}$ and $g : \mathbb{R}^d \times \mathbb{M}_0(\mathbb{R}^d) \to \mathbb{R}$ be functions such that the function $h_C : \mathbb{R}^d \times \mathbb{M}^* \times \mathcal{M}_0 \to \mathbb{R}$,*

$$h_C(t_1, y_1, \varphi) = \int_{\mathbb{R}^d \times \mathbb{M}^*} \mathbf{1}_C((t_1, t_2)) f(y_1, y_2) g(t_1, \varphi_{\mathrm{g}}) \, \varphi(dt_2 \times dy_2),$$

is $C_{\mathbb{P}}^!$-integrable for every $C \in \mathcal{B}^{2d}$. Then, for $C \in \mathcal{B}^{2d}$,

$$\alpha_{f,g}^{(2)}(C) = \int_{\mathbb{R}^d} \int_{\mathbb{M}^*} \int_{\mathbb{M}_0} \int_{\mathbb{R}^d \times \mathbb{M}^*} \mathbf{1}_C((t_1, t_2)) f(y_1, y_2) g(t_1, \varphi_{\mathrm{g}}) \tag{2.7}$$
$$\varphi(dt_2 \times dy_2) P_{(t_1, y_1)}^!(d\varphi) M_{t_1}(dy_1) \Lambda_{\mathrm{g}}(dt_1),$$

where M_t is the (Palm) mark distribution at position t as defined in (2.3).

Proof. By definition, we have

$$\alpha_{f,g}^{(2)}(C)$$
$$= \int_{\mathbb{M}_0} \int_{\mathbb{R}^d \times \mathbb{M}^*} \int_{\mathbb{R}^d \times \mathbb{M}^*} \mathbf{1}_C((t_1, t_2)) f(y_1, y_2) g(t_1, \varphi_{\mathrm{g}} - \delta_{t_1})$$
$$(\varphi - \delta_{(t_1, y_1)})(dt_2 \times dy_2) \varphi(dt_1 \times dy_1) \mathbb{P}^\Phi(d\varphi)$$
$$= \int_{\mathbb{M}_0} \int_{\mathbb{R}^d \times \mathbb{M}^*} h_C(t_1, y_1, \varphi - \delta_{(t_1, y_1)}) \, \varphi(dt_1 \times dy_1) \mathbb{P}^\Phi(d\varphi).$$

According to (2.2) and (2.3), for $B \in \mathcal{B}^d$, $L \in \mathcal{M}^*$, $Y \in \mathcal{M}_0$, we have

$$C_{\mathbb{P}}^!(B \times L \times Y) = \int_B \int_L P_{(t,y)}^!(Y) M_t(dy) \Lambda_{\mathrm{g}}(dt),$$

and, by algebraic induction,

$$\int_{\mathbb{M}_0} \int_{\mathbb{R}^d \times \mathbb{M}^*} h(t, y, \varphi - \delta_{(t,y)}) \varphi(dt \times dy) \mathbb{P}^{\Phi}(d\varphi) \tag{2.8}$$

$$= \int_{\mathbb{R}^d \times \mathbb{M}^* \times \mathbb{M}_0} h(t, y, \varphi) C_{\mathbb{P}}^!(dt \times dy \times d\varphi)$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{M}^*} \int_{\mathbb{M}_0} h(t, y, \varphi) P_{(t,y)}^!(d\varphi) M_t(dy) \Lambda_{\mathrm{g}}(dt)$$

for any $C_{\mathbb{P}}^!$-integrable function h. Applying this with $h = h_C$ we get

$$\alpha_{f,g}^{(2)}(C) = \int_{\mathbb{R}^d} \int_{\mathbb{M}^*} \int_{\mathbb{M}_0} h_C(t_1, y_1, \varphi) \, P_{(t_1,y_1)}^!(d\varphi) M_{t_1}(dy_1) \Lambda_{\mathrm{g}}(dt_1),$$

which completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 2.2.4. *The function g will have different meanings. If chosen to be a certain indicator function, it introduces an additional conditioning in the corresponding second-order mean mark. It can also take over the role of $f(y_1, y_2)$ if the mark at a point t is given as a deterministic function of Φ_{g} and t.*

As a direct consequence of Lemma 2.2.3 we have

Corollary 2.2.5. *Let $B_1, B_2 \in \mathcal{B}^d$.*

1. *If the function f satisfies $f(y_1, y_2) = \tilde{f}(y_1)$ for all $y_1, y_2 \in \mathbb{M}^*$ for some function $\tilde{f}$ (which we again denote by f), i.e. f is independent of its second argument, then (2.7) can be written as*

$$\alpha_{f,g}^{(2)}(B_1 \times B_2) = \int_{\mathbb{R}^d} \int_{\mathbb{M}^*} \int_{\mathbb{M}_0} \mathbf{1}_{B_1}(t_1) f(y_1) g(t_1, \varphi_{\mathrm{g}}) \varphi_{\mathrm{g}}(B_2)$$
$$P_{(t_1,y_1)}^!(d\varphi) M_{t_1}(dy_1) \Lambda_{\mathrm{g}}(dt_1).$$

2. *If, additionally, y_1 is given as a deterministic function y of t_1 and $\varphi_{\mathrm{g}} - \delta_{t_1}$, i.e. $y_1 = y(t_1, \varphi_{\mathrm{g}} - \delta_{t_1})$ for $(t_1, y_1) \in \mathrm{supp}(\varphi)$, then*

$$\alpha_f^{(2)}(B_1 \times B_2) = \alpha_{f,1}^{(2)}(B_1 \times B_2)$$
$$= \alpha_{1,f\circ y}^{(2)}(B_1 \times B_2)$$
$$= \int_{\mathbb{R}^d} \int_{\mathbb{M}^*} \int_{\mathbb{M}_0} \mathbf{1}_{B_1}(t_1) f(y(t_1, \varphi_{\mathrm{g}})) \varphi_{\mathrm{g}}(B_2) P_{(t_1,y_1)}^!(d\varphi) M_{t_1}(dy_1) \Lambda_{\mathrm{g}}(dt_1)$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{M}_0} \mathbf{1}_{B_1}(t_1) f(y(t_1, \varphi_{\mathrm{g}})) \varphi_{\mathrm{g}}(B_2) P_{t_1}^!(d\varphi_{\mathrm{g}}) \Lambda_{\mathrm{g}}(dt_1),$$

where $P_t^!$ is the Palm distribution of Φ_{g}. In this case, the calculation of $\alpha_f^{(2)}$ is reduced to the theory of unmarked point processes.

3. If, additionally, Φ_{g} is a Cox process with random intensity measure $\Xi(\cdot) = \int L(t) dt$, then

$$\alpha_f^{(2)}(B_1 \times B_2) = \int \int_{\mathbb{R}^d} \int_{\mathbb{M}_0} \mathbf{1}_{B_1}(t_1) f(y(t_1, \varphi_{\mathrm{g}})) \varphi_{\mathrm{g}}(B_2) \mathbb{P}^{\Phi_{\mathrm{g}} \mid L = \lambda}(d\varphi_{\mathrm{g}}) \lambda(t_1) dt_1 \mathbb{P}^L(d\lambda).$$

Notation: The value of $\mu_f^{(2)}(r)$, $r \in \mathbb{R}^d$, represents the mean value of $f(y_1, y_2)$ over all pairs of points $(t_1, y_1), (t_2, y_2)$ with $t_2 - t_1 = r$. This motivates using the notation

$$\mu_f^{(2)}(r) = \mathbb{E}\big[f(y(t), y(t+r)) \mid t, t+r \in \Phi_{\mathrm{g}} \big].$$

In the following, we consider an MPP Φ on the real axis with real-valued marks. As regards the function f, the two functions $e(y_1, y_2) = y_1$ and $v(y_1, y_2) = y_1^2$ are employed.

Definition 2.2.6 (Bi-directional second-order statistics). Let Φ be a stationary MPP on $\mathbb{R}$ with real-valued marks and $W \in \mathcal{B}$ such that (2.6) is satisfied for $f = e$ and $f = v$. Using the above notation, we define

$$E(r) = \mu_e^{(2)}(r), \ r \in \mathbb{R}, \qquad \textit{the bi-directional E-function,}$$
$$\text{and} \quad V(r) = \mu_v^{(2)}(r) - (\mu_e^{(2)}(r))^2, \ r \in \mathbb{R}, \quad \textit{the bi-directional V-function.}$$

Note that, in general, $E(r)$ and $V(r)$ are not continuous at $r = 0$, where the two-point statistics E and V pass into one-point statistics. Similarly to Schlather (2001), $E(r)$ and $V(r)$ can be interpreted as the conditional mean and variance of the mark of a point at an arbitrary time t, given the existence of another point r units of time away, $r \in \mathbb{R}$. Thus, a negative value of r refers to the existence of another point in the past. $E(0)$ and $V(0)$ are simply the unconditional mean and variance of a mark, respectively.

2.3 Estimation of conditional mean marks

We assume that the Lebesgue density $\rho_f^{(2)}(r)$ of $\alpha_f^{(2)}(C(\cdot))$ exists for $r \neq 0$. Then, it is common to apply a ratio estimator for $\mu_f^{(2)}(r)$ of the form $\hat{\mu}_f^{(2)}(r) = \hat{\rho}_f^{(2)}(r)/\hat{\rho}_1^{(2)}(r)$ (Stoyan & Stoyan, 2000), where

$$\hat{\rho}_f^{(2)}(r) = \frac{1}{\nu(L)} \sum_{(t_1, y_1), (t_2, y_2) \in \Phi}^{\neq} f(y_1, y_2) \mathbf{1}_L(t_1) K_h((t_2 - t_1) - r)$$

for $r \neq 0$, a kernel function K_h with bandwidth h, and some observation window L of the point process. Note that we abstain from an edge correction (e.g., Stoyan *et al.*, 1995) as the considered values of r are negligibly small, compared to the size of L. It can be shown that this estimator is asymptotically unbiased, that is, if the bandwidth h converges to zero (cf. Stoyan & Stoyan, 2000). Therein, it is also suggested to use the rectangular kernel instead of the Epanechnikov kernel because of a smaller estimation variance. Furthermore, we replace the bandwidth h by an adaptive bandwidth $h_r = \min\{h, |r|\}$ for the following reason: By using the bi-directional statistics $\mu_f^{(2)}$, we explicitly want to take account of the impact of the sign of the distance. So, it would not be reasonable to use a tuple of points with a negative distance for estimation of $\hat{\rho}_f^{(2)}(r)$ for a positive r, and vice versa. For $r = 0$, $\mu_f^{(2)}$ becomes a one-point statistic and we apply $\hat{\rho}_f^{(2)}(0) = \sum_{(t_1, y_1) \in \Phi} f(y_1, y_1) \mathbf{1}_L(t_1) / \nu(L)$. We end up with the following estimator of $\mu_f^{(2)}$:

$$
\hat{\mu}_f^{(2)}(r) = \frac{\hat{\rho}_f^{(2)}(r)}{\hat{\rho}_1^{(2)}(r)}
$$

$$
= \begin{cases}
\dfrac{\displaystyle\sum_{(t_1, y_1) \in \Phi} f(y_1, y_1) \mathbf{1}_L(t_1)}{\displaystyle\sum_{(t_1, y_1) \in \Phi} \mathbf{1}_L(t_1)}, & r = 0 \\[3em]
\dfrac{\displaystyle\sum_{\substack{(t_1, y_1),\, (t_2, y_2) \\ \in \Phi \cap (L \times \mathbb{R})}} f(y_1, y_2) \mathbf{1}_{(r - h_r,\, r + h_r)}(t_2 - t_1)}{\displaystyle\sum_{\substack{(t_1, y_1),\, (t_2, y_2) \\ \in \Phi \cap (L \times \mathbb{R})}} \mathbf{1}_{(r - h_r,\, r + h_r)}(t_2 - t_1)}, & r \neq 0.
\end{cases}
$$

3 Refined analysis of interactions within high-frequency transaction data through marked point process theory

This chapter is based on the manuscript Malinowski & Schlather (2011*b*).

3.1 Introduction

In the classical context of low-frequency data, asset prices are usually modeled as geometric Brownian motions or, more generally, as (jump) diffusion processes, that is, solutions to stochastic differential equations, with a possibly time-varying and random underlying volatility. Barndorff-Nielsen & Shephard (2001), for example, propose a sophisticated stochastic volatility model whose underlying volatility is given by a Lévy driven Ornstein-Uhlenbeck process. While inhomogeneity of volatility in those models is often seen to be caused by the flow of new information, the focus of this chapter is on volatility effects on a high-frequency scale caused by temporal proximity of past or future trades.

Excited by the seminal work of Engle & Russell (1998) on modeling financial data at its highest level of disaggregation as marked point processes (MPPs), a plenty of MPP models for high-frequency data have been developed in recent econometric literature (see Bauwens & Hautsch (2009) for a broad survey). As financial transactions occur irregularly spaced in time, a standard procedure in this setting is to consider time stamps as the points of a point process, marked by the according (log) prices or other characteristics. MPPs turned out to be a well-suited tool for modeling temporal (and spatial) dependencies between marks as well as interactions between marks and locations of the points.

Two main classes of MPP models for transaction data are commonly used: The class of dynamic duration and autoregressive conditional duration (ACD) models (e.g., Ghysels & Jasiak, 1998; Engle, 2000; Bauwens & Giot, 2001; Hautsch, 2004), see Engle & Russell (2009) or Pacurar (2008) for a survey, and the broad class of models based on doubly stochastic Poisson processes (DSPP) or Cox processes, where the underlying intensity can also be specified dynamically (e.g., Hawkes, 1971; Russell, 1999; Hautsch, 2004; Bauwens & Hautsch, 2006; Rydberg & Shephard, 2000; Frey, 2000; Centanni & Minozzo, 2006).

The main contribution of this chapter is the analysis of new statistics for temporal MPPs, namely the *bi-directional E- and V-function*. These functions can be interpreted as conditional expectations and variances, respectively. While the definition here is tailored to the temporal context, a slightly different definition is already known from MPPs in the field of spatial statistics (Schlather, 2001). The new statistics will be used to detect and to model interaction phenomena within high-frequency financial data and, based on this method, we

will discuss an extension of the commonly used ultra-high frequency GARCH model of Engle (2000) that includes certain effects of interaction. When applied to the UHF-GARCH model, our statistics serve as useful additional model fitting criteria, besides classical model fitting criteria.

Note that it is a well-known fact that the underlying (time-varying) variance of a price process generally increases in times of high trading intensity (e.g., Easley & O'Hara, 1992). Though, the bi-directional V-function will be able to distinguish between the influence of past and future transactions and therefore allows for a more profound insight into the latent volatility process and into interaction effects.

Our analysis is based on high-frequency transaction data. Due to increased automatization in financial markets and the fast development in computing power and storage capacity, (financial) databases today provide high-frequency data for a wide range of markets. Simultaneously, many econometric tools like ARCH, GARCH and related models have been developed enabling an analysis of the market's behavior at the fine scale of transaction data (e.g., Goodhart & O'Hara (1997), Ghysels & Jasiak (1998), Engle & Russell (1998, 2009), Engle (2000), Zhang *et al.* (2001), Racicot *et al.* (2008), and the references therein). Other approaches for modeling transaction price processes include, for example, the probit regression model of Hausman *et al.* (1992) or the approach of decomposition of price changes in Rydberg & Shephard (2003).

Compared to classical time-series analysis, high-frequency data pose some specific challenges. The most important feature is that financial transactions are not equally spaced so that the standard theory of time series, which is based on fixed time interval analysis, cannot be applied. One approach is aggregating returns to equally-spaced intervals but such aggregation will either lose information (if the new intervals are large) or create noise due to interpolation issues, or both. Aït-Sahalia *et al.* (2005), for example, show that, when microstructure noise is included into the model, it is reasonable to keep transaction data at their original ultimate frequency level. To avoid the disadvantages of temporal aggregation, various point process methods have been developed that are tailored to the irregular spacing of transaction level data. See Bauwens & Hautsch (2009) for a survey.

Further, inter-trade durations are usually clustered. That means, the autocorrelation function of the durations is significantly positive with a slow decay, which can be associated with long-memory properties of the process. Besides clustering, important properties are the discreteness of the price process and diurnal or periodical patterns, e.g., volatility of prices, traded volume and frequency of transactions exhibit a U-shaped pattern over the course of the day.

All those features can have substantial implications on temporal dependencies and on measuring volatility or other characteristics on a small scale. Especially, sequent inter-transaction returns are not free of correlation as it is often assumed for low-frequency data.

The remainder of this chapter is organized as follows: Section 3.2.1 starts with a detailed example that shows the usefulness and importance of the bi-directional E-function and its relatives. Section 3.2.2 provides an intuitive definition of the second-order statistics E and V as well as appropriate estimators. In Section 3.3, we briefly review the ACD

model and the commonly used UHF-GARCH model for high-frequency financial data that was introduced by Engle (2000) (Subsections 3.3.1 and 3.3.2). We apply our second-order statistics to the model but as analytical treatment is not possible for this kind of autoregressive processes, simulation techniques are employed. By applying the UHF-GARCH model to large transaction datasets of German stock index (DAX) enterprises (Subsection 3.3.3), we can show that the model's representation of interaction effects can be improved by some extension. Section 3.4 summarizes and concludes.

3.2 Novel measure of interaction

3.2.1 Motivation

In order to motivate our new statistics, we consider the autocovariance function that is well defined for any stationary process on a grid in $\mathbb{R}$ or on the whole real axis. In case that locations are neither on a grid nor independent of the values of the marks, we face at least two possibilities to extend the autocovariance function to this situation.

Let $\Phi_{\mathrm{g}} = \{t_i : i \in \mathbb{Z}\}$ be the process of trading times t_i with $t_i \neq t_j$ for $i \neq j$ and let the marks $y(t_i) = \log(p_i/p_{i-1})$ be the inter-transaction log returns, where p_i is the execution price of the ith transaction. Let $\Phi = \{(t_i, y(t_i)) : i \in \mathbb{Z}\}$ denote the resulting marked point process. We assume that Φ is stationary and that all of the following quantities exist. Note that, since points are irregularly spaced, we always have to condition on the existence of an event at time t when we consider the corresponding mark. For instance, the overall mean $\bar{y}$ of the marks is given as $\bar{y} = \mathbb{E}[y(t) \,|\, t \in \Phi_{\mathrm{g}}]$ for any fixed $t \in \mathbb{R}$.

Then the first possibility to generalize the autocovariance function is to consider

$$C(r) = \mathbb{E}\Big[(y(0) - \bar{y})(y(r) - \bar{y}) \,\Big|\, 0, r \in \Phi_{\mathrm{g}}\Big].$$

A second possibility is to remember the genuine definition of the autocovariance function, that is, we consider the covariance of two random variables, namely the marks $y(0)$ and $y(r)$, conditionally on the existence of points of Φ_{g} at 0 and r:

$$\mathrm{Cov}(r) = \mathbb{E}\Big[(y(0) - E(r)) \cdot (y(r) - E(-r)) \,\Big|\, 0, r \in \Phi_{\mathrm{g}}\Big],$$

where $E(r) = \mathbb{E}[y(0) \,|\, 0, r \in \Phi_{\mathrm{g}}]$.

In order to highlight the difference between these two quantities C and Cov, we consider the following simple process, where the inter-transaction durations $x_i = t_i - t_{i-1}$ and the marks are given as follows:

$$x_i = \frac{\mu}{x_{i-1}}\varepsilon_i, \qquad\qquad \varepsilon_i \sim_{\mathrm{i.i.d.}} \mathrm{Exp}, \quad \mathbb{E}\varepsilon_i = \mu, \qquad (3.1)$$

$$y(t_i) = x_i + c(\mathbf{1}_{\{B_i=1\}\cup\{B_{i-1}=1\}}), \qquad B_i \sim_{\mathrm{i.i.d.}} \mathrm{Bernoulli}(p).$$

This specification induces the durations to be negatively correlated at lag 1 and the mark process to contain few pairs of neighboring huge values—at least if we choose $p \ll 1$ and

$c \gg 1$. We will consider the particular choice $\mu = 1$ [minute], $p = 0.001$ and $c = 67$. Figure 3.1 shows a realization of this process.

By construction of the above process, marks with a fixed and sufficiently small distance are positively correlated, as the paired outliers overcompensate the duration's negative autocorrelation if c is sufficiently big. However, the quantity $C(r)$ turns out not to be able to discover this property but rather leads to a wrong conclusion (Figure 3.2). Although in many other econometric examples there might not be such a big difference between the two functions C and Cov, our simple example points out that it is important to choose the proper definition of conditional covariance and that the conditional mean function $E(r)$ therefore plays an important role in the analysis of marked point processes.

A very natural generalization of the conditional mean function is obtained by applying
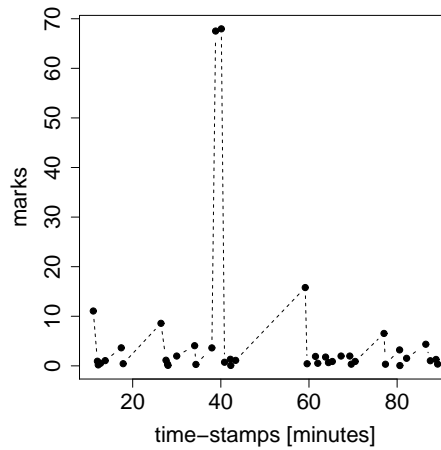

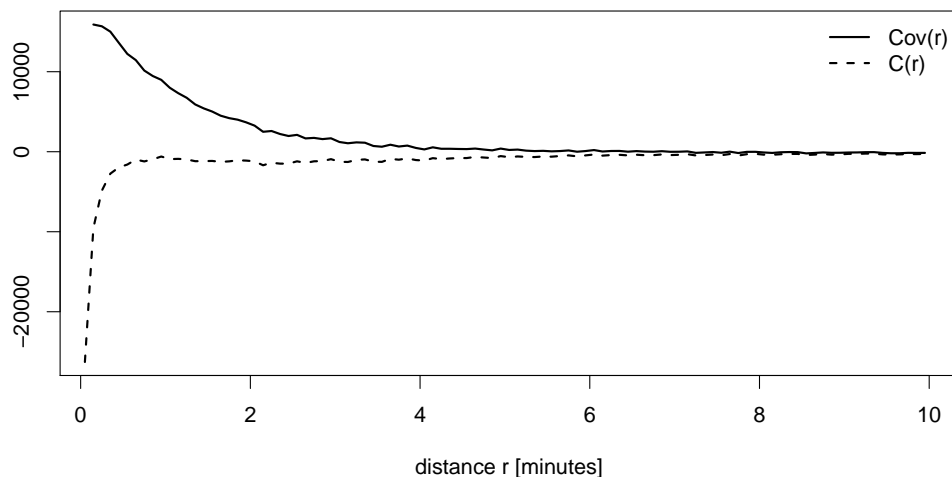**Figure 3.1:** Extract of a realization of the process in (3.1) with one pair of large values.

Figure 3.2: Comparison of the functions $C(r)$ and $\mathrm{Cov}(r)$ for the model given by (3.1).

higher moments or by conditioning on more than two points:

$$\mathbb{E}\left[\prod_{i=1}^{k} y(r_i) \,\middle|\, r_1, \ldots, r_k, r_{k+1}, \ldots, r_n \in \Phi_{\mathrm{g}}\right], \quad n > k. \tag{3.2}$$

These conditional moments can be employed for testing whether the process under consideration belongs to the so-called class of *random field models*, where the marks of the MPP are given by a realization of a random field being independent of the location of the points. In case of a Gaussian random field, the set of moment functions up to order $k = 2$ allows for necessary and sufficient conditions to characterize independence between marks and locations, hence also for tests on independence (Schlather *et al.*, 2004). This underlines the relevance of the above functions.

The conditional functions also serve as additional model fitting criteria: If they are not constant, the data should not be represented by models that do not include interactions between marks and locations.

Finally, they might be helpful for prediction when information about temporal locations of transactions is available.

3.2.2 E- and V-function

While a measure-theoretical deduction of the new statistic has already been given in Chapter 2, we give a more intuitive definition at this point; using the notation and the assumptions of the previous Subsection 3.2.1.

Definition 3.2.1. Let $\Phi = \{(t_i, y(t_i)) : i \in \mathbb{Z}\}$ be a stationary and simple MPP, which particularly includes stationarity of the mark distribution and the mark's moments. Then the *bi-directional E- and V-function* are defined as the conditional expectation and variance, respectively, of a trade's mark at time t, given the existence of a trade at time $t + r$:

$$E(r) = \mathbb{E}[y(t) \,|\, t, t + r \in \Phi_{\mathrm{g}}],$$
$$V(r) = \mathbb{E}[(y(t) - E(r))^2 \,|\, t, t + r \in \Phi_{\mathrm{g}}], \quad r \in \mathbb{R}.$$

Due to the stationarity of Φ, $E(r)$ and $V(r)$ do not depend on the value of $t \in \mathbb{R}$. Note that only the existence of a transaction at instance $t + r$ enters into the definition, but not the value of $y(t + r)$. Our definition explicitly utilizes the total order on the set of real numbers and therefore is a generalization of the corresponding second-order statistics in Schlather (2001), where isotropy is a central assumption and E and V are functions of the Euclidean distance between two points.

The two functions, E and V, are sensible tools for our purpose of analyzing interaction effects but they cover at least two different effects: Due to the clustering of transaction arrival times and the fact that volatility and trading intensity are positively correlated, we expect to observe a symmetrically increasing bi-directional V-function for small *absolute* values of r. Indeed, as we will see later, the V-function of real data exhibits an asymmetric behavior and thus covers additional effects at a fine temporal scale.

3.2.3 Estimation

We use the estimators proposed in Section 2.3. In particular, let $L \subset \mathbb{R}$ be an observation interval and let f stand for one of the functions $e(y) = y$ and $v(y) = y^2$. Let us denote the points of the marked point process within the interval L by (t_i, y_i), $i = 1, \ldots, N$, and let $t_i \neq t_j$ for $i \neq j$. Then we take

$$\frac{\sum_{1 \leq i,j \leq N} f(y_i)\mathbf{1}_{(r-h_r, r+h_r)}(t_j - t_i)}{\sum_{1 \leq i,j \leq N} \mathbf{1}_{(r-h_r, r+h_r)}(t_j - t_i)}$$

if $r \neq 0$ and $\frac{1}{N}\sum_{i=1}^{N} f(y_i)$ if $r = 0$ as an estimator for $E(r)$ and $V(r) + E(r)^2$, if $f = e$ and $f = v$, respectively. Here, $h_r = \min\{h, |r|\}$ is an adaptive bandwidth of the rectangular kernel function. Due to the massive dataset, edge correction (e.g., Stoyan *et al.*, 1995) is not needed.

3.3 GARCH models for high-frequency financial data

Engle (1982) introduced ARCH processes and therewith initiated a whole theory of modeling and forecasting time-varying financial market volatilities. Originally only applied to equally spaced data on a daily or larger scale, they were soon extended to more flexible GARCH models, even allowing for high-frequency data that are not equally spaced.

After having briefly reviewed the two parts of the UHF-GARCH model of Engle (2000), we apply the newly-defined E- and V-functions to this model.

3.3.1 The Autoregressive Conditional Duration model

Engle's UHF-GARCH model is based on the Autoregressive Conditional Duration (ACD) model of Engle & Russell (1998) that explicitly models the irregular spacing of transactions.

Let $x_i = t_i - t_{i-1}$, $i \in \mathbb{Z}$, be the duration between the $(i-1)^{\text{th}}$ and the ith transaction and let Ψ_i be the expectation of the ith duration conditional on all past durations, that is, $\Psi_i = \mathbb{E}(x_i \,|\, \mathcal{F}_{i-1}^x)$, where $\mathcal{F}_{i-1}^x = \sigma(x_j, j \leq i-1)$ is the natural filtration of the duration process. Then the ACD model consists of a parametrization of Ψ and the assumption that $x_i = \Psi_i \varepsilon_i$ with $\varepsilon_i \sim_{\text{i.i.d.}} F_\theta$ for some distribution function F_θ such that $\mathbb{E}\varepsilon_i = 1$. In the simplest form of an ACD model, the ε_i follow an exponential distribution and $\Psi_i = \omega^{\text{D}} + \alpha^{\text{D}} x_{i-1} + \beta^{\text{D}} \Psi_{i-1}$ with the parameters $\omega^{\text{D}}, \alpha^{\text{D}}, \beta^{\text{D}} \in \mathbb{R}$; this is called EACD(1,1). The more general form is the ACD(p,q) model (see Engle & Russell, 1998). In order to guarantee non-negativity of the durations, restrictions have to be imposed on the parameters (Nelson & Cao, 1992). Parameters can be estimated by quasi maximum likelihood methods (QMLE) by factorizing the joint density function of the observed durations into a product of univariate conditional densities. Moreover, the model can be estimated using standard GARCH estimation procedures by taking $\sqrt{x_i}$ as the dependent variable. We refer to Engle & Russell (1998) for more details.

3.3.2 The UHF-GARCH model and volatility modeling

In the following, the marks of the ACD point process are the inter-transaction log returns $r_i = \log(p_i/p_{i-1})$. For modeling the returns, Engle (2000) applies a GARCH specification, extended by additional variables. In particular, Engle (2000) incorporates the irregular spacing of transaction data by considering the returns divided by the square root of the corresponding durations. The conditional variance per transaction is defined by $h_i = \mathrm{Var}(r_i \,|\, \mathcal{F}_{i-1}^r \vee \mathcal{F}_i^x)$. As volatility is usually measured over fixed time intervals, Engle defines a volatility per time unit by taking $\sigma_i^2 = \mathrm{Var}(r_i/\sqrt{x_i} \,|\, \mathcal{F}_{i-1}^r \vee \mathcal{F}_i^x)$, which leads to the equation $h_i = x_i \sigma_i^2$. Then the standardized returns $y_i = r_i/\sqrt{x_i}$ are modeled as a GARCH(1,1) process whereby further explaining variables are added to the variance equation in order to represent economic effects:

$$\sigma_i^2 = \omega + \alpha\varepsilon_{i-1}^2 + \beta\sigma_{i-1}^2 + \gamma_1/x_i + \gamma_2 x_i/\Psi_i + \gamma_3\zeta_i + \gamma_4/\Psi_i, \tag{3.3}$$

where $\alpha, \beta, \gamma_j \in \mathbb{R}$, Ψ_i is the conditional expectation of the ith duration in the ACD model and ζ_i is a kind of long run volatility of the returns.

In the following, in order to keep ML estimation simple and to ease comparison of parameter estimates between different datasets, we use a simpler model of Engle (2000)—see Equation (39) therein—, that is, $\gamma_2 = \gamma_3 = \gamma_4 = 0$:

$$\begin{aligned}
y_i &= \mu_i + \varepsilon_i, \tag{3.4}\\
\mu_i &= p_1 y_{i-1} + p_2 \varepsilon_{i-1} + p_3 x_i,\\
\varepsilon_i &= \sigma_i z_i \quad \text{with } z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1),\\
\sigma_i^2 &= \omega + \alpha\varepsilon_{i-1}^2 + \beta\sigma_{i-1}^2 + \gamma/x_i
\end{aligned}$$

with the real-valued parameters $p_1, p_2, p_3, \omega, \alpha, \beta$ and γ.

Obviously, the conditional mean of y_i, given all past returns and all durations up to x_i, is $\mathbb{E}(y_i \,|\, \mathcal{F}_{i-1}^y \vee \mathcal{F}_i^x) = \mu_i$ and the conditional variance is given by $\mathrm{Var}(y_i \,|\, \mathcal{F}_{i-1}^y \vee \mathcal{F}_i^x) = \sigma_i^2$. The following propositions deal with the unconditional mean and variance. The proofs only involve elementary arguments and can be found in Malinowski (2009).

Proposition 3.3.1. *Let μ_i be as in (3.4) with $0 < p_1 < 1$. Then the unconditional mean of μ_i is*

$$\mathbb{E}\mu_i = \frac{1}{1-p_1} p_3 \mathbb{E}x_1$$

and the unconditional variance of μ_i is

$$\mathrm{Var}(\mu_i) = \sum_{n=0}^{\infty} p_1^n \sum_{k=0}^{n} p_3^2 C_x(|n - 2k|) + \frac{1}{1-p_1^2}(p_2 + p_1)^2 \mathbb{E}\varepsilon_1^2,$$

where $C_x(l)$ denotes the covariance of the duration at lag l.

Proposition 3.3.2. *The unconditional variance of ε_i in Model (3.4) with $0 < \alpha + \beta < 1$ and $\mathbb{E}|x_1^{-1}| < \infty$ is*

$$\mathrm{Var}(\varepsilon_i) = \mathbb{E}\varepsilon_i^2 = \left(\omega + \gamma\mathbb{E}x_1^{-1}\right)\frac{1}{1-(\alpha+\beta)}.$$

Combining these two results and the fact that z_i is stochastically independent of $\mu_i\sigma_i$, we can conclude that the unconditional variance of the returns depends linearly on the parameter vector (ω, γ, p_3^2), which will be useful in Section 3.3.3:

Corollary 3.3.3. *The unconditional variance of $y_i = \mu_i + \varepsilon_i$ is*

$$\mathrm{Var}(y_i) = \left(1 + \frac{1}{1-p_1^2}(p_2+p_1)^2\right)\left(\omega + \gamma\mathbb{E}x_1^{-1}\right)\frac{1}{1-(\alpha+\beta)} + \sum_{n=0}^{\infty}p_1^n\sum_{k=0}^{n}p_3^2 C_x(|n-2k|).$$

Similarly to Section 3.3.1, the parameter vector $\theta = (p_1, p_2, p_3, \omega, \alpha, \beta, \gamma)$ of the UHF-GARCH model can be estimated by QMLE methods using the log likelihood

$$\ell(\theta) = -\sum_{i=1}^{n}\left(\log(\sigma_i(\theta)) + \frac{(y_i - \mu_i(\theta))^2}{2\sigma_i(\theta)^2}\right) - \frac{n}{2}\log(2\pi). \tag{3.5}$$

We should emphasize at this point that the UHF-GARCH model is only one approach amongst others to describe financial market volatilities. Especially in terms of prediction, *realized volatility* (e.g., Andersen *et al.*, 2001; Barndorff-Nielsen & Shephard, 2002; Martens, 2002) is an important measure of variance and in recent econometric literature, different generalizations thereof (e.g., power variation) and related issues like microstructure noise have been investigated intensively (e.g., Barndorff-Nielsen, 2004; Hansen & Lunde, 2006; Bandi & Russell, 2008; Andersen *et al.*, 2011). Bollerslev & Wright (2001) was one of the first papers that used high-frequency data to obtain non-parametric volatility forecasts, based on realized volatility, that outperform certain parametric GARCH models. Racicot *et al.* (2008) adjusted the approach therein for the irregular spacing of data and provided a detailed comparison in terms of forecasting performance between different high-frequency variance models. Their results are in line with previous studies and indicate that GARCH models can have poor forecasting performance in an UHF context. However, in this chapter, our main focus is not on forecasting properties and we limit ourselves to the UHF-GARCH model.

Remark 3.3.4. *Note that the UHF-GARCH model in this subsection is indeed a stationary and simple MPP under some mild conditions: Fernandes & Grammig (2006), for example, provide the two conditions $|\beta^D| < 1$ and $\mathbb{E}|\beta^D + \alpha^D\varepsilon_i|^m < 1$ for some integer $m > 1$ that guarantee strict stationarity of the sequence of durations in the ACD(1,1) model and hence also stationarity of the unmarked point process Φ_g. Nelson (1990) gives conditions for strict stationarity of the classical GARCH(1,1) process that can be directly adapted to the UHF-GARCH(1,1) case. Accordingly, the assumptions of Proposition 3.3.2 and strict stationarity of the durations are sufficient conditions for strict stationarity of the sequence of conditional variances $\{\sigma_i\}_{i\in\mathbb{Z}}$. Stationarity of the series $\{\mu_i\}_{i\in\mathbb{Z}}$ follows by similar arguments and*

therewith strict stationarity of the series of returns. Combining those facts yields stationarity of the UHF-GARCH model in the marked point process sense.

3.3.3 Application of the ACD-UHF-GARCH model to data

In order to avoid results to be driven by irregularities in the data and to achieve consistency to the results of Engle (2000), we apply the same data pre-processing procedure which particularly includes an adjustment for diurnal patterns in duration and return series.

Engle's (2000) assessment of the estimated model is essentially done by a comparison of the autocorrelation structure in the original time series with the residuals' autocorrelation structure. To this end, the Ljung-Box test statistic and the partial autocorrelation function are considered.

Due to the immense datasets that are available nowadays, it becomes possible to additionally consider refined statistics as model fitting criteria, such as the E- and V-function.

Siemens data

We apply the model estimation procedure to large transaction datasets from stock trading in Germany, processed via the Xetra trading system in the period from 1997 till 2004. Data were provided by the "Karlsruhe capital market database" (KKMDB).

In the following, we present the results of our analysis using the example of only one enterprise from the German stock index DAX, but the qualitative results are valid for many of the other DAX stocks as well. We choose the Siemens stock (ISIN DE0007236101), and in order to avoid long term effects, we always consider a trading period of one year only (January 1st, 2004 to December 31th, 2004). Siemens is one of Europe's largest engineering conglomerate with 405,000 employees worldwide, a total revenue of 75 billion euros per year and an average pre tax income of 4.3 billion euros (averaged over the last 7 years). During that period, the debt ratio increased from 66 percent to 72 percent, the total amount of debt increased from 52 to 74 billion euros. According to traded volume and the Xetra liquidity measure, which is based on implied transaction costs, Siemens belongs to the five most liquid stocks on the German stock market.

For the MLE of the ACD(1,1) parameters, the initial parameter values are $(\omega_0^{\text{D}}, \alpha_0^{\text{D}}, \beta_0^{\text{D}}) = (1, 0.5, 0.5)$. After the first run of the optimization procedure, a second run is performed using the results of the first one as initial values and as scaling factors. By testing different sets of initial values and scaling factors, the optimization procedure turns out to be very robust with respect to bad choices of the starting parameters. The parameter estimates and standard deviations are

$$\omega^{\text{D}} = 0.021 \ (0.00033),$$
$$\alpha^{\text{D}} = 0.11 \ (0.00072),$$
$$\beta^{\text{D}} = 0.87 \ (0.001). \tag{3.6}$$

Analogously, the parameters of the UHF-GARCH model for the returns are estimated by maximizing the likelihood in (3.5), with $\theta_0 = (0.1, 0.1, 0.01, 0.4, 0.4, 0.4, 0.4)$ as initial

parameter vector. The reliability of this MLE is checked by re-estimating the parameters from a simulation according to the MLE results. This leads to satisfactory results. The parameter estimates of the extended GARCH(1,1) model for the returns are given in Table 3.1. The standard errors show that here, p_3 is statistically not distinguishable from 0. However, this is not always the case (cf. Table 3.3).

Analogously to Engle (2000), we compare the autocorrelation structure in pre-processed duration and return series to the autocorrelation in standardized residuals (Figure 3.3). The corresponding Ljung-Box statistics with 15 degrees of freedom are given in Table 3.2.

Furthermore, we compare the E- and V-function of the data to those of the model, which are obtained by simulating a new dataset according to the estimated parameters (Figure 3.4). Note that we only simulate the returns; the durations are taken from the pre-processed dataset. For conditions under which the estimated model satisfies the stationarity requirements in the definition of the E- and V-function, see Remark 3.3.4. We observe an almost symmetric graph of the V-function for the Siemens returns per time unit. Though, $V(-r)$ is slightly larger than $V(r)$ for small positive values of r which means that past transactions have a stronger influence on the actual variance than future transactions.

Recall that the UHF-GARCH model is seen as a model for the "returns per time unit", which are—up to now—defined as "returns per transaction" divided by the square root of the durations. Next, we reverse this transformation by re-multiplying both the simulated and the real "returns per time unit" with the square root of the durations, and also calculate the V-function of the "returns per transaction". (Figure 3.4 (bottom)).

Table 3.1: MLE results for the parameters of the model in (3.4)[a]

	mean equation			variance equation	
p_1	-0.026	(0.0031)	ω	0.17	(0.0016)
p_2	-0.18	(0.0029)	α	0.067	(0.00078)
p_3	$-7.4e-05$	(0.00028)	β	0.0091	(0.00026)
			γ_1	0.51	(0.0013)

[a] Standard errors given in parentheses.

Table 3.2: Comparison of Ljung-Box statistics between Engle's (2000) original model in (3.4) and its extension proposed in the sequel of this subsection.

		original model		extended model	
	duration	return	$(return)^2$	return	$(return)^2$
LB(15) of time series [a]	141943	26823	9622	72234	76994
LB(15) of residuals [b]	4198	2422	24749	1324	1036
reducement [c]	0.03	0.09	2.6	0.018	0.013

[a] LB(15) of Siemens series of pre-processed durations and returns.

[b] LB(15) of standardized residuals of the respective model.

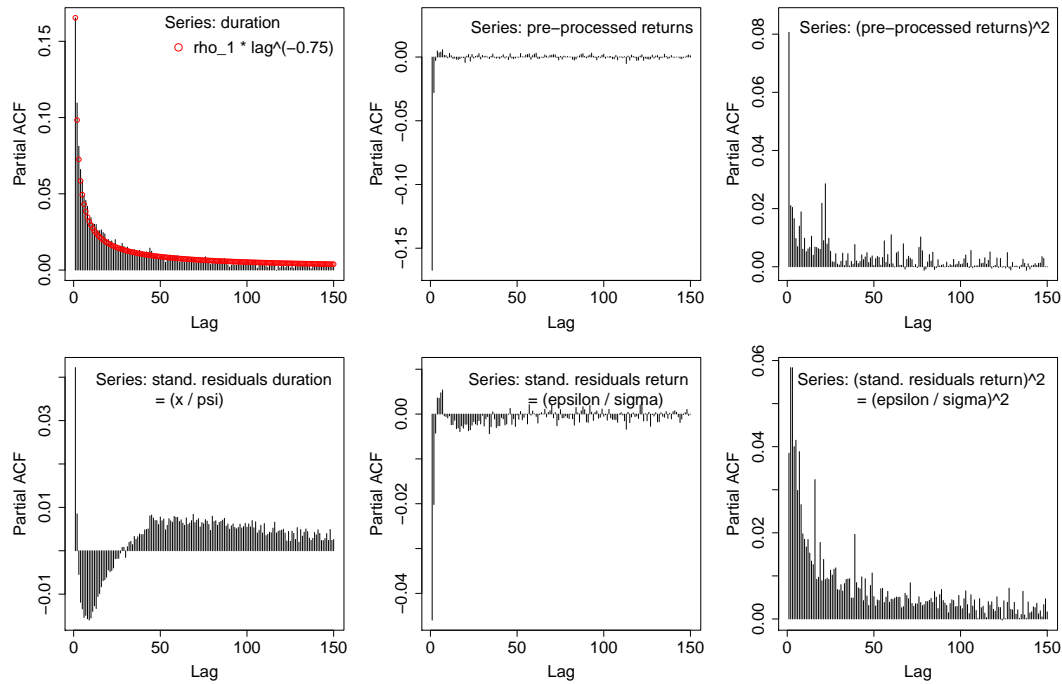[c] Factor by which the statistic is reduced by the model.

Figure 3.3: Top: Partial ACF of the Siemens series of pre-processed durations, returns and squared returns.
Bottom: Partial ACF of the standardized residuals of the model.

Results

Besides the finding that asymmetric interaction effects between inter-trade durations and volatility exist, the first important result of our analysis is that the E-function is constant for all datasets of the DAX. This implies that the knowledge of previous transaction times does not lead to a systematical shift in returns. Henceforth, we restrict to the V-function.

As a conclusion from Figures 3.3 and 3.4 and the part of Table 3.2 that refers to the original model of Engle (2000), we can state: The model reduces autocorrelation of sequent returns by a factor 0.3 but autocorrelation in squared returns is rather increased than decreased. The Ljung-Box statistics of durations and returns are reduced by one to two powers of ten. These results do not fully correspond with the findings in Engle (2000) where a comparatively small IBM dataset is analyzed. Particularly, autocorrelation in squared IBM returns is well represented by the model while autocorrelation in squared Siemens returns is even strengthened. Almost all DAX datasets exhibit the same effect as the Siemens dataset and its strength seems to depend on the parameter in the extended definition of "volatility per unit of time" in the following paragraph.

Furthermore, we conclude that the V-function of the simulated "returns per time unit" agrees with the Siemens data while the V-function of the simulated "returns per transaction" does not fit well to the one of the true returns.
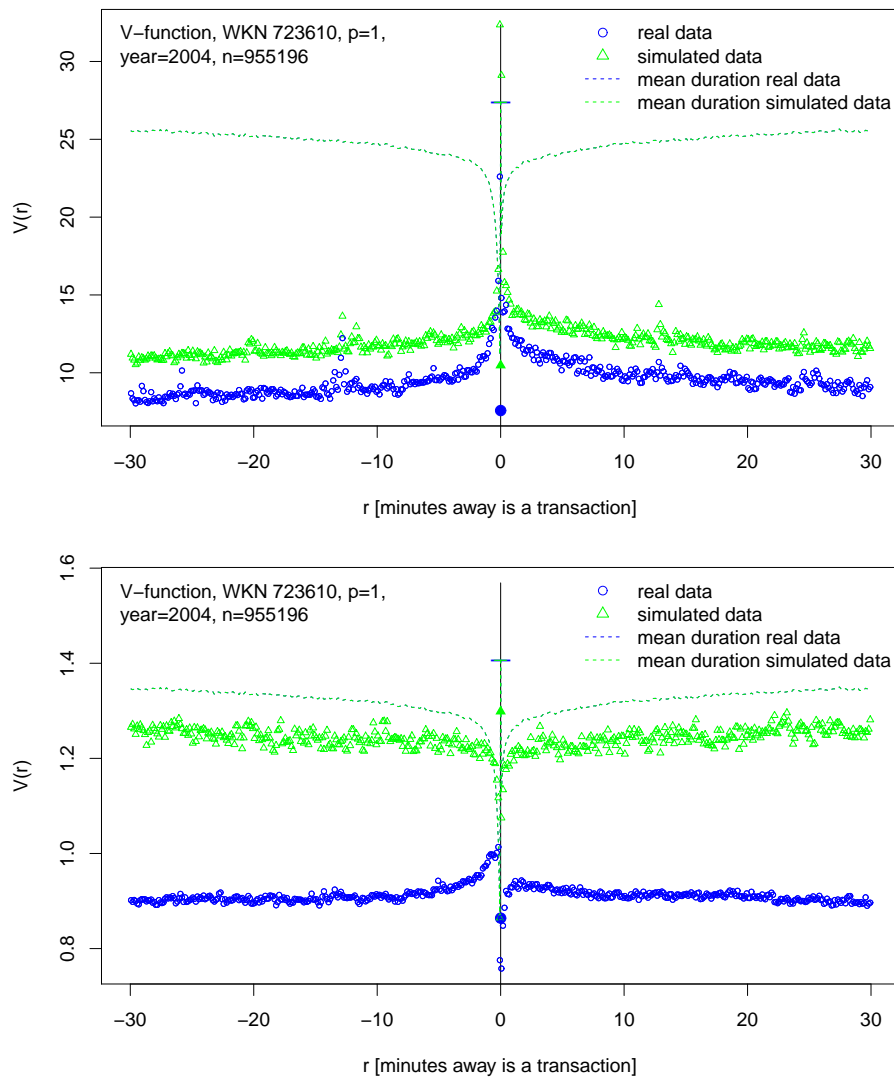
Figure 3.4: V-functions of Siemens and simulated return series.
Top: pre-processed real returns and simulated returns ("per time unit").
Bottom: pre-processed real returns multiplied by pre-processed durations to the power of $1/2$ and simulated returns multiplied by the same factors ("variances per transaction").

The extended model

We try to give an explanation for the above behavior and propose an extension of the model.
In fact, the definition $\sigma_i^2 = \mathrm{Var}(r_i/\sqrt{x_i} \,|\, \mathcal{F}_{i-1}^r \vee \mathcal{F}_i^x)$ of conditional "volatility per time unit" is reminiscent of a Brownian-motion-like price process, where returns over disjoint periods are independent and quadratic variation is proportional to the time t if the underlying volatility is constant. However, Sun & Engle (2007) already inferred from an empirical

analysis that this assumption is not necessarily adequate for high-frequency data. Further, market microstructure theory yields that the observed price does not only represent the efficient price, but also contains some microstructure noise, whose variance does not depend linearly on duration at all. For an introduction to the issues of microstructure noise and volatility measurement at high frequencies, see, for example, Andersen *et al.* (2009) or Bandi & Russell (2006, 2008). Due to microstructure noise, realized volatility of the observed price process, considered as a time-continuous process, does not necessarily converge to the volatility of the efficient price if sampling frequency is increased. In conformity to this, we observe a negative correlation in sequent returns and that volatility per transaction grows less than linearly in duration (Figures 3.3 and 3.5).

One possibility of further investigating this problem is given by Meddahi *et al.* (2006) who consider the discretization of the simple stochastic volatility model $d \log p_t = \sqrt{v_t} dW_t$ at irregularly spaced times, where p_t is the price process, W_t a Brownian motion, and $v_t \geq 0$ a stationary square-integrable process whose conditional mean $\mathbb{E}(v_{t+\Delta} \,|\, \mathcal{F}_t^v)$ is an exponentially fast mean-reverting function of Δ. The authors give an explicit form for the conditional volatility $\mathrm{Var}(r_i \,|\, \mathcal{F}_{i-1}^r \vee \mathcal{F}_i^x \vee \mathcal{F}_{i-1}^v)$, which shows that it is only approximately linear in the duration x_i and only if x_i is very small or very large. However, they consider the series $f_i = x_i^{-1} \mathrm{Var}(r_i \,|\, \mathcal{F}_{i-1}^r \vee \mathcal{F}_i^x \vee \mathcal{F}_{i-1}^v)$ and show that f_i is an autoregressive process with time-varying parameters preserving the natural property $f_i \to \mathbb{E}v_0$ for $x_i \to \infty$. Furthermore, the returns r_i follow a square-root stochastic autoregressive volatility process—a class of processes introduced by Meddahi & Renault (2004) that keep the structure under temporal aggregation.

Here, we follow the idea of Sun & Engle (2007) who suggest a multiplicative model for the "volatility per transaction" where the duration does not enter linearly but to the power of some p. The parameter p is interpreted as a relative speed of information arrivals. Moreover, the authors observe that the value of p depends on how liquidly a stock is traded and they give an explanation related to the issue of informed and uninformed trading.
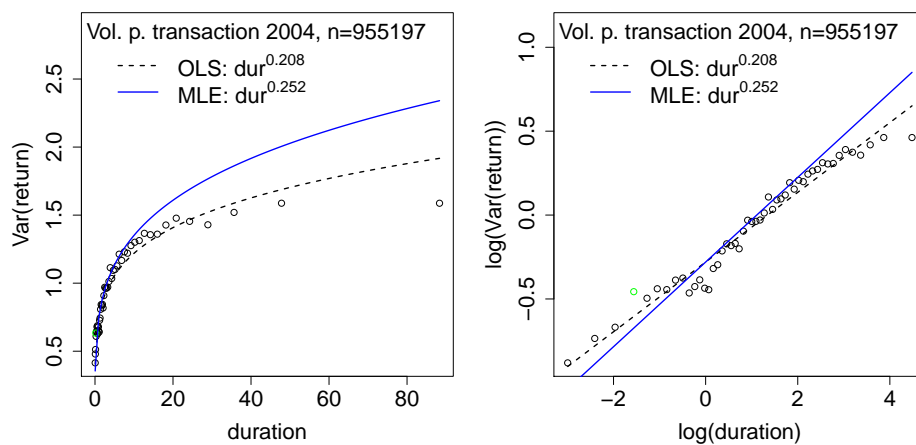

Figure 3.5: Regression of volatility per transaction on duration compared with the ML estimate of the exponent p.

We adopt this idea and give an extended definition of conditional "volatility per time unit" that incorporates the above microstructure effects:

$$\sigma_i^2 = \text{Var}(r_i/x_i^{p/2} \,|\, \mathcal{F}_{i-1}^r \vee \mathcal{F}_i^x)$$

with some unknown parameter $p \in [0, 1]$.

Similarly, we extend the variance equation of the UHF-GARCH model replacing in (3.4) the term γ/x_i by γ_2/x_i^q, where q is an additional parameter that introduces more flexibility in modeling the influence of durations on the volatility process. That is, we consider the model

$$\sigma_i^2 = \omega + \alpha\varepsilon_{i-1}^2 + \beta\sigma_{i-1}^2 + \gamma_2/x_i^q. \tag{3.7}$$

At this point, one could suppose that the conditional volatility σ_i^2 does not depend on the duration any more after having divided the returns by the durations to the power of $p/2$. However, this is not necessarily the case since this scaling only accounts for the conditional variance $\text{Var}(r_i \,|\, x_i)$ so that the term x_i^{-q} can still have a non-zero coefficient in the autoregressive variance model. Indeed, the parameter estimates for γ and γ_2 are always significantly positive for both Engle's IBM data and for our DAX data. Engle (2000) understands this as a support of the Easley & O'Hara (1992) hypothesis that long durations are an indicator for the absence of new information and therefore for a lower volatility of the price process. Regardless of what is the true mechanism behind this effect, we simply understand this as interaction between the point pattern and the corresponding returns. Note that without the terms γ/x_i and γ_2/x_i^q in (3.4) and (3.7), respectively, the V-function of the model would be a mere constant.

Reanalysis of Siemens data through the extended model

Conditionally on the model parameters $\theta = (p_1, p_2, p_3, \omega, \alpha, \beta, \gamma_2, q, p)$ and all durations x_i, the "return per transaction" r_i has a Gaussian distribution with mean $\mu_i(\theta)x_i^{p/2}$ and variance $\sigma_i^2(\theta)x_i^p$. The log likelihood (3.5) then becomes

$$\ell(\theta) = -\sum_{i=1}^n \left(\log(\sigma_i(\theta)x_i^{p/2}) + \frac{\left(r_i/x_i^{p/2} - \mu_i(\theta)\right)^2}{2\sigma_i^2(\theta)} \right) - \frac{n}{2}\log(2\pi) \tag{3.8}$$

and all parameters including p and q can be identified by QMLE.

In perfect accordance with the idea of Sun & Engle (2007) and the proposed model extension, it turns out that $\hat{p}^{\text{OLS}}$, a linear regression estimate of the logarithmized equation of the simple model $v_i = sx_i^p$, is a very good approximation of $\hat{p}^{\text{ML}}$ and hence a reasonable starting value for the ML procedure. Here, $v_i = \text{Var}(r_i \,|\, x_i)$ denotes the conditional "volatility per transaction". In particular, to obtain $\hat{p}^{\text{OLS}}$, transactions are grouped by their respective durations x and within each group j, $j = 1, \ldots, m$, the volatility $v(j)$ of the respective returns is calculated. Let $x(j)$ be a representative duration of group j, then the model $v(j) = sx^p(j)$ is fitted by a linear regression for the logarithmized equation $\log(v(j)) = \log(s) + p\log(x(j))$;

see Figure 3.5.

The MLE results for the extended model are given in Table 3.3. Note that we have allowed ω to become negative, still ensuring a positive value on the right-hand side (RHS) of the Variance Equation (3.7).

The autocorrelation within return series and standardized residuals is illustrated by Figure 3.6, the corresponding Ljung-Box statistics are given by Table 3.2. Figure 3.7 shows the comparison of V-functions.

The extended model ($p \neq 1$) explains autocorrelations within the modified return series fairly well. The major improvement in comparison to the case $p = 1$ is the relatively good fit of the V-function of the simulated "returns per transaction". Although the slope of the curves near the origin does not fit perfectly to the true data yet, at least the level of variance and the basic characteristics are the same.

Readjustment of model parameters for parallel shifts of the V-function

As can be seen from the V-function plot in Figures 3.4 and 3.7, the parameter estimates of the UHF-GARCH model may lead to a V-function which has the same appearance as the V-function of the real data, but which does not hit the actual level of variance.

However, in order to hit the actual V-function level and to ease comparison of the behavior of the model's and the data's V-function, we can slightly modify the estimated parameters
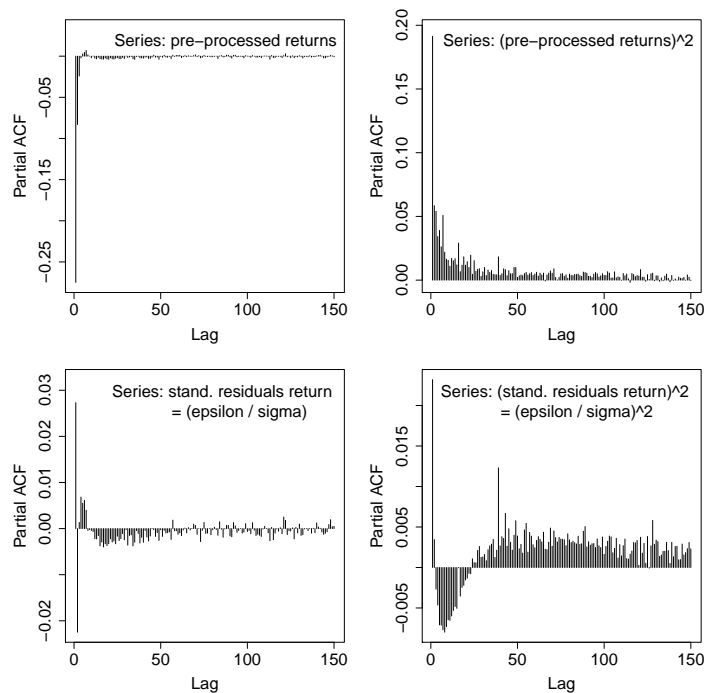

Figure 3.6: Top: Partial ACF of Siemens series after modified pre-processing ($p = 0.252$ instead of $p = 1$).
Bottom: Partial ACF of the standardized residuals of the model.

Table 3.3: MLE results for the parameters of the extended model[a]

	mean equation			variance equation	
p_1	-0.13	(0.0036)	ω	0.088	(0.0015)
p_2	-0.2	(0.0036)	α	0.09	(0.00075)
p_3	-0.0092	(0.00067)	β	0.87	(0.0012)
	exponent parameter		γ_2	0.013	(0.00021)
p	0.252	(0.00062)	q	0.65	(0.0022)

[a] Standard errors given in parentheses.

in such a manner that only the vertical location of the V-function is affected while the appearance is kept fix: As can be seen from Corollary 3.3.3, the parameter triple (ω, γ, p_3^2) enters linearly into the unconditional variance of the UHF-GARCH model. That means, if we use $(c\omega, c\gamma, \sqrt{c}p_3)$ instead of (ω, γ, p_3) with a positive shifting factor c, the level of variance will also be multiplied by the factor c. We choose c to be the minimizer of the squared distance between simulated and empirical V-function. For the DAX datasets, c mostly lies between 0.7 and 1.1.
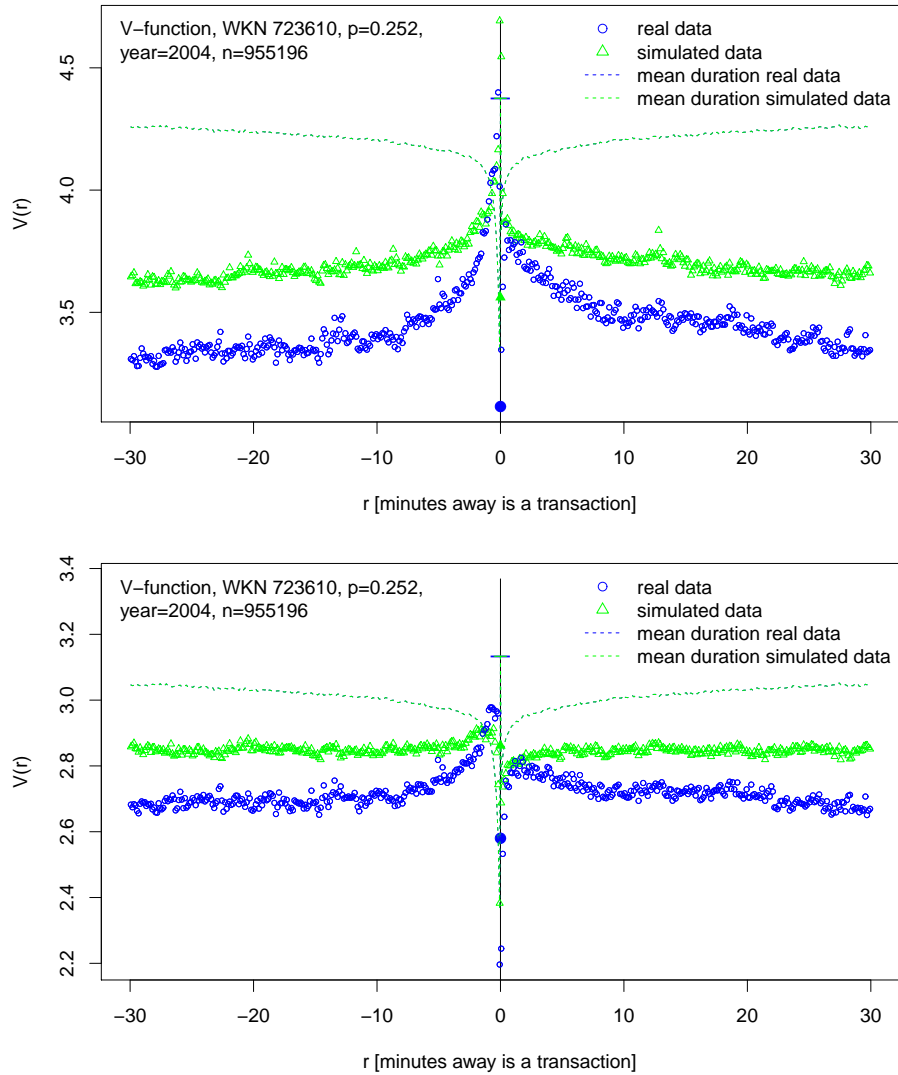
Figure 3.7: V-functions of Siemens and simulated return series, $p = 0.252$.
Top: pre-processed real returns and simulated returns ("per time unit").
Bottom: pre-processed real returns multiplied by pre-processed durations to the power of $p/2$ and simulated returns multiplied by the same factors ("variances per transaction").

3.4 Discussion

Leveraging the total order of $\mathbb{R}$, we have introduced novel second-order statistics for marked point processes on $\mathbb{R}$, in particular the bi-directional E- and V-function, in order to quantify interaction effects between the pattern of transaction times and the price process; so far basically in terms of conditional variances, but the approach can easily be extended to higher moments and to conditioning on two or more transaction times.

We have applied the methods to the ACD-UHF-GARCH model of Engle (2000). Motivated by implications from microstructure noise and by the results of Sun & Engle (2007), a modified method of tackling the irregular spacing of financial transactions is proposed. Here, we exclude the ACD model for the durations and only consider the GARCH model for the returns, but it should be an issue of future research to also include the ACD part into the analysis of E- and V-function. In particular, we add two new exponent parameters to the model (p and q). Especially p can be estimated very robustly by ML and for 80% of the datasets under consideration, $\hat{p}$ is between 0.16 and 0.27.

Engle (2000) used an IBM dataset that contained about 100 times fewer transactions than the DAX datasets employed in our analysis. While model estimation is also feasible with small datasets, the empirical V-function of such a dataset is dominated by noise and the impact of our extension w.r.t. to the newly-defined statistics is invisible. The reason for this is that the estimator of $V(r)$ only uses the set of pairs of points with a distance of approximately r, which is usually a very small set compared to the overall number of transactions. Though, large datasets reveal that both V-function and autocorrelations are better represented by the extended model.

There are further open issues for future research work: For instance, the V-function approach offers miscellaneous possibilities for quantifying the investigated effects, and Engle's GARCH model itself provides several variations with further explaining variables within the variance equation that shall represent economic effects. Moreover, the role of the autocovariance function Cov mentioned in Section 3.2.1 has to be investigated.

Although point processes are widely-used in high-frequency econometrics and it is well-known in literature that the underlying volatility of the price process tends to be higher in times of high trading intensity, there is only few literature that relates techniques from spatial statistics, where general theory for MPPs has been developed, to financial data. Interactions between durations and returns within high-frequency financial data are still not comprehensively understood. Nevertheless, this chapter provides some promising tools and techniques for tackling this issue and for including it into the framework of volatility modeling at an intradaily scale.

4 Second-order moment measures in different marked point process models

4.1 Introduction

In this chapter, the theoretical values of the bi-directional E- and V-function, introduced and applied in Chapters 2 and 3, are determined for different MPP models. Although not being directly used as models for transaction data in practice, some of these models are well-established in environmental applications and can be thought of as ingredients for sophisticated models in econometric applications. The models allow for a non-reversible behavior of the MPP.

One of the models being investigated with regard to mark-location interaction phenomena is based on the idea of intensity-dependent marking of log Gaussian Cox processes (LGCP). Myllymäki (2009) extensively discusses this issue and combines two sources of covariance between the marks: the intensity-dependency of marks and the spatial dependency of marks (known from geostatistical marking) that cannot be explained by fluctuation of intensity.

A more general class of models is received by replacing the underlying Gaussian random field by an arbitrary random intensity measure. While LGCP models with their advantageous theoretical properties allow for an explicit calculation of second-order characteristics, other specifications such as the generalized shot noise intensity discussed in Centanni & Minozzo (2006), do not. Therein, the authors replace the intensity-dependent marking by a marking based on inter-point distances. Although inter-point distances are directly connected to the intensity of points and the marking is therefore still intensity-dependent in a sense, effectively, the specification of marks is an autoregressive one.

The outline of this chapter is as follows: Section 4.2.1 considers Cox processes with a marking based on shot-noise processes. In Section 4.2.2, a generalization of the model in Myllymäki (2009) is presented by replacing the univariate intensity-generating field by a bivariate Gaussian field: one of the two components drives the intensity and the other component is used to construct the marks; then the random field's covariance structure directly translates into interaction between locations and marks. In the model in Section 4.2.3, the marks are generated by a function of the distance to previous points of the MPP, supplemented by some noise, which is a generalized version of the model in Centanni & Minozzo (2006). Section 4.2.4 is concerned with a model that describes the trading process at the order book level. Finally, Section 4.3 provides a construction principle for MPPs to realize a large subclass of all valid E- and V-functions.

4.2 Analytical treatment of E- and V-function in MPP models based on Poisson and Cox processes

We consider several types of marked point processes on the real axis, which are all based on Poisson and Cox processes but which differ in their specification of the marks. Via a random and non-constant intensity, the models are able to represent clustering effects in the data. The marks are related to the process of point locations or to its intensity process via different mechanisms.

4.2.1 Model 1: Cox processes marked by a random field with underlying shot noise volatility

Let $\tilde{\Phi}$ be an unmarked Cox process with random intensity L. In particular, we assume that L is a second-order stationary log-Gaussian random field, i.e. $L = \exp(Z)$ with $\{Z(t)\}_{t\in\mathbb{R}}$ being a Gaussian field with mean μ, covariance function C and $\sigma^2 = C(0)$. Let $\{W(t)\}_{t\in\mathbb{R}}$ and $\{U(t)\}_{t\in\mathbb{R}}$ be random fields of iid variables with $W \geq 0$, $\mathbb{E}U(0) = 0$ and $\operatorname{Var} U(0) = 1$. Conditionally on $\tilde{\Phi} = \tilde{\varphi}$, $W = w$ and $U = u$, let the mark y_i of a point t_i be given by

$$y_i = y(t_i,\tilde{\varphi}) = u(t_i)\sqrt{\int_{t<t_i} w(t)h(t_i - t)\tilde{\varphi}(dt)},$$

where $h(\cdot)$ is an integrable function on $\mathbb{R}^+$. Then we consider the MPP

$$\Phi = \{(t, y(t,\tilde{\Phi})) : t \in \tilde{\Phi}\}, \tag{4.1}$$

i.e., the ground process of point locations is $\Phi_{\mathrm{g}} = \tilde{\Phi}$. Note that, conditionally on Φ_{g}, the marks are independent random variables. However, their variances depend on the past of the process and are given by a shot noise process.

Remark 4.2.1. *For an MPP as in ([4.1](#)), the E-function as well as all functions $\mu_f^{(2)}$ with $f(y_1,y_2) = y_1^{(2n-1)}$, $n \in \mathbb{N}$, are constant and zero due to the moment condition on U and the independence between Φ_{g} and U.*

Theorem 4.2.2. *For an MPP as in ([4.1](#)), the V-function is given by*

$$V(r) = \mathbb{E}W \cdot \left(\mathbf{1}_{r<0}h(-r) + \int_0^\infty h(t)\exp(\mu + \tfrac{\sigma^2}{2} + C(t) + C(r+t))\,dt\right), \quad r \in \mathbb{R}.$$

Proof. Applying $f(y_1,y_2) = v(y_1,y_2) = y_1^2$ and conditioning on L, W and U, we are in the situation of Corollary [2.2.5](#), part [3](#), i.e.,

$$\alpha_v^{(2)}(B_1 \times B_2)$$
$$= \iiint \int_{B_1} \int_{\mathbb{M}_0} v(y(t_1,\varphi_{\mathrm{g}}))\varphi_{\mathrm{g}}(B_2)\mathbb{P}^{\Phi_{\mathrm{g}}\,|\,L=\lambda}(d\varphi_{\mathrm{g}})\lambda(t_1)dt_1\mathbb{P}^L(d\lambda)\mathbb{P}^W(dw)\mathbb{P}^U(du). \tag{4.2}$$

Applying Fubini's theorem yields

$$
\begin{aligned}
&\alpha_v^{(2)}(B_1 \times B_2) \\
&= \iiint \int_{B_1} \int_{\mathbb{M}_0} u(t_1)^2 \left(\int_{t<t_1} w(t)h(t_1-t)\varphi_{\mathrm g}(dt) \right) \varphi_{\mathrm g}(B_2) \\
&\qquad\qquad\qquad \mathbb{P}^{\Phi_{\mathrm g}\,|\,L=\lambda}(d\varphi_{\mathrm g})\lambda(t_1)dt_1 \mathbb{P}^L(d\lambda)\mathbb{P}^W(dw)\mathbb{P}^U(du) \\
&= \mathbb{E}U^2 \cdot \mathbb{E}W \cdot \int \int_{B_1} \int_{\mathbb{M}_0} \left(\int_{t<t_1} h(t_1-t)\varphi_{\mathrm g}(dt) \right) \varphi_{\mathrm g}(B_2) \\
&\qquad\qquad\qquad \mathbb{P}^{\Phi_{\mathrm g}\,|\,L=\lambda}(d\varphi_{\mathrm g})\lambda(t_1)dt_1 \mathbb{P}^L(d\lambda),
\end{aligned}
$$

where $\mathbb{E}U^2 = 1$ by assumption.

W.l.o.g., B_1 and B_2 are disjoint intervals in $\mathbb{R}$. If $\sup B_1 < \inf B_2$ then the inner integral w.r.t. $\mathbb{P}^{\Phi_{\mathrm g}\,|\,L=\lambda}$ factorizes, because of stochastic independence, into $\Lambda(B_2) = \int_{B_2} \lambda(t)dt$ and

$$
\int_{\mathbb{M}_0} \left(\int_{t<t_1} h(t_1-t)\varphi_{\mathrm g}(dt) \right) \mathbb{P}^{\Phi_{\mathrm g}|L=\lambda}(d\varphi_{\mathrm g}) = \int_{-\infty}^{t_1} h(t_1-t)\lambda(t)\,dt = \int_0^\infty h(t)\lambda(t_1-t)\,dt\,,
$$

where the first equality holds by the Campbell theorem, which is a special case of (2.8).

Otherwise, if $\inf B_1 > \sup B_2$, with $B_2 = [a,b]$, we have:

$$
\begin{aligned}
&\int_{\mathbb{M}_0} \left(\int_{t<t_1} h(t_1-t)\varphi_{\mathrm g}(dt) \right) \varphi_{\mathrm g}(B_2)\, \mathbb{P}^{\Phi_{\mathrm g}\,|\,L=\lambda}(d\varphi_{\mathrm g}) \\
&= \sum_{k=0}^\infty \mathbb{P}(\Phi_{\mathrm g}(B_2)=k) \cdot \int_{\mathbb{M}_0} \left(\int_{t<t_1} h(t_1-t)\varphi_{\mathrm g}(dt) \right) \varphi_{\mathrm g}(B_2)\, \mathbb{P}^{\Phi_{\mathrm g}|L=\lambda,\Phi_{\mathrm g}(B_2)=k}(d\varphi_{\mathrm g}) \\
&= \sum_{k=0}^\infty \exp(-\Lambda(B_2))\frac{\Lambda(B_2)^k}{k!} \cdot k \int_{\mathbb{M}_0} \int_{(b,t_1]\cup B_2\cup(-\infty,a)} h(t_1-t)\varphi_{\mathrm g}(dt)\, \mathbb{P}^{\Phi_{\mathrm g}|L=\lambda,\Phi_{\mathrm g}(B_2)=k}(d\varphi_{\mathrm g}) \\
&= \Lambda(B_2)\sum_{k=1}^\infty \exp(-\Lambda(B_2))\frac{\Lambda(B_2)^{k-1}}{(k-1)!} \\
&\qquad \cdot \left[\int_0^{t_1-b} h(t)\lambda(t_1-t)dt + \int_{t_1-b}^{t_1-a} kh(t)\frac{\lambda(t_1-t)}{\Lambda(B_2)}dt + \int_{t_1-a}^\infty h(t)\lambda(t_1-t)dt \right] \\
&= \Lambda(B_2)\left[\int_0^{t_1-b} h(t)\lambda(t_1-t)dt + \int_{t_1-b}^{t_1-a} (\Lambda(B_2)+1)h(t)\frac{\lambda(t_1-t)}{\Lambda(B_2)}dt + \int_{t_1-a}^\infty h(t)\lambda(t_1-t)dt \right] \\
&= \Lambda(B_2)\int_0^\infty h(t)\lambda(t_1-t)dt + \int_{t_1-b}^{t_1-a} h(t)\lambda(t_1-t)dt\,,
\end{aligned}
$$

where the third equation again follows by the Campbell theorem (2.8) and from the independence properties of the Poisson process.

Thus,

$$
\begin{aligned}
&\alpha_v^{(2)}(B_1 \times B_2) \\
&= \mathbb{E}W \int \int_{B_1} \left[\Lambda(B_2) \int_0^\infty h(t)\lambda(t_1 - t)dt + \mathbf{1}_{B_1 > B_2} \int_{B_2} h(t_1 - t)\lambda(t)dt \right] \lambda(t_1)dt_1 \, \mathbb{P}^L(d\lambda) \\
&= \mathbb{E}W \int \int_{B_1} \int_{B_2} \left(\int_0^\infty h(t)\lambda(t_1 - t) \, dt + \mathbf{1}_{B_1 > B_2} h(t_1 - t_2) \right) \lambda(t_2)dt_2 \, \lambda(t_1)dt_1 \, \mathbb{P}^L(d\lambda).
\end{aligned}
$$

Applying Fubini's theorem and using the distribution of L, we finally get

$$
\begin{aligned}
&\rho_v^{(2)}(t_1, t_2) \\
&= \mathbb{E}W \int \left(\int_0^\infty h(t)\lambda(t_1 - t) \, dt + \mathbf{1}_{t_1 > t_2} h(t_1 - t_2) \right) \lambda(t_2)\lambda(t_1) \, \mathbb{P}^L(d\lambda) \\
&= \mathbb{E}W \left(\int_0^\infty h(t) \int \lambda(t_1 - t)\lambda(t_2)\lambda(t_1) \, \mathbb{P}^L(d\lambda) \, dt + \mathbf{1}_{t_1 > t_2} h(t_1 - t_2) \int \lambda(t_2)\lambda(t_1) \, \mathbb{P}^L(d\lambda) \right) \\
&= \mathbb{E}W \left(\int_0^\infty h(t) \exp(3\mu + \tfrac{3}{2}\sigma^2 + C(t) + C(t_1 - t_2) + C(t_1 - t_2 - t)) \, dt \right. \\
&\qquad \left. + \mathbf{1}_{t_1 > t_2} h(t_1 - t_2) \exp(2\mu + \sigma^2 + C(t_1 - t_2)) \right).
\end{aligned}
$$

With $\rho_1^{(2)}(t_1, t_2) = \int \lambda(t_2)\lambda(t_1) \, \mathbb{P}^L(d\lambda) = \exp(2\mu + \sigma^2 + C(t_1 - t_2))$, it follows

$$
V(r) = \frac{\rho_v^{(2)}(0, r)}{\rho_1^{(2)}(0, r)} = \mathbb{E}W \left(\mathbf{1}_{r<0} h(-r) + \int_0^\infty h(t) \exp(\mu + \tfrac{\sigma^2}{2} + C(t) + C(r + t)) \, dt \right).
$$

$\square$

Example

We illustrate the MPP model in (4.1) by a simulation according to the following choice of parameters: The random intensity of the Cox process is given by $L = \exp(Z)$, where Z is a stationary Gaussian random field with mean $-\log(30)$, variance 0.1 and scale 200 [seconds], and the function h is given by $h(t) = \exp(-0.0002 \cdot t^2)$. The resulting mean duration between consecutive events is about 30 seconds, the length of the simulated dataset is 1000 times 8 hours. As for the random field W, we choose $W(t) = 1$ for $t \in \mathbb{R}$. By choosing U to be a field of iid $\mathcal{N}(0,1)$-variables, we have that, conditionally on the ground process of point locations, the marks are stochastically independent and Gaussian.

Figure 4.1 shows a 10-minutes window of the realization of the underlying shot noise variance process, together with the squared marks. In Figure 4.2, we compare the theoretical V-function according to Theorem 4.2.2 to the empirical one. We also consider the difference $\hat{V}(-r) - \hat{V}(r)$ for $r > 0$ compared to the function h that determines the shape of the shot

noise process. Note that for $r > 0$,

$$V(-r) - V(r)$$
$$= h(r) + \int_0^\infty h(t)\Big[\exp\big(\mu + \tfrac{\sigma^2}{2} + C(t) + C(-r+t)\big) - \exp\big(\mu + \tfrac{\sigma^2}{2} + C(t) + C(r+t)\big)\Big]dt$$

and that the integral on the RHS is close to zero for small values of r.

The ratio between "quasi-symmetric" effects, which are caused by the covariance structure of the Gaussian random field, and non-symmetric effects, which are caused by the shot-noise-type marking, on the V-function can be controlled by the ratio between the function h and the exponential of the first and second moments of the underlying Gaussian random field. If the jump size W of the shot noise process is chosen to be different from one, this merely results in a shift of the V-function by the constant factor $\mathbb{E}W$.

While marks at different locations are conditionally independent of each other and hence the conditional covariance function $\mathrm{Cov}(r) = \mu_c^{(2)}(r) - \mu_e^{(2)}(r)\mu_e^{(2)}(-r)$ equals zero almost
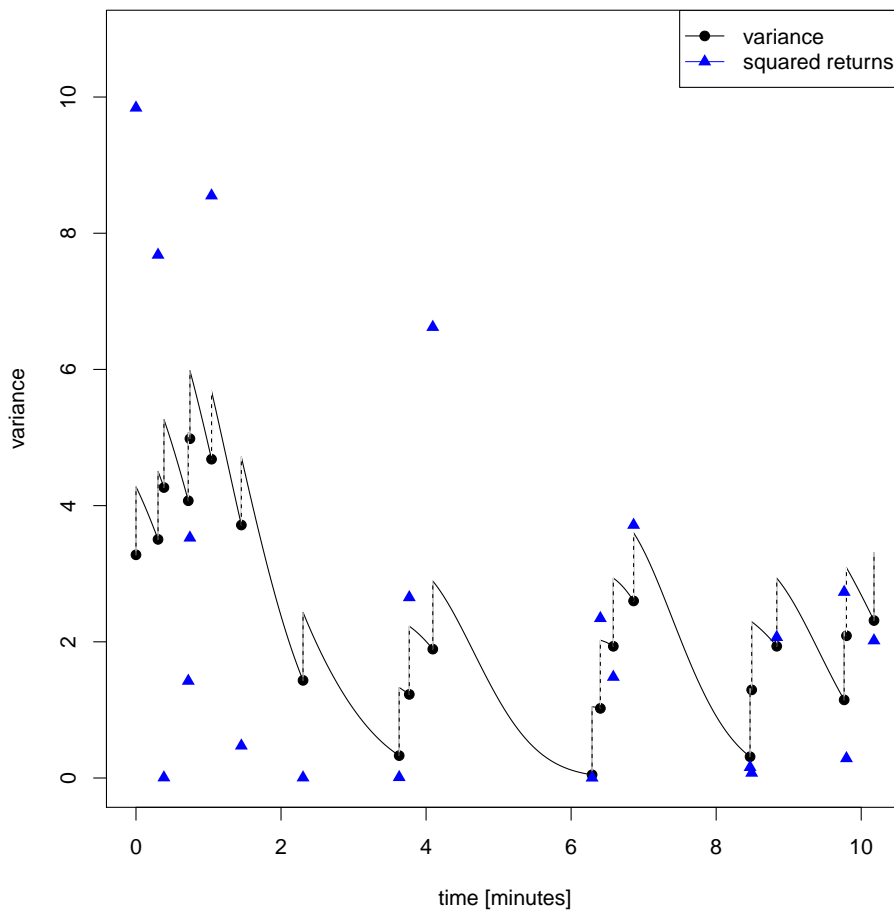

Figure 4.1: Realization of the variance process and squared returns.

everywhere, considering the absolute value of the marks yields a non-trivial covariance structure. Here, $c(y_1, y_2) = y_1 y_2$ and $e(y_1, y_2) = y_1$ and in intuitive notation, $\text{Cov}(r) = \mathbb{E}\big[(y(0) - E(r)) \cdot (y(r) - E(-r)) \,\big|\, 0, r \in \Phi_g\big]$. Figure 4.3 shows the estimates of the E-function and of the conditional covariance function for the modification of Φ, where the marks are replaced by their absolute values.
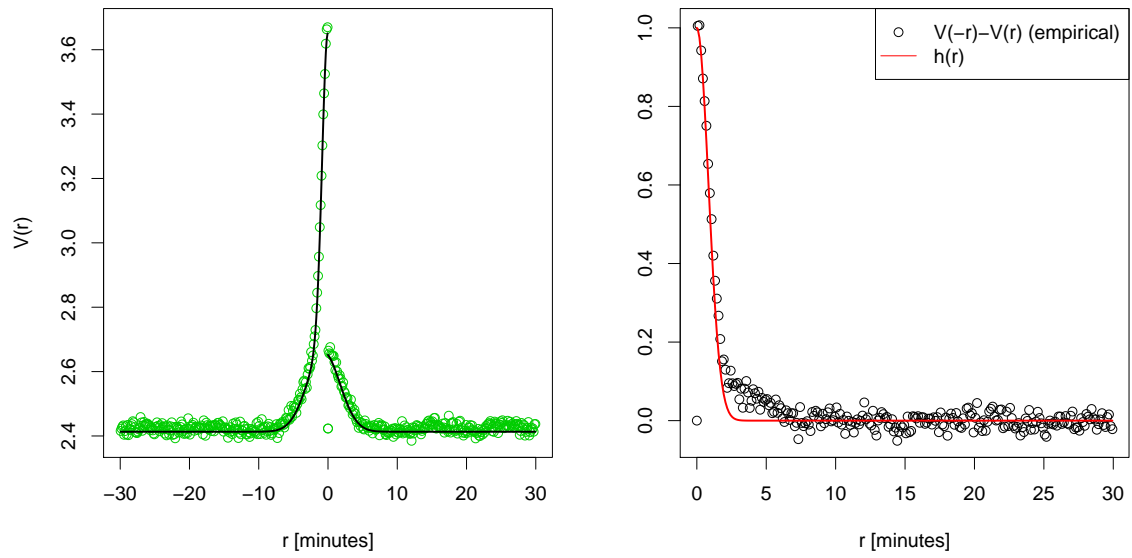

Figure 4.2: Comparison of theoretical and empirical V-function (left) and comparison of "net V-function" with $h(\cdot)$ (right).
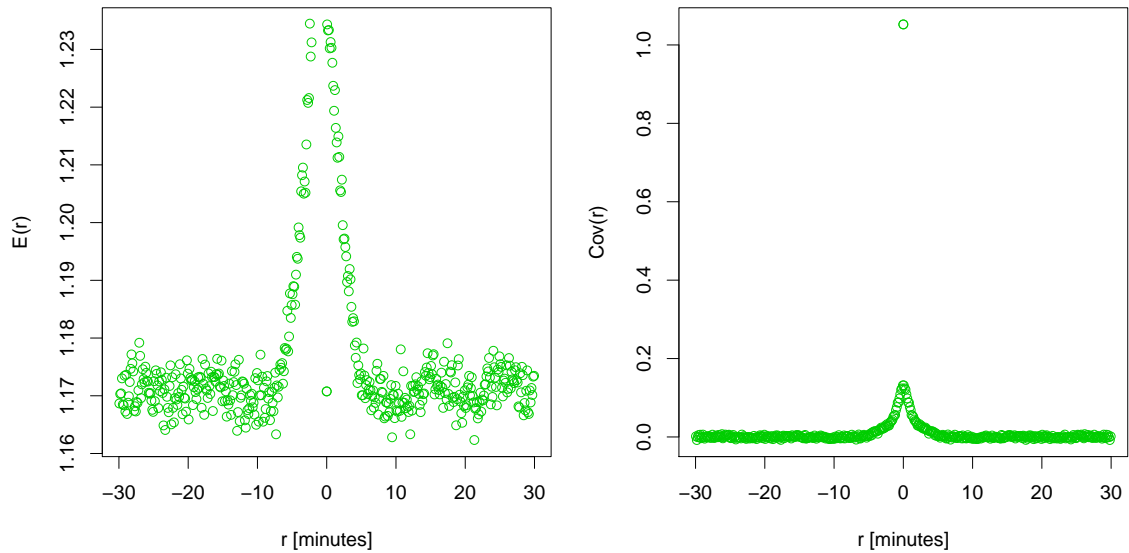

Figure 4.3: E-function and conditional covariance function for the absolute value of the marks.

4.2.2 Model 2: Cox processes marked by a bivariate Gaussian random field

In this model, the process of point locations is given by a log Gaussian Cox process. The marking is explicitly related to the associated Gaussian random field of the Cox process. One possibility of realizing such a marking is to take the intensity at position t (or a transformation thereof) directly as the mark of the point at t (e.g., Ho & Stoyan, 2008). Here, we follow a more general approach:

We consider a bivariate stationary Gaussian field $Z(t) = (Z_1(t), Z_2(t))$, $t \in \mathbb{R}$, where the Z_1-component will generate the intensity of the Cox process and the Z_2-component will drive the marks. This bivariate approach allows for flexible management of dependencies between intensity and marks via the bivariate covariance function

$$C(r) = \begin{pmatrix} C_{11}(r) & C_{12}(r) \\ C_{21}(r) & C_{22}(r) \end{pmatrix}, \; r \in \mathbb{R},$$

where $C_{ij}(r) = \mathrm{Cov}(Z_i(0), Z_j(r))$, $i, j \in \{1,2\}$. The mean of Z is denoted by (μ_1, μ_2).

In particular, the random intensity is given by $\exp(Z_1(\cdot))$ and—conditional on Z—the marks are essentially given by

$$\tilde{y}(t_i) = a + b \exp(Z_2(t_i)) + \varepsilon(t_i), \quad \varepsilon(t_i) \sim \mathcal{N}(0, c^2 + d^2 \exp(Z_2(t_i))) \tag{4.3}$$

$$\text{or by } \tilde{y}(t_i) = \exp(a + b Z_2(t_i) + \varepsilon(t_i)), \quad \varepsilon(t_i) \sim \mathcal{N}(0, c^2 + d^2 Z_2(t_i)^2), \tag{4.4}$$

where $\varepsilon(\cdot)$ is iid Gaussian noise, independent of Z. This construction yields a stationary MPP Φ where the unmarked ground process $\Phi_{\mathrm{g}} = \Phi_{\exp(Z_1)}$ is given by a LGCP with the random intensity measure $\Xi(B) = \int_B \exp(Z_1(t)) \, dt$. We refer to (4.3) and (4.4) as additive and multiplicative marking, respectively.

In the financial context, the marks might be supposed to represent inter-transaction returns. Then, adding a random sign to the marking in (4.3) and (4.4) yields a symmetric marginal distribution of the marks:

$$y(t_i) = U(t_i) \cdot \tilde{y}(t_i)$$

with $U(\cdot)$ being a field of iid variables with $\mathbb{P}(U(0) = \pm 1) = 0.5$, independent of Z and ε.

Actually, marks are only defined for the points t_i of the ground process, but y can also be regarded as a function on $\mathbb{R}$ because the random fields U, Z_2 and ε are defined on the whole space. Conditioning on U, Z_2 and ε, y is a deterministic function of the point location, only, and we are again in the situation of Corollary 2.2.5, part 3.

The E- and V-function of this model are given by the following theorem:

Theorem 4.2.3. *For an MPP Φ as described above, the E-function, the V-function V_{add} for the additive marking (4.3), and the V-function V_{mult} for the multiplicative marking (4.4) are given by*

$$E(r) = 0,$$
$$V_{\mathrm{add}}(r) = (a^2 + c^2) + b^2 \exp\big(2(\mu_2 + \sigma_2^2 + C_{21}(0) + \mathbf{1}_{r \neq 0} C_{21}(r))\big)$$
$$+ (2ab + d^2) \exp\big(\mu_2 + \sigma_2^2/2 + C_{21}(0) + \mathbf{1}_{r \neq 0} C_{21}(r)\big),$$

$$V_{\mathrm{mult}}(r) = \exp(2a + 2c^2)\,\mathbb{E}\exp\big(Z_1(0) + Z_1(r)\mathbf{1}_{r\neq0} + 2bZ_2(0) + 2d^2Z_2(0)^2\big)$$
$$\cdot \exp\left(-\left[(\mu_1 + \sigma_1^2/2)(1 + \mathbf{1}_{r\neq0}) + C_{11}(r)\mathbf{1}_{r\neq0}\right]\right),$$

where $\sigma_i^2 = C_{ii}(0) = \mathrm{Var}\,Z_i(\cdot)$, $i = 1,2$.

For the proof of Theorem 4.2.3, we use the following elementary fact:

Remark 4.2.4. *For a bivariate Gaussian random field Z as above, the vector*

$$\tilde{Z} = (Z_1(0), Z_1(r), Z_2(0), Z_2(r))^\top$$

has a multivariate Gaussian distribution with mean $(\mu_1, \mu_1, \mu_2, \mu_2)^\top$ and covariance matrix

$$\Sigma = \big(\mathrm{Cov}(\tilde{Z}_i, \tilde{Z}_j)\big)_{1\leq i,j\leq 4} = \begin{pmatrix} C_{11}(0) & C_{11}(r) & C_{12}(0) & C_{12}(r) \\ C_{11}(r) & C_{11}(0) & C_{12}(-r) & C_{12}(0) \\ C_{12}(0) & C_{12}(-r) & C_{22}(0) & C_{22}(r) \\ C_{12}(r) & C_{12}(0) & C_{22}(r) & C_{22}(0) \end{pmatrix}. \tag{4.5}$$

Thus, for $\gamma \in \mathbb{R}^4$, $\gamma^\top \tilde{Z} = \gamma_1 Z_1(0) + \gamma_2 Z_1(r) + \gamma_3 Z_2(0) + \gamma_4 Z_2(r) \sim \mathcal{N}(\mu', \sigma'^2)$ with

$$\mu' = (\gamma_1 + \gamma_2)\mu_1 + (\gamma_3 + \gamma_4)\mu_2$$
$$\textit{and}\quad \sigma'^2 = (\gamma_1^2 + \gamma_2^2)\sigma_1^2 + 2\gamma_1\gamma_2 C_{11}(r) + 2(\gamma_1\gamma_3 + \gamma_2\gamma_4)C_{12}(0)$$
$$+ (\gamma_3^2 + \gamma_4^2)\sigma_2^2 + 2\gamma_3\gamma_4 C_{22}(r) + 2\gamma_1\gamma_4 C_{12}(r) + 2\gamma_2\gamma_3 C_{12}(-r). \tag{4.6}$$

Note that $C_{21}(r) = C_{12}(-r)$ for $r \in \mathbb{R}$. Furthermore, stationarity of Z implies isotropy of the univariate one-dimensional marginal fields, but in general, the cross-covariance function does not have to be symmetric and therefore Z is not necessarily isotropic.

Proof of Theorem 4.2.3. Conditioning on U, Z and ε, we can apply Corollary 2.2.5, part 3, with $y(t_1, \varphi_{\mathrm{g}}) = y(t_1)$ and get

$$\alpha_f^{(2)}(B_1 \times B_2)$$
$$= \iiint \int_{\mathbb{M}_0} \int_{\mathbb{R}} \mathbf{1}_{B_1}(t_1) f(y(t_1)) \varphi_{\mathrm{g}}(B_2) \lambda(t_1) dt_1 \, \mathbb{P}^{\Phi_{\mathrm{g}}\mid\exp(Z_1)=\lambda}(d\varphi_{\mathrm{g}})$$
$$\mathbb{P}^{\varepsilon\mid\exp(Z_2)=\tilde{m}}(d\tilde{\varepsilon})\, \mathbb{P}^{(\exp(Z_1),\,\exp(Z_2))}(d\lambda, d\tilde{m})\, \mathbb{P}^U(du)$$
$$= \int \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{B_1}(t_1)\mathbf{1}_{B_2}(t_2) f(y(t_1)) \lambda(t_1)\lambda(t_2) dt_1 dt_2 \, \mathbb{P}^W(d\tilde{\varepsilon}, d\lambda, d\tilde{m}, du)$$

with $W = (\varepsilon, \exp(Z_1), \exp(Z_2), U)$.

Using Fubini's theorem, the Radon-Nikodym derivative of $\alpha_f^{(2)}$ w.r.t. the Lebesgue measure is given by

$$\rho_f^{(2)}(t_1, t_2) = \int f(y(t_1))\lambda(t_1)\lambda(t_2)\, \mathbb{P}^W(d\tilde{\varepsilon}, d\lambda, d\tilde{m}, du).$$

Since $(\lambda, \tilde{m})$ is a realization of $(\exp(Z_1), \exp(Z_2))$ and U and ε have zero mean, applying

Remark 4.2.4 yields

$$\rho_1^{(2)}(0,r) = \int \lambda(0)\lambda(r)\,\mathbb{P}^W(d\tilde{\varepsilon}, d\lambda, d\tilde{m}, du)$$
$$= \mathbb{E}\exp(Z_1(0) + Z_1(r)) = \exp(2\mu_1 + \sigma_1^2 + C_{11}(r)),$$

$$\rho_e^{(2)}(0,r) = \int e(y(0))\lambda(0)\lambda(r)\,\mathbb{P}^W(d\tilde{\varepsilon}, d\lambda, d\tilde{m}, du)$$
$$= \mathbb{E}(U(0)\tilde{y}(0)\exp(Z_1(0) + Z_1(r))) = 0,$$

$$\rho_v^{(2)}(0,r) = \int v(y(0))\lambda(0)\lambda(r)\,\mathbb{P}^W(d\tilde{\varepsilon}, d\lambda, d\tilde{m}, du)$$
$$= \mathbb{E}(U(0)^2\tilde{y}(0)^2\exp(Z_1(0) + Z_1(r)))$$
$$= \mathbb{E}(\tilde{y}(0)^2\exp(Z_1(0) + Z_1(r))).$$

For $\rho_v^{(2)}(0,r)$, we distinguish between the additive marking in (4.3) and the multiplicative marking in (4.4). In the first case we have

$$\rho_v^{(2)}(0,r) = \mathbb{E}([a + b\exp(Z_2(0)) + \varepsilon(0)]^2\exp(Z_1(0) + Z_1(r)))$$
$$= \mathbb{E}([a^2 + b^2\exp(2Z_2(0)) + \varepsilon(0)^2 + 2ab\exp(Z_2(0))]\exp(Z_1(0) + Z_1(r)))$$
$$= a^2\mathbb{E}\exp(Z_1(0) + Z_1(r)) + b^2\mathbb{E}\exp(Z_1(0) + Z_1(r) + 2Z_2(0))$$
$$\quad + \mathbb{E}((c^2 + d^2\exp(Z_2(0)))\exp(Z_1(0) + Z_1(r)))$$
$$\quad + 2ab\mathbb{E}\exp(Z_1(0) + Z_1(r) + Z_2(0))$$
$$= (a^2 + c^2)\exp(2\mu_1 + \sigma_1^2 + C_{11}(r))$$
$$\quad + b^2\exp(2\mu_1 + \sigma_1^2 + C_{11}(r) + 2\mu_2 + 2\sigma_2^2 + 2C_{21}(0) + 2C_{21}(r))$$
$$\quad + (2ab + d^2)\exp(2\mu_1 + \sigma_1^2 + C_{11}(r) + \mu_2 + \sigma_2^2/2 + C_{21}(0) + C_{21}(r)). \quad (4.7)$$

In the second case we have

$$\rho_v^{(2)}(0,r) = \mathbb{E}([\exp(a + bZ_2(0) + \varepsilon(0))]^2\exp(Z_1(0) + Z_1(r)))$$
$$= \mathbb{E}(\exp(2a + 2bZ_2(0) + 2\varepsilon(0) + Z_1(0) + Z_1(r)))$$
$$= \mathbb{E}(\exp(2a + 2bZ_2(0) + 2c^2 + 2d^2Z_2(0)^2 + Z_1(0) + Z_1(r)))$$
$$= \exp(2a + 2c^2)\mathbb{E}(\exp(Z_1(0) + Z_1(r) + 2bZ_2(0) + 2d^2Z_2(0)^2)). \quad (4.8)$$

For $r \neq 0$, $\mu_f(r)$ is given by $\rho_f^{(2)}(0,r)/\rho_1^{(2)}(0,r)$; for $r = 0$ we have $\mu_f(0) = \int_{\mathbb{R}} f(y)\,M(dy)$ according to Section 2.2. Applying stationarity of Φ, equation (2.8) with $h(t) = \mathbf{1}_{y(t)\in L}$, $L \in \mathcal{B}$, and Fubini's theorem yield

$$M(L) = (\mathbb{E}\Phi_g([0,1]))^{-1} \cdot \mathbb{E}\Phi([0,1] \times L)$$
$$= (\mathbb{E}\Lambda_g([0,1]))^{-1}\int\int_{[0,1]}\mathbf{1}_{y(t)\in L}\lambda(t)dt\,\mathbb{P}^W(d\tilde{\varepsilon}, d\lambda, d\tilde{m}, du)$$

$$= \exp(-(\mu_1 + \sigma_1^2/2)) \int \mathbf{1}_{y(0) \in L} \lambda(0) \mathbb{P}^W(d\tilde{\varepsilon}, d\lambda, d\tilde{m}, du),$$

and hence

$$\mu_f(0) = \exp(-(\mu_1 + \sigma_1^2/2)) \int f(y(0)) \lambda(0) \, \mathbb{P}^W(d\tilde{\varepsilon}, d\lambda, d\tilde{m}, du).$$

Analogously to the calculation in (4.7) and (4.8), we get

$$E(0) = 0,$$
$$V_{\text{add}}(0) = a^2 + c^2 + b^2 \exp(2(\mu_2 + \sigma_2^2 + C_{12}(0))) + (2ab + d^2) \exp(\mu_2 + \sigma_2^2/2 + C_{12}(0)),$$
$$V_{\text{mult}}(0) = \exp(-(\mu_1 + \sigma_1^2/2)) \exp(2a + 2c^2) \mathbb{E}(\exp(Z_1(0) + 2bZ_2(0) + 2d^2 Z_2(0)^2)),$$

which completes the proof. □

In Theorem 4.2.3, the V-function for the multiplicative marking is given in terms of an expectation w.r.t. the underlying random field and involves linear and quadratic terms in Z. The following remark provides a formula for the explicit calculation of these terms.

Remark 4.2.5. *Let Σ and $\tilde{Z}$ be as in Remark 4.2.4. Let $\Gamma \in M_{4 \times 4}(\mathbb{R})$ be such that $(\Sigma^{-1} - 2\Gamma)$ is invertible and $(\Sigma^{-1} - 2\Gamma)^{-1}$ is a valid covariance matrix. Then*

$$\mathbb{E} \exp(\gamma^\top \tilde{Z} + \tilde{Z}^\top \Gamma \tilde{Z}) \tag{4.9}$$
$$= |1 - 2\Sigma\Gamma|^{-1/2} \cdot \exp\left(-\frac{1}{2}\mu^\top \Sigma^{-1}\mu + \frac{1}{2}(\Sigma^{-1}\mu + \gamma)^\top (\Sigma^{-1} - 2\Gamma)^{-1}(\Sigma^{-1}\mu + \gamma)\right).$$

Proof. Using the notation $\xi = \mu + \Sigma\gamma$ we have

$$\mathbb{E} \exp(\gamma^\top \tilde{Z} + \tilde{Z}^\top \Gamma \tilde{Z})$$
$$= \int_{\mathbb{R}^4} \exp(\gamma^\top z + z^\top \Gamma z) \cdot (2\pi)^{-2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(z - \mu)^\top \Sigma^{-1}(z - \mu)\right) dz$$
$$= (2\pi)^{-2} |\Sigma|^{-1/2}$$
$$\qquad \cdot \int_{\mathbb{R}^4} \exp(z^\top \Gamma z) \exp\left(-\frac{1}{2}\left(z - \xi\right)^\top \Sigma^{-1}\left(z - \xi\right) + \frac{1}{2}\gamma^\top \Sigma\gamma + \gamma^\top \mu\right) dz$$
$$= (2\pi)^{-2} |\Sigma|^{-1/2}$$
$$\qquad \cdot \int_{\mathbb{R}^4} \exp\left(-\frac{1}{2}\left(z - (\Sigma^{-1} - 2\Gamma)^{-1}\Sigma^{-1}\xi\right)^\top \left(\Sigma^{-1} - 2\Gamma\right)\left(z - (\Sigma^{-1} - 2\Gamma)^{-1}\Sigma^{-1}\xi\right)\right.$$
$$\qquad\qquad + \frac{1}{2}\gamma^\top \Sigma\gamma + \gamma^\top \mu$$
$$\qquad\qquad - \frac{1}{2}\xi^\top \Sigma^{-1}\xi$$
$$\qquad\qquad + \left.\frac{1}{2}\left((\Sigma^{-1} - 2\Gamma)^{-1}\Sigma^{-1}\xi\right)^\top \underbrace{(\Sigma^{-1} - 2\Gamma)\left((\Sigma^{-1} - 2\Gamma)^{-1}\Sigma^{-1}\xi\right)}_{=1}\right) dz$$

$$= \int_{\mathbb{R}^4} (2\pi)^{-2} |(\Sigma^{-1} - 2\Gamma)^{-1}|^{-1/2}$$

$$\exp\left(-\frac{1}{2}\left(z - (\Sigma^{-1} - 2\Gamma)^{-1}\Sigma^{-1}\xi\right)^\top \left(\Sigma^{-1} - 2\Gamma\right)\left(z - (\Sigma^{-1} - 2\Gamma)^{-1}\Sigma^{-1}\xi\right)\right) dz$$

$$|\Sigma|^{-1/2} \cdot |(\Sigma^{-1} - 2\Gamma)^{-1}|^{1/2} \cdot \exp\left(\frac{1}{2}\mu^\top \Sigma\mu + \frac{1}{2}(\Sigma^{-1}\mu + \gamma)^\top (\Sigma^{-1} - 2\Gamma)^{-1}\Sigma^{-1}\xi\right)$$

$$= |1 - 2\Sigma\Gamma|^{-1/2} \cdot \exp\left(-\frac{1}{2}\mu^\top \Sigma^{-1}\mu + \frac{1}{2}(\Sigma^{-1}\mu + \gamma)^\top (\Sigma^{-1} - 2\Gamma)^{-1}\Sigma^{-1}\xi\right).$$

$\square$

Remark 4.2.6. *As the cross-covariance function $C_{21}(r)$—in contrast to the functions $C_{11}(r)$ and $C_{22}(r)$—does not have to be a symmetric function, the MPP in Theorem 4.2.3 is able to model interactions between the locations and the marks of the point process in terms of a non-symmetric V-function. The crucial point regarding this property is that we use a bivariate random field. In the univariate case, these dependencies vanish.*

Example

In order to illustrate the result of Theorem 4.2.3, we consider the theoretical and empirical V-function of a log Gaussian Cox process with the additive marking (4.3). Let Z_1 be a univariate Gaussian random field on $\mathbb{R}$ having a Gaussian covariance model

$$C_1(h) = \text{variance} \cdot \exp(-(h/\text{scale})^2).$$

We choose variance $= 0.1$, scale $= 180$ [seconds] and a mean of $= -\log(30)$, which causes the mean duration between two points of the Cox process to be approximately 30 seconds. Let the second component of Z be given by $Z_2(t) = Z_1(t-s) - Z_1(t+s)$ for some shifting constant $s > 0$, which will be set to 10 [seconds] in our case. Using the notation of Theorem 4.2.3, we have

$$\mu = (-\log 30, 0),$$
$$\sigma_1^2 = 0.1,$$
$$\sigma_2^2 = 2\sigma_1^2 - 2C_1(2s),$$
$$C_{12}(h) = \text{Cov}(Z_1(0), Z_2(h)) = C_1(h-s) - C_1(h+s).$$

For simplicity we let $a = c = d = 0$ and $b = 1$, i.e., the marks are simply

$$y(t_i) = U(t_i) \cdot \exp(Z_2(t_i)).$$

Figure 4.4 shows an extract of a realization of the corresponding MPP together with the bivariate driving random field. The theoretical V-function and the empirical V-function are shown in Figure 4.5. The length of the simulated dataset is 4000 hours.

As in the example in Section 4.2.1, the sign of the marks is independent of the process of point locations and has zero mean, which causes both the E-function and the conditional

covariance function $\mathrm{Cov}(r)$ to be zero almost everywhere. Hence, we consider these functions for the absolute value of the marks in Figure 4.6.
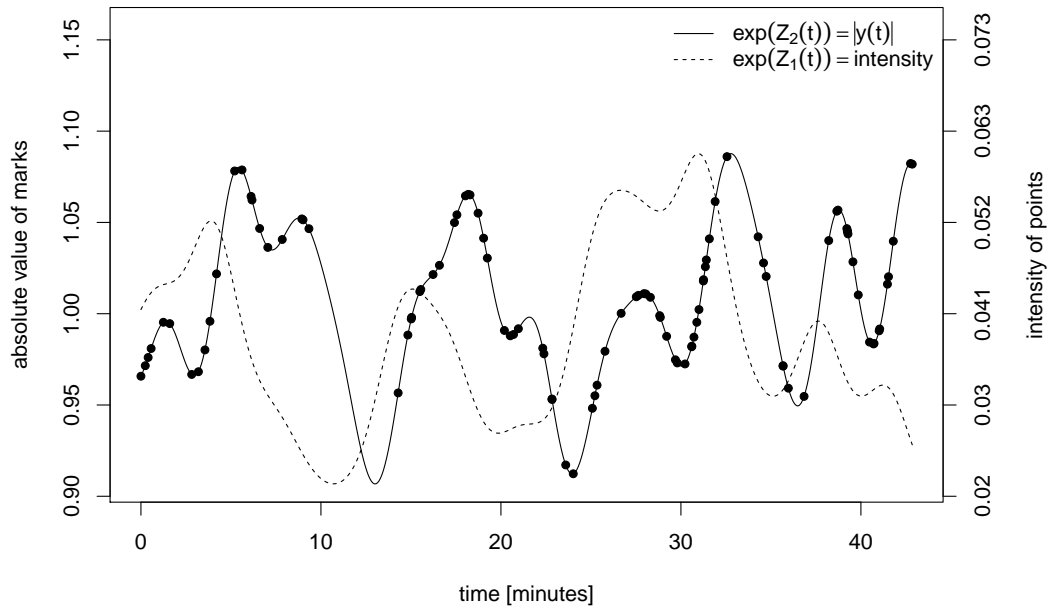

Figure 4.4: Realization of the above Cox process (45-minutes window). Only the absolute value of the marks is shown.
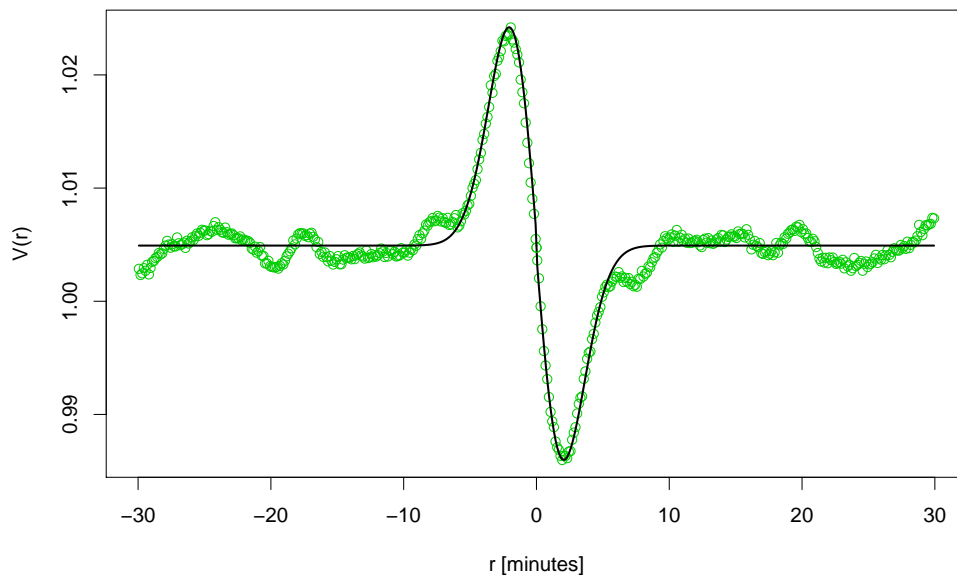

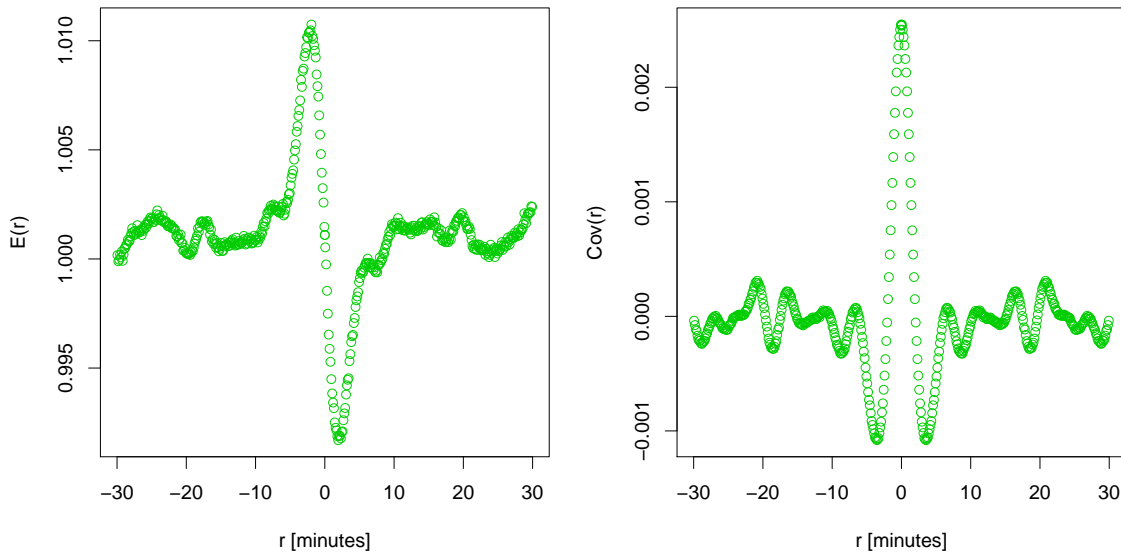**Figure 4.5:** Theoretical and empirical V-function.

Figure 4.6: Estimate of the E-function and the conditional covariance function for the absolute value of the marks.

4.2.3 Model 3: Marking based on inter-point distances

In this subsection, we consider Poisson point processes with a marking based on inter-point distances. To this end, we define a measurable function $\mathrm{ptd} : \mathbb{R} \times \mathbb{M}_0(\mathbb{R}) \to (0,\infty)$, which stands for *previous time stamp distance*, by

$$\mathrm{ptd}(t, \varphi_{\mathrm{g}}) = t - \sup\{s \in \mathrm{supp}(\varphi_{\mathrm{g}}) \,|\, s < t\},$$

i.e., $\mathrm{ptd}(t, \varphi_{\mathrm{g}})$ is the distance from t to the nearest previous point within the point pattern φ_{g}. Let $\tilde{y} : (0,\infty) \times \mathbb{R} \to \mathbb{R}$ be a measurable function that is differentiable in its first component. Let $U = \{U(t)\}_{t \in \mathbb{R}}$ be a stationary random field and let $y : \mathbb{R} \times \mathbb{M}_0(\mathbb{R}) \to \mathbb{R}$ be given by $y(t, \varphi_{\mathrm{g}}) = \tilde{y}(\mathrm{ptd}(t, \varphi_{\mathrm{g}}), U(t))$. Then we consider MPPs of the form

$$\Phi = \{(t, y(t, \tilde{\Phi})) : t \in \tilde{\Phi}\}, \tag{4.10}$$

where $\tilde{\Phi}$ is a stationary unmarked Poisson point process on $\mathbb{R}$.

As an example, one may imagine the specification in Centanni & Minozzo (2006), where U is some Gaussian white noise and the marking is given by

$$y(t, \varphi_{\mathrm{g}}) = \mu \, \mathrm{ptd}(t, \varphi_{\mathrm{g}}) + \alpha \int_{t-\mathrm{ptd}(t, \varphi_{\mathrm{g}})}^{t} \lambda(s) \, ds + \sigma U(t)$$

for some constants μ, α and σ.

Theorem 4.2.7. *For an MPP as in* ([4.10]), *the E- and V-function are given by*

$$E(r) = \int \left(\tilde{y}(0,u(0)) - \int_{a(r)}^0 \frac{d}{ds}(\tilde{y}(-s, u(0))) \exp(-\lambda s) \, ds \right) \mathbb{P}^U(du) \tag{4.11}$$

$$V(r) = \int \left((\tilde{y}(0,u(0)))^2 - \int_{a(r)}^0 \frac{d}{ds}((\tilde{y}(-s,u(0)))^2) \exp(-\lambda s) \, ds \right) \mathbb{P}^U(du) - E(r)^2 \tag{4.12}$$

with $a(r) = \begin{cases} r, & r < 0 \\ -\infty, & r \geq 0, \end{cases}$

given that the respective RHS integrals exist.

Remark 4.2.8. *This result can be generalized straightforward to doubly stochastic Poisson processes: Let L be the underlying random intensity function, which is assumed to be stationary and to have finite second moments. Let μ_L and C_L denote its mean and its covariance function, respectively, and let $\Lambda(\cdot) = \int L(s)ds$. Then, if the following RHS integrals exist, we have*

$$E(r) = \int \left(\tilde{y}(0,u(0)) - \int_{a(r)}^0 \frac{d}{ds}(\tilde{y}(-s, u(0))) \exp(-\Lambda([s,0])) \, ds \right)$$
$$\lambda(0)\lambda(r)^{\mathbf{1}_{r \neq 0}} \mathbb{P}^L(d\lambda) \, \mathbb{P}^U(du) \cdot \mu(r)^{-1},$$

$$V(r) = \int \left((\tilde{y}(0,u(0)))^2 - \int_{a(r)}^0 \frac{d}{ds}((\tilde{y}(-s,u(0)))^2) \exp(-\Lambda([s,0])) \, ds \right)$$
$$\lambda(0)\lambda(r)^{\mathbf{1}_{r \neq 0}} \mathbb{P}^L(d\lambda) \, \mathbb{P}^U(du) \cdot \mu(r)^{-1} - E(r)^2$$

with $a(r)$ as in Theorem [4.2.7] and $\mu(r) = \begin{cases} C_L(r) + \mu_L^2, & r \neq 0 \\ \mu_L, & r = 0. \end{cases}$

The proof is analog to that of Theorem [4.2.7] by conditioning on the intensity.

Proof of Theorem [4.2.7]. According to Corollary [2.2.5], part [3], and conditioning on U, it is

$$\alpha_f^{(2)}(B_1 \times B_2) = \int \int_{\mathbb{R}} \int_{\mathbb{M}_0} \mathbf{1}_{B_1}(t_1) f(y(t_1, \varphi_{\mathrm{g}})) \varphi_{\mathrm{g}}(B_2) \, \mathbb{P}^{\Phi_{\mathrm{g}}}(d\varphi_{\mathrm{g}}) \lambda dt_1 \, \mathbb{P}^U(du). \tag{4.13}$$

We define $A(t_1)$ to be the inner integral of ([4.13]):

$$A(t_1) = \int_{\mathbb{M}_0} f(y(t_1, \varphi_{\mathrm{g}})) \varphi_{\mathrm{g}}(B_2) \, \mathbb{P}^{\Phi_{\mathrm{g}}}(d\varphi_{\mathrm{g}}) = \mathbb{E}_{\Phi_{\mathrm{g}}}[f(y(t_1, \Phi_{\mathrm{g}})) \Phi_{\mathrm{g}}(B_2)].$$

W.l.o.g., we assume B_1 and B_2 to be disjoint intervals. We differentiate the two cases $\sup(B_1) < \inf(B_2)$ and $\inf(B_1) > \sup(B_2)$:

In the first case, since Φ_{g} is a Poisson process, $f(y(t_1, \Phi_{\mathrm{g}}))$ and $\Phi_{\mathrm{g}}(B_2)$ are stochastically

independent. For $\text{PTD} = \text{ptd}(t_1, \Phi_{\mathrm{g}})$ we have

$$\mathbb{P}(\text{PTD} \leq t_1 - s) = 1 - \mathbb{P}(\Phi_{\mathrm{g}}([s, t_1]) = 0) = 1 - \exp(-\lambda(t_1 - s)), \quad s \leq t_1. \tag{4.14}$$

The corresponding density is $h(t_1 - s) = \lambda \exp(-\lambda(t_1 - s))$. Thus,

$$A(t_1) = \lambda |B_2| \cdot \int_{-\infty}^{t_1} f(\tilde{y}(t_1 - s, u(t_1))) \exp(-\lambda(t_1 - s)) \lambda \, ds. \tag{4.15}$$

With $g_{t_1}(s) = f(\tilde{y}(t_1 - s, u(t_1))) \exp(-\lambda(t_1 - s))$ we have

$$\alpha_f^{(2)}(B_1 \times B_2) = \int \left(\int_{B_1} \int_{B_2} \left(\int_{-\infty}^{t_1} g_{t_1}(s) \lambda \, ds \right) \lambda^2 \, dt_2 dt_1 \right) \mathbb{P}^U(du). \tag{4.16}$$

In the second case, $f(y(t_1, \Phi_{\mathrm{g}}))$ and $\Phi_{\mathrm{g}}(B_2)$ are obviously not stochastically independent. Let $B_2 = [a,b]$. Then

$$A(t_1) = \sum_{i=1}^{\infty} \mathbb{E}_{\Phi_{\mathrm{g}}}[if(\tilde{y}(\text{PTD}, u(t_1))) \,|\, \Phi_{\mathrm{g}}(B_2) = i \wedge \text{PTD} > t_1 - b]$$
$$\cdot \exp(-\lambda |B_2|) \frac{(\lambda |B_2|)^i}{i!} \exp(-\lambda(t_1 - b))$$
$$+ \sum_{i=1}^{\infty} \mathbb{E}_{\Phi_{\mathrm{g}}}[if(\tilde{y}(\text{PTD}, u(t_1))) \,|\, \text{PTD} \leq t_1 - b]$$
$$\cdot \exp(-\lambda |B_2|) \frac{(\lambda |B_2|)^i}{i!} [1 - \exp(-\lambda(t_1 - b))]. \tag{4.17}$$

For $s \in B_2 = [a,b]$ we have

$$\mathbb{P}(\text{PTD} \leq t_1 - s \,|\, \Phi_{\mathrm{g}}(B_2) = i \wedge \text{PTD} > t_1 - b)$$
$$= 1 - \mathbb{P}(\Phi_{\mathrm{g}}([s,b]) = 0 \,|\, \Phi_{\mathrm{g}}(B_2) = i) = 1 - \left(\frac{\lambda(s - a)}{\lambda(b - a)} \right)^i,$$

which gives the conditional density $h_1(t_1 - s) = i \frac{(s-a)^{i-1}}{(b-a)^i}$. For $s \in [b, t_1]$ we have

$$\mathbb{P}(\text{PTD} \leq t_1 - s \,|\, \text{PTD} \leq t_1 - b) = 1 - \mathbb{P}(\Phi_{\mathrm{g}}([s, t_1]) = 0 \,|\, \text{PTD} \leq t_1 - b)$$
$$= 1 - \frac{\exp(-\lambda(t_1 - s))[1 - \exp(-\lambda(s - b))]}{1 - \exp(-\lambda(t_1 - b))} = \frac{1 - \exp(-\lambda(t_1 - s))}{1 - \exp(-\lambda(t_1 - b))},$$

which leads to a density $h_2(t_1 - s) = \lambda \frac{\exp(-\lambda(t_1 - s))}{1 - \exp(-\lambda(t_1 - b))}$. Plugging this into (4.17) and using $\sum_{i=1}^{\infty} i \exp(-c) \frac{c^i}{i!} = c$ for $c \geq 0$, an elementary calculation yields

$$A(t_1) = \int_a^b g_{t_1}(s) \lambda \left(1 + \lambda(s - a)\right) ds + \int_b^{t_1} g_{t_1}(s) \lambda \, ds \cdot \lambda(b - a)$$
$$= \int_a^b \int_z^b g_{t_1}(s) \lambda^2 \, ds \, dz + \int_a^b g_{t_1}(s) \lambda \, ds + \int_a^b \int_b^{t_1} g_{t_1}(s) \lambda^2 \, ds \, dz$$

$$= \int_{B_2} \left(g_{t_1}(t_2) + \int_{t_2}^{t_1} g_{t_1}(s)\lambda \, ds \right) \lambda dt_2,$$

where the second equality follows from changing the order of integration. Finally,

$$\alpha_f^{(2)}(B_1 \times B_2)$$
$$= \int \left(\int_{B_1} \int_{B_2} \left(g_{t_1}(t_2) + \int_{t_2}^{t_1} g_{t_1}(s)\lambda \, ds \right) \lambda^2 \, dt_2 dt_1 \right) \mathbb{P}^U(du). \tag{4.18}$$

Combining (4.16) and (4.18) and using Fubini's theorem, we obtain the Radon-Nikodym derivative of $\alpha_f^{(2)}$ with respect to the Lebesgue measure:

$$\rho_f^{(2)}(t_1, t_2) = \mathbf{1}_{t_1 < t_2} \int \left(\int_{-\infty}^{t_1} g_{t_1}(s)\lambda \, ds \right) \lambda^2 \, \mathbb{P}^U(du)$$
$$+ \mathbf{1}_{t_1 > t_2} \int \left(g_{t_1}(t_2) + \int_{t_2}^{t_1} g_{t_1}(s)\lambda \, ds \right) \lambda^2 \, \mathbb{P}^U(du). \tag{4.19}$$

Hence, re-substituting g_{t_1},

$$\rho_f^{(2)}(0, r)$$
$$= \mathbf{1}_{r>0} \int \left(\int_{-\infty}^{0} f(\tilde{y}(-s, u(0))) \exp(\lambda s)\lambda \, ds \right) \lambda^2 \, \mathbb{P}^U(du)$$
$$+ \mathbf{1}_{r<0} \int \left(f(\tilde{y}(-r,u(0))) \exp(\lambda r) + \int_{r}^{0} f(\tilde{y}(-s,u(0))) \exp(\lambda s)\lambda \, ds \right) \cdot \lambda^2 \, \mathbb{P}^U(du).$$

Since $f(\tilde{y}(-s, u(0)))$ was assumed to be differentiable with respect to s on $(-\infty, 0)$, partial integration yields

$$\rho_f^{(2)}(0, r) = \mathbf{1}_{r>0} \int \left(f(\tilde{y}(0,u(0))) - \int_{-\infty}^{0} \frac{d}{ds}(f(\tilde{y}(-s,u(0)))) \exp(\lambda s) \, ds \right) \cdot \lambda^2 \, \mathbb{P}^U(du)$$
$$+ \mathbf{1}_{r<0} \int \left(f(\tilde{y}(0, u(0))) - \int_{r}^{0} \frac{d}{ds}(f(\tilde{y}(-s,u(0)))) \exp(\lambda s) \, ds \right) \cdot \lambda^2 \, \mathbb{P}^U(du).$$

Applying $f \equiv 1$ yields $\rho_1^{(2)}(0,r) = \lambda^2$.

For $r \neq 0$, the assertion of the theorem follows from $\mu_f^{(2)}(r) = \rho_f^{(2)}(0, r)/\rho_1^{(2)}(0, r)$.

For $r = 0$, we have $\mu_f(0) = \int_{\mathbb{M}^*} f(m) \, M(dm)$ (cf. Section 2.2). Since PTD is exponentially distributed, we have

$$M(L) = \lambda^{-1} \int \int_{-\infty}^{0} \mathbf{1}_L(\tilde{y}(-s, u(0))) \exp(\lambda s)\lambda \, ds \, \mathbb{P}^U(du), \quad L \in \mathcal{B},$$

and hence

$$\mu_f(0) = \int_{\mathbb{R}} f(y) \, M(dy)$$

$$= \lambda^{-1} \int \left(\int_{-\infty}^{0} f(\tilde{y}(-s, u(0))) \exp(\lambda s) \lambda \, ds \right) \lambda \, \mathbb{P}^{U}(du)$$

$$= \int \left(f(\tilde{y}(0, u(0))) - \int_{-\infty}^{0} \frac{d}{ds} (f(\tilde{y}(-s, u(0)))) \exp(\lambda s) \, ds \right) \mathbb{P}^{U}(du). \qquad (4.20)$$

$\square$

Example

In order to illustrate the result of Theorem 4.2.7, we consider the theoretical and empirical E- and V-function of a log Gaussian Cox process satisfying the assumptions of Remark 4.2.8. Let $L = \exp(Z)$ be a log Gaussian random field with Z having an exponential covariance model $C(h) = \text{variance} \cdot \exp(-|h|/\text{scale})$. We choose $\text{variance} = 0.1, \text{scale} = 180 \,[\text{seconds}]$ and a mean of $-\log(30)$, which causes the mean duration between two points of the Cox process to be approximately 30 seconds. For the function $\tilde{y}$ we choose $\tilde{y}(\text{ptd}, u) = u \cdot (\text{ptd} + 1)^{-0.5}$. Let U be a field of iid variables with $\mathbb{P}(U(0) = \pm 1) = 0.5$. Then, $E(r) = 0$ for all $r \in \mathbb{R}$. We have $v(y(t, \varphi_{\text{g}})) = U(t)^2 \cdot (\text{ptd}(t, \varphi_{\text{g}}) + 1)^{-1} = (\text{ptd}(t, \varphi_{\text{g}}) + 1)^{-1}$

The empirical V-function (Figure 4.7) is estimated from a simulated realization of the marked Cox process with a length of 3200 hours. The corresponding (semi-)theoretical characteristics are determined by simulation of the underlying random fields L and U (50,000 realizations).
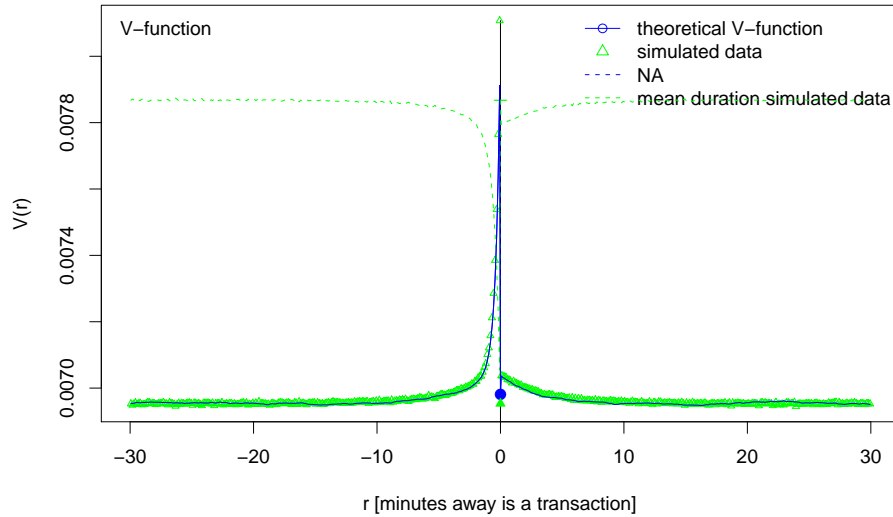

Figure 4.7: Theoretical V-function and empirical V-function of a simulation of the Cox process model (4.10).

4.2.4 Model 4: Order book model

In this subsection, we consider a continuous time model for the dynamics of a limit order book, proposed by Cont *et al.* (2010), in which the flow of orders is driven by independent Poisson processes.

The model and the notation used by Cont *et al.* (2010) can be summarized as follows:

- the prices at which limit orders can be placed are $\{1, \ldots, N\}$, which represents the multiples of a price tick

- all orders have unit size

- $X(t) = (X_1(t), \ldots, X_N(t))$ denotes the process of outstanding limit orders for each price, where $|X_p(t)|$ is the number of outstanding limit orders at price p, $1 \leq p \leq N$, and where a negative sign stands for bid (= buy) orders

- $p_A(t) = \inf\{p \in \{1, \ldots, N\} : X_p(t) > 0\} \wedge (N+1)$ is the lowest ask price,
  $p_B(t) = \sup\{p \in \{1, \ldots, N\} : X_p(t) < 0\} \vee 0$ is the highest bid price,
  $p_M(t) = \frac{1}{2}(p_A(t) + p_B(t))$ is the mid-price at time t,
  $s(t) = p_A(t) - p_B(t)$ is the bid-ask spread at time t

- let $\lambda, \theta : \{1, \ldots, N\} \to [0, \infty)$ be functions, given by $\lambda(i) = \frac{\lambda_0}{i^\alpha}$ and $\theta(i) = \frac{\theta_0}{i^\alpha}$ for some $\lambda_0, \theta_0, \alpha > 0$

- order arrival times are given by stochastically independent Poisson processes with the following intensities:
  ask orders at price level p arrive at rate $\lambda(p - p_B(t))$ for $p > p_B(t)$,
  bid orders at price level p arrive at rate $\lambda(p_A(t) - p)$ for $p < p_A(t)$,
  market sell orders arrive at rate μ,
  market buy orders arrive at rate μ,
  ask orders at price p are canceled at rate $\theta(p - p_B(t))|X_p(t)|$ for $p > p_B(t)$,
  bid orders at price p are canceled at rate $\theta(p_A(t) - p)|X_p(t)|$ for $p < p_A(t)$

Cont *et al.* (2010) show that under these assumptions, X is an ergodic Markov process and has a proper stationary distribution.

Let $\Phi_g = \{t_i : i \in \mathbb{Z}\}$ be the process of all market order time stamps, with $t_i < t_{i+1}$, $i \in \mathbb{Z}$. In this order book set-up, let the marks y_i be given by the absolute change in mid-prices: $y_i = p_M(t_i) - p_M(t_{i-1})$. Note that, if we assume the price level to be large compared to the price changes y_i, then the y_i are approximately proportional to the logreturns: $\log(p_M(t_i)/p_M(t_{i-1})) \approx (p_M(t_i) - p_M(t_{i-1}))/p_M$. Note that all prices are left-continuous functions of time. Let s_i denote a second mark of transaction i representing the spread immediately before the $(i-1)$-th transaction. Let $z_i \in \{1, -1\}$ indicate whether the $(i-1)$-th market order was a sell $(+1)$ or a buy (-1) order. Let d_i be the duration between the $(i-1)$-th and the i-th market order and let $x_{A,i}$ and $x_{B,i}$ be the values of the order book process at the respective ask and bid price immediately before the $(i-1)$-th market order. Then we consider the MPP Φ, given by the locations of all market orders, marked by the vector of

price change, spread, sign, duration and the amount of outstanding limit orders at ask and bid price: $\Phi = \{[t_i, (y_i, s_i, z_i, d_i, x_{A,i}, x_{B,i})] : i \in \mathbb{Z}\}$. Let us write $m_i = (y_i, s_i, z_i, d_i, x_{A,i}, x_{B,i})$ for the whole vector of marks.

In what follows, we aim to give an explicit representation of the E- and V-function of the MPP Φ w.r.t. the price change component y. Let $l : \mathbb{R} \times \mathbb{M}_0 \to (0, \infty)$ be a function that returns the calendar time of the last market order before time t, i.e., $l(t, \varphi) = \max\{t_i \in \varphi : t_i < t\}$. With this notation, we have $x_{A,i} = X_{p_A(l(t_i, \Phi))}(l(t_i, \Phi))$.

We make the additional assumption that whenever a sell limit order is placed at a price p_1 or a buy limit order is placed at a price p_2 and there exists a price $q_1 > p_1$ or a price $q_2 < p_2$ such that $X_{q_1}(t) = 0$ or $X_{q_2}(t) = 0$, respectively, then automatically additional limit orders are placed that fill these gaps. This additional assumption guarantees that neither p_A nor p_B can be moved by more than one tick by a single market order. We also exclude the cancellation of limit orders, i.e., we set $\theta \equiv 0$.

Moreover, we consider a modification of the E- and V-function, which only takes into account those market orders, for which the spread is minimal, i.e., one tick. We define $\tilde{E} = \frac{d\tilde{\alpha}_e^{(2)}}{d\tilde{\alpha}_1^{(2)}}$ and $\tilde{V} = \frac{d\tilde{\alpha}_v^{(2)}}{d\tilde{\alpha}_1^{(2)}}$ with

$$\tilde{\alpha}_f^{(2)}(B_1 \times B_2) = \int_{\mathbb{M}_0} \sum_{(t_1, m_1), (t_2, m_2) \in \varphi}^{\neq} \mathbf{1}_{B_1}(t_1) \mathbf{1}_{B_2}(t_2) f(y_1) \mathbf{1}_{s_1 = 1} \, \mathbb{P}^{\Phi}(d\varphi).$$

Theorem 4.2.9. *Let Φ be an MPP as described above. We pretend that, for a market order $(t_i, m_i) \in \Phi$, the events $\{s_i = 1\}$ (i.e, the bid-ask spread before the $(i-1)$-th market order is a single tick), $\{X_{p_A(l(t_i, \Phi))}(l(t_i, \Phi)) = 1 \,|\, s_i = 1\}$ and $\{X_{p_B(l(t_i, \Phi))}(l(t_i, \Phi)) = -1 \,|\, s_i = 1\}$ are stochastically independent of the unmarked process Φ_{g} of all market order time-stamps. Then the modified V-function $\tilde{V}$ is given by*

$$\tilde{V}(r) = \frac{1}{2} M^{(x_A)|s=1}(\{1\}) \left(1 - \frac{\lambda_{\mathrm{g}}}{2(2\lambda_0 + \lambda_{\mathrm{g}})} + \mathbf{1}_{r < 0} \cdot \frac{\lambda_0}{2\lambda_0 + \lambda_{\mathrm{g}}} \exp(r \cdot (2\lambda_0 + \lambda_{\mathrm{g}}))\right),$$

where $\lambda_{\mathrm{g}} = 2\mu$ is the intensity of the Poisson process of market orders and $M^{(x_A)|s=1}$ is the Palm mark distribution of the x_A-component conditional on the s-component being equal to 1.

Note that the above independence assumptions are clearly not satisfied by the process under consideration. However, we take this as an approximation to the truth in order to allow for an analytical calculation of the modified V-function. Due to the symmetry of the model, the E-function equals 0 almost everywhere.

Proof of Theorem 4.2.9. Conditioning on $\Phi(B_2 \times \mathbb{M}^*)$ and applying Corollary 2.2.5, part 1,

yields

$$
\tilde{\alpha}_f^{(2)}(B_1 \times B_2)
$$

$$
= \int_{\mathbb{N}} \int_{\mathbb{M}_0} \sum_{(t_1,m_1),(t_2,m_2)\in\varphi}^{\neq} \mathbf{1}_{B_1}(t_1)\mathbf{1}_{B_2}(t_2)f(y_1)\mathbf{1}_{s_1=1}\, \mathbb{P}^{\Phi|\Phi_\mathrm{g}(B_2)=n}(d\varphi)\, \mathbb{P}^{\Phi_\mathrm{g}(B_2)}(dn)
$$

$$
= \int_{\mathbb{N}} \int_{\mathbb{R}} \int_{\mathbb{M}^*} \int_{\mathbb{M}_0} \mathbf{1}_{B_1}(t_1)f(y_1)\mathbf{1}_{s_1=1}\varphi_\mathrm{g}(B_2)\, P_{(t_1,y_1)}^{!'}(d\varphi)\, M'_{t_1}(dy_1)\, \Lambda'_\mathrm{g}(dt_1)\, \mathbb{P}^{\Phi_\mathrm{g}(B_2)}(dn),
$$

where for fixed $n \in \mathbb{N}$, $P^{!'}$, M' and Λ'_g refer to the process $\Phi' = \Phi|\{\Phi_\mathrm{g}(B_2) = n\}$. Because of the independence property of a Poisson process, we have $\Lambda'_\mathrm{g}(\cdot \cap (\mathbb{R}\backslash B_2)) = \Lambda_\mathrm{g}(\cdot \cap (\mathbb{R}\backslash B_2))$ and since Φ_g is stationary, $\Lambda_\mathrm{g}(dx) = \lambda_\mathrm{g}dx$. With $M^{(\cdot)}$ denoting the marginal distribution of the respective mark component, we have

$$
\tilde{\alpha}_f^{(2)}(B_1 \times B_2)
$$

$$
= \int_{\mathbb{R}} \int_{\mathbb{N}} \int_{\mathbb{M}^{*,(y)}} \int_{\mathbb{M}_0} \mathbf{1}_{B_1}(t_1)f(y_1)n\, P_{(t_1,y_1|s=1)}^{!'}(d\varphi) M_{t_1}^{'(y)|s=1}(dy_1) M_{t_1}^{'(s)}(\{1\})\mathbb{P}^{\Phi_\mathrm{g}(B_2)}(dn)\lambda_\mathrm{g}dt_1
$$

$$
= \lambda_\mathrm{g} \int_{\mathbb{R}} \mathbf{1}_{B_1}(t_1) \int_{\mathbb{N}} n M_{t_1}^{'(s)}(\{1\}) \int_{\mathbb{M}^{*,(y)}} f(y_1)\, M_{t_1}^{'(y)|s=1}(dy_1)\, \mathbb{P}^{\Phi_\mathrm{g}(B_2)}(dn)\, dt_1. \tag{4.21}
$$

Since the last transaction before t_1 has a spread of 1, the change of mid-price (which is given by y_1), can only take one of the five values 0, $\pm\frac{1}{2}$, ± 1:

- It equals 0 if the market order at time $l(t_1, \varphi)$ does not move the ask and the bid price. This is the case for each of the following situations:
    - $z(l(t_1,\varphi)) = -1$ and $X_{p_A(l(t_1,\Phi))}(l(t_1,\Phi)) > 1$
    - $z(l(t_1,\varphi)) = +1$ and $X_{p_B(l(t_1,\Phi))}(l(t_1,\Phi)) < -1$
    - $z(l(t_1,\varphi)) = -1$ and $X_{p_A(l(t_1,\Phi))}(l(t_1,\Phi)) = 1$ and a new sell limit order at $p_A(l(t_1,\Phi))$ arrives until time t_1
    - $z(l(t_1,\varphi)) = +1$ and $X_{p_B(l(t_1,\Phi))}(l(t_1,\Phi)) = -1$ and a new buy limit order at $p_B(l(t_1,\Phi))$ arrives until time t_1

- It equals $\frac{1}{2}$ if $X_{p_A(l(t_1,\Phi))}(l(t_1,\Phi)) = 1$ and $z(l(t_1,\varphi)) = -1$ and no new limit order at $p_A(l(t_1,\Phi))$ arrives between time $l(t_1,\Phi)$ and t_1. In other words, the last limit order at price p_A is consumed, this increases the ask price by one tick and hence the mid-price is increased by half a tick.

- It equals 1 if $X_{p_A(l(t_1,\Phi))}(l(t_1,\Phi)) = 1$ and $z(l(t_1,\varphi)) = -1$ and a new buy limit order at $p_A(l(t_1,\Phi))$ arrives between time $l(t_1,\Phi)$ and t_1. I.e., both the bid price and the ask price are increased by one tick, and so is the mid-price.

- Analogously for the values $-\frac{1}{2}$ and -1.

We consider the inner part of the RHS of equation (4.21):

$$\int_{\mathbb{N}} n M_{t_1}^{'(s)}(\{1\}) \int_{\mathbb{M}^{*,(y)}} f(y_1)\, M_{t_1}^{'(y)|s=1}(dy_1)\, \mathbb{P}^{\Phi_{\mathrm{g}}(B_2)}(dn). \qquad (4.22)$$

Let $f(y) = v(y) = y^2$. Then, $f(y)$ takes values in $\{0, \frac{1}{4}, 1\}$ only. Since the order book model is symmetric w.r.t. the ask and bid side, we have

$$M_t^{'(x_A)}(\{1\}) = M_t^{'(x_B)}(\{-1\}),$$
$$M_t^{'(z)}(\{1\}) = M_t^{'(z)}(\{-1\}) = \frac{1}{2}$$

for $t \in \mathbb{R}$. Since buy and sell market orders arrive with the same rate and independently of the processes of limit orders, and since the duration component is also independent of the mark components s, z, x_A and x_B, M_t' can be decomposed in the following way:

$$M_t'(dm) = M_t^{'(y)|s,z,x_A,x_B,d}(dy) M_t^{'(x_A,x_B)|s}(d(x_A, x_B)) M_t^{'(s)}(ds) M_t^{'(z)}(dz) M_t^{'(d)}(dd). \qquad (4.23)$$

For notational convenience, we define the events

$$A(t) = \{X_{p_A(l(t,\Phi))}(l(t,\Phi)) = 1, z(l(t,\Phi)) = -1, s(l(t,\Phi)) = 1\},$$
$$B(t) = \{X_{p_B(l(t,\Phi))}(l(t,\Phi)) = -1, z(l(t,\Phi)) = 1, s(l(t,\Phi)) = 1\}.$$

Also note that the arrival rate of new ask and bid orders between $l(t,\Phi)$ and t at price $p_A(l(t,\Phi))$ and $p_B(l(t,\Phi))$, given the events $A(t)$ and $B(t)$, respectively, is $\lambda(1) = \lambda_0$. Hence, the probability that neither a new ask limit order nor a new bid limit order arrive at these prices within a period of length d is $\exp(-2\lambda_0 d)$.

W.l.o.g. we assume $B_2 = [a_2, b_2]$ to be an interval. If $t_1 < \min(B_2)$, then the mark at location t_1 and $\Phi_{\mathrm{g}}(B_2)$ are obviously independent and hence $M_{t_1}' = M_{t_1}$. Since the unconditional MPP Φ is stationary, the mark distribution does not depend on the location, i.e., $M_{t_1} = M$. Then

$$(4.22) = \int_{\mathbb{N}} n M^{(s)}(\{1\}) \mathbb{P}^{\Phi_{\mathrm{g}}(B_2)}(dn) \int_{\mathbb{M}^{*,(y)}} f(y_1)\, M^{(y)|s=1}(dy_1)$$

$$= M^{(s)}(\{1\}) \Lambda_{\mathrm{g}}(B_2) \int_{\mathbb{M}^{*,(y)}} f(y_1)\, M^{(y)|s=1}(dy_1)$$

$$= M^{(s)}(\{1\}) \lambda_{\mathrm{g}} |B_2|$$
$$\cdot \int_{\mathbb{M}^{*,(y,x_A,x_B,d)}} f(y_1)\, M^{(y)|s=1,x_A,x_B,d}(dy_1)\, M^{(x_A,x_B)|s=1}(d(x_A,x_B))\, M^{(d)}(dd)$$

$$= M^{(s)}(\{1\}) \lambda_{\mathrm{g}} |B_2| \cdot 2 M^{(x_A)|s=1}(\{1\}) M^{(z)}(\{-1\})$$
$$\int_0^\infty \left(\frac{1}{4} \exp(-2\lambda_0 d) + 1 \cdot (1 - \exp(-2\lambda_0 d)) \frac{1}{2} \right) \lambda_{\mathrm{g}} \exp(-\lambda_{\mathrm{g}} d) dd$$

$$= M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})\lambda_{\mathrm{g}}^2|B_2|$$
$$\int_0^\infty \left(1 - \frac{1}{2}\exp(-2\lambda_0 d)\right)\exp(-\lambda_{\mathrm{g}}d)dd$$
$$= M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})|B_2|\left(\lambda_{\mathrm{g}} - \frac{\lambda_{\mathrm{g}}^2}{2(2\lambda_0 + \lambda_{\mathrm{g}})}\right).$$

If $t_1 > \max(B_2)$, we condition on $\Phi_{\mathrm{g}}(B_2)$ and $\Phi_{\mathrm{g}}([b_2, t_1])$. Then

$$(4.22) = \int_{\mathbb{N}\times\mathbb{N}} nM_{t_1}^{\prime(s)}(\{1\})$$
$$\int_{\mathbb{M}^{*,(y)}} f(y_1)\, M_{t_1}^{\prime(y)|s=1}(dy_1)\, \mathbb{P}^{[\Phi_{\mathrm{g}}(B_2),\,\Phi_{\mathrm{g}}([b_2,t_1])]}(dn\times dj),$$

where now M_{t_1}' refers to the process $\Phi' = \Phi\,|\,\{\Phi_{\mathrm{g}}(B_2) = n,\, \Phi_{\mathrm{g}}([b_2,t_1]) = j\}$. We decompose the Palm mark distribution M_{t_1}' analogously to (4.23). Conditional on $\Phi_{\mathrm{g}}(B_2) = n$, $\Phi_{\mathrm{g}}([b_2,t_1]) = j$, the distribution $M_{t_1}^{\prime(d)}$ corresponds to that of a maximum of independent, uniformly distributed random variables. Recall that, by assumption, $M_{t_1}^{\prime(x_A)|s=1}(\{1\}) = M^{(x_A)|s=1}(\{1\})$, $M_{t_1}^{\prime(x_B)|s=1}(\{-1\}) = M^{(x_B)|s=1}(\{-1\})$, and $M_{t_1}^{\prime(s)}(\{1\}) = M^{(s)}(\{1\})$. Then,

$$(4.22) = \int_{\mathbb{N}}\int_{\mathbb{N}} nM_{t_1}^{\prime(s)}(\{1\})$$
$$\int_{\mathbb{M}^{*,(y,x_A,x_B,d)}} f(y_1)\, M_{t_1}^{\prime(y)|s=1,x_A,x_B,d}(dy_1)\, M_{t_1}^{\prime(x_A,x_B)|s=1}(d(x_A,x_B))\, M_{t_1}^{\prime(d)}(dd)$$
$$\mathbb{P}^{\Phi_{\mathrm{g}}([b_2,t_1])}(dj)\, \mathbb{P}^{\Phi_{\mathrm{g}}(B_2)}(dn)$$
$$= \sum_{n=1}^\infty n\exp(-\lambda_{\mathrm{g}}|B_2|)\frac{(\lambda_{\mathrm{g}}|B_2|)^n}{n!}$$
$$\left(\sum_{j=1}^\infty \exp(-\lambda_{\mathrm{g}}(t_1 - b_2))\frac{(\lambda_{\mathrm{g}}(t_1 - b_2))^j}{j!}\right.$$
$$\int_0^{t_1-b_2}\left(\frac{1}{4}\exp(-2\lambda_0 d) + (1 - \exp(-2\lambda_0 d))\frac{1}{2}\right)$$
$$2M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})\cdot j\frac{(t_1 - b_2 - d)^{j-1}}{(t_1 - b_2)^j}\,dd$$
$$+\exp(-\lambda_{\mathrm{g}}(t_1 - b_2))$$
$$\int_{t_1-b_2}^{t_1-a_2}\left(\frac{1}{4}\exp(-2\lambda_0 d) + (1 - \exp(-2\lambda_0 d))\frac{1}{2}\right)$$
$$\left.2M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})\cdot n\frac{(t_1 - a_2 - d)^{n-1}}{(b_2 - a_2)^n}\,dd\right)$$

$$= \exp(-\lambda_g(t_1 - b_2)) M^{(s)}(\{1\}) M^{(x_A)|s=1}(\{1\}) M^{(z)}(\{-1\})$$

$$\sum_{n=1}^{\infty} n \exp(-\lambda_g |B_2|) \frac{(\lambda_g |B_2|)^n}{n!}$$

$$\cdot \left(\int_0^{t_1-b_2} \sum_{j=1}^{\infty} \frac{(\lambda_g(t_1 - b_2))^j}{j!} \left(1 - \frac{1}{2}\exp(-2\lambda_0 d)\right) j \frac{(t_1 - b_2 - d)^{j-1}}{(t_1 - b_2)^j} \, dd \right.$$

$$\left. + \int_{t_1-b_2}^{t_1-a_2} \left(1 - \frac{1}{2}\exp(-2\lambda_0 d)\right) n \frac{(t_1 - a_2 - d)^{n-1}}{(b_2 - a_2)^n} \, dd \right)$$

$$= \exp(-\lambda_g(t_1 - b_2)) M^{(s)}(\{1\}) M^{(x_A)|s=1}(\{1\}) M^{(z)}(\{-1\})$$

$$\cdot \left(\lambda_g |B_2| \int_0^{t_1-b_2} \lambda_g \sum_{j=1}^{\infty} \frac{(\lambda_g(t_1 - b_2 - d))^{j-1}}{(j-1)!} \left(1 - \frac{1}{2}\exp(-2\lambda_0 d)\right) \, dd \right.$$

$$\left. + \int_{t_1-b_2}^{t_1-a_2} \lambda_g \sum_{n=1}^{\infty} n \exp(-\lambda_g |B_2|) \frac{(\lambda_g(t_1 - a_2 - d))^{n-1}}{(n-1)!} \left(1 - \frac{1}{2}\exp(-2\lambda_0 d)\right) \, dd \right)$$

$$= \exp(-\lambda_g(t_1 - b_2)) M^{(s)}(\{1\}) M^{(x_A)|s=1}(\{1\}) M^{(z)}(\{-1\})$$

$$\cdot \left(\lambda_g |B_2| \int_0^{t_1-b_2} \lambda_g \exp(\lambda_g(t_1 - b_2 - d)) \left(1 - \frac{1}{2}\exp(-2\lambda_0 d)\right) \, dd \right.$$

$$+ \int_{t_1-b_2}^{t_1-a_2} \lambda_g \exp(-\lambda_g |B_2|) \exp(\lambda_g(t_1 - a_2 - d)) \, (\lambda_g(t_1 - a_2 - d) + 1)$$

$$\left. \left(1 - \frac{1}{2}\exp(-2\lambda_0 d)\right) \, dd \right)$$

$$= M^{(s)}(\{1\}) M^{(x_A)|s=1}(\{1\}) M^{(z)}(\{-1\}) \lambda_g$$

$$\cdot \left(\lambda_g |B_2| \int_0^{t_1-b_2} \left(\exp(-\lambda_g d) - \frac{1}{2}\exp(-(2\lambda_0 + \lambda_g)d)\right) \, dd \right.$$

$$\left. + \int_{t_1-b_2}^{t_1-a_2} (\lambda_g(t_1 - a_2 - d) + 1) \left(\exp(-\lambda_g d) - \frac{1}{2}\exp(-(2\lambda_0 + \lambda_g)d)\right) \, dd \right)$$

$$= M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})\,\lambda_{\mathrm{g}}$$
$$\cdot\left(\lambda_{\mathrm{g}}|B_2|\left(\frac{1-\exp(-\lambda_{\mathrm{g}}(t_1-b_2))}{\lambda_{\mathrm{g}}}-\frac{1-\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-b_2))}{2(2\lambda_0+\lambda_{\mathrm{g}})}\right)\right.$$
$$+\left[-(t_1-a_2-d)\exp(-\lambda_{\mathrm{g}}d)\right.$$
$$\left.\left.+\frac{1}{2}\left[\frac{\lambda_{\mathrm{g}}(t_1-a_2-d)}{2\lambda_0+\lambda_{\mathrm{g}}}+\frac{2\lambda_0}{(2\lambda_0+\lambda_{\mathrm{g}})^2}\right]\exp(-(2\lambda_0+\lambda_{\mathrm{g}})d)\right]_{t_1-b_2}^{t_1-a_2}\right)$$

(by partial integration)

$$= M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})\,\lambda_{\mathrm{g}}$$
$$\cdot\left(\lambda_{\mathrm{g}}|B_2|\left(\frac{1-\exp(-\lambda_{\mathrm{g}}(t_1-b_2))}{\lambda_{\mathrm{g}}}-\frac{1-\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-b_2))}{2(2\lambda_0+\lambda_{\mathrm{g}})}\right)\right.$$
$$+(b_2-a_2)\exp(-\lambda_{\mathrm{g}}(t_1-b_2))$$
$$+\frac{2\lambda_0}{2(2\lambda_0+\lambda_{\mathrm{g}})^2}\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-a_2))$$
$$-\frac{\lambda_{\mathrm{g}}(b_2-a_2)}{2(2\lambda_0+\lambda_{\mathrm{g}})}\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-b_2))$$
$$\left.-\frac{2\lambda_0}{2(2\lambda_0+\lambda_{\mathrm{g}})^2}\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-b_2))\right)$$
$$= M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})\,\lambda_{\mathrm{g}}$$
$$\cdot\left(\lambda_{\mathrm{g}}|B_2|\left(\frac{1}{\lambda_{\mathrm{g}}}-\frac{1}{2(2\lambda_0+\lambda_{\mathrm{g}})}\right)\right.$$
$$\left.+\frac{2\lambda_0}{2(2\lambda_0+\lambda_{\mathrm{g}})^2}\left[\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-a_2))-\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-b_2))\right]\right)$$
$$= M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})$$
$$\cdot\left(|B_2|\left(\lambda_{\mathrm{g}}-\frac{\lambda_{\mathrm{g}}^2}{2(2\lambda_0+\lambda_{\mathrm{g}})}\right)\right.$$
$$\left.+\frac{\lambda_0\lambda_{\mathrm{g}}}{(2\lambda_0+\lambda_{\mathrm{g}})^2}\left[\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-a_2))-\exp(-(2\lambda_0+\lambda_{\mathrm{g}})(t_1-b_2))\right]\right).$$

Hence, if $t_1<\min(B_2)$,

$$\tilde{\alpha}_v^{(2)}(B_1\times B_2)$$
$$= M^{(s)}(\{1\})M^{(x_A)|s=1}(\{1\})M^{(z)}(\{-1\})\cdot\lambda_{\mathrm{g}}\left(\lambda_{\mathrm{g}}-\frac{\lambda_{\mathrm{g}}^2}{2(2\lambda_0+\lambda_{\mathrm{g}})}\right)\int_{B_1}\int_{B_2}dt_2dt_1$$

and if $t_1 > \min(B_2)$,

$$
\begin{aligned}
&\tilde{\alpha}_v^{(2)}(B_1 \times B_2) \\
&= M^{(s)}(\{1\}) M^{(x_A)|s=1}(\{1\}) M^{(z)}(\{-1\}) \\
&\quad \cdot \lambda_{\mathrm{g}} \int_{B_1} \int_{B_2} \Big(\lambda_{\mathrm{g}} - \frac{\lambda_{\mathrm{g}}^2}{2(2\lambda_0 + \lambda_{\mathrm{g}})} + \frac{\lambda_0 \lambda_{\mathrm{g}}}{2\lambda_0 + \lambda_{\mathrm{g}}} \exp(-(2\lambda_0 + \lambda_{\mathrm{g}})(t_1 - t_2)) \Big) dt_2 dt_1.
\end{aligned}
$$

For $f \equiv 1$, we have (4.22) $= M^{(s)}(\{1\}) \mathbb{E}(\Phi_{\mathrm{g}}(B_2)) = M^{(s)}(\{1\}) \lambda_{\mathrm{g}} |B_2|$ and hence

$$
\tilde{\alpha}_1^{(2)}(B_1 \times B_2) = M^{(s)}(\{1\}) \int_{B_1} \int_{B_2} \lambda_{\mathrm{g}}^2 \, dt_2 dt_1.
$$

Then

$$
\begin{aligned}
\tilde{V}(r) &= \frac{d\tilde{\alpha}_v^{(2)}}{d\tilde{\alpha}_1^{(2)}}(0, r) \\
&= M^{(x_A)|s=1}(\{1\}) M^{(z)}(\{-1\}) \\
&\quad \left(1 - \frac{\lambda_{\mathrm{g}}}{2(2\lambda_0 + \lambda_{\mathrm{g}})} + \mathbf{1}_{r<0} \cdot \frac{\lambda_0}{2\lambda_0 + \lambda_{\mathrm{g}}} \exp(r \cdot (2\lambda_0 + \lambda_{\mathrm{g}})) \right).
\end{aligned}
$$

$\square$

The fact that the function $\tilde{V}$ is constant on the positive half-axis merely reflects the independence property of the Poisson point process. Conditioning on past transactions, however, leads to an exponential increase of the variance in the above model.

4.3 Valid bi-directional E- and V-functions

Having calculated the theoretical E- and V-function for different MPP models in Section 4.2, we now turn towards the question which functions belong to the class of valid bi-directional E- and V-functions of stationary MPPs on $\mathbb{R}$. To this end, we provide a modification of a construction principle proposed in Schlather (2001, Section 5), which is based on bivariate Gaussian variables and Matérn hard-core processes.

Theorem 4.3.1. *Let R be a positive number and let E^*, V^* and C^* be real-valued functions on $\mathbb{R}$ with*

$$
E^*(r)^2 \leq V^*(r), \tag{4.24}
$$

$$
C^*(-r) = C^*(r), \tag{4.25}
$$

$$
\left| C^*(r) - E^*(r)E^*(-r) \right| \leq \sqrt{[V^*(r) - E^*(r)^2] \cdot [V^*(-r) - E^*(-r)^2]} \tag{4.26}
$$

for $r \in \mathbb{R}$. Let

$$m \in \left(\inf_{0 < r < R} \frac{E^*(-r) + E^*(r)}{2}, \sup_{0 < r < R} \frac{E^*(-r) + E^*(r)}{2} \right). \tag{4.27}$$

Then there exists a stationary MPP Φ on $\mathbb{R}$ such that for almost all $r \in [-R, R]$,

$$\mu_e^{(2)}(r) = E^*(r), \quad \mu_v^{(2)}(r) = V^*(r), \quad \mu_c^{(2)}(r) = C^*(r),$$

and for almost all r with $|r| > R$,

$$\mu_e^{(2)}(r) = \mu_e^{(1)} = m, \quad \mu_c^{(2)}(r) = m^2,$$

where c is the function that maps a tuple of marks (y_1, y_2) to its product $y_1 y_2$.

Proof. The proof is similar to that of Schlather (2001, Theorem 5.1). Condition (4.27) implies the existence of a probability measure P with $\text{supp}(P) = [0, R]$ such that

$$\int_0^R \frac{E^*(-r) + E^*(r)}{2} P(dr) = m.$$

Let Π_1 be a stationary MPP on $\mathbb{R}$ with iid marks $(\alpha, \xi, \eta_1, \eta_2)$, where α is a random sign with $\mathbb{P}(\alpha = \pm 1) = 0.5$, $\xi \sim P$, $\eta_1, \eta_2 \sim \mathcal{N}(0, 1)$, and all mark components are independent of each other. By a thinning procedure, we obtain a second stationary MPP Π_2. In particular, we only keep those points of Π_1 that do not have a neighbor within a distance less than or equal to $4R$, i.e.,

$$\Pi_2 = \left\{ [t, (\alpha, \xi, \eta_1, \eta_2)] \in \Pi_1 : [t - 4R, t + 4R] \cap \Pi_{1,\text{g}} = \{t\} \right\}.$$

Then, Φ is constructed as follows: The set of point locations of Φ is

$$\left\{ t, t + \alpha\xi : [t, (\alpha, \xi, \eta_1, \eta_2)] \in \Pi_2 \right\}.$$

In other words, each point of Π_2 gets a neighbor at the random distance $\alpha\xi$. The marks of each of these pairs of points are given by a bivariate Gaussian vector with a covariance matrix constructed from the functions E^*, V^* and C^*. For $r \in \mathbb{R}$, let

$$\sigma_1^2(r) = V^*(r) - E^*(r)^2,$$
$$\sigma_2^2(r) = V^*(-r) - E^*(-r)^2,$$
$$\tau(r) = C^*(r) - E^*(r)E^*(-r)$$

and let $A(r)$ be a symmetric (2×2)-matrix such that $A(r)^2 = \begin{pmatrix} \sigma_1^2(r) & \tau(r) \\ \tau(r) & \sigma_2^2(r) \end{pmatrix}$. Note that $A(r)^2$ is a valid covariance matrix if and only if conditions (4.24)–(4.26) are satisfied. To each pair

of point locations $(t, t + \alpha\xi)$, we then assign the mark vector

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = A(\alpha\xi) \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} E^*(\alpha\xi) \\ E^*(-\alpha\xi) \end{pmatrix},$$

which yields a bivariate Gaussian distribution with covariance matrix $A(\alpha\xi)^2$. In summary,

$$\Phi = \left\{ [t, (1,0)A(\alpha\xi)\left(\begin{smallmatrix}\eta_1\\\eta_2\end{smallmatrix}\right)], [t + \alpha\xi, (0,1)A(\alpha\xi)\left(\begin{smallmatrix}\eta_1\\\eta_2\end{smallmatrix}\right)] : [t, (\alpha, \xi, \eta_1, \eta_2)] \in \Pi_2 \right\}.$$

By construction, the distance between two points of Φ is contained in $[-R, R]$ if and only if the two points represent a pair $(t, t + \alpha\xi)$ (with $t \in \Pi_{2,g}$); the marks of points with absolute distance larger than R are stochastically independent. We consider the tuple $(y_1, y_2, \alpha, \xi, \eta_1, \eta_2)$ as a random vector on some space $(\Omega, \mathcal{A}, \mathbb{P})$ and denote the expectation w.r.t. $\mathbb{P}$ by $\mathbb{E}$. Then, for $r \in [-R, R]$, we have

$$\mu_e^{(2)}(r) = \mathbb{E}[y_1 \,|\, \alpha\xi = r] = \mathbb{E}[y_2 \,|\, \alpha\xi = -r] = E^*(r),$$
$$\mu_v^{(2)}(r) = \mathbb{E}[y_1^2 \,|\, \alpha\xi = r] = \mathbb{E}[y_2^2 \,|\, \alpha\xi = -r] = \sigma_1^2(r) + \mathbb{E}[y_1 \,|\, \alpha\xi = r]^2 = V^*(r),$$
$$\mu_c^{(2)}(r) = \mathbb{E}[y_1 y_2 \,|\, \alpha\xi = r] = \tau(r) + E^*(r)E^*(-r) = C^*(r)$$

for P'-almost all r, where $P'(\,\cdot\,) = \frac{1}{2}(P(\,\cdot\,) + P(-\,\cdot\,))$. For $|r| > R$, we have

$$\mu_e^{(2)}(r) = \mathbb{E}y_1 = \mathbb{E}[E^*(\alpha\xi)] = \int_0^R \frac{E^*(-r) + E^*(r)}{2} \, P(dr) = m,$$
$$\mu_v^{(2)}(r) = \mathbb{E}y_1^2 = \mathbb{E}[E^*(\xi^2)] = \int_0^R E^*(r^2) \, P(dr),$$
$$\mu_c^{(2)}(r) = (\mathbb{E}y_1)^2 = m^2$$

for almost all r, which closes the proof. $\qquad\square$

Note that V^* and C^* are the uncentered second-order moment measures of Φ. For the centered V-function $V(r) = \mu_v^{(2)}(r) - \mu_e^{(2)}(r)^2$, there are analog conditions to (4.24)–(4.26). The construction principle in Theorem 4.3.1 is based on isolating pairs of points by a "hard-core thinning" of a Poisson point process. The radius R gives the range up to which the E- and V-function can be freely controlled—only subject to the (modified) conditions (4.24)–(4.26). Similar principles can be used to control the general second-order moment measure $\mu_f^{(2)}$ or higher-order moment measures, i.e., interaction of three or more points.

5 Intrinsically weighted means of marked point processes

This chapter is based on the manuscript Malinowski *et al.* (2012*a*), submitted to *Advances in Applied Probability.* Note that the definitions of mean mark in this chapter are slight extensions compared to the definitions in Chapter 2 in that they include a weighting component that might assign a different weight to each point of the MPP. While the focus of the previous two chapters was rather on conditional variances of the marks, in what follows, arbitrary moments of the marks are considered with particular regard to non-ergodic or non-stationary MPPs.

5.1 Introduction

Marked point processes (MPPs) provide an adequate framework for modeling irregularly scattered events in space or time in that they incorporate the joint distribution of the observed values and the point locations (e.g., Karr, 1991; Møller & Waagepetersen, 2003; Schlather *et al.*, 2004; Daley & Vere-Jones, 2008; Myllymäki & Penttinen, 2009; Diggle *et al.*, 2010). Due to the variety of possible forms of dependence between marks and locations in an MPP framework, already the notion of the mean, which is usually considered as being the simplest summary statistic, rises tantalizing and challenging questions.

An introductory example for the type of MPP averages being considered within this chapter is the trading process in financial markets. Transactions of assets are typically characterized by the two quantities price and volume; a benchmark quantity that is of major interest especially for institutional investors is the so-called volume-weighted average price (VWAP) (e.g., Madhavan, 2002; Bialkowski *et al.*, 2008). The VWAP of n transactions with prices p_i and traded volumes v_i, $i = 1, \ldots, n$, is defined as $p_{\mathrm{VWAP}} = \sum (p_i v_i) / \sum v_i$.
We embed this example in the following general MPP framework: We consider stationary MPPs on $\mathbb{R}^d$ of the form

$$\Phi = \{(t_i, y_i, z_i) : i \in \mathbb{N}\},$$

where $t_i \in \mathbb{R}^d$ is the point location, $y_i \in \mathbb{R}$ is the first mark and $z_i \in [0, \infty)$ is a second mark of the ith point of Φ. Formally, the mark at t_i is given by the vector $(y_i, z_i) \in \mathbb{R} \times [0, \infty)$. Let $\Phi_{\mathrm{g}} = \{t : (t, y, z) \in \Phi\}$ denote the ground process of point locations of Φ and let us denote the marks at a location $t \in \Phi_{\mathrm{g}}$ by $y(t)$ and $z(t)$. The non-negativity assumption on the z-component simplifies technical assumptions when employing this mark component as weights in weighted averages of the first mark component $y(t)$, or of $f(y(t))$ for some

function $f : \mathbb{R} \to \mathbb{R}$. In intuitive notation we write the corresponding weighted mean as

$$\mu_f^{(1)} = \mathbb{E}[z(t)f(y(t)) \,|\, t \in \Phi_{\mathrm{g}}], \tag{5.1}$$

where we assume that the z-component is normalized such that $\mathbb{E}[z(t) \,|\, t \in \Phi_{\mathrm{g}}] = 1$. Here, the conditioning on "$t \in \Phi_{\mathrm{g}}$" is understood in the sense of the Palm mark distribution (Stoyan *et al.*, 1995, Chap. 4). Since the weights $z(t)$ are provided by the MPP itself and may depend on both the marks $y(t)$ and the point locations $t \in \Phi_{\mathrm{g}}$, we refer to $\mu_f^{(1)}$ as *intrinsically weighted mean mark* of Φ. The formal definition of $\mu_f^{(1)}$ and related quantities will be given at the beginning of Section 5.2.

When a system of randomly distributed objects is modeled by means of MPPs, there can exist different sensible choices of intrinsic weights $z(t)$ leading to different weighted mean marks that are relevant for one and the same process, but for different statistical questions:

- *Average height of trees:* Consider n forests of about equal size, each of which is sampled on an area with fixed size and shape. Then the unweighted average of the height of all trees provides a measure of the entire timber stand, which is relevant for forest inventory applications. This amounts to $z(t) = 1$ in (5.1). Additionally, the average height of a typical *forest* (as opposed to a typical tree) might be of interest, independently of how dense the trees occur in the different forests. Then, a nested definition of mean seems to be adequate where we first average within each forest and then between all forests. This is equivalent to using a weighted average over all trees with $z(t)$ being proportional to the inverse of the number of trees in the forest that location t belongs to.

- *Density of insects on plants (cf. Begon et al., 1990):* Consider n plants and a population of insects distributed over the plants. Let k_i, $i = 1, \ldots, n$, be the number of insects on the ith plant. In this set-up there are different well-established definitions of density referring to different ecological effects. The ordinary density of insects, also called *resource-weighted density*, is $(k_1 + \ldots + k_n)/n$ and quantifies the average availability of resources. In contrast, the *organism-weighted density* is the density that an average insect experiences. Each individual on plant i experiences a density of k_i insects per plant, i.e., the organism-weighted density is $(k_1^2 + \ldots + k_n^2)/(k_1 + \ldots + k_n)$. In MPP notation, each insect is represented by a point, marked by the total number of insects on the plant on which the insect is located. Then the organism-weighted density corresponds to the ordinary mean mark ($z(t) = 1$), whereas the resource-weighted density is the average of all plant-wise averages of the marks, i.e., $z(t) = (nk_i)^{-1} \sum_{i=1}^{n} k_i$ if t belongs to plant i.

- *Sampling of continuous-space processes:* Taking measurements of continuous-space or continuous-time processes usually aims at estimating or predicting the underlying process and the mean of interest is therefore the spatial or temporal mean over the whole domain of the process. Since measurement locations are not necessarily independent of the underlying process, knowledge of the pattern of point locations might already provide information about the values of the process. This situation is commonly

referred to as *biased* or *preferential sampling* and different weighting approaches exist to correct for this form of biases (e.g., Isaaks & Srivastava, 1989). Although most statistical methods only use stationarity, ergodicity is often implicitly assumed. In case of non-ergodicity, which means that different realizations can have a different stochastic behavior, we are faced with an additional dimension of biasedness: Within each ergodic subclass, the pattern of point locations can be independent of the underlying process, while there might be a strong dependence between the pattern of measurement locations and the process itself if multiple realizations are considered. For a simple example, consider a Gaussian random field with a random mean m combined with a Poisson point process of measurement locations whose intensity of points is a function of m.

While ergodicity of MPPs is necessary for a straightforward interpretation of the mark distribution as the distribution of a typical point and, at least implicitly, is required by many applications for consistent estimation, in this chapter, we investigate the behavior of moment-based summary statistics in case of non-ergodic MPPs and intend to point out problems of ambiguity in this context. When the different forests and plants in the above examples are perceived as a set of MPP realizations and exhibit different ecological characteristics, non-ergodicity has to be included. Examples for non-ergodic MPPs that evolve in time can easily be found in the financial world: For subsequent days of asset trading, the process of executed transactions can be considered as different realizations of a possibly non-ergodic MPP. To treat non-ergodic MPPs adequately, we propose intrinsically weighted mean marks as a special case of (5.1) in which the weights are constant within each ergodicity class but allow for compensating for differences between the different ergodicity classes.

The remainder of this chapter is organized as follows: In Section 5.2 we recall and generalize moment-based characteristics for MPPs which also form the central tool for the analysis of interactions in MPPs. We study their behavior and interpretation for non-ergodic processes and, following the idea of the above examples, propose alternative definitions of moment-based summary statistics in Section 5.3. Different estimators for the above characteristics and their asymptotic properties are discussed in Section 5.4; the chapter closes with a comparison of the point process set-up with estimation of continuous-space processes, which typically occur within geostatistical applications. Sections 5.7 and 5.8 review some basic results from ergodic theory and contain the proof of Section 5.4, respectively.

5.2 MPP moment-measures and measurement of interaction effects

Throughout this chapter, $\Phi = \{(t_i, y_i, z_i) : i \in \mathbb{N}\}$ is a stationary marked point process on $\mathbb{R}^d$ with marks $(y(t_i), z(t_i)) = (y_i, z_i) \in \mathbb{R} \times [0, \infty)$, and $\Phi_{\mathrm{g}} = \{t : (t, y, z) \in \Phi\}$ is its ground process of point locations. In particular, the point configuration Φ_{g} is locally finite. Let us remark that the following definitions of MPP statistics can directly be generalized to MPPs on Polish spaces (cf. Kallenberg, 1986) whose marks are also in a Polish space.

One of the most basic mark summary statistic is the weighted mean mark $\mu_f^{(1)}$, which we introduced in (5.1) as a conditional mean, conditional on the event $\{t \in \Phi_{\mathrm{g}}\}$. Since for fixed

$t \in \mathbb{R}^d$, this is a zero-probability event, the classical formal definition is

$$\mu_f^{(1)} = \frac{\mathbb{E}\sum_{(t,y,z)\in\Phi} zf(y)\mathbf{1}_B(t)}{\mathbb{E}\sum_{(t,y,z)\in\Phi} z\mathbf{1}_B(t)} \tag{5.2}$$

for any Borel set $B \subset \mathbb{R}^d$ with $\nu(B) > 0$, where ν denotes the Lebesgue measure. Here we implicitly exclude the degenerate case $z(t) \equiv 0$. Due to the stationarity of Φ, this definition does not depend on the choice of B.

Proposition 5.2.1. *Both definitions of $\mu_f^{(1)}$, (5.1) and (5.2), coincide.*

Proof. We use standard arguments of marked point process theory, similar to Daley & Vere-Jones (2008, Chap. 13). Let $g : \mathbb{R}^d \times \mathbb{R} \times [0,\infty) \to \mathbb{R}$ be an integrable function w.r.t. the intensity measure $\Lambda(B \times L) = \mathbb{E}(\Phi(B \times L))$, $B \in \mathcal{B}(\mathbb{R}^d)$, $L \in \mathcal{B}(\mathbb{R} \times [0,\infty))$. Stationarity allows to decompose Λ into the product $\lambda\nu \times M$, where M is the (Palm) mark distribution and $\lambda > 0$ is the intensity of the ground process Φ_{g}. Campbell's theorem states $\mathbb{E}\sum_{(t,y,z)\in\Phi} g(t,y,z) = \lambda \iint g(t,y,z)M(d(y,z))\nu(dt)$ and plugging in $g(t,y,z) = zf(y)\mathbf{1}_B(t)$ yields

$$\mathbb{E} \sum_{(t,y,z)\in\Phi} zf(y)\mathbf{1}_B(t) = \lambda\nu(B) \int zf(y)\, M(d(y,z)) = \lambda\nu(B)\mathbb{E}\left[Zf(Y)\right],$$

for a random vector (Y,Z) distributed according to M. Thus, the right-hand side (RHS) of (5.2) is $\mathbb{E}\left[Zf(Y)\right]/[\mathbb{E}Z]$, which equals (5.1) under the assumption $\mathbb{E}[z(t)\,|\,t \in \Phi_{\mathrm{g}}] = 1$. $\qquad\square$

The most relevant example of f in practical application is $f(y) = y^n$ for $n = 1, 2, \dots$ Then, if $z(t) = 1$ for $t \in \Phi_{\mathrm{g}}$, $\mu_f^{(1)}$ simply represents the n-th moment of the (Palm) mark distribution. Note that in case the MPP represents measurements of an underlying continuous process, the mean mark can substantially differ from the mean of the underlying process due to stochastic dependence between the sampling locations and the process itself.

While the above statistic $\mu_f^{(1)}$ reflects (average) properties of single points, second-order characteristics (in intuitive notation $\mathbb{E}[f(y(t_1), y(t_2))\,|\,t_1, t_2 \in \Phi_{\mathrm{g}}, t_1 \neq t_2]$) provide a framework to investigate dependency structures within MPPs. We use the superscripts [(1)] and [(2)] to indicate whether first- or second-order measures are meant.

Definition 5.2.2. For any non-negative function f on $\mathbb{R} \times \mathbb{R}$, we define a σ-finite measure on $\mathbb{R}^d \times \mathbb{R}^d$ by

$$\alpha_f^{(2)}(C) = \mathbb{E} \sum_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi}^{\neq} z_1 f(y_1, y_2)\mathbf{1}_C((t_1, t_2)), \quad C \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d), \tag{5.3}$$

which we call *weighted second moment measure*. Here, "$\neq$" indicates that the sum runs over all pairs of points with $(t_1, y_1) \neq (t_2, y_2)$.

With the notation

$$C(B, I) = \begin{cases} \{(t_1, t_2) : t_1 \in B, t_2 \in t_1 + I\}, & d = 1, \\ \{(t_1, t_2) : t_1 \in B, t_2 \in t_1 + \{x \in \mathbb{R}^d : \|x\| \in I\}\}, & d > 1, \end{cases}$$
$$C(t, I) = C([\mathbf{0}, t], I),$$
$$C(I) = C([\mathbf{0}, \mathbf{1}], I),$$

for $B \in \mathcal{B}(\mathbb{R}^d)$, $t \in \mathbb{R}^d$, $I \in \mathcal{B}(\mathbb{R})$, which slightly differs from the definition in Chapter 2 w.r.t. the I-component,

$$\alpha_f^{(2)}(C(I)), \quad I \in \mathcal{B}(\mathbb{R}), \tag{5.4}$$

defines a σ-finite measure on $\mathbb{R}$. Well-known examples of second-order mark characteristics for stationary and isotropic MPPs are Cressie's mark variogram and covariance function (Cressie, 1993), Stoyan's k_{mm}-function (Stoyan, 1984), and Isham's mark correlation function (Isham, 1985), which can all be expressed in terms of (5.3) or (5.4) with a constant z-component and which are all functions of the Euclidean distance between two points of a stationary and isotropic process. Schlather (2001) provided a unifying notation for the above characteristics and further introduced new functions, E and V, where $E(r)$ and $V(r)$ represent the mean and variance of a mark, respectively, given that there exists a further point at distance $r > 0$. For the one-dimensional case, e.g., for temporal processes, in Chapter 3, we have extended those characteristics to the non-isotropic set-up, where a negative value of r means that the point that is conditioned on is in the past. The above second-order characteristics only involve the three functions $f(y_1, y_2) = y_1 y_2$, $f(y_1, y_2) = y_1$ and $f(y_1, y_2) = y_1^2$.

Definition 5.2.3. (cf. Schlather, 2001). For a general non-negative function f on $\mathbb{R} \times \mathbb{R}$, we define

$$\mu_f^{(2)}(I) = \frac{\alpha_f^{(2)}(C(I))}{\alpha^{(2)}(C(I))}, \quad I \in \mathcal{B}(\mathbb{R}), \tag{5.5}$$

if $\alpha^{(2)}(C(I)) > 0$. Here, $\alpha^{(2)}$ is short notation for $\alpha_f^{(2)}$ with $f \equiv 1$. We call $\mu_f^{(2)}$ the *(weighted) second-order mean mark*.

In the following, we always assume that I is chosen such that $\alpha^{(2)}(C(I)) > 0$. Note that the distinction between $d = 1$ and $d > 1$ in the definition of the set $C(B, I)$ allows to capture a possibly anisotropic behavior of $\mu_f^{(2)}$ in the one-dimensional case. In particular,

$$\alpha_f^{(2)}(C(I)) = \begin{cases} \mathbb{E}_\Phi \sum_{(t_1, y_1, z_1), (t_2, y_2, z_2) \in \Phi, \ t_1 \in [0,1]}^{\neq} z_1 f(y_1, y_2) \mathbf{1}_{t_2 - t_1 \in I}, & d = 1 \\ \mathbb{E}_\Phi \sum_{(t_1, y_1, z_1), (t_2, y_2, z_2) \in \Phi, \ t_1 \in [0,1]}^{\neq} z_1 f(y_1, y_2) \mathbf{1}_{\|t_2 - t_1\| \in I}, & d > 1. \end{cases}$$

For higher dimensions, it is also possible to assign different directions of isotropy, but the technical burden increases considerably as $\mu_f^{(2)}$ will not be a function of a scalar argument anymore. For further notational convenience, we assume that the derivative of $\alpha_f^{(2)}$ w.r.t. the

Lebesgue measure exists, which is then referred to as *product density* and denoted by $\rho_f^{(2)}$.

Due to the stationarity of Φ, we have $\rho_f^{(2)}(t_1, t_2) = \rho_f^{(2)}(0, t_2 - t_1)$ for almost all $(t_1, t_2) \in \mathbb{R}^{2d}$ and hence $\alpha_f^{(2)}(C) = \int_C \rho_f^{(2)}(0, h_2 - h_1) d(h_1 \times h_2)$, $C \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$. Let $\rho_f^{C,(2)}(r)$, $r \in \mathbb{R}$, denote the derivative of $\alpha_f^{(2)}(C(\cdot))$ w.r.t. the *one-dimensional* Lebesgue measure. Obviously, $\alpha_f^{(2)}(C(\cdot))$ is dominated by $\alpha^{(2)}(C(\cdot))$, which ensures that the limit of $\mu_f^{(2)}(I)$ for $|I| \to 0$ exists and can be expressed in terms of Radon-Nikodym derivatives. For $r \neq 0$ we define

$$\mu_f^{(2)}(r) = \left. \frac{\partial \alpha_f^{(2)}(C(\cdot))}{\partial \alpha^{(2)}(C(\cdot))} \right|_{\cdot = r} = \frac{\rho_f^{C,(2)}(r)}{\rho_1^{C,(2)}(r)}. \tag{5.6}$$

Note that for $d = 1$, we have $\rho_f^{C,(2)}(r) = \rho_f^{(2)}(0, r)$. With a slight abuse of notation, we refer to both definitions (5.5) and (5.6) as $\mu_f^{(2)}$, i.e., formally, $\mu_f^{(2)}$ is a function on $\mathcal{B}(\mathbb{R}^d) \cup \mathbb{R} \backslash \{0\}$. For $r \neq 0$ and f only depending on its first argument, $\mu_f^{(2)}(r)$ can be interpreted as the (weighted) expectation of a mark at location t subject to the conditioning that Φ has a point at location t and at location $t + r\mathbf{e}_1$, i.e., $\mu_f^{(2)}(r) = \mathbb{E}[z(t) f(y(t)) \,|\, t, t + r\mathbf{e}_1 \in \Phi_g]$, where $\mathbf{e}_1$ denotes the vector $(1, 0, \ldots, 0)^T \in \mathbb{R}^d$. For $\mu_f^{(2)}(I)$, this interpretation becomes slightly ambiguous: Considering an event at time t, there may be multiple other points located within the set $t + I$ and in case that interactions of higher order are present, these will be reflected by the second-order statistic $\mu_f^{(2)}(I)$ as well. More precisely, by the definitions in (5.5) and (5.6),

$$\mu_f^{(2)}(I) = \alpha^{(2)}(C(I))^{-1} \int_I \mu_f^{(2)}(r) \, d\alpha^{(2)}(C(r)), \tag{5.7}$$

i.e., $\mu_f^{(2)}(I)$ is a weighted average of conditional expectations $\mu_f^{(2)}(r)$ with weights proportional to the expected number of pairs of points with distance dr.

Remark 5.2.4. *(a) The extension to moment measures of higher order is straightforward and allows to condition on arbitrary point constellations. In practice, however, mostly first- and second-order statistics are considered.*

(b) The non-negativity condition on f can be weakened by considering the restriction of $\mu_f^{(2)}(\cdot)$ to some bounded set $J \in \mathcal{B}(\mathbb{R})$. Then it is sufficient for f that $\alpha_h^{(2)}(C(J)) < \infty$ is satisfied for $h = f_+ = \max\{f, 0\}$ or for $h = f_- = -\min\{f, 0\}$.

(c) Another generalization allows to include further conditioning on the marks and is related to Section 6.2, where $\mu_f^{(2)}$ is being adapted to extreme value statistics. For f_{cond} a non-negative function on $\mathbb{R} \times \mathbb{R}$ we consider

$$\mu_{f, f_{\text{cond}}}^{(2)}(I) = \frac{\alpha_{f \cdot f_{\text{cond}}}^{(2)}(C(I))}{\alpha_{f_{\text{cond}}}^{(2)}(C(I))} = \frac{\mu_{f \cdot f_{\text{cond}}}^{(2)}(I)}{\mu_{f_{\text{cond}}}^{(2)}(I)}. \tag{5.8}$$

Choosing f_{cond} to be an indicator function $f_{\text{cond}}(y_1, y_2) = \mathbf{1}_A(y_1)\mathbf{1}_B(y_2)$ implicitly conditions the marks on the events A and B, respectively.

Remark 5.2.5. *For $d > 1$, $\mu_f^{(2)}$ is a function of the Euclidean distance between two points, whereas for $d = 1$, $\mu_f^{(2)}$ is a function of the signed distance. In the latter case, $\mu_f^{(2)}(\cdot)$ is in general not symmetric: Consider a temporal process consisting of pairs of points (t_1, t_2) with $t_1 < t_2$ and with small intra- but large inter-pair distances. Assume that the marks of different pairs are stochastically independent and that for each pair of points $f(y_1, y_2) > f(y_2, y_1)$ holds. Then $\mu_f^{(2)}(r) > \mu_f^{(2)}(-r)$ holds for all $r > 0$ which are small enough and which can occur as intra-pair distances.*

For notational convenience, we will write $\mu_f^{(i)}$ to indicate that a statement is valid for $\mu_f^{(1)}$ and $\mu_f^{(2)}$. Recall that $\mu_f^{(1)}$ is a single number, whereas $\mu_f^{(2)} = \mu_f^{(2)}(\cdot)$ is a function of distance.

5.3 New moment measures for non-ergodic MPPs

Ergodicity makes spatial averages over suitably increasing observation windows of a single realization converge to the corresponding expectation over the state space:

$$|W|^{-1} \int_W X(T_x \Phi) \, dx \xrightarrow{\text{a.s.}} \mathbb{E}(X(\Phi)), \quad \text{for } |W| \to \infty \text{ suitably,}$$

for any integrable function X on the space of all locally finite counting measures. Here, T_x denotes the shift of the whole random point pattern Φ by $x \in \mathbb{R}^d$. In essence, ergodicity enables consistent estimation of MPP moment measures by observing a single realization on a suitably increasing domain. In this section, though, we consider the opposite situation, namely where Φ is a non-ergodic process. Although the above definitions of $\mu_f^{(i)}$ are valid irrespective of ergodicity properties, their meaning is slightly more complex in the non-ergodic case, and alternative possibilities of defining conditional mark expectations arise canonically, as we see in the following.

The following proposition directly relates to the fact that a non-ergodic MPP can be seen as a hierarchical model, which, in a first step, draws an ergodic *source of randomness* out of which the final realization is drawn in a second step.

Proposition 5.3.1. *Let Φ be a non-ergodic MPP with probability law P. By $\mathbb{M}_0$ and $\mathcal{M}_0$ we denote the space of all locally finite counting measures on $\mathbb{R}^d \times \mathbb{R} \times [0, \infty)$ and the usual σ-algebra, respectively. See Section 5.7 for more details. Then*

$$\mu_f^{(1)} = \frac{\mathbb{E}_Q\left[\mu_{f,\Phi|Q}^{(1)} \cdot \alpha_{\Phi|Q}^{(1)}(B)\right]}{\alpha^{(1)}(B)}, \quad \mu_f^{(2)}(\cdot) = \frac{\mathbb{E}_Q\left[\mu_{f,\Phi|Q}^{(2)}(\cdot)\alpha_{\Phi|Q}^{(2)}(C(\cdot))\right]}{\alpha^{(2)}(C(\cdot))}, \quad (5.9)$$

where $Q \sim \lambda$ is a random variable with values in the space $\mathcal{P}_{\text{erg}}$ of all ergodic MPP probability laws, distributed according to some probability measure λ, such that $P(M) =$

$\int_{\mathcal{P}_{\mathrm{erg}}} Q^*(M)\lambda(dQ^*)$, $M \in \mathcal{M}_0$. If $\mu_f^{(2)}$ is evaluated for a fixed distance $r \in \mathbb{R}$, $\alpha^{(2)}(C(r))$ has to be replaced by $\rho_1^{C,(2)}(r)$ in (5.9).

Proof. The ergodic decomposition theorem (cf. Theorem 5.7.5) guarantees the existence and uniqueness of a decomposition $P(\cdot) = \int_{\mathcal{P}_{\mathrm{erg}}} Q^*(\cdot)\lambda(dQ^*)$ and a corresponding mixing random variable $Q \sim \lambda$. Conditioning Φ on Q, we can decompose the moment measures $\alpha_f^{(i)}$ and obtain

$$
\mu_f^{(2)}(r) = \frac{\partial \mathbb{E}_Q \alpha_{f,\Phi|Q}^{(2)}(C(\cdot))}{\partial \alpha^{(2)}(C(\cdot))}\bigg|_{\cdot = r} = \frac{\partial \mathbb{E}_Q \alpha_{f,\Phi|Q}^{(2)}(C(\cdot))/\partial\nu(\cdot)\big|_{\cdot = r}}{\partial \alpha^{(2)}(C(\cdot))/\partial\nu(\cdot)\big|_{\cdot = r}}
$$
$$
= \frac{\mathbb{E}_Q \rho_{f,\Phi|Q}^{C,(2)}(0,r)}{\rho_1^{C,(2)}(0,r)} = \frac{\mathbb{E}_Q\big[\mu_{f,\Phi|Q}^{(2)}(r) \cdot \rho_{1,\Phi|Q}^{C,(2)}(0,r)\big]}{\rho_1^{C,(2)}(0,r)},
$$

where ν denotes the Lebesgue measure. For $\mu_f^{(2)}(I)$ and $\mu_f^{(1)}$, the decomposition is analogous. $\square$

Example 5.3.2. *The so-called log-Gaussian Cox process (Møller et al., 1998) is ergodic if and only if the underlying stationary Gaussian random field Z is ergodic. A sufficient condition for Z being ergodic is that the covariance function decays to zero. Amongst others, Myllymäki & Penttinen (2009) and Diggle et al. (2010) use log-Gaussian Cox processes, combined with an intensity-dependent marking, as parametric models for preferential sampling applications.*

Proposition 5.3.1 shows that in case of non-ergodicity, $\mu_f^{(i)}$ is an average of its ergodic subclasses counterparts, in which each class Q^* is implicitly weighted by the respective intensity $\alpha_{\Phi|Q=Q^*}^{(i)}$. If all ergodic subprocesses $[\Phi|Q = Q^*]$ have the same intensity measure, the weights cancel out and we have $\mu_f^{(i)} = \mathbb{E}_Q \mu_{f,\Phi|Q}^{(i)}$. Though, if $\alpha_{\Phi|Q}^{(i)}$ is not constant for all Q, a single ergodicity class Q^* with a small probability of occurrence $P(Q = Q^*)$ might be able to drive the value of the moment measure $\mu_f^{(i)}$ by means of a large value of $\alpha_{\Phi|Q=Q^*}^{(i)}$. More precisely, let $\mathcal{Q}$ be a family of ergodic MPP distributions s.t. $P(Q \in \mathcal{Q}) = 1$. Assume that we can choose the values of $\mu_{f,\Phi|Q=Q^*}^{(2)}(r)$ independently of the location pattern and vice versa. Then for given values of $\mu_{f,\Phi|Q=Q^*}^{(2)}(r)$, $Q^* \in \mathcal{Q}$, we can choose the second-order product densities $\rho_{\Phi|Q}^{(2)}(0,r)$ in such a way that $\mu_f^{(2)}(r)$ takes any value m for which $P^Q(\mu_{f,\Phi|Q}^{(2)}(r) > m) > 0$ and $P^Q(\mu_{f,\Phi|Q}^{(2)}(r) < m) > 0$. In light of this observation, the demand for a new characteristic $\tilde{\mu}_f^{(i)}$ arises naturally, that summarizes the properties of all ergodicity classes, irrespectively of how the processes of point locations differ between the different ergodicity classes. We meet these requirements by a definition that excludes the implicit weighting proportional to the ith order intensities:

Definition 5.3.3. Let λ and Q be the ergodic decomposition mixture measure and mixture variable, respectively, of Φ, and let $\mathbb{E}_Q|\mu_{f,\Phi|Q}^{(i)}| < \infty$. Then we call

$$\tilde{\mu}_f^{(i)} = \mathbb{E}_Q\mu_{f,\Phi|Q}^{(i)} = \int_{\mathcal{P}_{\mathrm{erg}}} \mu_{f,\Phi|Q=Q^*}^{(i)} \; \lambda(dQ^*). \tag{5.10}$$

the *(equally-weighted) average ith-order mean mark* of Φ.

Relating to the introductory forest example, the classical definition of the mean mark in (5.2) corresponds to the average height of all trees, irrespectively of differences w.r.t. the tree densities between the different forests, while the new definition in (5.10) refers to the average height of a typical forest.

Remark 5.3.4. *Comparing the new definition with (5.9) yields that $\tilde{\mu}_f^{(i)}$ coincides with $\mu_f^{(i)}$ if $\alpha_{\Phi|Q}^{(i)}$ is λ-a.s. constant. This is particularly the case if Φ is ergodic.*

Lemma 5.3.5. *For any $I \in \mathcal{B}(\mathbb{R})$ we have*

$$\tilde{\mu}_f^{(2)}(I) = \mathbb{E}_Q\left[\alpha_{\Phi|Q}^{(2)}(C(I))^{-1} \int_I \mu_{f,\Phi|Q}^{(2)}(r) \; d\alpha_{\Phi|Q}^{(2)}(C(r))\right].$$

If, for λ-almost all measures Q^*, $\mu_{f,\Phi|Q=Q^*}^{(2)}(r)$ is uniformly bounded by some positive constant $c(Q^*)$ and $\mathbb{E}_Qc(Q) < \infty$, for $I \in \mathcal{B}(\mathbb{R})$ and $r \in \mathbb{R}$, we have

$$\lim_{I\to\{r\}} \tilde{\mu}_f^{(2)}(I) = \tilde{\mu}_f^{(2)}(r).$$

Proof. The first assertion follows directly from applying the representation (5.7) to the ergodic subprocesses $[\Phi|Q = Q^*]$. Since $\lim_{I\to\{r\}} \mu_f^{(2)}(I) = \mu_f^{(2)}(r)$ by construction, the second assertion is merely an application of Lebesgue's dominated convergence theorem. $\qquad\square$

From Lemma 5.3.5 we see that the nested conditional mean $\tilde{\mu}_f^{(2)}(r)$ is a Radon-Nikodym derivative of $\alpha_f^{(2)}(C(\cdot))$ w.r.t. $\alpha^{(2)}(C(\cdot))$ if and only if the expectation of $\alpha_{\Phi|Q}^{(2)}(C(\cdot))\mu_{f,\Phi|Q}^{(2)}(\cdot)$ factorizes. This contrasts the ordinary conditional mean $\mu_f^{(2)}(r)$, which is already defined as a Radon-Nikodym derivative of $\alpha_f^{(2)}(C(\cdot))$ w.r.t. $\alpha^{(2)}(C(\cdot))$.

The ergodic decomposition and an analog to Definition 5.3.3 can be applied to any expectation-based functional of an MPP including the Palm mark distribution itself. While the classical definition of the mean mark represents a typical point, irrespectively of the different ergodicity classes, the two-stage-expectation $\tilde{\mu}_f^{(i)}$ refers to the mean of a typical realization. We provide more details on the meaning of the differences between $\mu_f^{(i)}$ and $\tilde{\mu}_f^{(i)}$ and between different estimators in the next section.

5.4 Estimation principles for the new MPP moment-measures

5.4.1 The ergodic case

For ergodic processes Φ, the pointwise ergodic theorem for MPPs (Proposition 5.7.4) yields that

$$\mathbb{E}\left[\sum_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi}^{\neq} z_1 f(y_1,y_2)\mathbf{1}_{(t_1,t_2)\in C(I)}\right]$$

$$= \lim_{n\to\infty}\left[n^{-d}\sum_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\varphi}^{\neq} z_1 f(y_1,y_2)\mathbf{1}_{(t_1,t_2)\in C(n\mathbf{1},I)}\right]$$

for almost all realizations φ of Φ, which builds the basis for the estimators being discussed in this section. For readability reasons, and since we will be only dealing with second-order statistics from now on, we drop the superscript $^{(2)}$ in all the estimators of $\mu_f^{(2)}$.

Applying the standard estimator for MPP moment measures to a realization of Φ observed on the set $[\mathbf{0},\mathbf{T}]$, $\mathbf{T}\in(0,\infty)^d$, we obtain

$$\hat{\mu}_f(I,\Phi,\mathbf{T}) = \frac{\hat{\alpha}_f(I,\Phi,\mathbf{T})}{\hat{\alpha}_1(I,\Phi,\mathbf{T})}, \tag{5.11}$$

where $\hat{\alpha}_f(I,\Phi,\mathbf{T}) = \sum_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi}^{\neq} z_1 f(y_1,y_2)\mathbf{1}_{(t_1,t_2)\in C(\mathbf{T},I)}$.

Lemma 5.4.1. *If Φ is ergodic, $\hat{\mu}_f(I,\Phi,\mathbf{T})$ is consistent for $\mu_f^{(2)}(I)$. Here, "$\mathbf{T}\to\infty$" is understood componentwise. If Φ is non-ergodic, $\hat{\mu}_f(I,\Phi,\mathbf{T})$ is consistent if and only if $\mu_{f,\Phi|Q=Q^*}^{(2)}(I)$ is constant w.r.t. Q^*.*

Proof. By Proposition 5.7.4, the tuple consisting of the numerator and the denominator of (5.11), each normalized by the volume of $[0,\mathbf{T}]$, converges a.s. to $(\alpha_f^{(2)}(C(I)),\alpha^{(2)}(C(I)))$ if Φ is ergodic. The first assertion thus follows from the continuous mapping theorem. In the non-ergodic case, clearly only $\mu_{f,\Phi|Q=Q^*}^{(2)}(I)$ can be estimated consistently for Q^* being the respective ergodicity class. Though, if $\mu_{f,\Phi|Q=Q^*}^{(2)}(I)$ is constant w.r.t. Q^* we have $\mu_f^{(2)}(I) = \mu_{f,\Phi|Q=Q^*}^{(2)}(I)$ for any $Q^*\in\mathcal{P}_{\mathrm{erg}}$. $\qquad\square$

To establish asymptotic normality of $\hat{\mu}_f(I,\Phi,\mathbf{T})$, we introduce some idealized assumptions. In particular, we assume stochastic independence between the point locations and the marks of the MPP. For simplicity, we restrict to the case where f only depends on its first argument and the MPP is a process on $\mathbb{R}$.

Condition 5.4.2 (*m-dependent Random Field Model*)**.** *Let $\tilde{\Phi}$ be a stationary unmarked point process on $\mathbb{R}$, for which neighboring points have some minimum distance $d_0 > 0$. Let $\{Y(t) : t\in\mathbb{R}\}$ be an independent stationary process with finite second moments and a covariance function C that has finite range, i.e., $C(h) = 0$ for all $|h| > h_0$ for some $h_0 > 0$.*

Then, with $m = [d_0/h_0]$, we say that an MPP Φ is an m-dependent Random Field Model, if $\Phi \overset{d}{=} \{(t_i, Y(t_i), 1) \mid t_i \in \tilde{\Phi}\}$.

The following theorem transfers a central limit theorem (CLT) for arrays of m-dependent random variables to the MPP context. It also covers a thinning of the MPP in which the threshold increases with the observation window. The result allows to derive asymptotically exact confidence intervals for the estimator of $\mu_f^{(2)}(I)$ and will be specialized in Chapter 6 in the context of extreme value analysis for MPPs.

Theorem 5.4.3 (CLT for m-dependent Random Field Models). *Let Φ be an ergodic MPP that satisfies Condition 5.4.2. For $f : \mathbb{R} \to [0,\infty)$ and $u \geq 0$, let $f_u, f_{\mathrm{cond},u} : \mathbb{R} \to [0,\infty)$ be given by $f_u(y) = (f(y) - u)_+ = (f(y) - u)\mathbf{1}_{f(y)>u}$ and $f_{\mathrm{cond},u}(y) = \mathbf{1}_{f(y)>u}$. Let*

$$\hat{\alpha}_{f_u}^*(I, \Phi, T) = \sum_{(t_1,y_1),(t_2,y_2) \in \Phi}^{\neq} \left(f_u(y_1) - \mu_{f_u, f_{\mathrm{cond},u}}^{(2)}(I) \right) \cdot f_{\mathrm{cond},u}(y_1) \cdot \mathbf{1}_{(t_1,t_2) \in C(\mathbf{T},I)}$$

be a centered version of $\hat{\alpha}_{f_u}(I, \Phi, T)$, where $\mu_{f_u, f_{\mathrm{cond},u}}^{(2)}(I)$ is defined as in (5.8). Let $(u_T)_{T \geq 0}$ be a family of non-negative, non-decreasing numbers such that the following conditions are satisfied:

$$\lim_{T \to \infty} \mathbb{E}\left[f_{u_T}(Y(0))^i \mid f(Y(0)) > u_T \right] < \infty \quad (i = 1, \ldots, 4),$$

$$\frac{T^{-1}\hat{\alpha}_1(I, \Phi, T) - \lambda}{\mathbb{E}_\Phi \hat{\alpha}_{f_{\mathrm{cond},u_T}}(I, \Phi, 1)} \to 0 \quad a.s. \ (T \to \infty).$$

Then, for $I \in \mathcal{B}(\mathbb{R})$ and $T \to \infty$, we have

$$\frac{\hat{\alpha}_{f_{u_T}}^*(I, \Phi, T)}{\sqrt{\hat{\alpha}_{f_{\mathrm{cond},u_T}}(I, \Phi, T)}} \Rightarrow \mathcal{N}(0, s_{u_\infty}),$$

where

$$s_{u_\infty} = \lim_{T \to \infty} \left\{ (\lambda_{u_T} T)^{-1} \mathrm{Var}\left[\hat{\alpha}_{f_{u_T}}^*(I, \Phi, T) \right] \right\},$$

$$\lambda_u = \mathbb{E}_\Phi \left[\hat{\alpha}_{f_{\mathrm{cond},u}}(I, \Phi, 1) \right], \quad u \geq 0.$$

The proof is given in Section 5.8. Note that the asymptotic variance s_{u_∞} can be given in a more explicit form for suitable choices of f and suitable distributional assumptions on the underlying random field Y. A related CLT result was provided by Heinrich & Molchanov (1999) for random measures associated to germ-grain models.

5.4.2 The non-ergodic case

If Φ is non-ergodic, consistent estimation of summary statistics generally requires multiple realizations of the process. Let P and λ denote the probability law and the ergodic mixture

measure of Φ, respectively. Then, drawing iid realizations of Φ corresponds to drawing ergodicity classes according to the mixture measure λ. Though, a finite collection of realizations merely approximates the mixing measure λ and we can only expect consistency if both n and $\mathbf{T}$ tend to infinity simultaneously. To see why $n \to \infty$ is not sufficient, consider an MPP with infinitely many ergodicity classes $Q_1, Q_2, \ldots$ and with $\mathbb{E}_{\Phi|Q=Q_i}\Phi([0,1]) = 2^{-i}$. Then, for fixed $\mathbf{T}$, the probability of observing at least one point in a realization that belongs to class i tends to zero as $i \to \infty$. Hence, the classes Q_i, for i large, are only captured by the estimator if $\mathbf{T}$ also tends to infinity.

Considering iid realizations $\Phi_1, \ldots, \Phi_n$ of Φ, different possibilities arise of how to put together the respective estimators. Let $\mathbf{w} = (w_1, \ldots, w_n)$ denote a vector of weight functions $w_i : \mathbb{M}_0 \times [0,\infty)^d \to [0,\infty)$. We assume that for λ-almost all ergodic MPP laws Q^* there exist constants $w_i^*(Q^*) \geq 0$ with $w^*(Q^*) = \sum_{i=1}^n w_i^*(Q^*) > 0$ to which the weights converge stochastically within the respective ergodicity class, i.e.,

$$P_{\Phi|Q=Q^*}\left(|w_i(\Phi, \mathbf{T}) - w_i^*(Q^*)| > \varepsilon\right) \longrightarrow 0 \quad (\mathbf{T} \to \infty) \tag{5.12}$$

for all $\varepsilon > 0$. Then we consider estimators of the form

$$\hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}) = \hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), \mathbf{T})$$

$$= \left(\sum w_i(\Phi_i, \mathbf{T})\right)^{-1} \sum_{i=1}^n w_i(\Phi_i, \mathbf{T})\hat{\mu}_f(I, \Phi_i, \mathbf{T}). \tag{5.13}$$

Note that the functions w_i might also depend on I. With $w_1 = \ldots = w_n = n^{-1}$, we obtain as a special case

$$\hat{\mu}_f^n(I) = \hat{\mu}_f^n(I, (\Phi_1, \ldots, \Phi_n), \mathbf{T}) = n^{-1} \sum_{i=1}^n \hat{\mu}_f(I, \Phi_i, \mathbf{T}). \tag{5.14}$$

As summarized by equation (5.9), $\mu_f^{(2)}(I)$ is the weighted expectation of its subclass counterparts $\mu_{f,\Phi|Q}^{(2)}(I)$, weighted proportional to the second-order point intensity $\alpha_{\Phi|Q}^{(2)}(C(I))$. Hence, in order that (5.13) consistently estimates $\mu_f^{(2)}(I)$ for all possible combinations of $\mu_{f,\Phi|Q=Q^*}^{(2)}(I)$ and $\alpha_{\Phi|Q=Q^*}^{(2)}(C(I))$, the weights have essentially to be chosen as

$$w_i(\Phi_i, \mathbf{T}) = \hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)/v_{\mathbf{T}} = \sum_{t_1, t_2 \in \Phi_{i,\mathrm{g}}}^{\neq} \mathbf{1}_{(t_1, t_2) \in C(\mathbf{T}, I)}/v_{\mathbf{T}}, \tag{5.15}$$

where $v_{\mathbf{T}}$ is the volume of the cube $[\mathbf{0}, \mathbf{T}]$. By Proposition 5.7.4, $\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)/v_{\mathbf{T}}$ converges to $\alpha_{\Phi|Q=Q_i}^{(2)}(C(I))$ a.s. as $\mathbf{T} \to \infty$, where Q_i is the realized ergodicity class of Φ_i. With $\mathbf{w}$ being the vector of weights from (5.15), we define

$$\hat{\mu}_f^{\alpha}(I, (\Phi_1, \ldots, \Phi_n), \mathbf{T}) = \hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), \mathbf{T}), \tag{5.16}$$

which, in a sense, represents the family of *all* pairs of points with a distance contained in I from all realizations. This choice of weights satisfies the above stochastic convergence condition (5.12) and is sufficient but not necessary for consistency. The following theorem gives a weaker set of conditions that is still sufficient for consistency

Theorem 5.4.4. *Let Φ_i, $i \in \mathbb{N}$, be iid copies of a possibly non-ergodic MPP Φ and let Q_{j_i} denote the respective ergodicity classes. For weight functions $\tilde{w}_i : \mathbb{M}_0 \times [0, \infty)^d \to [0, \infty)$ and iid random factors W_i with $\mathbb{E}|W_i| < \infty$, $i \in \mathbb{N}$, let $w_i(\Phi_i, \mathbf{T}) = W_i \cdot \tilde{w}_i(\Phi_i, \mathbf{T})$ and $\mathbf{w} = (w_1(\Phi_1, \mathbf{T}), \dots, w_n(\Phi_n, \mathbf{T}))$. Then, $\hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w})$ is consistent for $\mu_f^{(2)}(I)$ if the following conditions hold:*

$$W_i > 0 \qquad a.s., \tag{5.17}$$

$$\mathrm{Var}\, \tilde{w}_i(\Phi_i, \mathbf{T}) \leq c_1 \qquad \text{for some } c_1 > 0, \tag{5.18}$$

$$n^{-1}\mathbb{E}\sum_{i=1}^{n} \tilde{w}_i \geq c_2 > 0 \qquad \forall n \geq n_0 \text{ for some } n_0 \in \mathbb{N}, \tag{5.19}$$

$$\mathbb{E}\left[W_i \tilde{w}_i(\Phi_i, \mathbf{T})\right] = \mathbb{E}\left[W_i\right] \cdot \mathbb{E}\left[\tilde{w}_i(\Phi_i, \mathbf{T})\right], \tag{5.20}$$

$$\mathbb{E}\left[W_i \cdot \hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)\mu_{f, \Phi|Q=Q_{j_i}}^{(2)}(I)\right] = \mathbb{E}\left[W_i\right] \cdot \mathbb{E}\left[\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)\mu_{f, \Phi|Q=Q_{j_i}}^{(2)}(I)\right], \tag{5.21}$$

$$\mathbb{P}\left\{\max_{i=1}^{n} \left|\frac{\tilde{w}_i(\Phi_i, \mathbf{T})\sum_{j=1}^{n}\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_j)}{\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)\sum_{j=1}^{n}\tilde{w}_j(\Phi_j, \mathbf{T})}\right| > c_3\right\} \to 0 \quad (n, \mathbf{T} \to \infty). \tag{5.22}$$

Proof. We consider

$$\left|\frac{\sum_{i=1}^{n} w_i(\Phi_i, \mathbf{T})\hat{\mu}_f(I, \Phi_i, \mathbf{T})}{\sum_{i=1}^{n} w_i(\Phi_i, \mathbf{T})} - \mu_f^{(2)}(I)\right|$$

$$\leq \left|\frac{\sum_{i=1}^{n} w_i(\Phi_i, \mathbf{T})\left[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) - \mu_{f, \Phi|Q=Q_{j_i}}^{(2)}(I)\right]}{\sum_{i=1}^{n} w_i(\Phi_i, \mathbf{T})}\right| \tag{5.23}$$

$$+ \left|\frac{\sum_{i=1}^{n} W_i \tilde{w}_i(\Phi_i, \mathbf{T})\mu_{f, \Phi|Q=Q_{j_i}}^{(2)}(I)}{\sum_{i=1}^{n} W_i \tilde{w}_i(\Phi_i, \mathbf{T})} - \mu_f^{(2)}(I)\right|. \tag{5.24}$$

By Lemma 5.4.1, $\hat{\mu}_f(I, \Phi_i, \mathbf{T})$ is consistent (for $\mathbf{T} \to \infty$) within the respective ergodicity class. Thus, (5.23) converges to 0 in probability if $\mathbf{T} \to \infty$. Using the short notation $\alpha_i = \hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)$ and $\tilde{w}_i = \tilde{w}_i(\Phi_i, \mathbf{T})$, we have

$$(5.24) = \left|\sum_{i=1}^{n} \frac{W_i \alpha_i\left[\mu_{f, \Phi|Q=Q_{j_i}}^{(2)}(I) - \mu_f^{(2)}(I)\right]}{\sum_{j=1}^{n} W_j \alpha_j} \cdot \frac{\tilde{w}_i \sum_{j=1}^{n} W_j \alpha_j}{\alpha_i \sum_{j=1}^{n} W_j \tilde{w}_j}\right|$$

$$\leq \max_{i=1}^{n} \left\{ \left| \frac{\tilde{w}_i \sum_{j=1}^{n} \alpha_j}{\alpha_i \sum_{j=1}^{n} \tilde{w}_j} \right| \right\} \cdot \left| \frac{\sum_{j=1}^{n} W_j \alpha_j}{\sum_{j=1}^{n} \alpha_j} \right| \cdot \left| \frac{\sum_{j=1}^{n} \tilde{w}_j}{\sum_{j=1}^{n} W_j \tilde{w}_j} \right|$$
$$\cdot \left| \frac{\sum_{i=1}^{n} W_i \alpha_i [\mu_{f,\Phi|Q=Q_{j_i}}^{(2)}(I) - \mu_f^{(2)}(I)]}{\sum_{i=1}^{n} W_i \alpha_i} \right|.$$

Since by assumption, $(n^{-1}\mathbb{E}\sum_{i=1}^{n}\tilde{w}_i)_{n\in\mathbb{N}}$ is eventually bounded away from 0 and the variance of the $\tilde{w}_i$ is uniformly bounded, the law of large numbers yields that $\sum_{j=1}^{n}\tilde{w}_j/\mathbb{E}\sum_{j=1}^{n}\tilde{w}_j$ and $\sum_{j=1}^{n}\tilde{W}_j w_j/\mathbb{E}\sum_{j=1}^{n}W_j\tilde{w}_j$ converge to 1 in probability. Additionally using that $\mathbb{E}[W_j\tilde{w}_j] = \mathbb{E}W_j\mathbb{E}\tilde{w}_j$, for $n \to \infty$, we get the convergence

$$\frac{\sum_{j=1}^{n}\tilde{w}_j}{\sum_{j=1}^{n}W_j\tilde{w}_j} = \frac{\sum_{j=1}^{n}\tilde{w}_j/\mathbb{E}\sum_{j=1}^{n}\tilde{w}_j}{\sum_{j=1}^{n}W_j\tilde{w}_j/\mathbb{E}\sum_{j=1}^{n}W_j\tilde{w}_j} \cdot \frac{\mathbb{E}\sum_{j=1}^{n}\tilde{w}_j}{\mathbb{E}\sum_{j=1}^{n}W_j\tilde{w}_j} \xrightarrow{p} \frac{1}{\mathbb{E}W_1}$$

as $n \to \infty$. Similarly, for $n \to \infty$ and $n, \mathbf{T} \to \infty$, we have

$$\frac{\sum_{j=1}^{n}W_j\alpha_j}{\sum_{j=1}^{n}\alpha_j} \xrightarrow{p} \frac{\mathbb{E}[W_1\alpha_1]}{\mathbb{E}\alpha_1},$$

$$\frac{\sum_{i=1}^{n}W_i\alpha_i\mu_{f,\Phi|Q=Q_{j_i}}^{(2)}(I)}{\sum_{i=1}^{n}W_i\alpha_i} \xrightarrow{p} \frac{\mathbb{E}\left[\alpha_{\Phi|Q=Q_{j_i}}^{(2)}(C(I)) \cdot \mu_{f,\Phi|Q=Q_{j_i}}^{(2)}(I)\right]}{\mathbb{E}\left[\alpha_{\Phi|Q=Q_{j_i}}^{(2)}(C(I))\right]} = \mu_f^{(2)}(I),$$

respectively. Together with (5.22) we obtain that (5.24) converges to 0 in probability, which completes the proof. $\square$

Note that if $\tilde{w}_i = \tilde{w}$ for all $i \in \mathbb{N}$ for some weight function $\tilde{w}$ with $\mathbb{E}|\tilde{w}(\Phi,\mathbf{T})| < \infty$, the $\tilde{w}_i(\Phi_i,\mathbf{T})$ are iid and conditions (5.18), (5.19) and (5.20) become obsolete.

Now we turn to the estimation of $\tilde{\mu}_f^{(2)}(I)$. By construction (cf. Definition 5.3.3), $\hat{\mu}_f^n(I)$ consistently estimates $\tilde{\mu}_f^{(2)}(I)$; in contrast to $\hat{\mu}_f^\alpha(I)$, it reflects a random pair of points with distance I within a randomly chosen ergodicity class. Again, also other choices of weights are feasible for consistent estimation of $\tilde{\mu}_f^{(2)}(I)$, apart from the choice $w_i(\Phi_i,\mathbf{T}) = 1$. By replacing $\hat{\alpha}^{(2)}(C(\mathbf{T},I),\Phi_i)$ by the constant 1 in Theorem 5.4.4, we get the following corollary.

Corollary 5.4.5. *Under the assumptions of Theorem 5.4.4 with $\hat{\alpha}^{(2)}(C(\mathbf{T},I),\Phi_i)$ being replaced by the constant 1, $\hat{\mu}_f^{n,\mathrm{wght}}(I,\mathbf{w})$ is a consistent estimator for $\tilde{\mu}_f^{(2)}(I)$.*

Remark 5.4.6. *If Φ is ergodic, $\hat{\mu}_f^{n,\mathrm{wght}}(I,\mathbf{w})$ is consistent for $\mu_f^{(2)}(I)$ (as $\mathbf{T} \to \infty$) for any choice of weights $\mathbf{w}$ that satisfies (5.12). Note that in this case, consistency is independent of n, which can be fixed to any finite value.*

Proof. If Φ is ergodic, the mixing measure λ is the one-point distribution δ_P and condition (5.12) simply means stochastic convergence of the weights w.r.t. P. The assertion directly follows from the continuous mapping theorem. $\square$

5.4.3 Variance minimization

In what follows, we seek for an optimal consistent estimator for $\tilde{\mu}_f^{(2)}(I)$ in the sense of minimal variance. We introduce some additional assumptions on the mark-location dependence for analytical tractability. For simplicity, we set $\tilde{w}_i(\Phi_i, \mathbf{T}) = 1$, i.e., we consider $w_i(\Phi_i, \mathbf{T}) = W_i$. Let $\mathcal{A}_n^*$ denote the σ-algebra generated by the unmarked ground processes $\Phi_{1,\mathrm{g}}, \ldots, \Phi_{n,\mathrm{g}}$, i.e., $\mathcal{A}_n^* = \sigma(\{\{\omega : \Phi_{i,\mathrm{g}}(\omega)(B) = k\} : k \in \mathbb{N}, B \in \mathcal{B}, i = 1, \ldots, n\})$. We assume that $\mathbb{E}[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) \,|\, \mathcal{A}_n^*]$ is a.s. constant. We further assume that $\mathcal{A}_n^*$ is maximal w.r.t. this property and that $\mathrm{Var}\,[\hat{\mu}_f(I, \Phi, \mathbf{T})| \,\mathcal{A}_n^*]$ is independent of the random ergodicity class Q.

Proposition 5.4.7. *With the above notation and assumptions, the variance minimizing weights for $\hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), \mathbf{T})$ that satisfy (5.17)–(5.22) with $\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)$ being replaced by 1 are given by*

$$w_i(\Phi_i, \mathbf{T}) = W_i = \mathrm{Var}\,[\hat{\mu}_f(I, \Phi_i, \mathbf{T})| \,\mathcal{A}_n^*]^{-1}.$$

Note that an analog variance minimizing procedure via random factors W_i could also be included into the estimator $\hat{\mu}_f^\alpha$ of $\mu_f^{(2)}(I)$.

Proof of Proposition 5.4.7. For general $\mathcal{A}_n^*$-measurable weights $w_i(\Phi_i, \mathbf{T})$, $i = 1, \ldots, n$, we have

$$\mathrm{Var}\left[\hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), \mathbf{T})\right]$$
$$= \mathbb{E}\left[\frac{1}{(\sum w_i(\Phi_i, \mathbf{T}))^2} \sum_{i=1}^n w_i(\Phi_i, \mathbf{T})^2 \,\mathrm{Var}\,[\hat{\mu}_f(I, \Phi_i, \mathbf{T})| \,\mathcal{A}_n^*]\right]$$
$$+ \mathrm{Var}\left[\frac{1}{\sum w_i(\Phi_i, \mathbf{T})} \sum_{i=1}^n w_i(\Phi_i, \mathbf{T}) \mathbb{E}\,[\hat{\mu}_f(I, \Phi_i, \mathbf{T})| \,\mathcal{A}_n^*]\right]$$
$$= \mathbb{E}\left[\sum_{i=1}^n w_i^{\mathrm{rel}}(\Phi_i, \mathbf{T})^2 \,\mathrm{Var}\,[\hat{\mu}_f(I, \Phi_i, \mathbf{T})| \,\mathcal{A}_n^*]\right] \quad + \quad 0 \qquad (5.25)$$

with $w_i^{\mathrm{rel}}(\Phi_i, \mathbf{T}) = w_i(\Phi_i, \mathbf{T}) / \sum_{i=1}^n w_i(\Phi_i, \mathbf{T})$. Since any weighted average $\sum v_i^2 x_i$ with $x_i > 0$ and $\sum v_i = 1$ is minimized by $v_i = x_i^{-1} / \sum x_i^{-1}$ (Lagrange method), the unconditional variance (5.25) is minimized by choosing

$$w_i(\Phi_i, \mathbf{T}) = W_i = \mathrm{Var}\,[\hat{\mu}_f(I, \Phi_i, \mathbf{T})| \,\mathcal{A}_n^*]^{-1}.$$

The W_i are $\mathcal{A}_n^*$-measurable by definition of the conditional variance and satisfy (5.17)–(5.22) with $\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)$ being replaced by 1. Maximality of $\mathcal{A}_n^*$ ensures optimality of the weights. $\square$

If there exist interaction effects in the MPP that are of higher than second order, the assumption on $\mathbb{E}[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) \,|\, \mathcal{A}_n^*]$ might not be satisfied anymore and weighting according to the above conditional variances should be handled with care. Clusters of point locations which tend to increase the conditional variance of $\hat{\mu}_f$ given the ground process, can additionally

influence the mean of other marks in excess of the bivariate interaction measured by $\mu_f^{(2)}(I)$. Then, a bias will be introduced by using the above random weights. More generally, the more is known about the relation between $\hat{\mu}_f(I, \Phi, \mathbf{T})$ and the ground process Φ_g, the more can be gained from using different (random) weights while preserving consistency of the estimator. Without any assumption, only deterministic or independent weights are feasible and then $w_i(\Phi_i, \mathbf{T}) = 1$ is naturally the best choice, i.e., the use of $\hat{\mu}_f^n(I)$.

We consider two simple examples of optimal weighting in the following. Here we assume that the z-components of the marks are 1 for all points. Recall that $\hat{\mu}_f(I, \Phi, \mathbf{T}) = \hat{\alpha}_f(I, \Phi, \mathbf{T})/\hat{\alpha}_1(I, \Phi, \mathbf{T})$, that the denominator is $\mathcal{A}_n^*$-measurable, and that $\hat{\alpha}_f(I, \Phi_i, \mathbf{T})$ is a sum consisting of $\hat{\alpha}_1(I, \Phi_i, \mathbf{T})$ random summands.

Remark 5.4.8. *In general, the summands of $\hat{\alpha}_f(I, \Phi_i, \mathbf{T})$ are not iid since each value $f(y_1)$ can occur multiple times in the sum—in fact, it occurs as many times as there are points t_2 s.t. $t_2 - t_1 \in I$. However, if conditionally on $\mathcal{A}_n^*$, the summands were iid with variance v, the conditional variance $\mathrm{Var}\left[\hat{\mu}_f(I, \Phi, \mathbf{T}) \,|\, \mathcal{A}_n^*\right]$ would be $v/\hat{\alpha}_1(I, \Phi_i, \mathbf{T})$.*

In the following scenarios, we assume f to depend on its first argument, only.

Example 5.4.9. *Let Φ have marks that are stochastically independent of the process of point locations and let these point locations be fully regularly spaced in every realization. Let $v_\mathbf{T}$ and $N = N(\mathbf{T})$ denote the volume of $[\mathbf{0}, \mathbf{T}]$ and the random number of points in $[\mathbf{0}, \mathbf{T}]$, respectively, and assume that the $f(y_i)$, $i \in \mathbb{Z}$, are iid with variance v. Then, asymptotically, $\mathrm{Var}[\hat{\mu}_f(I, \Phi, \mathbf{T})|\mathcal{A}_n^*] \sim v/N$ and the resulting weights are $w_i(\Phi_i) = N_i/v$, where N_i denotes the number of points within the i-th realization.*

Proof. For $|I|$ and $\mathbf{T}$ large, we have $\hat{\alpha}_1(I, \Phi, \mathbf{T}) \sim N \cdot N|I|/v_\mathbf{T}$ and each distinct summand in $\hat{\alpha}_f(I, \Phi, \mathbf{T})$ occurs $N|I|/v_\mathbf{T} \sim \hat{\alpha}_1(I, \Phi, \mathbf{T})/N$ times. Thus we have $\hat{\alpha}_f(I, \Phi, \mathbf{T}) \sim \hat{\alpha}_1(I, \Phi, \mathbf{T}) \sum_{i=1}^N f(y_i)/N$ and $\mathrm{Var}[\hat{\alpha}_f(I, \Phi, \mathbf{T})|\mathcal{A}_n^*] \sim \hat{\alpha}_1(I, \Phi, \mathbf{T})^2 v/N$ $\qquad\qquad\Box$

Since $N(\mathbf{T})$ is usually much smaller than $\hat{\alpha}_1(I, \Phi, \mathbf{T})$, the variance $\mathrm{Var}[\mu_f(I, \Phi, \mathbf{T})|\mathcal{A}_n^*]$ in Example 5.4.9 is larger than the one in the hypothetical example in Remark 5.4.8.

In the following example, we consider arbitrary point locations but still assume independence between marks and locations.

Example 5.4.10. *Let $\tilde{\Phi}$ be a one-dimensional, stationary unmarked point process and Y a stationary continuous-time process which is independent of $\tilde{\Phi}$ and such that $f(Y)$ has finite second moments. We consider the MPP $\Phi = \{(t, Y(t), 1) : t \in \tilde{\Phi}\}$. Then*

$$\mathrm{Var}\left[\hat{\mu}_f(I, \Phi, T) \,|\, \mathcal{A}_n^*\right]$$
$$= \frac{\sum_{t_1 \in \Phi_\mathrm{g} \cap [0, T]} \sum_{s_1 \in \Phi_\mathrm{g} \cap [0, T]} \mathrm{Cov}\left[f(Y(t_1)), f(Y(s_1))\right] n(t_1, \Phi_\mathrm{g}, I) n(s_1, \Phi_\mathrm{g}, I)}{\left[\sum_{t_1 \in \Phi_\mathrm{g} \cap [0, T]} n(t_1, \Phi_\mathrm{g}, I)\right]^2},$$

where $n(t_1, \Phi_\mathrm{g}, I) = \sum_{t_2 \in \Phi_\mathrm{g} \setminus \{t_1\}} \mathbf{1}_{t_2 - t_1 \in I}$.

The proof of this assertion is obtained as a side result of the proof of Theorem 6.2.1 in Section 6.7, with u being replaced by $-\infty$ and with the logarithm being replaced by a general function f.

5.4.4 Remarks

Remark 5.4.11. *The weighting of multiple realizations and the intrinsically weighted means coincide in the following sense: Let $\Phi_1, \ldots, \Phi_n$ be iid copies of an MPP $\Phi = \{(t_i, y_i, 1) : i \in \mathbb{N}\}$, for which the second mark component equals 1 for all points. Then the weighting of realizations via $w_i(\Phi_i, \mathbf{T})$ in the estimator (5.13) can alternatively be captured by the second mark component. For $i = 1, \ldots, n$, let $\tilde{\Phi}_i = \{(t, y, w_i^{rel}(\Phi_i, \mathbf{T})) : (t, y, 1) \in \Phi_i\}$, where $w_i^{rel}(\Phi_i, \mathbf{T}) = w_i(\Phi_i, \mathbf{T})/\sum_{k=1}^n w_k(\Phi_k, \mathbf{T})$. Let Ψ_n be the concatenation of the processes $\tilde{\Phi}_1, \ldots, \tilde{\Phi}_n$, each restricted to the observation window $[\mathbf{0}, \mathbf{T}]$ and concatenated with a buffer of $\max(I)$ and such that all points of Ψ_n are contained in $[\mathbf{0}, \mathbf{T}_n]$ for some $\mathbf{T}_n \in \mathbb{R}^d$. Then, with $\mathbf{w} = (w_i(\Phi_i, \mathbf{T}))_{i=1}^n$, we have*

$$\hat{\mu}_f(I, \Psi_n, \mathbf{T}_n) = \hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), \mathbf{T}).$$

We close this section with a note on the estimation of $\mu_f^{(2)}(r)$ and $\tilde{\mu}_f^{(2)}(r)$, $r \in \mathbb{R}$.

Remark 5.4.12. *For most MPPs used in applications, finding two points of an MPP with a fixed distance r within a bounded observation window, has probability zero. Then the simplest approach is to apply any of the estimators (5.11), (5.13), (5.14) or (5.16), with I being a small interval containing r, e.g., $[r - \delta, r + \delta]$ for some $\delta > 0$. This is equivalent to use (Nadaraya-Watson) kernel regression with the rectangular kernel, applied to the tuples $\{(z_1 f(y_1), \mathrm{dist}(t_2 - t_1)) : (t_1, y_1, z_1), (t_2, y_2, z_2) \in \Phi\}$, where $\mathrm{dist}(x) = x$ if $x \in \mathbb{R}^1$ and $\mathrm{dist}(x) = \|x\|$ if $x \in \mathbb{R}^d$ with $d > 1$.*
An obvious generalization is to replace the rectangular kernel by a general kernel K_h with bandwidth h. For the basic estimator (5.11), this yields

$$\hat{\mu}_f(r, \Phi, T) = \frac{\sum_{(t_1, y_1, z_1), (t_2, y_2, z_2) \in \Phi \; t_1 \in [0, T]}^{\neq} z_1 f(y_1) K_h(r - \mathrm{dist}(t_2 - t_1))}{\sum_{(t_1, y_1), (t_2, y_2) \in \Phi, \; t_1 \in [0, T]}^{\neq} K_h(r - \mathrm{dist}(t_2 - t_1))},$$

likewise for the other estimators. If the support of K_h covers the whole real line, the denominator is always strictly larger than zero, which simplifies implementation, but also allows $\hat{\mu}_f(r, \Phi, T)$ to be driven by pairs of points whose distance differs largely from r.

5.5 Application to continuous-space processes

Picking up the introductory example on continuous-space processes, taking measurements from such a process with measurement locations that are possibly irregularly spaced but independent of the underlying process, leads to a subclass of MPPs. At the same time, particularly developed in the geostatistical context, there exist numerous methods of inference for continuous-space processes, including methods to account for biased and preferential sampling. We compare the concept of intrinsically weighted means of MPPs to statistical methods for continuous-space processes in the following.

One of the classical problems in geostatistical applications (e.g., Chiles & Delfiner, 1999) is prediction of averages from measurements $\{(t_i, Y(t_i)) : i = 1, \ldots, n\}$, where $\{Y_t : t \in T\}$,

$T \subset \mathbb{R}^d$, is a latent second-order stationary random field. When predicting global moments of Y, redundancies in the data can be excluded via the spatial correlation structure, e.g., the best linear unbiased estimator (BLUE) for $\mathbb{E}Y$ is well-known to be $(\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1} \cdot \mathbf{1}'\Sigma^{-1}\mathbf{Y}$, where $\mathbf{1} = (1, \ldots, 1)'$, $\mathbf{Y} = (Y(t_1), \ldots, Y(t_n))'$ and $\Sigma = \mathrm{Cov}(Y(t_i), Y(t_j))_{i,j=1}^n$ (e.g., Chiles & Delfiner, 1999, p.179). Basically, each data point is weighted by the corresponding column sum of the precision matrix Σ^{-1}. Note that even when the t_i are on a fully regular grid, the weights do not necessarily equal n^{-1}. More generally, any estimator that is linear in a transformation g of the data, including estimators for the correlation structure itself, allows for assigning a different weight to each data point; then the estimator takes the form $\sum_{i=1}^n z_i g(Y(t_i))$ or $\sum_{i,j=1}^n z_{ij} g(Y(t_i), Y(t_j))$; similarly for higher-order moments. The weights z_i and z_{ij} are supposed to capture the spatial or temporal pattern of measurement locations when statistical inference from irregularly spaced data is carried out. Similar weighting procedures are used for declustering and debiasing methods (cf. Journel, 1983; Isaaks & Srivastava, 1989; Deutsch, 1989; Bourgault, 1997; Emery & Ortiz, 2005; Olea, 2007).

Assertion 5.5.1. *Identifying the geostatistical weights z_i with the z-component of the MPP $\Phi = \{(t_i, y_i, z_i) : i \in \mathbb{N}\}$, the estimator $\sum_{i=1}^n z_i g(Y(t_i))$ of $\mathbb{E}g(Y)$ coincides with the canonical estimator for the weighted mean mark $\mu_f^{(1)}$, defined by (5.2).*

The geostatistical guiding principle of choosing optimal weights for aggregation of measurements adheres to the idea that a) there exists an underlying random field and b) that this field can be measured at any location without causally influencing the other measurements. It is important to note that this is far from being satisfied for processes in which the measurements reflect physical objects that interact with each other. Trees in a forest, for example, compete for resources and if another tree had been added at some point, the measured characteristics of the surrounding trees would have likely changed. In the context of high-frequency transaction data, measurements are taken in form of transactions, which, in turn, not only influence the future price process but even the instantaneous price. Hence, the imagination of an underlying random field is not well suited, at least not for the microscopic scale of transactions. Though, with increasing distance, interaction effects between single objects of an MPP may become negligible compared to the variability of the mark values and the random field assumption might be sensible on a larger scale. This perspective motivates combining classical mean mark estimators for MPPs of the form $\Phi = \{(t_i, y_i, 1) : i \in \mathbb{N}\}$ with a geostatistical weighting. Partitioning the observation window in smaller parts, we assign a z-component to Φ such that $z_i = z_j$ whenever t_i and t_j belong to the same cell of the partition. This leads to a classical unweighted average within each cell and therewith maintains the information contained in the small-scale pattern of the point locations. Between the different cells, we allow for a weighting in the geostatistical sense and therewith allow to smooth out large-scale irregularities in the distribution of point locations. We denote the resulting estimator by $\hat{\mu}_f^{(1),\mathrm{geo}}$.

Assertion 5.5.2. *Considering a realization of Φ as a collection of realizations of a possibly non-ergodic MPP on smaller observation windows corresponding to the above partition, the form of $\hat{\mu}_f^{(1),\mathrm{geo}}$ coincides with that of $\hat{\mu}_f^n$ and $\hat{\mu}_f^{n,\mathrm{wght}}$, which estimate the average mean mark $\tilde{\mu}_f^{(1)}$ (see Definition 5.3.3) instead of the classical mean mark $\mu_f^{(1)}$.*

The application of such a weighting scheme is of particular interest when the underlying process jumps between different regimes that differ substantially from each other, e.g., w.r.t. the intensity of point locations. In summary, applying the geostatistical idea of declustering in the MPP context in a sense corresponds to the concept of non-ergodic modeling.

To avoid possible confusion, we conclude this section with a final remark.

Remark 5.5.3. *For certain choices of f, the random field counterpart of $\mu_f^{(2)}$ is well-defined. For $f(y_1, y_2) = y_1 y_2$, for instance, the counterpart is the ordinary (non-centered) covariance function. If f only depends on one of the two marks of a pair of points, $\mu_f^{(2)}$ implicitly conditions on the existence of other points and there is no sensible way of interpreting a suchlike statistic in a random field context, where there exist values at all points of the index space. Nevertheless, the geostatistical idea of variance-minimizing weights can be applied to $\mu_f^{(2)}$ by a simple mean squared error approach.*

5.6 Discussion

The MPP summary statistics considered in this chapter are (weighted) mean marks. In practice, the choice of weights is not always clear, for example when data from different stochastic sources are combined. In Section 5.5, we point out that, if there was an underlying continuous-time process from which the data were generated by a random sampling procedure, then the *mean of interest* would rather be the temporal average over the whole index space instead of the average over all sampling locations. The weights might then be chosen to compensate for the irregular distribution of point locations. Though, the assumption of a continuous-time background process is problematic if the points represent physical objects that influence each other. Then, the mean of interest might include the randomness of the point pattern, as it is reflected by the MPP moment measures $\alpha_f^{(2)}$.

Related questions arises when multiple realizations of a non-ergodic MPP are considered: Should the definition of mean include possibly different intensities of points between different ergodicity classes or not? A non-ergodic MPP can be seen as a hierarchical model and expectation functionals w.r.t. the point process can naturally be replaced by two-step expectations by averaging within each ergodicity class first and then aggregating the different classes (cf. Section 5.3). This alternative definition filters out the differences w.r.t. the point location patterns between different ergodicity classes. Which definition of mean should be chosen eventually depends on the purpose of the characteristic at hand and on the intended interpretation.

The above considerations contribute to the understanding of the relation between geostatistical weighting methods, weighted averages for estimation of MPP moment measures, and non-ergodic modeling of MPPs, but the topic is still not comprehensively understood. A related, more general question, which to our knowledge is still open, is the following: Is it possible to distinguish, by means of suitable summary statistics, genuine MPPs, in which objects physically interact with each other, from MPPs that result from a possibly dependent sampling of a random field? Concerning the distinction between independent and dependent sampling within the latter class of MPPs, feasibility of this approach crucially depends on

the distributional assumptions imposed on the marks. In the context of max-stable random fields, for example, this distinction requires much more effort than in the classical Gaussian world (cf. Section 6.5 and Schlather *et al.* (2004)).

5.7 Some basics of ergodic theory

Ergodicity is a mixing property that can be defined in the very general context of dynamical systems. An MPP on $\mathbb{R}^d$ together with the group of $\mathbb{R}^d$-indexed shift operators is a special case of a dynamical system.

We denote by $\mathbb{M}_0$ the set of all locally finite counting measures on $\mathbb{R}^d \times \mathbb{R}$, and by $\mathcal{M}_0$ the smallest σ-algebra on $\mathbb{M}_0$ that makes all mappings $\mathbb{M}_0 \to \mathbb{N}_0 \cup \infty$, $\varphi \mapsto \varphi(S)$, measurable. Formally, an MPP Φ is a measurable mapping from some probability space $(\Omega, \mathcal{A}, P)$ into $(\mathbb{M}_0, \mathcal{M}_0)$ and we can identify $(\Omega, \mathcal{A})$ with $(\mathbb{M}_0, \mathcal{M}_0)$ in the usual way. Let $\mathbf{T} = \{T_x : x \in \mathbb{R}^d\}$ with

$$(T_x \varphi)(B \times L) = \varphi((B + x), L), \qquad B \in \mathcal{B}^d, L \in \mathcal{B}. \tag{5.26}$$

Recall that Φ is said to be stationary if the induced probability measure P^Φ is $\mathbf{T}$-invariant. Further, a stationary MPP Φ is called ergodic if $P^\Phi(A)$ is either zero or one for all $\mathbf{T}$-invariant sets $A \in \mathcal{M}_0$. Let $\mathcal{A}_0 \subset \mathcal{M}_0$ be the sub-σ-algebra of all $\mathbf{T}$-invariant sets in $\mathcal{M}_0$, i.e., $A = T^{-1}A$ for all $A \in \mathcal{A}_0$ and $T \in \mathbf{T}$.

The following theorem is commonly termed *pointwise* or *individual ergodic theorem* in literature and establishes almost sure convergence of a certain average of values of a random variable X.

Definition 5.7.1 (Daley & Vere-Jones (2008, Def. 12.2.I))**.** A sequence of bounded convex Borel sets $W_n \subset \mathbb{R}^d$, $n \in \mathbb{N}$, with $W_n \subset W_{n+1}$ for $n \in \mathbb{N}$, is called *convex averaging sequence in $\mathbb{R}^d$* if the maximal radius of a ball contained in W_n goes to infinity if n increases.

Theorem 5.7.2 (Daley & Vere-Jones (2008, Prop. 12.2.II))**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathbf{T} = \{T_x : x \in \mathbb{R}^d\}$ a group of measure-preserving transformations acting on $(\Omega, \mathcal{A}, P)$ such that the mapping $(T_x, \omega) \mapsto T_x \omega$ is jointly measurable, i.e., $(\mathcal{B}(\mathbf{T}) \otimes \mathcal{A}, \mathcal{A})$-measurable. (Multiplication in $\mathbf{T}$ is given by $T_x T_y = T_{x+y}$.) Let $\{W_n\}_{n \in \mathbb{N}}$ be a convex averaging sequence in $\mathbb{R}^d$ and $\mathcal{A}_0$ the σ-algebra of $\mathbf{T}$-invariant events. Then for all real-valued integrable functions X on $(\Omega, \mathcal{A}, P)$*

$$\bar{X}_n = \frac{1}{\nu(W_n)} \int_{W_n} X(T_x \omega)\, \nu(dx) \xrightarrow{a.s.} \mathbb{E}(X \mid \mathcal{A}_0), \qquad n \to \infty.$$

If X is additionally L_p-integrable, then $\mathbb{E}(X \mid \mathcal{A}_0)$ is also the L_p-limit of $\bar{X}_n$.

Remark 5.7.3. *If P is ergodic (i.e., $P(A) \in \{0, 1\}\ \forall A \in \mathcal{A}_0$) then $\mathbb{E}(X \mid \mathcal{A}_0)$ reduces to the constant $\mathbb{E}X$. Loosely speaking, this means that a suitable average over transformations of a single realization converges to the expectation over the state space Ω.*

While Theorem 5.7.2 refers to a general probability space with a general group of transformations acting on it, the following Proposition relates this results to the context of MPPs on

$\mathbb{R}^d$, in which the transformations T_x, $x \in \mathbb{R}^d$, are given by shifts of the whole point pattern by the vector x. Here, the point is that the index $x \in \mathbb{R}^d$ has a direct geometric meaning when T_x is applied to a realization φ of Φ. This yields convergence of spatial averages within a single realization of the MPP to the state space mean.

The proof of the following Proposition is based on a simple sandwich argument, which can also be used for other consistency statements. We include the proof here, because to our knowledge, it is not available in this form in pertinent literature. A similar assertion can be found in (Daley & Vere-Jones, 2008, Thm. 12.2.IV).

Proposition 5.7.4. *Let Φ be stationary and ergodic and* **T** *as in Theorem 5.7.2. Let $f : \mathbb{R}^d \times \mathbb{R} \times \mathbb{M}_0 \to \mathbb{R}$ be a non-negative function that satisfies $f(t - x, y, T_x\varphi) = f(t, y, \varphi)$ for all $t, x \in \mathbb{R}^d, y \in \mathbb{R}$, and that is integrable w.r.t. to the marked Campbell measure $C(B \times L \times M) = \mathbb{E}\big[\Phi((B \cap [0,1]^d) \times L)\mathbf{1}_M(\Phi)\big]$, $B \in \mathcal{B}^d$, $L \in \mathcal{B}$, $M \in \mathcal{M}_0$. We define random variables $X, X_n : \mathbb{M}_0 \to \mathbb{R}$ by*

$$X(\varphi) = \sum_{(t,y)\in\varphi,\ t\in[0,1]^d} f(t, y, \varphi),$$

$$X_n(\varphi) = \frac{1}{n^d} \sum_{(t,y)\in\varphi,\ t\in[0,n]^d} f(t, y, \varphi).$$

Then X_n converges to $\mathbb{E}^\Phi X$ almost surely if $n \to \infty$.

Proof. An extension of the classical Campbell theorem (e.g., Daley & Vere-Jones, 2008, lem. 13.1.II) guarantees that $\mathbb{E}|X| < \infty$ if f is integrable w.r.t. the Campbell measure. The $W_n = [0, n]^d$ obviously form an averaging sequence and

$$X_n(\varphi) = \frac{1}{\nu(W_n)} \sum_{(t,y)\in\varphi,\ t\in W_n} f(t, y, \varphi) \int_{\mathbb{R}^d} \mathbf{1}_{[t,\,t+1]}(x)\,\nu(dx)$$

$$= \frac{1}{\nu(W_n)} \int_{\mathbb{R}^d} \sum_{(t,y)\in\varphi,\ t\in W_n\cap[x-1,\,x]} f(t, y, \varphi)\,\nu(dx), \tag{5.27}$$

where $x \pm 1$ for $x \in \mathbb{R}^d$ is defined component-wise. Note that the integrand on the RHS equals 0 whenever $W_n \cap [x - 1,\, x] = \varnothing$, which means that x is not contained in $W_n \oplus [0, 1]^d$, which is, on its part, a subset of W_{n+1}. Thus, we can shrink the region of integration to W_{n+1} without changing the integral. If we then drop the condition '$t \in W_n$' under the summation sign, we enlarge the whole expression since f is non-negative, i.e.

$$X_n(\varphi) \le \frac{1}{\nu(W_n)} \int_{W_{n+1}} \sum_{(t,y)\in\varphi,\ t\in[x-1,\,x]} f(t, y, \varphi)\,\nu(dx)$$

$$= \frac{1}{\nu(W_n)} \int_{W_{n+1}} \sum_{(t,y)\in T_{x-1}\varphi,\ t\in[0,1]^d} f(t, y, T_{x-1}\varphi)\,\nu(dx)$$

$$= \frac{\nu(W_{n+1})}{\nu(W_n)} \frac{1}{\nu(W_{n+1})} \int_{W_{n+1}-1} X(T_x\varphi)\,\nu(dx), \tag{5.28}$$

where the second equation uses that $f(t - x, y, T_x\varphi) = f(t, y, \varphi)$ and the last equation uses that ν is shift-invariant. Since the ratio $\nu(W_{n+1})/\nu(W_n)$ converges to 1, Theorem 5.7.2 yields that the RHS of (5.28) converges to $\mathbb{E}^{\Phi}(X \mid \mathcal{A}_0)$ for almost all $\varphi \in \mathbb{M}_0$. Since Φ was assumed to be ergodic, this conditional expectation equals $\mathbb{E}^{\Phi}X$.

Similarly, if we restrict integration in (5.27) to the set W_{n-1}, we reduce the value of the integral. Since $W_{n-1} \oplus [-1, 0]^d \subset W_n$, we can again drop the condition '$t \in W_n$' under the summation sign and by the same argument as before, we have

$$X_n(\varphi) \geq \frac{1}{\nu(W_n)} \int_{W_{n-1}} \sum_{(t,y)\in\varphi, \ t\in[x, x+1]} f(t, y, \varphi) \, \nu(dx) \ \overset{n\to\infty}{\longrightarrow} \ \mathbb{E}^{\Phi}X$$

for almost all $\varphi \in \mathbb{M}_0$. Thus, we have a sandwich relation for $X_n(\varphi)$ and can conclude that $X_n \to \mathbb{E}^{\Phi}X$ a.s. $\qquad\square$

Note that the convex averaging sequence $\{[0, n]^d\}_{n\in\mathbb{N}}$ in Proposition 5.7.4 can be replaced by any sequence $\{W \oplus nV\}_{n\in\mathbb{N}}$ with W a bounded Borel set and $V \subset \mathbb{R}^d$ a convex and bounded set with $\nu(V) > 0$ and $0 \in V$.

In case that Φ is not ergodic, the following results provide a representation of Φ as a mixture of a set of ergodic MPPs. To this end, let $\mathcal{P}$ ($\mathcal{P}_{\mathrm{erg}}$ resp.) denote the set of all probability measures on $(\mathbb{M}_0, \mathcal{M}_0)$ induced by stationary (and ergodic) MPPs and let Π_{erg} be the smallest σ-algebra making all mappings $\mathcal{P}_{\mathrm{erg}} \to [0, 1]$, $P \mapsto P(A)$, $A \in \mathcal{M}_0$, measurable. We say that $\mathbf{T}$ fulfills the condition (LocCompGrp) if $\mathbf{T}$ is a locally compact, second-countable Hausdorff group of jointly measurable, surjective transformations.

From Farrell (1962), we can extract the very general result

Theorem 5.7.5. *Let $(\Omega, \mathcal{A})$ be a measurable space with Ω a complete separable metric space and $\mathcal{A}$ its Borel-σ-algebra. Let $\mathbf{T}$ be a set of measurable transformations of Ω satisfying the condition (LocCompGrp) and let $P \in \mathcal{P}$. Here, $\mathcal{P}$ ($\mathcal{P}_{\mathrm{erg}}$ resp.) is the set of all $\mathbf{T}$-invariant (and ergodic) probability measures on $(\Omega, \mathcal{A})$. Then there is a unique probability measure λ_P on $(\mathcal{P}_{\mathrm{erg}}, \Pi_{\mathrm{erg}})$ and a $\mathcal{P}_{\mathrm{erg}}$-valued random variable Q_P s.t.*

$$P(A) = \int_{\mathcal{P}_{\mathrm{erg}}} Q(A) \, \lambda_P(dQ) = \int_{\Omega} Q_P(\omega)(A) \, P(d\omega), \qquad A \in \mathcal{A},$$

i.e., λ_P is the distribution of Q_P.

In the context of MPPs on $\mathbb{R}^d$, the group $\mathbf{T}$ of shifts, as defined in (5.26), obviously fulfills the condition (LocCompGrp), and since $\mathbb{M}_0$ is a complete separable metric space and $\mathcal{M}_0$ its Borel-σ-algebra (e.g., Kallenberg, 1986), Theorem 5.7.5 can directly be applied, which yields a decomposition of the non-ergodic MPP Φ with law P:

$$P(M) = \int_{\mathcal{P}_{\mathrm{erg}}} Q(M) \, \lambda(dQ), \qquad M \in \mathcal{M}_0.$$

Note that each Q induces a new ergodic MPP $\Phi_Q : \Omega \to \mathbb{M}_0$ which is given implicitly by $P(\Phi_Q \in M) = Q(M)$, $M \in \mathcal{M}_0$. By the second representation in Theorem 5.7.5, we can also

consider Q as a random variable on $(\mathbb{M}_0, \mathcal{M}_0, P)$ with distribution $\lambda = \lambda_P$. Thus, Φ and Q^Φ have a joint distribution and the conditional distribution of Φ given Q is well-defined:

$$P(\cdot \mid Q = Q^*) = Q^*(\cdot).$$

5.8 Proof of Theorem 5.4.3

The following lemma will be used in the proof of Theorem 5.4.3 and generalizes the classical individual ergodic theorem (Daley & Vere-Jones, 2008, Prop. 12.2.II) to a situation in which the thinning of the point process depends on the size of the observation window.

Lemma 5.8.1. *Let Φ be a stationary and ergodic MPP on $\mathbb{R}$ with real-valued marks and let $(u_T)_{T \geq 0}$ be a family of non-negative non-decreasing numbers such that*

$$\frac{T^{-1}\hat{\alpha}_1(I, \Phi, T) - \lambda}{\mathbb{E}_\Phi \hat{\alpha}_{f_{\mathrm{cond}, u_T}}(I, \Phi, 1)} \to 0 \quad a.s. \ (T \to \infty). \tag{5.29}$$

Then, for $T \to \infty$, we have the almost sure convergence

$$\frac{\hat{\alpha}_{f_{\mathrm{cond}, u_T}}(I, \Phi, T)}{T \mathbb{E}_\Phi \hat{\alpha}_{f_{\mathrm{cond}, u_T}}(I, \Phi, 1)} \longrightarrow 1.$$

Note that the almost sure convergence $(\lambda T)^{-1}\hat{\alpha}_1(I, \Phi, T) \to 1$ as $T \to \infty$ follows from the classical individual ergodic theorem (e.g., Daley & Vere-Jones, 2008, Prop. 12.2.II).

Proof of Lemma 5.8.1. With $g_u(y) = 1 - f_{\mathrm{cond}, u}(y)$, $y \in \mathbb{R}$, we obtain the almost sure convergence

$$\frac{\hat{\alpha}_{g_{u_T}}(I, \Phi, T)}{T \mathbb{E}_\Phi \hat{\alpha}_{g_{u_T}}(I, \Phi, 1)} \to 1$$

from (Daley & Vere-Jones, 2008, Prop. 12.2.VII) and the subsequent remarks. Further, $\lambda = \mathbb{E}_\Phi \hat{\alpha}_1(I, \Phi, 1) = \mathbb{E}_\Phi \hat{\alpha}_{f_{\mathrm{cond}, u_T}}(I, \Phi, 1) + \mathbb{E}_\Phi \hat{\alpha}_{g_{u_T}}(I, \Phi, 1)$. Hence,

$$\frac{\hat{\alpha}_{f_{\mathrm{cond}, u_T}}(I, \Phi, T)}{T \mathbb{E}_\Phi \hat{\alpha}_{f_{\mathrm{cond}, u_T}}(I, \Phi, 1)} = \frac{\hat{\alpha}_1(I, \Phi, T) - \hat{\alpha}_{g_{u_T}}(I, \Phi, T)}{T \mathbb{E}_\Phi \hat{\alpha}_{f_{\mathrm{cond}, u_T}}(I, \Phi, 1)}$$

$$= \frac{\lambda \frac{\hat{\alpha}_1(I, \Phi, T)}{\lambda T} - \mathbb{E}_\Phi \hat{\alpha}_{g_{u_T}}(I, \Phi, 1) \frac{\hat{\alpha}_{g_{u_T}}(I, \Phi, T)}{T \mathbb{E}_\Phi \hat{\alpha}_{g_{u_T}}(I, \Phi, 1)}}{\mathbb{E}_\Phi \hat{\alpha}_{f_{\mathrm{cond}, u_T}}(I, \Phi, 1)}$$

and the RHS converges to 1 as long as $\mathbb{E}_\Phi \hat{\alpha}_{1, f_{\mathrm{cond}, u_T}}(I, \Phi, 1)$ converges to 0 at a slower rate (in the sense of (5.29)) than $\frac{\hat{\alpha}_1(I, \Phi, T)}{\lambda T}$ and $\frac{\hat{\alpha}_{g_{u_T}}(I, \Phi, T)}{T \mathbb{E}_\Phi \hat{\alpha}_{g_{u_T}}(I, \Phi, 1)}$ approach 1. $\qquad\square$

Proof of Theorem 5.4.3. We have

$$\frac{\hat{\alpha}^*_{f_{u_T}}(I,\Phi,T)}{\sqrt{\hat{\alpha}_{f_{\text{cond},u_T}}(I,\Phi,T)}} = \frac{\hat{\alpha}^*_{f_{u_T}}(I,\Phi,T)}{\sqrt{[\lambda_{u_T}T]}} \frac{\sqrt{[\lambda_{u_T}T]}}{\sqrt{\lambda_{u_T}T}} \frac{\sqrt{\lambda_{u_T}T}}{\sqrt{\hat{\alpha}_{f_{\text{cond},u_T}}(I,\Phi,T)}}$$

and by Lemma 5.8.1, the last factor converges to 1. (Here, for $a \geq 0$, $[a]$ denotes the smallest integer $\geq a$.) Hence, for convergence of the LHS it is sufficient to show that $\hat{\alpha}^*_{f_{u_T}}(I,\Phi,T)/\sqrt{[\lambda_{u_T}T]}$ converges to a Gaussian variable. According to Kallenberg (1986, Lemma 2.1, Lemma 2.3), we can write Φ as a sum of Dirac measures $\delta_{(T_i,Y_i)}$, $i \in \mathbb{N}$, with random vectors (T_i,Y_i) and $T_1 \leq T_2 \leq \ldots$ If only a finite observation window $[0,T]$ is considered, the number of summands $N(T)$ is also finite but random. Then we introduce a modified version of $\hat{\alpha}^*_{f_u}(I,\Phi,T)$, in which the sum is cut after a fixed number $N_{\max} \in \mathbb{N}$ of terms:

$$\hat{\alpha}^{*,N_{\max}}_{f_u}(I,\Phi,T) = \sum_{i=1}^{N(T)} \sum_{j=1}^{N(T)} \left(f_u(Y_i) - \mu^{(2)}_{f_u,f_{\text{cond},u}}(I) \right) \cdot f_{\text{cond},u}(Y_i) \cdot \mathbf{1}_{T_j - T_i \in I}$$

$$\cdot \mathbf{1}_{\left[\sum_{i'=1}^{i-1} \sum_{j'=1}^{N(T)} f_{\text{cond},u}(Y_i)\mathbf{1}_{T_{j'}-T_{i'} \in I} + \sum_{j'=1}^{j} f_{\text{cond},u}(Y_i)\mathbf{1}_{T_{j'}-T_i \in I} \leq N_{\max} \right]} \cdot$$

Then we have

$$\frac{\hat{\alpha}^*_{f_{u_T}}(I,\Phi,T)}{\sqrt{[\lambda_{u_T}T]}} = \frac{\hat{\alpha}^{*,[\lambda_{u_T}T]}_{f_{u_T}}(I,\Phi,\infty)}{\sqrt{[\lambda_{u_T}T]}} + \frac{\hat{\alpha}^*_{f_{u_T}}(I,\Phi,T) - \hat{\alpha}^{*,[\lambda_{u_T}T]}_{f_{u_T}}(I,\Phi,\infty)}{\sqrt{[\lambda_{u_T}T]}} \tag{5.30}$$

and the first summand of the RHS contains a non-random number of summands (namely $[\lambda_{u_T}T]$). By the minimum distance assumption in condition (m-dependent Random Field Model), each mark Y_i occurs at most $|I|/d_0$ times in $\hat{\alpha}^{*,[\lambda_{u_T}T]}_{f_{u_T}}(I,\Phi,\infty)$. By the finite-range assumption on the covariance function of the underlying random field, the sequence $(Y_i)_{i\in\mathbb{N}}$ is $[h_0/d_0]$-dependent. Hence, the sequence of summands in $\hat{\alpha}^{*,[\lambda_{u_T}T]}_{f_{u_T}}(\Phi,I,\infty)$ is $[|I|h_0/d_0^2]$-dependent. By assumption, the first four moments of the excesses $Z_i = [f_{u_T}(Y_i) \,|\, f(Y_i) > u_T]$ exist and converge to some constant in $(0,\infty)$ as $T \to \infty$. Then the sequence of summands in $\hat{\alpha}^{*,[\lambda_{u_T}T]}_{f_{u_T}}(\Phi,I,\infty)$ satisfies the assumptions of Berk's CLT for triangular arrays of m-dependent random variables (Berk, 1973) and thus, for $T \to \infty$, $\hat{\alpha}^{*,[\lambda_{u_T}T]}_{f_{u_T}}(I,\Phi,\infty)/\sqrt{[\lambda_{u_T}T]}$ approaches a Gaussian distribution with zero mean and variance

$$u_\infty = \lim_{T\to\infty} \text{Var}\left[\hat{\alpha}^{*,[\lambda_{u_T}T]}_{f_{u_T}}(I,\Phi,\infty) \right] / ([\lambda_{u_T}T]).$$

Next, we show that the second summand in (5.30) converges to 0 in probability. We use the notation $\Delta\alpha_{f_u} = \hat{\alpha}^*_{f_u}(I,\Phi,T) - \hat{\alpha}^{*,[\lambda_{u_T}T]}_{f_u}(I,\Phi,\infty)$ and $\Delta\alpha_1 = \hat{\alpha}_{f_{\text{cond},u}}(I,\Phi,T) -$

$\hat{\alpha}_{f_{\text{cond},u}}^{[\lambda_{u_T}T]}(I,\Phi,\infty)$ and consider

$$\mathbb{P}(|\Delta\alpha_{f_{u_T}}| \geq \varepsilon\sqrt{[\lambda_{u_T}T]})$$
$$= \mathbb{P}\Big(|\Delta\alpha_{f_{u_T}}| \geq \varepsilon\sqrt{[\lambda_{u_T}T]} \,\Big|\, |\Delta\alpha_1| \geq \varepsilon[\lambda_{u_T}T]\Big) \cdot \mathbb{P}(|\Delta\alpha_1| \geq \varepsilon[\lambda_{u_T}T])$$
$$+ \mathbb{P}\Big(|\Delta\alpha_{f_{u_T}}| \geq \varepsilon\sqrt{[\lambda_{u_T}T]} \,\Big|\, |\Delta\alpha_1| < \varepsilon[\lambda_{u_T}T]\Big) \cdot \mathbb{P}(|\Delta\alpha_1| < \varepsilon[\lambda_{u_T}T])$$
$$\leq \mathbb{P}(|\Delta\alpha_1| \geq \varepsilon[\lambda_{u_T}T]) + \mathbb{P}\Big(|\Delta\alpha_{f_{u_T}}| \geq \varepsilon\sqrt{[\lambda_{u_T}T]} \,\Big|\, |\Delta\alpha_1| < \varepsilon[\lambda_{u_T}T]\Big). \qquad (5.31)$$

Note that $\hat{\alpha}_{f_{\text{cond},u_T}}^{[\lambda_{u_T}T]}(I,\Phi,\infty) = [\lambda_{u_T}T]$ and hence

$$\mathbb{P}(|\Delta\alpha_1| \geq \varepsilon[\lambda_{u_T}T]) = \mathbb{P}\left(\Big|\hat{\alpha}_{f_{\text{cond},u_T}}(I,\Phi,T)/[\lambda_{u_T}T] - 1\Big| \geq \varepsilon\right) \to 0 \quad \text{for } T \to \infty. \quad (5.32)$$

To estimate the the last summand in (5.31), we use again that the sequence $(Y_i)_{i\in\mathbb{N}}$ is $[h_0/d_0]$-dependent and that the number of points in any interval of length $|I|$ is bounded by $c = |I|/d_0$. This means that each term $f_{u_T}(Y_i)$ occurs at most c times in the sum $\Delta\alpha_{f_{u_T}}$. Obviously, the variance of $\Delta\alpha_{f_{u_T}}$, or more generally all even centered moments of $\Delta\alpha_{f_{u_T}}$, become maximal, if this boundary is bailed, i.e., if for a given total number $\Delta\alpha_1$ of summands, only $[\Delta\alpha_1/c]$ different Y_i are involved. With $Z_i^* = Z_i - \mathbb{E}Z_i = [f_{u_T}(Y_i) \,|\, f(Y_i) > u_T] - e(u_T)$, where $e(u) = \mathbb{E}\left[f_u(Y(0)) \,|\, f(Y(0)) > u\right]$, we get

$$\mathbb{P}(|\Delta\alpha_{f_{u_T}}| \geq \varepsilon\sqrt{[\lambda_{u_T}T]} \,|\, |\Delta\alpha_1| < \varepsilon[\lambda_{u_T}T])$$
$$= \mathbb{P}\left(|\Delta\alpha_{f_{u_T}}|^4 \geq \varepsilon^4[\lambda_{u_T}T]^2 \,\Big|\, |\Delta\alpha_1| < \varepsilon[\lambda_{u_T}T]\right)$$
$$\leq \mathbb{P}\left(\Big|\textstyle\sum_{i=1}^{[\varepsilon[\lambda_{u_T}T]c^{-1}]} cZ_i^*\Big|^4 \geq \varepsilon^4[\lambda_{u_T}T]^2\right)$$
$$\leq c^4 \sum_{i,j,k,l=1}^{[\varepsilon[\lambda_{u_T}T]c^{-1}]} \mathbb{E}(Z_i^*Z_j^*Z_k^*Z_l^*) \cdot (\varepsilon^4[\lambda_{u_T}T]^2)^{-1}$$
$$\leq c^4 \cdot [\varepsilon[\lambda_{u_T}T]c^{-1}] \cdot \left(\frac{h_0}{d_0}\right)^3 \mathbb{E}\left[(Z_1^*)^4\right] \cdot (\varepsilon^4[\lambda_{u_T}T]^2)^{-1}$$
$$= (\lambda_{u_T}T)^{-1}\varepsilon^{-3}\left(c\frac{h_0}{d_0}\right)^3 \mathbb{E}\left[(Z_1^*)^4\right](1 + o(1)) \longrightarrow 0, \quad (T \to \infty).$$

Plugging this and (5.32) into (5.31) yields that $\Delta\alpha_{f_{u_T}}/\sqrt{[\lambda_{u_T}T]} \to 0$ in probability. $\qquad\square$

6 Marked point process adjusted tail dependence analysis for high-frequency financial data

This chapter is based on the manuscript Malinowski *et al.* (2012*b*) and can partly be seen as an application of Chapter 5 in that large high-frequency datasets are modeled as a non-ergodic MPP.

6.1 Introduction

The irregular spacing of financial data recorded at intra-day frequency level has been inspiring an extensive usage of (marked) point process methods in econometric and financial applications. The seminal paper of Engle (2000) and the contributions of Engle & Lunde (2003), Bowsher (2007) and Bauwens & Hautsch (2009) are well-known examples. Yet, at the same time, movements of asset prices are commonly modeled via continuous-time stochastic processes—an approach that suggests to perceive transaction data as non-evenly spaced measurements of an underlying continuous-time process (e.g., Aït-Sahalia *et al.* 2005; Hansen & Lunde 2006). As long as the pattern of point locations is stochastically independent of the underlying process, global parameter estimation is well-established in literature (e.g., Hamilton 1993; Chiles & Delfiner 1999; Wackernagel 2003) including various contributions on declustering and debiasing of non-evenly spaced measurements (e.g., Journel 1983; Deutsch 1989; Isaaks & Srivastava 1989; Bourgault 1997; Emery & Ortiz 2005; Olea 2007). In contrast, preferential sampling introduces dependencies between locations and measured values; ignoring these dependencies and applying standard tools can lead to severe biases. Financial transaction data additionally involve another source of dependence: Since executing a transaction impacts on both the instantaneous and on the future price process, this form of taking measurements even causally influences the sample path. In suchlike situations, the random process imagination becomes less meaningful and the focus should be put on the marked points themselves rather than on a possibly underlying continuous-time process. Here, MPPs provide a commonly used framework that can capture arbitrary forms of dependency between point locations and marks (e.g., Karr 1991; Møller & Waagepetersen 2003; Schlather *et al.* 2004; Daley & Vere-Jones 2008; Myllymäki & Penttinen 2009; Bauwens & Hautsch 2009; Diggle *et al.* 2010). While the effect of single transactions is relevant on a rather small scale, on larger scales, continuous-time representations of asset prices might be advantageous; a general challenge is to model these processes across different scales (e.g., Duval & Hoffmann, 2011).

Another main issue of financial modeling and quantitative asset pricing is the assessment of financial risk. The rapid growth and globalization of financial markets together with the financial crises during the last decades have led to a strong demand for risk management

systems, which is also reflected by the regulatory accords issued by the Basel Committee on Banking Supervision. While some of the early models for asset returns, e.g., the capital asset pricing model (CAPM), are solely based on variances, risk measures that reflect the shareholders' preferences more adequately include the value at risk (VaR) and the expected shortfall, which is an average VaR and belongs to the class of coherent risk measures (cf. Artzner *et al.*, 1999). In contrast to VaR, the expected shortfall takes into account the whole tail of the distribution, yet it boils down the tail behavior to a single number and estimation of these quantities requires to restrict to the rare extreme events. Here, extreme value theory (EVT) provides suitable tools for precise estimation, based on the fundamental univariate limit theorem by Fisher & Tippett (1928). EVT further builds the theoretical framework for analyzing and modeling the joint extremal behavior of multiple assets, say, which is of particular interest with regard to crashes and large portfolio losses (e.g., Embrechts *et al.*, 1997; Embrechts, 2000).

By means of MPP moment measures, the definition of moments for inter-transaction returns is straightforward and can naturally be extended to the concept of conditional moments; here, conditioning refers to the existence of other transactions at a fixed temporal distance (cf. Chapter 3 and Schlather *et al.*, 2004) and introduces a notion of dependence between returns and the pattern of transaction times. While variances and covariances rather refer to the center part of the return distribution, i.e., normal market conditions, quantification of risk w.r.t. extreme market behavior particularly needs to include the returns' tail behavior. This chapter tries to bring together the two concepts, MPPs and EVT, in order to quantify interactions between the tail behavior and the pattern of transaction times.

In this context we have to question the assumptions of stationarity and ergodicity even after having corrected for seasonalities and trends. Many financial and economic processes exhibit structural breaks due to abrupt changes in the underlying economic mechanisms and conditions. Andreou & Ghysels (2009) reviews the literature on structural changes in financial time series and also covers stochastic volatility models for (intra-daily) asset return data. The authors emphasize both that there is empirical evidence for the existence of breaks in financial markets and that ignoring these structural breaks can have severe implications on statistical inference. Commonly, structural breaks are captured by means of regime-switching (in particular Markov-switching) models (Goldfeld & Quandt, 1973; Cosslett & Lee, 1985), in which the model parameters are allowed to take on different values in each of the different regimes. The regimes are given by an (unobserved) finite-state Markov chain. Originally only being applied to measures of economic output such as gross domestic product (Hamilton, 1989), applications of regime-switching models nowadays also include the modeling of inflation and interest rates, exchange rates, asset returns with different volatility regimes, and of other financial and economic quantities. We refer to Hamilton (2008), Piger (2009) and Lange & Rahbek (2009) for surveys on regime-switching and Markov-switching models and to Zucchini & MacDonald (2009) and Langrock (2011) for examples of applied literature on hidden Markov models. While structural breaks in the above references usually refer to substantial changes in the economy's structure and thus to larger temporal horizons, the concept of non-linear modeling has also found its way to the context of intra-day data. Zhang *et al.* (2001) introduced a threshold autoregressive conditional duration (TACD) model for financial transaction data that allows for multiple regimes and therewith improves

several aspects of the original ACD model (Engle & Russell, 1998). Their results on NYSE trading data suggest that the dynamics of the trading process differ between fast, normal and slow trading periods. Bauwens & Hautsch (2009, p. 963) names further contributions in which this model is extended and in which other mixture models are proposed. Based on a linear joint model for price changes and durations (PCD model), McCulloch & Tsay (2001) provide a nonlinear hierarchical model for financial transaction data which they consider as a compromise between fitting a PCD model for each trading day separately and fitting only a single PCD model for the concatenation of all trading days. They find that "there are some special days on which the behavior of the stock seems different from the others". In a similar spirit, Cartea & Jaimungal (2010) use a hidden Markov model combined with a joint model for price changes and durations in order to describe the intra-day dynamics of stock markets, particularly with regard to algorithmic trading. Here, regime switches represent different intra-day stages of the market.

In light of the above contributions on structural changes and nonlinear modeling, it seems both reasonable and natural to perceive financial transaction data as a concatenation of structurally different realizations of finite clock time length from a possibly non-ergodic MPP. Our analyses with high-frequency transaction data further motivate using the non-ergodic modeling approach developed in Chapter 5.

Since in general, the asymptotic distribution of point estimates for MPP characteristics is not analytically tractable, variance estimates and confidence intervals either have to be based on rather strong mixing or independence assumptions, or non-parametric techniques such as subsampling and bootstrapping can be used. When estimating the tail dependence index for stationary time series, which is closely related to estimating the ordinary tail index, Laurini & Tawn (2008) and Ledford & Tawn (2003) state that confidence intervals based on iid assumptions will be too small when the extremes are dependent. They propose a block bootstrapping method to obtain proper variance estimates for their estimators. In our MPP set-up and in view of the massive amount of data when intra-daily financial data are used, the subsampling approach of Politis & Sherman (2001) can be expected to yield reliable results. We will compare subsampling-based variance estimates to those obtained from assuming independence between point locations and marks of the MPP.

The rest of this chapter is organized as follows: In Section 6.2, we first review some basic concepts from EVT and MPP theory. Then the definition of conditional mean marks for MPPs is tailored to the extreme value context, where the interest is rather in the tail behavior and tail dependence than in moments. Section 6.5 also introduces extremal coefficients for point processes and discusses the extreme value analog of a Gaussian random field model. Though, the focus of this chapter is on the tail index of the mark distribution and its interaction with point locations. We propose an alternative to subsampling in order to assess the variability of the corresponding estimators and we provide a CLT-type result for MPPs to obtain the estimators' asymptotic distribution. In Sections 6.3 and 6.4, the methods are applied to simulated data and to real high-frequency transaction data from the German stock index DAX, respectively. We close with a summary and discussion of the results in Section 6.6. Technical details and proofs are given in Section 6.7.

6.2 Methods

Throughout this chapter, $\Phi = \{(t_i, y_i) : i \in \mathbb{N}\}$ is a stationary (not necessarily ergodic) marked point process on $\mathbb{R}$ with real-valued marks $y(t_i) = y_i$. Here, the t_i can simply be regarded as time points of transactions. Note that this is a special case of the framework in Chapter 5, from where we also adopt the definitions and notation.

6.2.1 Extremes

A random variable X is said to be extreme-value distributed if it is non-degenerate and there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$, $n \in \mathbb{N}$, such that, for independent copies Y_i, $i \in \mathbb{N}$, of a random variable Y, $\max_{i=1}^{n}(Y_i - d_n)/c_n$ converges to X in distribution as $n \to \infty$. The distribution of Y is then said to be in the max-domain of attraction (MDA) of X. The fundamental Fisher-Tippett-Gnedenko theorem states that there exist only three possible distributions for X, out of which the Fréchet distribution $\exp(-x^{-\alpha})$, $x > 0$, $\alpha > 0$, is the only heavy-tailed case and hence relevant for large parts of financial data. The quantity α^{-1} is commonly referred to as tail-index or extreme-value parameter and controls the thickness of the tail. For a random variable Y (w.l.o.g., $Y \geq 0$) in the MDA of a Fréchet(α) distribution, it is well-known that

$$\xi(u) = \mathbb{E}(\log Y - u | \log Y > u) \to \xi = \alpha^{-1} \qquad \text{for } u \to \infty \qquad (6.1)$$

(Embrechts *et al.*, 1997, Sec. 6.4.2). This result is also the basis for the well-known Hill estimator (Hill, 1975) of the tail index.

6.2.2 (Conditional) tail index for MPPs

Using the first- and second-order moment characteristics $\mu_f^{(i)}$ and $\tilde{\mu}_f^{(i)}$ defined in Chapter 5 we are now able to define a tail-index for MPPs and to extend it to conditional tail indices.

An MPP analog of $\xi(u)$ as in (6.1), is given by

$$\frac{\mathbb{E} \sum_{(t,y) \in \Phi, \, t \in [0,1]} (\log y - u) \cdot \mathbf{1}_{\log y > u}}{\mathbb{E} \sum_{(t,y) \in \Phi, \, t \in [0,1]} \mathbf{1}_{\log y > u}}.$$

which equals $\mu_{f,f_{\text{cond}}}^{(1)}$ for $f(y) = \log y - u$ and $f_{\text{cond}}(y) = \mathbf{1}_{\log y > u}$, and where $\mu_{f,f_{\text{cond}}}^{(1)}$ is defined analogously to $\mu_{f,f_{\text{cond}}}^{(2)}$ in (5.8). Taking limits for $u \to \infty$ gives a definition of the marks' tail index in an MPP setting, based on its mean excess representation.

We define *conditional tail indices* by including an additional conditioning on the existence of a further point, i.e.,

$$\xi(I, u) = \mu_{f_u, f_{\text{cond},u}}^{(2)}(I)$$

with $f_u(y) = \log y - u$ and $f_{\text{cond},u}(y) = \mathbf{1}_{\log y > u}$ for $y > 0$. For $y \leq 0$, we set $f_u(y) = f_{\text{cond},u}(y) = 0$. As in (5.6), we may define $\xi(r, u) = \mu_{f_u, f_{\text{cond},u}}^{(2)}(r)$ and consider ξ as a function

on $(\mathcal{B}(\mathbb{R}) \cup \mathbb{R}) \times \mathbb{R}^+$.

Let $\tilde{\xi}(I, u)$ and $\tilde{\xi}(r, u)$ denote the corresponding average second-order mean mark according to Definition 5.3.3, i.e., the two-stage expectation, averaging within each ergodicity class first, and then pooling the different classes. Finally, we let u tend to infinity and call $\xi(\cdot) = \lim_{u \to \infty} \xi(\cdot, u)$ and $\tilde{\xi}(\cdot) = \lim_{u \to \infty} \tilde{\xi}(\cdot, u)$ *conditional tail indices* of a mark of Φ, conditional on the existence of a further point at a certain distance.

6.2.3 Estimation

Following Section 5.4, the quantities $\mu^{(2)}_{f, f_{\mathrm{cond}}}(I)$ and $\tilde{\mu}^{(2)}_{f, f_{\mathrm{cond}}}(I)$ can naturally be estimated through

$$
\begin{aligned}
\hat{\mu}^{n, \mathrm{wght}}_{f, f_{\mathrm{cond}}}(I, \mathbf{w}) &= \hat{\mu}^{n, \mathrm{wght}}_{f, f_{\mathrm{cond}}}(I, \mathbf{w}, (\Phi_1, \dots, \Phi_n), T) \\
&= \left(\sum w_i(\Phi_i, T) \right)^{-1} \sum_{i=1}^{n} w_i(\Phi_i, T) \hat{\mu}_{f, f_{\mathrm{cond}}}(I, \Phi_i, T),
\end{aligned}
\tag{6.2}
$$

with

$$
\hat{\mu}_{f, f_{\mathrm{cond}}}(I, \Phi, T) = \frac{\hat{\alpha}_{f, f_{\mathrm{cond}}}(I, \Phi, T)}{\hat{\alpha}_{1, f_{\mathrm{cond}}}(I, \Phi, T)},
\tag{6.3}
$$

$$
\hat{\alpha}_{f, f_{\mathrm{cond}}}(I, \Phi, T) = \sum_{(t_1, y_1), (t_2, y_2) \in \Phi}^{\neq} f(y_1) f_{\mathrm{cond}}(y_1) \mathbf{1}_{(t_1, t_2) \in C(T, I)},
$$

where the weights $w_i(\Phi_i, T)$ are required to converge stochastically to some constant within each ergodicity class. For instance, with $w_i(\Phi_i, T) = 1$ and $w_i(\Phi_i, T) = T^{-1} \hat{\alpha}_{f, f_{\mathrm{cond}}}(I, \Phi_i, T)$, respectively, $\mu^{(2)}_{f, f_{\mathrm{cond}}}(I)$ and $\tilde{\mu}^{(2)}_{f, f_{\mathrm{cond}}}(I)$ can be estimated consistently.

Estimation of tail behavior generally requires a trade-off between tail relevance and the amount of data. For estimation of $\tilde{\xi}$, we have to choose a suitable threshold u and then take the estimator of $\tilde{\xi}(I, u)$ as an approximation of $\tilde{\xi}(I)$. Plugging in $f_u(y) = \log y - u$ and $f_{\mathrm{cond}, u}(y) = \mathbf{1}_{\log y > u}$ into (6.3), the canonical estimator of $\xi(I, u)$, based on a single realization of Φ, is

$$
\hat{\xi}(I, u, \Phi, T) = \frac{\sum_{(t_1, y_1), (t_2, y_2) \in \Phi}^{\neq} (\log y_1 - u) \mathbf{1}_{\log y_1 > u} \mathbf{1}_{(t_1, t_2) \in C(T, I)}}{\sum_{(t_1, y_1), (t_2, y_2) \in \Phi}^{\neq} \mathbf{1}_{\log y_1 > u} \mathbf{1}_{(t_1, t_2) \in C(T, I)}}.
$$

Having n realizations of Φ, we will consider the estimator

$$
\hat{\xi}^{n, \mathrm{wght}}(I, u, \mathbf{w}) = \hat{\xi}^{n, \mathrm{wght}}(I, u, \mathbf{w}, (\Phi_1, \dots, \Phi_n), T) = \hat{\mu}^{n, \mathrm{wght}}_{f_u, f_{\mathrm{cond}, u}}(I, \mathbf{w}),
\tag{6.4}
$$

where the RHS is given by (6.2). If all weights are chosen equal to 1, the estimator is already consistent for $\tilde{\xi}(I, u)$.

In order to employ a variance minimizing weighting similar to Section 5.4.3, we first need to introduce a σ-algebra that contains all information about the point locations of

$\Phi_1, \ldots, \Phi_n$ and about the locations of points whose log marks exceed the threshold u. Let $\mathcal{I}_u = \{\varnothing, [u, \infty), [u, \infty)^c, \mathbb{R}\} \subset \mathcal{B}(\mathbb{R})$ be the σ-algebra generated by the interval $[u, \infty)$ and let $\mathcal{M}^*$ be the smallest σ-algebra in $\mathcal{M}_0$, making all mappings $\Phi_{B \times L} : \mathbb{M}_0 \to \mathbb{N}_0$, $\varphi \mapsto \varphi(B \times L)$, $B \in \mathcal{B}(\mathbb{R}), L \in \mathcal{I}_u$, measurable, i.e.

$$\mathcal{M}^* = \sigma\left(\{\Phi_{B \times L}^{-1}(k) : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R}), L \in \mathcal{I}_u\}\right).$$

Here, $\mathcal{M}_0$ is the canonical σ-algebra associated to Φ (cf. Definition 2.1.1). Then we set $\mathcal{A}_u^* = \Phi^{-1}(\mathcal{M}^*) \subset \mathcal{A}$. Since $\sigma(\Phi^{-1}(\mathcal{E})) = \Phi^{-1}(\sigma(\mathcal{E}))$ holds true for any subset $\mathcal{E} \subset \mathcal{M}_0$, we have

$$\mathcal{A}_u^* = \Phi^{-1}(\mathcal{M}^*) = \sigma(\Phi^{-1}(\{\Phi_{B \times L}^{-1}(k) : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R}), L \in \mathcal{I}_u\}))$$
$$= \sigma\left(\{\{\omega \in \Omega : \Phi(\omega)(B \times L) = k\} : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R}), L \in \mathcal{I}_u\}\right),$$
$$\Phi_g^{-1}(\mathcal{M}_0(\mathbb{R})) = \sigma(\Phi_g^{-1}(\{\Phi_B^{-1}(k) : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R})\}))$$
$$= \sigma\left(\{\{\omega \in \Omega : \Phi(\omega)(B \times \mathbb{R}) = k\} : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R})\}\right).$$

Thus, $\Phi_g^{-1}(\mathcal{M}_0(\mathbb{R})) \subset \mathcal{A}_u^*$ and the ground process Φ_g is $(\mathcal{A}_u^*, \mathcal{M}_0(\mathbb{R}))$-measurable. By similar arguments, also the $\mathbb{N}_0$-valued random variable $\hat{\alpha}_{1, f_{\mathrm{cond}}, u}(I, \Phi, T)$ is $\mathcal{A}_u^*$-measurable.

Under some assumptions on the mark-location dependence, similar to those stated at the beginning of Section 5.4.3, we can improve the estimator's variance by choosing different weights while retaining consistency. In particular, we assume $\mathbb{E}[\hat{\xi}(I, u, \Phi_i, T) \,|\, \mathcal{A}_u^*]$ to be constant a.s. Then, according to Proposition 5.4.7, the optimal weights in (6.4) are given by the inverse of the conditional variances, i.e., $w_i = \mathrm{Var}[\hat{\xi}(I, u, \Phi_i, T) \,|\, \mathcal{A}_u^*]^{-1}$, provided that they are stochastically independent of the mixing random variable Q that corresponds to the ergodic decomposition of Φ according to Theorem 5.7.5. To get a feeling for the behavior of these conditional variances, we derive explicit expressions under some idealized assumptions. To treat the estimator's variance analytically, we assume that the point locations are independent of the marks and that the marks satisfy suitable mixing conditions. The worthiness of the following results for practical applications is discussed in the adjacent Section 6.2.4.

Condition (Independent-noise-marking).   *Let Y_i, $i \in \mathbb{Z}$, be iid variables in the MDA of a standard Fréchet distribution. We say that an MPP Φ satisfies the condition (Independent-noise-marking), if $\Phi \overset{d}{=} \{(t_i, Y_i) \,|\, t_i \in \tilde{\Phi}\}$ for some stationary unmarked point process $\tilde{\Phi}$ on $\mathbb{R}$, for which neighboring points have some minimum distance $d_0 > 0$ and which is independent of the Y_i.*

Condition (GRFM-trans).   *Let $\tilde{\Phi}$ be as in condition (Independent-noise-marking), and let $\{Y(t) : t \in \mathbb{R}\}$ be an independent random process which arises from a stationary Gaussian process Z by a monotone transformation of the margins, i.e., $Y = g(Z)$, such that the marginals of Y are in the Fréchet MDA. The covariance function C of Z is assumed to have a finite range, i.e., $C(h) = 0$ for all $|h| > h_0$ for some $h_0 > 0$. Then, we say that an MPP Φ is a Gaussian random field model with transformed margins, for short: Φ satisfies the*

condition (GRFM-trans), if $\Phi \overset{d}{=} \{(t_i, Y(t_i)) \mid t_i \in \tilde{\Phi}\}$.

We can formulate the following theorem. The proof is given in Section 6.7.

Theorem 6.2.1. *For a stationary MPP as in (GRFM-trans),*

$$\mathrm{Var}[\hat{\xi}(I, u, T) \mid \mathcal{A}_u^*] = v_u \left[\frac{\sum_{t_1 \in \Phi_{\mathrm{g}} \cap [0, T]} n(t_1, \Phi_{\mathrm{g}}, I, u)^2}{\left[\sum_{t_1 \in \Phi_{\mathrm{g}} \cap [0, T]} n(t_1, \Phi_{\mathrm{g}}, I, u)\right]^2} + \varepsilon_u \right],$$

where $v_u = \mathrm{Var}[\log Y(0) \mid \log Y(0) > u]$, $n(t_1, \Phi_{\mathrm{g}}, I, u) = \mathbf{1}_{\log Y(t_1) > u} \cdot \sum_{t_2 \in \Phi_{\mathrm{g}} \setminus \{t_1\}} \mathbf{1}_{t_2 - t_1 \in I}$ and ε_u is an $\mathcal{A}_u^*$-measurable random variable with $|\varepsilon_u| \to 0$ a.s. and in L_1, as $u \to \infty$.

Since condition (Independent-noise-marking) is a special case of condition (GRFM-trans), we directly get the following corollary:

Corollary 6.2.2. *For an MPP Φ satisfying the condition (Independent-noise-marking), the assertion of Theorem 6.2.1 holds with $\varepsilon_u = 0$.*

If u is large enough, the term ε_u in Theorem 6.2.1 can be neglected and the resulting optimal weights in $\hat{\xi}^{n,\mathrm{wght}}(I, u, \mathbf{w})$ are

$$w_i(\Phi_i, I, u) = \left[v_u \cdot \frac{\sum_{t_1 \in \Phi_{i,\mathrm{g}} \cap [0, T]} n(t_1, \Phi_{i,\mathrm{g}}, I, u)^2}{\left[\sum_{t_1 \in \Phi_{i,\mathrm{g}} \cap [0, T]} n(t_1, \Phi_{i,\mathrm{g}}, I, u)\right]^2} \right]^{-1}. \tag{6.5}$$

For the continuous case $\xi(r, u)$, we apply the analog of $\hat{\mu}_{f, f_{\mathrm{cond}}}(r, \Phi, T)$, which is given in Remark 5.4.12 and in which the indicator function $\mathbf{1}_I$ is replaced by a general kernel K_h. Then, the above formulae for the conditional variance and the weights are still valid if we replace $n(t_1, \Phi_{\mathrm{g}}, I, u)$ by

$$n(t_1, \Phi_{\mathrm{g}}, r, u) = \mathbf{1}_{\log Y(t_1) > u} \cdot \sum_{t_2 \in \Phi_{\mathrm{g}} \setminus \{t_1\}} K_h(r - (t_2 - t_1)).$$

6.2.4 Confidence intervals

In the following, we first derive the asymptotic distribution of $\hat{\xi}^{n,\mathrm{wght}}(I, u, \mathbf{w})$ under the above assumptions (Independent-noise-marking) and (GRFM-trans). If these assumptions are violated, the reliability of the resulting confidence intervals (CIs) can be assessed, e.g., via the non-parametric subsampling approach according to Politis & Sherman (2001).

Confidence intervals based on (Independent-noise-marking) and (GRFM-trans)

The estimator $\hat{\xi}^{n,\mathrm{wght}}(I, u, \mathbf{w})$ involves two levels of aggregation of independent or weakly dependent random terms: the outer summation over different realizations and the inner summation over all points of a particular realization.

We consider the inner level of aggregation first and assume Φ to be ergodic throughout this paragraph. We have $\hat{\xi}(I, u) = \hat{\alpha}_{f_u, f_{\mathrm{cond}, u}}(I, \Phi, T)/\hat{\alpha}_{1, f_{\mathrm{cond}, u}}(I, \Phi, T)$, which is an average of a random number of (dependent) summands. We get the following CLT-type result, which is a direct application of Theorem 5.4.3.

Theorem 6.2.3. *Let Φ be an MPP as in (Independent-noise-marking) or (GRFM-trans) and let $(u_T)_{T \geq 0}$ be a family of non-negative non-decreasing numbers such that the limit $u_\infty = \lim_{T \to \infty} u_T \in [0, \infty]$ exists and*

$$\frac{T^{-1}\hat{\alpha}_{1,1}(I, \Phi, T) - \lambda}{\mathbb{E}_\Phi \hat{\alpha}_{1, f_{\mathrm{cond}, u_T}}(I, \Phi, 1)} \to 0 \ a.s., \qquad as \ T \to \infty,$$

where λ is the intensity of point locations. Let

$$\hat{\alpha}^*_{f_u, f_{\mathrm{cond}, u}}(I, \Phi, T) = \sum_{(t_1, y_1), (t_2, y_2) \in \Phi}^{\neq} \left(f_u(y_1) - \mu^{(2)}_{f_u, f_{\mathrm{cond}, u}}(I) \right) \cdot f_{\mathrm{cond}, u}(y_1) \cdot \mathbf{1}_{(t_1, t_2) \in C(T, I)}$$

be a centered version of $\hat{\alpha}_{f_u, f_{\mathrm{cond}, u}}(I, \Phi, T)$.
  Then, for $I \in \mathcal{B}(\mathbb{R})$ and $T \to \infty$, we have

$$\frac{\hat{\alpha}^*_{f_{u_T}, f_{\mathrm{cond}, u_T}}(I, \Phi, T)}{\sqrt{\hat{\alpha}_{1, f_{\mathrm{cond}, u_T}}(I, \Phi, T)}} \Rightarrow \mathcal{N}(0, s_{u_\infty}),$$

where

$$s_{u_\infty} = \lim_{T \to \infty} v_{u_T} \cdot \left[\frac{\mathbb{E}_{\Phi_{\mathrm{g}}} \sum_{t_1 \in \Phi_{\mathrm{g}} \cap [0,1]} n(t_1, \Phi_{\mathrm{g}}, I, u_T)^2}{\mathbb{E}_{\Phi_{\mathrm{g}}} \hat{\alpha}_{1, f_{\mathrm{cond}, u_T}}(I, \Phi, 1)} + \mathbb{E}_{\Phi_{\mathrm{g}}} \varepsilon_{u_T} \right],$$

$$v_u = \mathrm{Var}[\log Y(0) \mid \log Y(0) > u], \qquad u \in [0, \infty),$$

and ε_u is given by Theorem 6.2.1 or Corollary 6.2.2. If the family $(u_T)_{T \geq 0}$ is eventually constant, then u_T can be replaced by the limiting constant $u_\infty \in [0, \infty)$. Furthermore, for u large (and $T > 0$ arbitrary),

$$\mathrm{Var} \frac{\hat{\alpha}_{f_u, f_{\mathrm{cond}, u}}(I, \Phi, T)}{\hat{\alpha}_{1, f_{\mathrm{cond}, u}}(I, \Phi, T)} = \mathrm{Var} \frac{\hat{\alpha}^*_{f_u, f_{\mathrm{cond}, u}}(I, \Phi, T)}{\hat{\alpha}_{1, f_{\mathrm{cond}, u}}(I, \Phi, T)} \approx v_u \mathbb{E}_{\Phi_{\mathrm{g}}} \left\{ \frac{\sum_{t_1 \in \Phi_{\mathrm{g}} \cap [0, T]} n(t_1, \Phi_{\mathrm{g}}, I, u)^2}{\hat{\alpha}_{1, f_{\mathrm{cond}, u}}(I, \Phi, T)^2} \right\}.$$
$$(6.6)$$

For a proof, the reader is referred to Section 6.7.

Concerning the outer level of aggregation in $\hat{\xi}^{n, \mathrm{wght}}(I, u, \mathbf{w})$, again by a CLT argument, the finite sample distribution is approximately Gaussian. By assumption, $\mathbb{E}[\hat{\xi}(I, u, \Phi_i, T) \mid \mathcal{A}^*_u]$ is a.s. constant and the weights in (6.5) are $\mathcal{A}^*_u$-measurable. Thus, the variance is obtained by a straightforward calculation using Theorem 6.2.1 (cf. (6.7)–(6.9) below). The weights in (6.5) can be considered as local weights in that they depend on the interval I. If $\xi(\cdot, u)$ is to be estimated for different distances (e.g. for a set of intervals $I_1, \ldots, I_J$), the use of local weights

might lead to instabilities and we may also consider global weights $w_k = w(\Phi_k, \cup_j I_j, u)$. In this case we have

$$
\begin{aligned}
&\mathrm{Var}\left[\hat{\xi}^{n,\mathrm{wght}}(I, u, \mathbf{w})\right] \\
&= \mathrm{Var}\left[\frac{1}{\sum_k w(\Phi_k, \cup_j I_j, u)} \sum_{k=1}^n w(\Phi_k, \cup_j I_j, u) \frac{\hat{\alpha}_{f_u, f_{\mathrm{cond},u}}(\Phi_k, I, T)}{\hat{\alpha}_{1, f_{\mathrm{cond},u}}(\Phi_k, I, T)}\right] \\
&= \mathbb{E}\,\mathrm{Var}\left[\frac{1}{\sum_k w(\Phi_k, \cup_j I_j, u)} \sum_{k=1}^n w(\Phi_k, \cup_j I_j, u) \frac{\hat{\alpha}_{f_u, f_{\mathrm{cond},u}}(\Phi_k, I, T)}{\hat{\alpha}_{1, f_{\mathrm{cond},u}}(\Phi_k, I, T)}\;\middle|\;\mathcal{A}_u^*\right] \\
&\approx v_u \cdot \mathbb{E}\left[\left[\frac{1}{\sum_k w(\Phi_k, \cup_j I_j, u)}\right]^2\right. \\
&\qquad \left.\sum_{k=1}^n \frac{\left[\sum_{t_1 \in \Phi_{k,\mathrm{g}} \cap [0,T]} n(t_1, \Phi_{k,\mathrm{g}}, \cup_j I_j, u)\right]^2}{\sum_{t_1 \in \Phi_{k,\mathrm{g}} \cap [0,T]} n(t_1, \Phi_{k,\mathrm{g}}, \cup_j I_j, u)^2} \frac{\sum_{t_1 \in \Phi_{k,\mathrm{g}} \cap [0,T]} n(t_1, \Phi_{k,\mathrm{g}}, I, u)^2}{\left[\sum_{t_1 \in \Phi_{k,\mathrm{g}} \cap [0,T]} n(t_1, \Phi_{k,\mathrm{g}}, I, u)\right]^2}\right].
\end{aligned}
$$
$$(6.7)$$

If local weights $w_k = w(\Phi_k, I, u)$ are used instead, (6.7) simplifies to

$$
\mathrm{Var}\left[\hat{\xi}^{n,\mathrm{wght}}(I, u, \mathbf{w})\right] \approx v_u \cdot n \cdot \mathbb{E}\left[\left[\frac{1}{\sum_k w(\Phi_k, I, u)}\right]^2\right].
\tag{6.8}
$$

If equal weights are used,

$$
\mathrm{Var}\left[\hat{\xi}^n(I, u)\right] \approx v_u \cdot n^{-1} \cdot \mathbb{E}\left[\frac{\sum_{t_1 \in \Phi_{k,\mathrm{g}} \cap [0,T]} n(t_1, \Phi_{k,\mathrm{g}}, I, u)^2}{\left[\sum_{t_1 \in \Phi_{k,\mathrm{g}} \cap [0,T]} n(t_1, \Phi_{k,\mathrm{g}}, I, u)\right]^2}\right].
\tag{6.9}
$$

In any of the three cases, the resulting CI is given by the Gaussian approximation. We will refer to the CIs based on this approach as *model-based confidence intervals.*

Subsampling-based confidence intervals

While the assumptions (Independent-noise-marking) and (GRFM-trans) allow for a theoretical calculation of the tail index estimator's asymptotic variance, subsampling provides a fully non-parametric way of estimating the uncertainty of the estimator. For a broad survey on bootstrapping and subsampling methods, we refer to Politis *et al.* (1999). For a general statistic $s(\Phi, T)$ for which $T\,\mathrm{Var}[s(\Phi, T)] \to V$ for some $V > 0$ as $|T| \to \infty$, Politis & Sherman (2001) showed that, under some mixing assumptions, V is consistently estimated through

$$
\hat{V} = [(1-c)T]^{-1} \int_{[0,(1-c)T]} cT \cdot \mathbb{E}\left[s(\Phi_{-y}, cT) - \overline{s(\Phi, cT)}\right]^2 dy
$$

if $c = c_T \to 0$ and $cT \to \infty$ as $T \to \infty$. Here, $\overline{s(\Phi, cT)} = [(1-c)T]^{-1} \int_{[0,(1-c)T]} s(\Phi_{-y}, cT) dy$ and Φ_x denotes the translation of the point pattern Φ by x.
The Riemann sum approximation of $\hat{V}$, is then given by

$$\hat{V}^* = n^{-1} \sum_{i=1}^{n} \frac{T}{n} \left[s(\Phi, [\tfrac{i-1}{n}T, \tfrac{i}{n}T]) - \bar{s} \right]^2$$

with $\bar{s} = n^{-1} \sum_{i=1}^{n} s(\Phi, [\tfrac{i-1}{n}T, \tfrac{i}{n}T])$. Hence,

$$\mathrm{Var}[s(\Phi, T)] \approx T^{-1} \hat{V}^* = n^{-2} \sum_{i=1}^{n} \left[s(\Phi, [\tfrac{i-1}{n}T, \tfrac{i}{n}T]) - \bar{s} \right]^2.$$

With regard to $\hat{\xi}^n(I, u)$, which already is an average of n realizations, an additional splitting of the observation window is not needed if n is sufficiently large. Then, $\mathrm{Var}\,\hat{\xi}^n(I, u)$ is naturally estimated through $n^{-2} \sum_{i=1}^{n} [\hat{\xi}(I, u, \Phi_i) - \bar{\xi}]^2$, where $\bar{\xi} = n^{-1} \sum_{i=1}^{n} \hat{\xi}(I, u, \Phi_i)$. Confidence intervals can again be based on the quantiles of the normal distribution since $\hat{\xi}^n(I, u)$ is asymptotically Gaussian (for $n \to \infty$) by the classical CLT. We will refer to these CIs as *subsampling-based confidence intervals*.

6.3 Simulation study

The model

Doubly stochastic Poisson processes (DSPPs), also called Cox processes, are well-established in the modeling of high-frequency financial data (e.g., Lando, 1998; Centanni & Minozzo, 2006; Hautsch, 2011). Here, we consider an DSPP-based MPP model, combined with an intensity-dependent marking (e.g., Ho & Stoyan, 2008; Myllymäki & Penttinen, 2009). Let $Z(t) = (Z_1(t), Z_2(t))$, $t \in \mathbb{R}$, be a bivariate stationary Gaussian field, where Z_1 generates the intensity and Z_2 drives the marks. This approach allows for a flexible management of dependencies between intensity and marks via the matrix-valued cross-covariance function $C(r) = \begin{pmatrix} C_{11}(r) & C_{12}(r) \\ C_{21}(r) & C_{22}(r) \end{pmatrix}$, $r \in \mathbb{R}$, where $C_{ij}(r) = \mathrm{Cov}(Z_i(0), Z_j(r))$, $i, j \in \{1,2\}$. The mean of Z is denoted by (m_1, m_2). In particular, the random intensity of point locations is given by $\exp(Z_1(\cdot))$, i.e., the unmarked ground process $\Phi_g = \{t_i : i \in \mathbb{N}\}$ is a log Gaussian Cox process (LGCP) with random intensity measure $\Lambda(B) = \int_B \exp(Z_1(t)) dt$. In addition, let $(Y_i)_{i \in \mathbb{N}}$ be a sequence of independent random Fréchet variables with $Y_i \sim F_{\alpha(t_i)}$, where $F_\alpha(x) = \exp(-x^{-\alpha})$ denotes the Fréchet distribution function with parameter α, and $\alpha(\cdot)$ is given by $\alpha(t) = \alpha_0 + \alpha_1 \exp(-Z_2(t))$, $\alpha_0, \alpha_1 \geq 0, t \in \mathbb{R}$. Let further $(S_i)_{i \in \mathbb{N}}$ be a sequence of iid random signs with $\mathbb{P}(S_1 = 1) = \mathbb{P}(S_1 = -1) = 0.5$. Then, conditionally on Z, let the marks be given by $y_i = m(t_i) = S_i Y_i$, i.e., their absolute values are Fréchet-distributed with an intensity-dependent tail parameter and their signs are random and independent of Z. In order that the first moments of the marks be finite, we assume $\alpha_0 \geq 1$.

Since $F_\alpha^{-1}(V) \sim F_\alpha$ for $V \sim U[0,1]$, we can consider the marks $y_i = m(t_i)$ as a deterministic function of time if we condition on Z, on the random field $V = \{V(t)\}_{t \in \mathbb{R}}$ of iid $U[0,1]$

variables and on the random field $S = \{S(t)\}_{t \in \mathbb{R}}$ of iid signs:

$$m_{v,s,\lambda}(t) = \big[m(t) \,|\, V(\cdot) = v(\cdot),\ S(\cdot) = s(\cdot),\ \exp(Z_2)(\cdot) = \lambda_2(\cdot) \big]$$
$$= s(t) \cdot F^{-1}_{\alpha_0 + \alpha_1/\lambda_2(t)}(v(t)).$$

If the two components of Z are positively correlated, the Fréchet parameter α tends to be small when the intensity of points is high. This will lead to increased conditional tail indices $\xi(r, u)$ for small temporal distances r. (Recall that the ordinary tail index of a Fréchet(α)-distribution is α^{-1}.)

Theoretical value of $\xi(r, u)$

Since $[\Phi | Z, V, S]$ is a Poisson point process with deterministic marks, we can treat the conditional tail index $\xi(r, u)$ partially analytically using an extended Campbell theorem and the fact that the reduced Palm measure of a Poisson process coincides with the probability measure P^Φ (e.g., Daley & Vere-Jones, 2008, Prop. 13.1.IV and Prop. 13.1.VII, resp.). For the second order moment measure $\alpha_{f, f_{\mathrm{cond}}}$ (cf. (5.3)), this yields

$$\alpha^{(2)}_{f^\ell, f_{\mathrm{cond}},}(C(I)) = \iint \int_{[0,1]} \mathbb{E}_{\Phi | Z, V, S}\big[(f^\ell f_{\mathrm{cond}})(m_{v,s,\lambda}(t_1)) \Phi_{\mathrm{g}}(I + t_1) \big]$$
$$\cdot \lambda_1(t_1) dt_1 \, \mathbb{P}^{(\exp(Z_1), \exp(Z_2))}(d\lambda_1, d\lambda_2) \mathbb{P}^V(dv) \mathbb{P}^S(ds)$$

$$= \iint \int_{[0,1]} (f^\ell f_{\mathrm{cond}})(m_{v,s,\lambda}(t_1)) \int_{I + t_1} \lambda_1(r) dr$$
$$\cdot \lambda_1(t_1) dt_1 \, \mathbb{P}^{(\exp(Z_1), \exp(Z_2))}(d\lambda_1, d\lambda_2) \mathbb{P}^V(dv) \mathbb{P}^S(ds), \quad \ell \in \{0, 1\}.$$

Due to the Cox-process-based construction of Φ, the measures $\alpha^{(2)}_{f^\ell, f_{\mathrm{cond}}}(C(\cdot))$, $\ell \in \{0, 1\}$, are dominated by the Lebesgue measure ν on $\mathbb{R}$ and with Fubini's theorem we get

$$\frac{\partial \alpha^{(2)}_{f^\ell, f_{\mathrm{cond}}}(C(r))}{\partial \nu(r)} = \iint (f^\ell f_{\mathrm{cond}})(m_{v,s,\lambda}(t_1)) \lambda_1(r + t_1) \lambda_1(t_1)$$
$$\mathbb{P}^{(\exp(Z_1), \exp(Z_2))}(d\lambda_1, d\lambda_2) \mathbb{P}^V(dv) \mathbb{P}^S(ds).$$

Hence,

$$\xi(r, u) = \frac{\partial \alpha^{(2)}_{f_u, f_{\mathrm{cond}, u}}(C(r))}{\partial \alpha^{(2)}_{1, f_{\mathrm{cond}, u}}(C(r))} = \frac{\partial \alpha^{(2)}_{f_u, f_{\mathrm{cond}, u}}(C(r))}{\partial \nu(r)} \cdot \left(\frac{\partial \alpha^{(2)}_{1, f_{\mathrm{cond}, u}}(C(r))}{\partial \nu(r)} \right)^{-1}$$

$$= \frac{\iint (f_u f_{\mathrm{cond}, u})(m_{v,s,\lambda}(0)) \lambda_1(r) \lambda_1(0) \, \mathbb{P}^{(\exp(Z_1), \exp(Z_2))}(d\lambda_1, d\lambda_2) \mathbb{P}^V(dv) \mathbb{P}^S(ds)}{\iint f_{\mathrm{cond}, u}(m_{v,s,\lambda}(0)) \lambda_1(r) \lambda_1(0) \, \mathbb{P}^{(\exp(Z_1), \exp(Z_2))}(d\lambda_1, d\lambda_2) \mathbb{P}^V(dv) \mathbb{P}^S(ds)},$$

$$(6.10)$$

where $f_u(y) = \log y - u$ and $f_{\mathrm{cond},u}(y) = \mathbf{1}_{\log y > u}$ for $y > 0$ and $f_u(y) = f_{\mathrm{cond},u}(y) = 0$ for $y \leq 0$, as before. Note that the right-hand side (RHS) of (6.10) is not an integral w.r.t. the law of the MPP anymore, but only w.r.t. the law of the random fields that drive the intensity of points and the marking. Although it is analytically intractable, Monte-Carlo simulation of Z, V and S provides an approximation $\hat{\xi}^{\mathrm{MC}}(r, u)$ of $\xi(r, u)$ for the above model. This enables a direct comparison of the true conditional tail index with the estimated one based on realizations of the full point process.

Results

Since, by construction, locations and marks are dependent, the confidence intervals derived in Section 6.2.4, which are based on the assumption (GRFM-trans), are only approximate and possibly underestimate the true variance. By this simulation study, we can determine the actual level of the confidence intervals.

The particular set-up is the following: For Z_1, we choose a mean value of $m_1 = -\log(0.5)$ and the exponential covariance model $C(h) = 0.1 \exp(-|h|/4)$. Perceiving distances as being measured in minutes, this choice causes the average distance between consecutive observations to be approximately 0.5 minutes and interaction effects to range up to 10 minutes, which roughly corresponds to the respective numbers in real transaction data (cf. Section 6.4). The second component of Z is a linear combination of shifts of Z_1: $Z_2(\cdot) = Z_1(\cdot) + \sum_{i=1}^n c_i[Z_1(\cdot) - Z_1(\cdot - s_i)]$ with $c_i, s_i \in \mathbb{R}$. This determines a particular form for the cross-covariance function of (Z_1, Z_2). While $c_1 = \ldots = c_n = 0$ implies
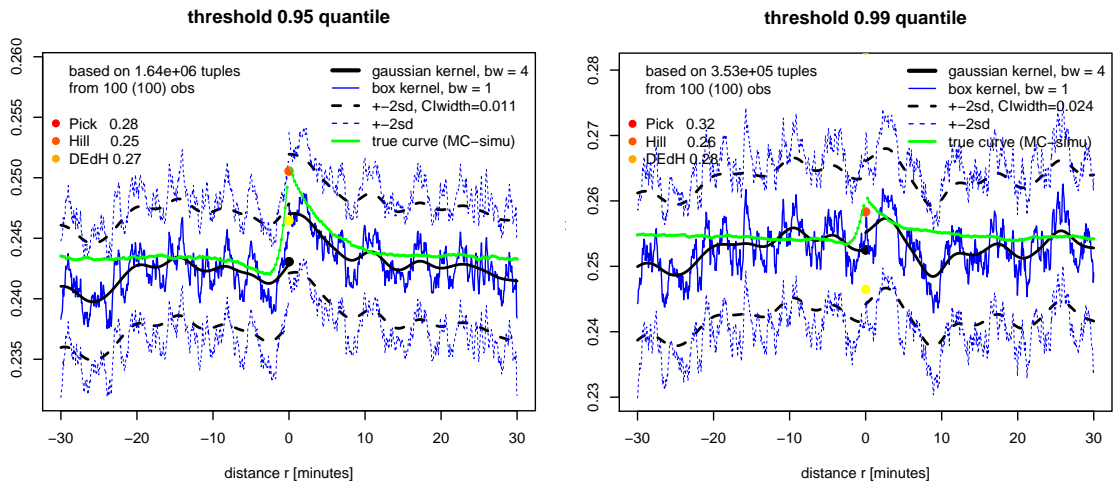

Figure 6.1: Estimation of $\xi(r, u)$ for u the 95% (left) and the 99% sample quantile (right) together with pointwise approximated 95%-CIs. The Gaussian kernel with a bandwidth of 4 and the rectangular kernel with a bandwidth of 1 are used. An approximation of the theoretical values $\xi(r, u)$ (green line) is based on Monte-Carlo simulation of the RHS of (6.10) with 10^6 realizations of the random fields Z, V and S.

completely symmetric interaction effects between marks and locations, positive values of c and s introduce asymmetry: $Z_2(t)$ is positively correlated with $Z_1(t)$ but $Z_2(t)$ is particularly large if Z_1 is small at the locations $t - s_i$. Since the Fréchet parameters of the marks are given by $\alpha(t) = \alpha_0 + \alpha_1 \exp(-Z_2(t))$, the larger the value of Z_2, the heavier the tail of the mark distribution. Hence, this specification of Z_2 with positive values of s_i induces a heavy tail at time t if the intensity of points at t is large *or* if there is an increase in intensity immediately before t. We choose $n = 100$ for smoothness reasons and $(c_1, \ldots, c_{100}) = (0.100, 0.099, \ldots, 0.001)$ and $(s_1, \ldots, s_{100}) = (2, 4, \ldots, 200)/60$. Further, we let $\alpha_0 = 3$ and $\alpha_1 = 0.1$.

The model is simulated on a 24,000 hour interval, which roughly corresponds to 3,000 days of trading, i.e., the point process contains approximately 3 million points. Figure 6.1 summarizes the behavior of the estimator $\hat{\xi}^n(r, u)$ based on such a realization, where u is the 95% and the 99% sample quantile and n is chosen to be 100, which means that the simulated dataset is split into 100 parts of a length roughly corresponding to one month. Note that in the simulation, there is no instationarity or regime-switching included; hence, the non-ergodic
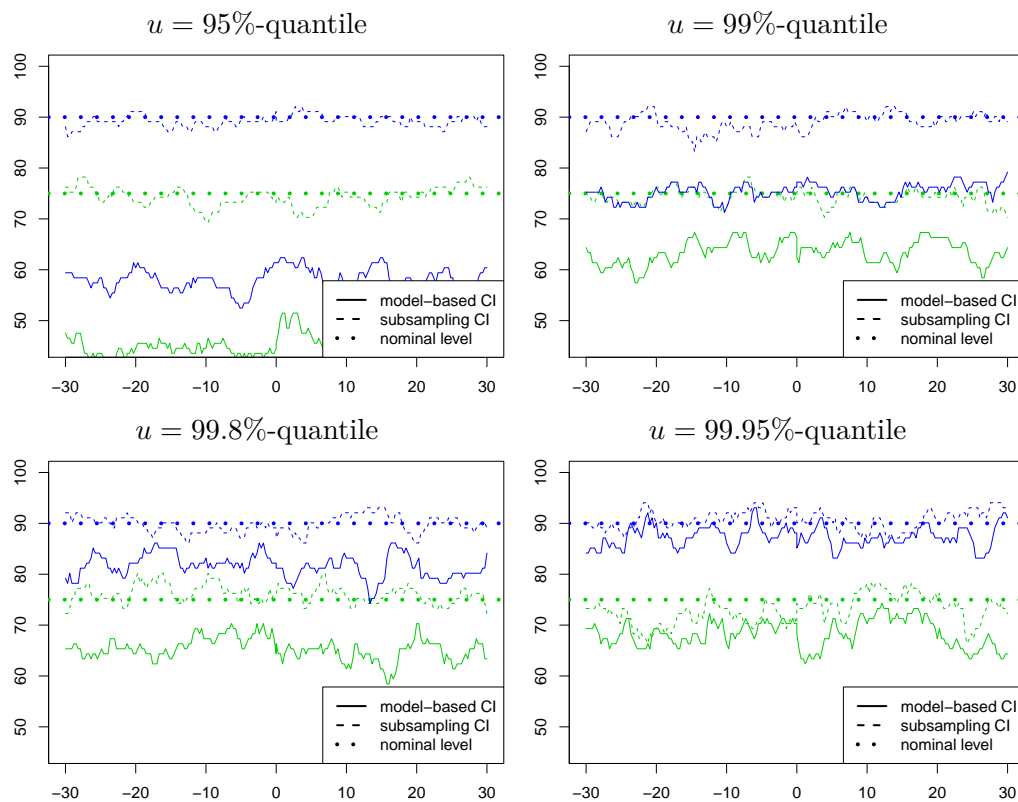

Figure 6.2: Empirical level of the model-based CIs according to (6.6) in Theorem 6.2.3 (continuous lines) and of subsampling-based CIs from Section 6.2.4 (dashed lines) for different thresholds u (from left to right, then top to bottom: 95%, 99%, 99.8% and 99.95%-quantile) and for the two nominal levels 75% and 90% (dotted lines).

modeling does not play an important role, here. As kernels for the estimator $\hat{\xi}(r, u)$, we use the Gaussian and the rectangular kernel with a bandwidth of 4 and 1, respectively. The approximated pointwise 95%-confidence intervals according to (6.6) in Theorem 6.2.3 (model-based CIs) are included. For the approximation $\hat{\xi}^{\mathrm{MC}}(r, u)$ of $\xi(r, u)$, one million realizations of the random fields Z, V and S on $[-30, 30]$ are generated. Note that, once Z is simulated, it is sufficient to simulate $V(\cdot) \sim U[u^*, 1]$ with $u^* = \inf_{t \in [-30,30]} F_{\alpha(t)}(\exp(u))$. For smaller values of V, $f_{\mathrm{cond}, u}(m(t))$ is zero and the corresponding points would not enter the estimator.

In order to validate the confidence intervals, a realization of the above model is simulated and the confidence intervals are calculated. Then, the model is simulated another 100 times and for each grid point $r \in [-30, 30]$, those realizations are counted whose respective values $\hat{\xi}^n(r, u)$ fall into the afore calculated CI. Then, the roles are interchanged 100 times such that each realization once becomes the center of the CI. As nominal levels, we choose 75% and 90%. Figure 6.2 shows the results for different values of the threshold u. It displays that for a relatively low threshold ($u = 95\%$-quantile), the variance of $\hat{\xi}^n(r, u)$ is considerably underestimated leading to an empirical level that is up to 25 percentage points below the nominal level. With increasing threshold, this error decreases. Already for the 99.8%-quantile, the confidence intervals hit the nominal level in average. The subsampling-based confidence intervals hit the nominal level for all thresholds, as expected. Note that here, the subsampling estimate of the variance of $\hat{\xi}^n(r, u)$ is based on all 101 simulated realizations, while each model-based CI is based on a single realization only. Hence, the variability of the subsampling variance estimate is smaller than for the model-based approach; though, this does not affect the comparison of coverage rates.

Figure 6.3 shows the estimator $\hat{\xi}^n(r, u)$, applied to the union of all 101 simulated datasets, together with an approximation to the true function $\xi(r, u)$, obtained from Monte-Carlo
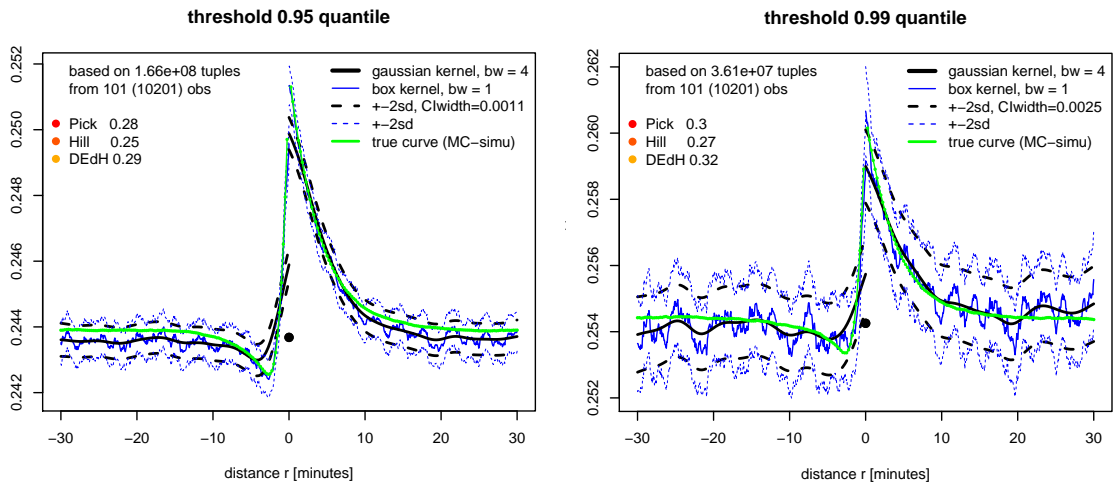

Figure 6.3: Analogously to Figure 6.1, but averaged over all 101 realizations.

simulation. The MC-estimate is fairly smooth since it is based on 10^8 random field realizations on the interval $[-30, 30]$. The width of the confidence intervals is roughly one tenth of that in Figure 6.1 since we have 101 realizations instead of one.

6.4 Application to transaction data from the German stock exchange

We apply the conditional tail index estimator $\hat{\xi}(r, u)$ to large transaction datasets from stock trading in Germany, processed via the Xetra trading system between 1997 and 2004. We consider blocks of size one year separately in order to exclude possible long-term effects. The same data pre-processing as in Engle (2000) is applied in order to account for diurnal patterns in the duration and return series. Further, the original returns are transformed to *returns per time unit* (cf. Engle, 2000 and Malinowski & Schlather, 2011b).

In correspondence with the various contributions on structural changes and nonlinear modeling of financial processes mentioned in the introduction, also in transaction datasets we can observe periods of trading that behave differently from the major part of the trading time. In the case that tail characteristics of the logreturns are of interest, only the set of extremal transactions is considered, which may further strengthen the effects of structural breaks. Hence, we consider a realization φ of the process Φ, observed on a certain interval B, as a concatenation of multiple realizations $\varphi_1, \ldots, \varphi_n$ of a possibly non-ergodic MPP, observed on smaller intervals $B_1, \ldots, B_n$, respectively, where the B_i are pairwise disjoint and $B = \cup_i B_i$; the objective is then to estimate the conditional tail index $\tilde{\xi}(r, u)$. This means that each trading period is considered to belong to one randomly chosen regime (ergodicity class) out of a possibly infinite number of different regimes. Here, we assume additionally independence between the concatenated parts, which is in general only an approximation to the truth. However, if the clock time length D of each period is large compared to the average inter-event distance within each realization of the point process, dependence between events from different parts can be expected to be fairly small and the error of this independence assumption is negligible. This argument might be formalized via some weak mixing conditions guaranteeing that the estimators of $\mu_f^{(i)}$, applied to the small realizations, become asymptotically independent (as $D \to \infty$). By the same arguments, edge effects due to finite observation windows can be neglected (e.g., Stoyan *et al.*, 1995).

As regards the concrete choice of D, McCulloch & Tsay (2001), for instance, assume that the length of each realization corresponds to one trading day, but also other choices of D might be adequate, depending on the statistical questions at hand.

In order to test the results for being significant, we consider the following null model: Within each subsample of length D, the marks of the MPP are randomly permuted while the pattern of point locations is kept fixed. Applying the above estimator to multiple realizations of this null model yields a set of reference curves that correspond to the scenario "no interaction effects".

6.4.1 Results

We apply four different levels of disaggregation, in particular, we split the data into blocks of length one year ($n = 1$), one month ($n = 12$), one week ($n = 52$) and one day ($n \approx 250$).

It turns out that a choice $n > 1$ yields more stable results and smaller estimated variances, compared to $n = 1$. However, going below a length of one week (i.e., $n > 52$) does not seem to be sensible since, particularly through the selection of extreme transactions, data become sparse and many of the small blocks would not contain *any* observation exceeding the threshold. Moreover, the estimation results do not differ significantly between moderate choices of n, i.e., n between 12 and 52. In the following, we therefore restrict to the results for partitioning into blocks of length one week, i.e. $n = 52$.

In section 6.2.3, variance-minimizing weights were introduced that maintain the consistency property of $\hat{\xi}^{n,\mathrm{wght}}(r, u, \mathbf{w})$ for $\tilde{\xi}(r,u)$ under some suitable independence assumptions. Though, it turns out that non-equal weighting (i.e., the use of $\hat{\xi}^{n,\mathrm{wght}}(r, u, \mathbf{w})$ instead of $\hat{\xi}^n(r, u)$) only marginally improves the estimator's variance since all weights turn out to be fairly similar in this particular dataset $(\exp(\mathrm{entropy}(\mathbf{w})) \approx n - 2)$. Hence, in order to be able to compare the estimated variance to a subsampling-based variance estimate, we restrict to the unweighted estimator in the following.

With regard to the choice of the threshold u, we remark the following: Actually, we are interested in the limit of $\tilde{\xi}(r, u)$ for $u \to \infty$, where in the general non-ergodic case, $\tilde{\xi}(r, u)$ is a weighted average over its ergodic counterparts. This averaging operator and the limit for u can be interchanged:

$$\lim_{u \to \infty} \tilde{\xi}(r, u) = \lim_{u \to \infty} \int_{\mathcal{P}_{\mathrm{erg}}} \xi^{\Phi|Q=Q^*}(r, u) \; \lambda(dQ^*) = \int_{\mathcal{P}_{\mathrm{erg}}} \lim_{u \to \infty} \xi^{\Phi|Q=Q^*}(r, u) \; \lambda(dQ^*),$$

so that using different sequences $(u_n) \to \infty$ for different ergodicity classes does not matter from a theoretical point of view. However, in a finite sample context, a finite threshold has to be chosen and the threshold values for the different ergodicity classes should be comparable. Hence, we might want to choose u to be the $(1 - \delta)$-quantile of the mark distribution. Since the quantiles are unknown and have to be estimated by their empirical counterparts, different approaches arise to choose the thresholds: (a) choose u to be the global $(1 - \delta)$-quantile of the set of all $\log y_i$ from all realizations or (b) calculate the $(1 - \delta)$-quantile for each realization separately. Especially for high quantiles, approach (a) causes many realizations to be excluded from the average because they do not contain marks exceeding the threshold. On the other hand, if—by chance—a realization does not contain extreme values, it is reasonable to exclude it from extreme value parameter estimation. Therefore, we apply a compromise: For every realization, we choose u to be the $(1 - \delta)$-quantile of all $\log y_i$ belonging to the "natural" larger temporal unit, e.g., if a realization corresponds to one day, the larger temporal unit is one week; accordingly for realizations consisting of one week or one month.

Exemplarily, Figure 6.4 shows the conditional tail index estimator for a one-year period (2004) of transaction data of the Deutsche Telekom AG stock (ISIN DE0005557508) with a total of 898,000 transactions. Here, exemplarily, only the lower tail, i.e., negative log returns, are considered. While in the above simulation study, the tails were symmetric by construction, in real data we might expect the tails to behave differently. However, the basic characteristics of the conditional tail index $\xi(r)$ turn out to be the same for negative and positive log returns in our transaction datasets. Note that this is contrary to larger scale

return data (e.g., daily data), for which the negative returns usually exhibit heavier tails than the positive returns.

Figure 6.4 exhibits that the tail index is significantly increased for small values of r (from -5 to $+10$ minutes). The confidence intervals indicate the precision of the estimates. Qualitatively the same results can be shown for most of the other stocks of the German stock index that have sufficiently long records. The increase of the tail index $\xi(r, u)$ at the origin is not completely symmetric, but the decay for $r > 0$ (conditioning on the future) is slower than for negative values of r. Assuming that causal influence can only be carried out by past events, this might sound counterintuitive at first sight. Though, transactions are generally clustered, which causes a large overlap between the data that enter into $\hat{\xi}^n(r, u)$ and those entering $\hat{\xi}^n(-r, u)$. Furthermore, an extreme log return possibly induces further immediate transactions due to reactions of other market participants. Hence, for small positive values of r, $\hat{\xi}^n(r, u)$ might be even larger than $\hat{\xi}^n(-r, u)$, although there is no causal influence from future transactions to current log returns. The gray curves in Figure 6.4 stem from applying the estimator to multiple realizations of the null model. The fact that the estimated curve for the original data projects beyond the range of the null model curves confirms that mark-location interactions w.r.t. the tail index exist.

Another observation that can be made from Figure 6.4 and that also holds true for the other German stock index datasets, is that the model-based confidence intervals approach the subsampling-based intervals as the threshold increases to a sufficiently high level. Figure 6.5 shows the ratio of average model-based standard deviation to average subsampling standard deviation, averaged over all distances r and upper and lower tails. Each boxplot
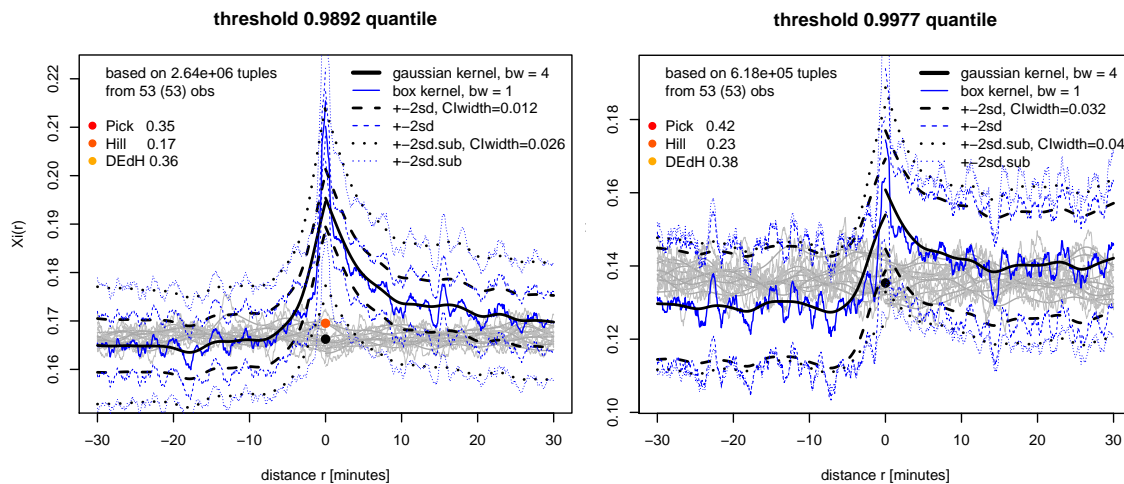

Figure 6.4: Estimation of $\xi(r, u)$ for a one-year period of transaction data of the Deutsche Telekom AG stock (ISIN DE0005557508) with a total of 898,000 transactions. u being the 98.9%- (left) and the 99.77%-quantile (right). Pointwise 95%-CIs based on Theorem 6.2.3 and subsampling-based CIs in dashed and dotted lines, resp. (Compare also Figure 6.1.) The gray lines stem from realizations of a null model.
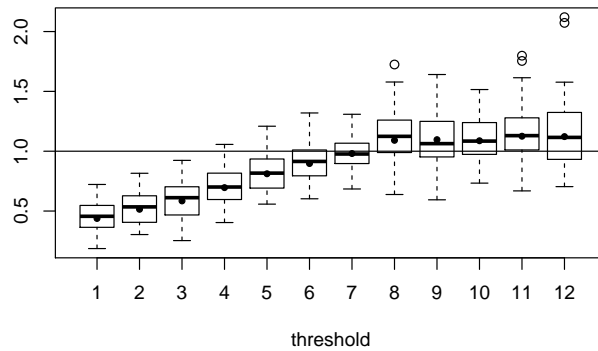
Figure 6.5: Ratio of model-based standard deviation to subsampling standard deviation for different thresholds: The i-th boxplot corresponds to the $\left(1 - 0.05 \cdot (\frac{1}{2})^{i-1}\right)$-quantile. Each boxplot contains the values for the 12 largest datasets ($\geq 500{,}000$ transactions).

represents a different threshold and contains the values for the 12 largest datasets ($\geq 500{,}000$ transactions). For low thresholds, the true variance, represented by the subsampling variance, is substantially underestimated by the model-based variance. For large thresholds, the ratio is slightly larger than 1 in average, thus introducing a bit of conservatism in the resulting confidence intervals. In summary, the asymptotic confidence intervals derived in Section 6.2.4, based on the assumption (GRFM-trans), work reasonably well for our transaction datasets and yield reliable results for thresholds above the 99.5%-quantile.

6.5 Max-stable random field model

In the previous part of this chapter, *tail dependence* was understood as interaction of the tail index with neighboring point locations in an MPP context. In classical extreme value literature, tail dependence refers to the joint extremal behavior of the components of a random vector or to the spatial dependence within a max-stable process, and there exist different concepts of measuring tail dependence (e.g., Smith, 1990, Embrechts *et al.*, 1997, Sec. 8.1, Schlather & Tawn, 2003). In what follows, we combine the above concept of measuring mark-location interactions in MPPs with the classical notion of tail dependence for multivariate distributions. We will assume that the mark distributions belong to the maximum domain of attraction (MDA) of a max-stable distribution.

A simple MPP Φ on T, $T \subset \mathbb{R}$ or $T \subset \mathbb{R}^d$, is said to belong to the class of *random field models* (Mase, 1996), if it is in distribution equal to unbiased sampling of a process, i.e.

$$\Phi \stackrel{d}{=} \{(t, Z_t) : t \in \Phi_{\mathrm{g}}\}$$

for some random process $(Z_t)_{t \in T}$ being independent of the unmarked ground process Φ_{g}. In this situation, mark-location interactions clearly do not exist. Hence, it is generally worthwhile to obtain conditions that guarantee an MPP to be a random field model and

that can be verified easily.

In the case where Z is a Gaussian random process, Schlather *et al.* (2004) provide an obvious set of necessary and sufficient conditions. Since a multivariate Gaussian distribution is determined by its mean and covariance structure, those conditions involve no higher than second moments.

The dependence structure of extreme events, though, is generally much more complex than in the Gaussian case and cannot necessarily be described by a finite set of parameters. For a multivariate extreme value distribution (MEVD), the dependence structure is fully characterized by the so-called *spectral measure* on the unit sphere (e.g., Resnick, 2008). In applications, tail dependencies play an important role since they can seriously increase the risk of extreme events, e.g., in an asset portfolio.

In Example 6.5.4, we will illustrate that, due to the complexity of the dependence structure of a MEVD, it is not possible to establish a random field model criterion comparable to the Gaussian case.

A set of simple characteristics summarizing the dependence structure of a multivariate extreme-value distribution are the extremal coefficients (Smith, 1990). For $Z = (Z_1, \ldots, Z_n)$ following a MEVD with identically distributed margins, there are $2^n - 1$ extremal coefficients θ_I, I a non-empty subset of $\{1, \ldots, n\}$, defined by $\mathbb{P}(\max_{j \in I} Z_j < z) = \mathbb{P}(Z_1 < z)^{\theta_I}$. Even the full set of extremal coefficients does not completely determine the spectral measure H. Hence, we cannot expect a condition only based on extremal coefficients to be sufficient to decide whether an MPP belongs to the class of max-stable random field models, even though we restrict ourselves to MPPs with standard Fréchet marks. In the following, we aim at constructing an MPP example with all extremal coefficients being independent of further points of the MPP while the multivariate mark distribution is not independent of the pattern of locations. We use a family of distributions that is determined by the full set of extremal coefficients and that is obtained as limits of certain max-linear combinations (cf. Schlather, 2002; Strokorb & Schlather, 2012).

6.5.1 Extremal coefficients for MPPs

First, we introduce extremal coefficients and conditional extremal coefficients for MPPs: Analogously to (5.3), let $\alpha_f^{(n)}$ be the n-th order f-moment measure of Φ with a non-negative function f, i.e.

$$\alpha_f^{(n)}(C) = \mathbb{E} \sum_{(t_1, y_1), \ldots, (t_n, y_n) \in \Phi}^{\neq} f(y_1, \ldots, y_n) \mathbf{1}_C((t_1, \ldots, t_n)), \quad C \in \mathcal{B}(\mathbb{R}^n).$$

Again, the Radon-Nikodym theorem provides a density of $\alpha_f^{(n)}$ w.r.t. $\alpha^{(n)}$. For the special choice $f(y_1, \ldots, y_n) = \mathbf{1}_{y_1 \leq m_1} \cdot \ldots \cdot \mathbf{1}_{y_n \leq m_n}$, this density is commonly referred to as the n-point (Palm) mark distribution, which we denote as

$$F_{(t_1, \ldots, t_n)}(m_1, \ldots, m_n) = \frac{\partial \alpha_f^{(n)}(\cdot)}{\partial \alpha^{(n)}(\cdot)}((t_1, \ldots, t_n)) \tag{6.11}$$

for pairwise disjoint locations $t_1, \ldots, t_n$. Its actual meaning might become clearer by writing it as a conditional distribution:

$$F_{(t_1,\ldots,t_n)}(m_1, \ldots, m_n) = \mathbb{P}(y(t_i) \leq m_i, i = 1, \ldots, n \,|\, t_1, \ldots, t_n \in \Phi_g). \qquad (6.12)$$

For more details on Palm mark distributions, the reader is referred to Stoyan *et al.* (1995, Chap. 4) and Kallenberg (1986, p. 164).

In what follows, we assume the univariate margins of $F_{(t_1,\ldots,t_n)}$ to be identical for $\alpha^{(n)}$-almost all tuples $(t_1, \ldots, t_n)$, and we assume $F_{(t_1,\ldots,t_n)}$ to be in the MDA of some extreme value distribution G with Fréchet margins. Then, following Marshall & Olkin (1983, Prop. 3.1), for example, and with $\mathbf{m} = (m_1, \ldots, m_n)$,

$$\lim_{u \to \infty} \frac{1 - F_{(t_1,\ldots,t_n)}(um_1, \ldots, um_n)}{1 - F_{(t_1,\ldots,t_n)}(u, \infty, \ldots, \infty)} = -\log G(\mathbf{m}).$$

In particular, and by Taylor expanding $\log F_{(t_1,\ldots,t_n)}$ at 1, we have

$$\lim_{u \to \infty} \frac{\log F_{(t_1,\ldots,t_n)}(u, \ldots, u)}{\log F_{(t_1,\ldots,t_n)}(u, \infty, \ldots, \infty)} = \lim_{u \to \infty} \frac{1 - F_{(t_1,\ldots,t_n)}(u, \ldots, u)}{1 - F_{(t_1,\ldots,t_n)}(u, \infty, \ldots, \infty)} = -\log G(\mathbf{1}).$$

Hence, the limits in the following definition are well-defined.

Definition 6.5.1. For $F_{(t_1,\ldots,t_n)}$ having identical margins and being in the MDA of some MEVD with Fréchet margins, we define extremal coefficients

$$\theta_{t_1,\ldots,t_n} = \lim_{z \to \infty} \frac{\log F_{(t_1,\ldots,t_n)}(z, \ldots, z)}{\log F_{(t_1,\ldots,t_n)}(z, \infty, \ldots, \infty)}, \qquad (6.13)$$

Furthermore, for $m \in \mathbb{N}$ and $s_1, \ldots, s_m \in \mathbb{R}^d$, we define conditional extremal coefficients

$$\theta_{t_1,\ldots,t_n}^{s_1,\ldots,s_m} = \lim_{z \to \infty} \frac{\log F_{(t_1,\ldots,t_n,s_1,\ldots,s_m)}(z, \ldots, z, \infty, \ldots, \infty)}{\log F_{(t_1,\ldots,t_n,s_1,\ldots,s_m)}(z, \infty, \ldots, \infty)}. \qquad (6.14)$$

Remark 6.5.2. *(a)   The marginal distributions of the max-limiting distribution of (6.11) do not necessarily coincide for different sets of locations t_i, even if two sets of locations differ by one point, only.*
(b)   It is essential to require that (6.11) be in the MDA of some extreme value distribution for all $n \in \mathbb{N}$. If this was only required for $n = 1$,

$$F_{(t_1,t_2)}(\,\cdot\,, \infty) = \mathbb{P}(y(t_1) \leq \cdot \,|\, t_1, t_2 \in \Phi_g)$$

would not necessarily be in the MDA of any EVD: Via a hard-core construction as in the following Example 6.5.4, for instance, we can easily define a process for which $F_{(t_1,t_2)}(\,\cdot\,, \infty) = \mathbb{P}(Z \leq \cdot \,|\, Z < 1)$ if $\|t_1 - t_2\| < 1$ and $F_{(t_1,t_2)}(\,\cdot\,, \infty) = \mathbb{P}(Z \leq \cdot \,|\, Z \geq 1)$ if $\|t_1 - t_2\| \geq 1$, for Z being a standard Fréchet variable. Then, for $\|t_1 - t_2\| < 1$, the max-limiting distribution is clearly degenerate.

Remark 6.5.3. *If Φ is stationary, the mark distribution $F_{(t_1,\ldots,t_n)}$ and hence also $\theta_{t_1,\ldots,t_n}$ only depend on the distance vectors $t_2 - t_1, \ldots, t_n - t_1$.*

Let Φ be additionally isotropic. Note that, in general, this does not imply that $F_{(t_1,\ldots,t_n)}$ only depends on the Euclidean distances $\|t_2 - t_1\|, \ldots, \|t_n - t_1\|$. To see this, consider a hard-core MPP, in which each cluster of point locations forms the same non-equilateral triangle. Then, stationarity means that all clusters have the same stochastic behavior. Now, if we mirror some of the triangles and assign a different mark distribution to the respective clusters, the Euclidean distances between the three points do not change but the mark distribution does change. Note that this is not contradictory to the isotropy assumption since mirroring a point configuration can in general not be replaced by a rotation.

However, if the point process is defined on the real axis, as it is the case for temporal data, mirroring and rotating a point pattern yield the same result. Hence, in this case, isotropy implies that $\theta_{t_1,\ldots,t_n}$ only depends on the distances $|t_2 - t_1|, \ldots, |t_n - t_1|$. Note that even if all mark distributions are multivariate Gaussian, isotropy is not necessarily implied by stationarity, as opposed to the case of a continuous-time Gaussian process.

In the following Example 6.5.4, we will construct an MPP, for which all multivariate mark distributions, described by (6.11), are max-stable distributions. By marginal transformation and hence w.l.o.g., we can assume that all mark distributions have standard Fréchet margins. Particularly, this means that all univariate mark distributions are independent of the positions of further points and that the limits in (6.13) and (6.14) can be dropped. Then

$$\theta^{s_1,\ldots,s_m}_{t_1,\ldots,t_n} = \frac{\log F_{(t_1,\ldots,t_n,s_1,\ldots,s_m)}(z,\ldots,z,\infty,\ldots,\infty)}{\log F_{t_1}(z)}$$

for any $z > 0$; the analog holds for (6.13).

6.5.2 Extremal coefficients and random field models

Obviously, the condition

$$\theta^{s_1,\ldots,s_m}_{t_1,\ldots,t_n} = \theta_{t_1,\ldots,t_n} \quad \text{for } \alpha^{(n+m)}\text{-almost all } t_1,\ldots,t_n,s_1,\ldots,s_m \in \mathbb{R}^d, \ n,m \in \mathbb{N} \qquad (6.15)$$

is a necessary condition for Φ to be a random field model. The following counterexample however shows that *even* for the restricted class of MPPs with standard Fréchet margins, (6.15) is not a sufficient condition. In a way, this contrast the Gaussian random field model case for which Schlather *et al.* (2004) provide equivalent conditions based on conditional means and covariances.

Example 6.5.4. *For $n \in \mathbb{N}$, let S_n denote the unit sphere in $\mathbb{R}^{n+1}$. Let Φ^* be a stationary marked Poisson point process on $\mathbb{R}^d$, $d \geq 3$, with iid marks $M = (\alpha,\beta,\xi_1,\xi_2,Y)$, where $\alpha \sim U(S_{d-1})$, $\beta \sim U(S_{d-2})$, $\xi_1 \sim U((0,1])$, $\xi_2 \sim U((1,3])$ and Y are independent random variables, and where Y is a vector of five independent standard Fréchet variables.*

Let $\tilde{\Phi}$ consist of all points $[t,M] \in \Phi^*$ for which the ball $B_{12}(t)$ around t with radius 12 contains no further points of Φ^*. Then, we obtain an MPP Φ by adding two points to each $[t_1,M] \in \tilde{\Phi}$ in the following way: Let $t_2 = t_1 + \xi_1\alpha$ and let t_3 be uniformly distributed on an

orbit consisting of all points from which t_1 and t_2 have the same fixed distance. More precisely: $t_3 = (t_1 + t_2)/2 + \xi_2 \cdot \iota(\beta)$, where ι is an inner-product-preserving embedding of $\mathbb{R}^{d-1} \supset S_{d-2}$ into $\mathbb{R}^d$ such that the center point of $\iota(S_{d-2})$ equals 0 and the orthogonal complement of $\iota(S_{d-2})$ in $\mathbb{R}^d$ contains $t_2 - t_1$. Note that $\|t_1 - t_2\| \in (0,1]$, $\|t_1 - t_3\| = \|t_2 - t_3\| \in (1,4)$, and that the distance between points from different clusters is ≥ 4. The corresponding marks m_1, m_2, m_3 are supposed to follow a multivariate Fréchet distribution. In particular let

$$A = \begin{pmatrix} .5 & 0 & .5 & 0 & 0 \\ 0 & .5 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad and \quad B = \begin{pmatrix} .4 & 0 & .4 & .2 & 0 \\ 0 & .5 & .4 & .1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and let $y = (y_1, y_2, y_3)^T = A \dot{\vee} Y \cdot \mathbf{1}_{\xi_2 \leq 2} + B \dot{\vee} Y \cdot \mathbf{1}_{\xi_2 > 2}$, where $A \dot{\vee} Y$ denotes the vector of max-linear combinations $\max_j A_{ij} Y_j$, $i = 1, 2, 3$. Note that the extremal coefficients of $A \dot{\vee} Y$ have the simple form $\theta_I = \sum_{j=1}^{J} \max_{i \in I} A_{ij}$ (by direct calculation), where $J = 5$ in our case. A and B are chosen such that the respective max-linear combinations have the same extremal coefficients but not the same multivariate distribution.

In summary, t_2, t_3 and y are deterministic functions of $[t_1, M]$ and

$$\Phi = \left\{ [t_1, y_1], [t_2, y_2], [t_3, y_3] : [t_1, M] \in \tilde{\Phi} \right\}.$$

By construction, the marks of two points of Φ with a distance greater than 1 are stochastically independent. Within each cluster consisting of three points, there is exactly one pair of points with a distance less or equal than 1. For each cluster, y_3 is independent of (y_1, y_2) since the third row and the fifths column of A and B equal the unit vectors e_5 and e_3, respectively. Different clusters are also stochastically independent. Hence, every extremal coefficient $\theta_{t_1,\dots,t_n}$ with $3 \leq n \leq \infty$ can be expressed as a sum of bivariate extremal coefficients and it suffices to consider those conditional extremal coefficients that condition on the existence of only one additional point. Henceforth, we consider all coefficients θ_{t_1,t_2} and θ_{t_1,t_2}^s, $t_1, t_2, s \in \mathbb{R}^d$.

Due to stationarity of Φ and the rotation invariant construction of the local point clusters, we may define $\theta(h) = \theta_{t_1,t_2}$ if $\|t_2 - t_1\| = h$ and $\theta(h, d_1, d_2) = \theta_{t_1,t_2}^s$ if additionally $\|s - t_1\| = d_1$ and $\|s - t_2\| = d_2$. Then the above construction yields that

$$\theta(h, d_1, d_2) = \begin{cases} 1.5, & h \in [0,1] \\ 2, & h \in (1, \infty). \end{cases}$$

Note that, even for $h \in [0,1]$, no further distinction of cases is needed since the matrices A and B are chosen such that they lead to the same extremal coefficients. Hence, $\theta(h, d_1, d_2)$ only depends on h and we have that $\theta(h) = \theta(h, d_1, d_2)$ for all valid d_1 and d_2.

However, for $\|t_2 - t_1\| \in (0,1]$, the conditional mark distributions

$$F_{(t_1,t_2,s)}(m_1, m_2, \infty)$$

obviously depend on the distance $\|s - t_1\|$ in the same way in which the choice between the

matrices A and B depends on the distance to the third point of each cluster.

The example shows that max-stable random field models cannot be characterized by only considering conditional extremal coefficients. Intuitively, the extremal coefficients only report the multivariate distribution function on the diagonal, which is not sufficient information to reconstruct the MEVD. Though, defining more general extremal dependence measures with different thresholds m_i for the different components $y(t_i)$ (e.g., Fasen *et al.*, 2010), simply leads to a re-parametrization of the distribution function and it would be a merely trivial result to characterize a max-stable random field model via conditions of the type

$$F_{(t_1,\ldots,t_n)}(m_1,\ldots,m_n) = F_{(t_1,\ldots,t_n,s_1,\ldots,s_m)}(m_1,\ldots,m_n,\infty,\ldots,\infty)$$
$$\forall n,m \in \mathbb{N},\ t_i,s_i \in \mathbb{R}^d,\ m_i \in \mathbb{R}.$$

6.6 Discussion

Irregularly spaced financial data, particularly log returns between consecutive transactions of electronically traded assets, can naturally be perceived as MPPs, which are therefore well-established in financial and econometric literature. At the same time, modeling extreme financial events is of pivotal interest, for example, for insurance or risk management purposes. This chapter tries to bring together these two concepts.

Based on existing second-order moment measures for marked point processes, we propose an MPP analog of the extreme value index (tail index) and extremal coefficients as well as conditional versions thereof to detect whether these mark characteristics depend on the point pattern in its neighborhood. MPP analogs for other summary statistics of (multivariate) extreme value distributions can be defined in a similar way. While the tail index is a univariate property, conditional extremal coefficients, for instance, allow for detection of interaction of multivariate mark distributions with the pattern of point locations. Suchlike characteristics can also help to detect, whether the sampling of a continuous-space process is independent of the process itself or not.

Since the above summary statistics are defined as a mean of certain mark functionals, the question arises, which mean is actually of interest in a practical situation. If there is an underlying continuous-time process from which the data are generated by a random sampling procedure, then the *mean of interest* will be reflected by the temporal average over the whole index space instead of the average over the sampling locations. Weighting procedures can then be used to compensate for the irregular distribution of point locations. Though, the assumption of a continuous-time background process seems to be problematic in the framework of financial transaction data since the observed values interact with each other and with the point pattern. In Section 5.5, we already suggested to proceed differently for different scales, i.e., to consider the data as a genuine MPP on the very small scale, but to assume an underlying random field on larger scales and to correct for the irregular distribution of locations by a weighting procedure based on the idea of variance minimization. This is closely related to including non-ergodicity into the model and to replacing expectation functionals w.r.t. the point process by two-step expectations that average within each ergodicity class

first and then aggregate the different classes.

When being faced with real data, assuming ergodicity or not is entirely discretionary since we are always restricted to finite observation windows. As regards financial transaction data, we argue that it is sensible to perceive the data as a concatenation of multiple realizations of a possibly non-ergodic MPP. Furthermore, regardless of whether the process is ergodic or not, employing the estimators derived for the non-ergodic set-up can improve the statistical properties of the estimate.

By applying the conditional tail index estimator to real transaction data, we see that the tail index of log returns is significantly increased if there are other transactions within few minutes before. We assess finite sample properties of the respective estimators, in particular the variability, exploiting that, due to the thresholding, consecutive events that exceed the threshold become stochastically independent under some weak assumptions. The variance estimates based on the assumption (GRFM-trans) and the subsampling-based variance estimates turn out to coincide for sufficiently high thresholds. Though, for general processes, variance estimates of $\hat{\xi}^{n,\mathrm{wght}}(I, u, \mathbf{w})$ based on (Independent-noise-marking) or (GRFM-trans) can be highly biased.

The detection of an increase of risk caused by the existence of other transactions, might by itself be a valuable finding for risk management purposes or automated trading algorithms— referring to a very fine temporal scale. Yet, it also indicates that treating this type of data as measurements of a continuous-time process might be suboptimal because this does not capture physical interaction between the observed events.

6.7 Proofs

The following lemma shows in which way the Gaussian dependence structure of the underlying random field enters into the proof of Theorem 6.2.1.

Lemma 6.7.1. *A random field Y on $\mathbb{R}^d$ as in assumption (GRFM-trans) has the following property: For all $t, s \in \mathbb{R}^d$, $t \neq s$, the conditional distribution*

$$F_u(x, y) = \mathbb{P}(\log Y(t) \leq x, \log Y(s) \leq y \,|\, \log Y(t) > u, \log Y(s) > u)$$

becomes a product distribution in the limit $u \to \infty$. In copula language, the upper tail dependence copula of any two-dimensional marginal distribution converges to the independent copula.

Proof. Follows directly from Juri & Wüthrich (2003, Thm. 5.3) and the fact that a copula is invariant under monotone transformation of the margins. □

Proof of Theorem 6.2.1 and extension. With regard to the proof of Theorem 6.2.3, we consider the more general case of $\frac{\hat{\alpha}_{fu, f_{\mathrm{cond}}, u}(I, \Phi, T)}{\hat{\alpha}_{1, f_{\mathrm{cond}}, u}(I, \Phi, T)^\ell}$ with $\ell \geq 0$.

With $e(u) = \mathbb{E}\left[\log Y(0) - u \mid \log Y(0) > u\right]$ we have

$$
\mathbb{E}\left[\frac{\hat{\alpha}_{f_u,f_{\text{cond},u}}(I,\Phi,T)}{\hat{\alpha}_{1,f_{\text{cond},u}}(I,\Phi,T)^\ell} \,\Big|\, \mathcal{A}_u^*\right]
$$

$$
= \hat{\alpha}_{1,f_{\text{cond},u}}(I,\Phi,T)^{-\ell} \cdot \mathbb{E}\left[\sum_{(t_1,y_1),(t_2,y_2)\in\Phi,\; t_1\in[0,T]}(\log y_1 - u)\cdot\mathbf{1}_{\log y_1 > u}\cdot\mathbf{1}_{t_2-t_1\in I}\,\Big|\,\mathcal{A}_u^*\right]
$$

$$
= \hat{\alpha}_{1,f_{\text{cond},u}}(I,\Phi,T)^{-\ell}\cdot\sum_{t_1\in\Phi_{\text{g}}\cap[0,T]}\mathbf{1}_{\log Y(t_1)>u}\cdot\#\{t_2\in\Phi_{\text{g}}:t_2-t_1\in I\}
$$
$$
\cdot\,\mathbb{E}\left[(\log Y(t_1)-u)|\mathcal{A}_u^*\right]
$$

$$
= e(u)\cdot\hat{\alpha}_{1,f_{\text{cond},u}}(I,\Phi,T)^{1-\ell}. \tag{6.16}
$$

Furthermore,

$$
\text{Var}\left[\frac{\hat{\alpha}_{f_u,f_{\text{cond},u}}(I,\Phi,T)}{\hat{\alpha}_{1,f_{\text{cond},u}}(I,\Phi,T)^\ell}\,\Big|\,\mathcal{A}_u^*\right]
$$

$$
= \mathbb{E}\left[(\hat{\alpha}_{f_u,f_{\text{cond},u}}/\hat{\alpha}_{1,f_{\text{cond},u}}^\ell)^2\,|\,\mathcal{A}_u^*\right] - \left(\mathbb{E}[\hat{\alpha}_{f_u,f_{\text{cond},u}}/\hat{\alpha}_{1,f_{\text{cond},u}}^\ell\,|\,\mathcal{A}_u^*]\right)^2
$$

$$
= \hat{\alpha}_{1,f_{\text{cond},u}}^{-2\ell}\cdot\mathbb{E}\left[\hat{\alpha}_{f_u,f_{\text{cond},u}}^2\,|\,\mathcal{A}_u^*\right] - e(u)^2\hat{\alpha}_{1,f_{\text{cond},u}}^{2-2\ell} \tag{6.17}
$$

with

$$
\mathbb{E}\left[\hat{\alpha}_{f_u,f_{\text{cond},u}}(I,\Phi,T)^2\,|\,\mathcal{A}_u^*\right]
$$

$$
= \mathbb{E}\Big[\sum_{t_1\in\Phi_{\text{g}}\cap[0,T]}\sum_{s_1\in\Phi_{\text{g}}\cap[0,T]}(\log Y(t_1)-u)(\log Y(s_1)-u)\cdot\mathbf{1}_{\log Y(t_1)>u}\mathbf{1}_{\log Y(s_1)>u}
$$
$$
\cdot\,\#\{t_2\in\Phi_{\text{g}}:t_2-t_1\in I\}\cdot\#\{s_2\in\Phi_{\text{g}}:s_2-s_1\in I\}\,\Big|\,\mathcal{A}_u^*\Big]
$$

$$
= \sum_{t_1\in\Phi_{\text{g}}\cap[0,T]}\sum_{s_1\in\Phi_{\text{g}}\cap[0,T]}n(t_1,\Phi_{\text{g}},I,u)n(s_1,\Phi_{\text{g}},I,u)
$$
$$
\cdot\,\mathbb{E}\left[(\log Y(t_1)-u)(\log Y(s_1)-u)\,|\,\mathcal{A}_u^*\right]
$$

$$
= \sum_{t_1\in\Phi_{\text{g}}\cap[0,T]}\sum_{s_1\in\Phi_{\text{g}}\cap[0,T]}n(t_1,\Phi_{\text{g}},I,u)n(s_1,\Phi_{\text{g}},I,u)
$$
$$
\cdot\,\Big[\mathbb{E}\left[\log Y(0)-u|\mathcal{A}_u^*\right]^2 + \text{Cov}\left[\log Y(t_1),\log Y(s_1)\,|\,\mathcal{A}_u^*\right]\Big]
$$

$$
= \sum_{t_1\in\Phi_{\text{g}}\cap[0,T]}\sum_{s_1\in\Phi_{\text{g}}\cap[0,T]}n(t_1,\Phi_{\text{g}},I,u)n(s_1,\Phi_{\text{g}},I,u)
$$
$$
\cdot\,\text{Cov}\left[\log Y(t_1),\log Y(s_1)\,|\,\log Y(t_1)>u,\log Y(s_1)>u\right]
$$
$$
+ e(u)^2\cdot\hat{\alpha}_{1,f_{\text{cond},u}}(I,\Phi,T)^2. \tag{6.18}
$$

Due to the finite range h_0 of the covariance function of Y and the minimum distance d_0 between point locations

$$
\mathbb{E}\left[\hat{\alpha}_{f_u,f_{\text{cond},u}}(I,\Phi,T)^2\,|\,\mathcal{A}_u^*\right]
$$
$$
= v_u\sum_{t_1\in\Phi_{\text{g}}\cap[0,T]}n(t_1,\Phi_{\text{g}},I,u)^2 + e(u)^2\hat{\alpha}_{1,f_{\text{cond},u}}(I,\Phi,T)^2 + \varepsilon_u v_u\hat{\alpha}_{1,f_{\text{cond},u}}(I,\Phi,T)\frac{h_0}{d_0}
$$

$$
\tag{6.19}
$$

for some $\mathcal{A}_u^*$-measurable random variable ε_u with $|\varepsilon_u| \leq 1$. It follows directly from Lemma 6.7.1 that the conditional covariance terms for $t_1 \neq s_1$ vanish for $u \to \infty$. Hence, $\varepsilon_u \to 0$ a.s. and since ε_u is dominated by 1, also $\mathbb{E}|\varepsilon_u| \to 0$, as $u \to \infty$. Note that, if additionally condition (Independent-noise-marking) is satisfied, all covariance terms in (6.18) vanish due to the iid assumption, up to those for which $t_1 = s_1$. Hence, ε_u equals 0 in this case.

Plugging (6.19) into (6.17) yields

$$\text{Var}\left[\hat{\alpha}_{f_u, f_{\text{cond}}, u}(I, \Phi, T)/\hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^\ell \mid \mathcal{A}_u^*\right]$$
$$= v_u \cdot \frac{\sum_{t_1 \in \Phi_{\text{g}} \cap [0,T]} n(t_1, \Phi_{\text{g}}, I, u)^2}{\hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^{2\ell}} + \varepsilon_u v_u \hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^{1-2\ell}\frac{h_0}{d_0} \tag{6.20}$$

and the proof is complete. $\qquad\qquad\square$

Proof of Theorem 6.2.3. Since the marginals of the underlying random field Y are assumed to be in the Fréchet MDA, $\log Y(0)$ is in the Gumbel MDA. It is well known that being in the MDA of the generalized extreme value distribution with shape parameter ξ is equivalent to the distribution of excesses over high thresholds converging to the generalized Pareto distribution (GPD) with the same shape parameter ξ (e.g., Embrechts *et al.*, 1997, Thm. 3.4.13). In the Gumbel case, where ξ equals 0, the corresponding GPD reduces to the exponential distribution and therefore all moments of the excesses $Z_i = [\log Y(0) - u \mid \log Y(0) > u]$ exist and converge to some constant in $(0, \infty)$ as $u \to \infty$.
Then, application of Theorem 5.4.3 yields the weak convergence of

$$\hat{\alpha}^*_{f_{u_T}, f_{\text{cond}, u_T}}(I, \Phi, T)/\sqrt{\hat{\alpha}_{1, f_{\text{cond}}, u_T}(I, \Phi, T)}$$

to a centered Gaussian variable.

For the asymptotic variance of $\hat{\alpha}^*_{f_u, f_{\text{cond}}, u}(I, \Phi, T)/\hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^\ell$, $\ell \in \{0.5, 1\}$, note that under the random field model assumption, $\xi(I, u) = e(u)$. Hence, applying the decomposition of variance w.r.t. $\mathcal{A}_u^*$ and replacing $f_u(y) = \log y - u$ by $\tilde{f}_u(y) = \log y - u - e(u)$ in (6.16)–(6.20), the terms $e(u)$ in these equations vanish and it follows directly that

$$\text{Var}\left[\frac{\hat{\alpha}^*_{f_u, f_{\text{cond}}, u}(I, \Phi, T)}{\hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^\ell}\right] = \text{Var}\left[\frac{\hat{\alpha}_{f_u, f_{\text{cond}}, u}(I, \Phi, T)}{\hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^\ell}\right] - \text{Var}\left[e(u) \cdot \hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^{1-\ell}\right],$$
$$\tag{6.21}$$

for arbitrary u and T.
With (6.16) we get

$$\text{Var}\,\mathbb{E}\left[\frac{\hat{\alpha}_{f_u, f_{\text{cond}}, u}(I, \Phi, T)}{\hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^\ell} \,\Big|\, \mathcal{A}_u^*\right] = \text{Var}\left[e(u)\hat{\alpha}_{1, f_{\text{cond}}, u}(I, \Phi, T)^{1-\ell}\right],$$

which equals 0 if and only if $\ell = 1$. Together with (6.20), we obtain

$$\mathrm{Var}\left[\frac{\hat{\alpha}_{f_u,f_{\mathrm{cond}},u}(I,\Phi,T)}{\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)^\ell}\right]$$

$$= \mathbb{E}\,\mathrm{Var}\left[\frac{\hat{\alpha}_{f_u,f_{\mathrm{cond}},u}(I,\Phi,T)}{\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)^\ell}\,\Big|\,\mathcal{A}_u^*\right] + \mathrm{Var}\,\mathbb{E}\left[\frac{\hat{\alpha}_{f_u,f_{\mathrm{cond}},u}(I,\Phi,T)}{\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)^\ell}\,\Big|\,\mathcal{A}_u^*\right]$$

$$= v_u \mathbb{E}\left[\frac{\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)^2}{\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)^{2\ell}}\right]$$

$$+ \mathbb{E}\left[\varepsilon_u v_u \hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)^{1-2\ell}\frac{h_0}{d_0}\right] + e(u)^2\,\mathrm{Var}\left[\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)^{1-\ell}\right] \qquad (6.22)$$

with some random function ε_u satisfying $|\varepsilon_u| \leq 1$ and $\varepsilon_u \to 0$ a.s. and in L_1 for $u \to \infty$ (see the proof of Theorem 6.2.1).

Then, with (6.21), (6.22) and $\ell = 0.5$, the asymptotic variance of $\frac{\hat{\alpha}^*_{f_{u_T},f_{\mathrm{cond}},u_T}(I,\Phi,T)}{\sqrt{\hat{\alpha}_{1,f_{\mathrm{cond}},u_T}(I,\Phi,T)}}$ (for $T \to \infty$) is

$$s_{u_\infty} = \lim_{T\to\infty} v_{u_T}\cdot\mathbb{E}_{\Phi_{\mathrm{g}}}\left[\frac{\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u_T)^2}{\hat{\alpha}_{1,f_{\mathrm{cond}},u_T}(I,\Phi,T)} + \varepsilon_{u_T}\frac{h_0}{d_0}\right],$$

where the expectation can also be applied to numerator and denominator separately due to the a.s. convergence of $\hat{\alpha}_{1,f_{\mathrm{cond}},u_T}(I,\Phi,T)/\mathbb{E}\hat{\alpha}_{1,f_{\mathrm{cond}},u_T}(I,\Phi,T)$ (cf. Lemma 5.8.1). For fixed u, the asymptotic variance of $\frac{\hat{\alpha}^*_{f_u,f_{\mathrm{cond}},u}(I,\Phi,T)}{\sqrt{\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)}}$ is

$$s_u = \lim_{T\to\infty} v_u\left[\mathbb{E}_{\Phi_{\mathrm{g}}}\left[\frac{\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)^2}{\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)} + \varepsilon_u\frac{h_0}{d_0}\right]\right] \qquad (6.23)$$

$$= v_u\left[\frac{\mathbb{E}_{\Phi_{\mathrm{g}}}\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,1]} n(t_1,\Phi_{\mathrm{g}},I,u)^2}{\mathbb{E}_{\Phi_{\mathrm{g}}}\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,1)} + \frac{h_0}{d_0}\mathbb{E}_{\Phi_{\mathrm{g}}}\varepsilon_u\right], \qquad (6.24)$$

where the second equation follows by again applying Proposition 5.7.4 to both numerator and denominator, and by noting that the ratio on the RHS of (6.23) is bounded by $(|I|/d_0)^2$.

The last assertion concerning the variance of $\frac{\hat{\alpha}_{f_u,f_{\mathrm{cond}},u}(I,\Phi,T)}{\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)}$ and $\frac{\hat{\alpha}^*_{f_u,f_{\mathrm{cond}},u}(I,\Phi,T)}{\hat{\alpha}_{1,f_{\mathrm{cond}},u}(I,\Phi,T)}$ also follows from (6.21) and (6.22) with $\ell = 1$. $\qquad \square$

7 Representations of max-stable processes based on single extreme events

While conventional second-order dependency measures, such as the covariance, are well-established for continuous-space processes as well as for the MPP framework, second-order characteristics for the extremes are still in development. In the field of extreme value statistics, (spatial) dependency of the extremes is modeled via max-stable processes; inference for general max-stable processes is a major subject of current research. One of the two general classes of estimation principles is the class of *peaks-over-threshold* methods, which are particularly promising from an MPP point of view in that they can be expected to be also applicable to data of MPP type. This chapter provides some theoretical foundations for new statistical methods for the max-domain of attraction of max-stable processes, and closes with a short outlook for how the results can be applied.

The chapter is based on the manuscript Engelke *et al.* (2012*b*).

7.1 Introduction

The joint extremal behavior at multiple locations of some random process $\{\eta(t) : t \in T\}$, T an arbitrary index set, can be captured via its limiting *max-stable process*, assuming the latter exists and is non-trivial everywhere. Then, for independent copies η_i of η, $i \in \mathbb{N}$, the functions $b_n : T \to \mathbb{R}$, $c_n : T \to (0,\infty)$ can be chosen such that the convergence

$$\zeta(t) = \lim_{n \to \infty} c_n(t)\Big(\max_{i=1}^{n} \eta_i(t) - b_n(t)\Big), \quad t \in T, \tag{7.1}$$

holds in the sense of finite-dimensional distributions. The process ζ is said to be *max-stable* and η is in its max-domain of attraction (MDA). The theory of max-stable processes is mainly concerned with the dependence structure while the marginals are usually assumed to be known. Even for finite-dimensional max-stable distributions, the space of possible dependence structures is uncountably infinite-dimensional and parametric models are required to find a balance between flexibility and analytical tractability (de Haan & Ferreira, 2006; Resnick, 2008).

A general construction principle for max-stable processes was provided by de Haan (1984) and Smith (1990): Let $\sum_{i \in \mathbb{N}} \delta_{(U_i,S_i)}$ be a Poisson point process (PPP) on $(0,\infty) \times \mathcal{S}$ with intensity measure $u^{-2}du \cdot \nu(ds)$, where $(\mathcal{S},\mathfrak{S})$ is an arbitrary measurable space and ν a positive measure on $\mathcal{S}$. Further, let $f : \mathcal{S} \times T \to [0,\infty)$ be a non-negative function with $\int_{\mathcal{S}} f(s,t)\nu(ds) = 1$ for all $t \in T$. Then the process

$$\zeta(t) = \max_{i \in \mathbb{N}} U_i f(S_i, t), \quad t \in T, \tag{7.2}$$

is max-stable and has standard Fréchet margins with distribution function $\exp(-1/x)$ for $x \geq 0$. In this chapter, two specific choices for f and $(\mathcal{S}, \mathfrak{S}, \nu)$ are considered. First, let $\{W(t) : t \in T\}$ be a non-negative stochastic process with $\mathbb{E}W(t) = 1$, $t \in T$, and $W(t_0) = 1$ a.s. for some point $t_0 \in T$. The latter condition means that $W(t)$ simply describes the multiplicative increment of W w.r.t. the location t_0. For $(\mathcal{S}, \mathfrak{S}, \nu)$ being the canonical probability space for the sample paths of W and with $f(w,t) = w(t)$, $w \in \mathcal{S}$, $t \in T$, we refer to

$$\zeta(t) = \max_{i \in \mathbb{N}} U_i W_i(t), \quad t \in T, \tag{7.3}$$

as the *incremental representation* of ζ, where $\{W_i\}_{i \in \mathbb{N}}$ are independent copies of W. Since T is an arbitrary index set, the above definition covers multivariate extreme value distributions, i.e. $T = \{t_1, \ldots, t_k\}$, as well as max-stable random fields, i.e. $T = \mathbb{R}^d$.

For the second specification, let $\{F(t) : t \in \mathbb{R}^d\}$ be a stochastic process with sample paths in the space $C(\mathbb{R}^d)$ of non-negative continuous functions and

$$\mathbb{E} \int_{\mathbb{R}^d} F(t) dt = 1. \tag{7.4}$$

With $S_i = (T_i, F_i)$, $i \in \mathbb{N}$, in $\mathcal{S} = \mathbb{R}^d \times C(\mathbb{R}^d)$, intensity measure $\nu(dt \times dg) = dt \mathbb{P}_F(dg)$ and $f((t,g), s) = g(s - t)$, $(t,g) \in \mathcal{S}$, we obtain the class of *mixed moving maxima (M3) processes*

$$\zeta(t) = \max_{i \in \mathbb{N}} U_i F_i(t - T_i), \quad t \in \mathbb{R}^d. \tag{7.5}$$

These processes are max-stable and stationary on $\mathbb{R}^d$ (see for instance Wang & Stoev (2010)). The function F is called *shape function of ζ* and can also be deterministic (e.g., in case of the Smith process). In Smith's "rainfall-storm" interpretation (Smith, 1990), U_i and T_i are the strength and center point of the ith storm, respectively, and $U_i F_i(t - T_i)$ represents the corresponding amount of rainfall at location t. In this case, $\zeta(t)$ is the process of extremal precipitation.

Based on the two representations (7.3) and (7.5), we will provide convergence results for processes in the MDA of ζ, which lead to new and efficient methods of inference.

When i.i.d. realizations $\eta_1, \ldots, \eta_n$ of η in the MDA of a max-stable process ζ are observed, a classical approach for parametric inference on ζ is based on generating (approximate) realizations of ζ out of the data $\eta_1, \ldots, \eta_n$ via componentwise block maxima and applying maximum likelihood (ML) estimation afterwards. A clear drawback of this method is that it ignores all information on large values that is contained in the order statistics below the within-block maximum. Further, ML estimation needs to evaluate the multivariate densities while for many max-stable models only the bivariate densities are known in closed form. Thus, composite likelihood approaches have been proposed (Padoan *et al.*, 2010; Davison & Gholamrezaee, 2012).

In univariate extreme-value theory, the second standard procedure estimates parameters by fitting a certain PPP to the *peaks-over-thresholds* (POT), i.e., to the empirical process of exceedances over a certain critical value (Leadbetter, 1991; Embrechts *et al.*, 1997). Also in the multivariate framework we can expect to profit from using all extremal data via

generalized POT methods instead of aggregated data. In contrast to the ML approach, here, we assume that ζ admits one of the two representations (7.3) and (7.5) and we aim at extracting realizations of the processes W and F, respectively, from *single extreme events*. Here, the specification of a single extreme event will depend on the respective representation. In Engelke *et al.* (2012a), this concept is applied to derive estimators for the class of Brown-Resnick processes (Brown & Resnick, 1977; Kabluchko *et al.*, 2009), which have the form (7.3) by construction. With $a(n)$ being a sequence with $\lim_{n\to\infty} a(n) = \infty$, the convergence in distribution

$$\left(\frac{\eta(t_1)}{\eta(t_0)}, \ldots, \frac{\eta(t_k)}{\eta(t_0)} \;\middle|\; \eta(t_0) > a(n) \right) \Rightarrow (W(t_1), \ldots, W(t_k)), \tag{7.6}$$

for $t_0, t_1, \ldots, t_k \in T$, $k \in \mathbb{N}$, is established for η being in the MDA of a Brown-Resnick process and with W being the corresponding log-Gaussian random field.

Basrak & Segers (2009) and Meinguet & Segers (2010) consider multivariate time series $(X_t)_{t\in\mathbb{Z}}$ and time series in general Banach spaces, respectively, rescaled and conditioned on $\|X_0\|$ being large. They provide equivalent conditions for the existence of the corresponding tail processes and its spectral decomposition, whereas here, we explicitly calculate the limiting processes under more specific assumptions.

In this direction, we generalize the convergence result (7.6) in two different aspects. Arbitrary non-negative processes $\{W(t) : t \in T\}$ with $\mathbb{E}W(t) = 1$, $t \in T$, are considered, and convergence of the conditional increments of η in the sense of finite-dimensional distributions as well as weak convergence in continuous function spaces is shown (Theorems 7.2.1 and 7.2.10). Since also M3 processes might admit an incremental representations (7.3) we provide formulae for switching between the two representations in Section 7.3. Section 7.4 gives an exemplary outlook on how our results can be applied for statistical inference.

7.2 Extracting the incremental process

Throughout this section, we consider a slightly more general version of representation 7.3, which does not require that $W(t_0) = 1$ a.s. for some $t_0 \in T$. In fact, we suppose that $\{\zeta(t) : t \in T\}$, T an arbitrary index set, is normalized to standard Fréchet margins and admits a representation

$$\zeta(t) = \max_{i \in \mathbb{N}} U_i V_i(t), \quad t \in T. \tag{7.7}$$

Here, $\sum_{i\in\mathbb{N}} \delta_{U_i}$ is a PPP on $(0,\infty)$ with intensity $u^{-2}du$, which we call *Fréchet point process* in the following. The $\{V_i\}_{i\in\mathbb{N}}$ are independent copies of a non-negative stochastic process $\{V(t) : t \in T\}$ with $\mathbb{E}V(t) = 1$, $t \in T$. For any fixed $t_0 \in T$, we have

$$\zeta(t) \overset{d}{=} \max_{i \in \mathbb{N}} U_i \left(\mathbf{1}_{P_i=0} V_i^{(1)}(t) + \mathbf{1}_{P_i=1} V_i^{(2)}(t) \right), \quad t \in T, \tag{7.8}$$

where $\{P_i\}_{i \in \mathbb{N}}$ are i.i.d. Bernoulli variables with parameter $p = \mathbb{P}(V(t_0) = 0)$ and the $V_i^{(1)}$ and $V_i^{(2)}$ are independent copies of the process $\{V(t) : t \in T\}$, conditioned on the events $\{V(t_0) > 0\}$ and $\{V(t_0) = 0\}$, respectively.

Note that for $k \in \mathbb{N}$, $t_0, \ldots, t_k \in T$, the vector $\zeta = (\zeta(t_0), \ldots, \zeta(t_k))$ follows a $(k+1)$-variate extreme-value distribution and its distribution function G can therefore be written as

$$G(\mathbf{x}) = \exp(-\mu([\mathbf{0}, \mathbf{x}]^C)), \quad \mathbf{x} \in \mathbb{R}^{k+1}, \tag{7.9}$$

where μ is a measure on $E = [0, \infty)^{k+1} \setminus \{\mathbf{0}\}$, the so-called *exponent measure* of G (Resnick, 2008, Prop. 5.8), and $[\mathbf{0}, \mathbf{x}]^C = E \setminus [\mathbf{0}, \mathbf{x}]$.

The following convergence result provides the theoretical foundation for statistical inference based on the incremental process V.

Theorem 7.2.1. *Let $\{\eta(t) : t \in T\}$ be non-negative and in the MDA of some max-stable process ζ that admits a representation (7.7) and suppose that η is normalized such that (7.1) holds with $c_n(t) = 1/n$ and $b_n(t) = 0$ for $n \in \mathbb{N}$ and $t \in T$. Let $a(n) \to \infty$ as $n \to \infty$. For $k \in \mathbb{N}$ and $t_0, \ldots, t_k \in T$ we have the convergence in distribution on $\mathbb{R}^{k+1}$*

$$\left(\frac{\eta(t_0)}{a(n)}, \frac{\eta(t_1)}{\eta(t_0)}, \ldots, \frac{\eta(t_k)}{\eta(t_0)} \,\middle|\, \eta(t_0) > a(n) \right) \Rightarrow \left(Z, \Delta\tilde{\mathbf{V}}^{(1)} \right), \quad n \to \infty,$$

where the distribution of $\Delta\tilde{\mathbf{V}}^{(1)}$ is given by

$$\mathbb{P}(\Delta\tilde{\mathbf{V}}^{(1)} \in d\mathbf{z}) = (1-p)\mathbb{P}(\Delta\mathbf{V}^{(1)} \in d\mathbf{z})\mathbb{E}(V^{(1)}(t_0)|\Delta\mathbf{V}^{(1)} = \mathbf{z}), \tag{7.10}$$

$\mathbf{z} \geq \mathbf{0}$. Here, $\Delta\mathbf{V}^{(1)}$ denotes the vector of increments $\left(\frac{V^{(1)}(t_1)}{V^{(1)}(t_0)}, \ldots, \frac{V^{(1)}(t_k)}{V^{(1)}(t_0)} \right)$ with respect to t_0, and Z is an independent Pareto variable.

Remark 7.2.2. *Note that, if ζ has standard Fréchet margins, any process η satisfying the convergence in (7.1) can be normalized such that the norming functions in (7.1) are $c_n(t) = 1/n$ and $b_n(t) = 0$, $n \in \mathbb{N}$, $t \in T$ (Resnick, 2008, Prop. 5.10).*

Proof of Theorem 7.2.1. For $\mathbf{X} = (\eta(t_0), \ldots, \eta(t_k))$, which is in the MDA of the random vector $\zeta = (\zeta(t_0), \ldots, \zeta(t_k))$, it follows from Resnick (2008, Prop. 5.17) that

$$\lim_{m \to \infty} m\mathbb{P}(\mathbf{X}/m \in B) = \mu(B), \tag{7.11}$$

for all elements B of the Borel σ-algebra $\mathcal{B}(E)$ of E bounded away from $\{\mathbf{0}\}$ with $\mu(\partial B) = 0$, where μ is defined by (7.9). For $s_0 > 0$ and $\mathbf{s} = (s_1, \ldots, s_k) \in [0, \infty)^k$, we consider the sets $A_{s_0} = (s_0, \infty) \times [0, \infty)^k$, $A = A_1$ and $B_{\mathbf{s}} = \{\mathbf{x} \in [0, \infty)^{k+1} : (x^{(1)}, \ldots, x^{(k)}) \leq x^{(0)}\mathbf{s}\}$ for $\mathbf{s}$ satisfying $\mathbb{P}(\Delta\tilde{\mathbf{V}}^{(1)} \in \partial[\mathbf{0}, \mathbf{s}]) = 0$. Since $B_{\mathbf{s}}$ satisfies $B_{\mathbf{s}} = cB_{\mathbf{s}}$ for any $c > 0$, we obtain

$$\mathbb{P}\left(\eta(t_0) > s_0 a(n), \, (\eta(t_1)/\eta(t_0), \ldots, \eta(t_k)/\eta(t_0)) \leq \mathbf{s} \,\middle|\, \eta(t_0) > a(n) \right)$$

$$= \frac{a(n)\mathbb{P}(\mathbf{X}/a(n) \in B_{\mathbf{s}} \cap A \cap A_{s_0})}{a(n)\mathbb{P}(\mathbf{X}/a(n) \in A)} \longrightarrow \frac{\mu(B_{\mathbf{s}} \cap A \cap A_{s_0})}{\mu(A)}, \quad (n \to \infty), \tag{7.12}$$

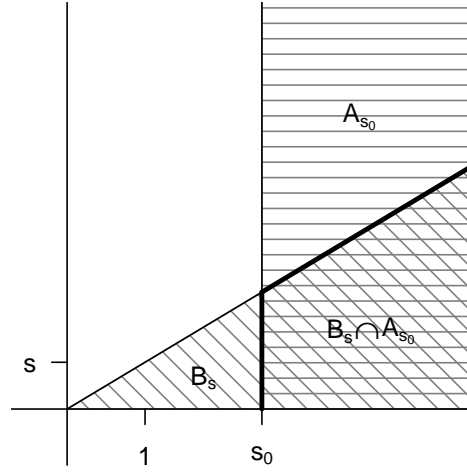**Figure 7.1:** The sets A, $B_{\mathbf{s}}$ and $B_{\mathbf{s}} \cap A_{s_0}$ for $k = 1$.

where the convergence follows from (7.11), if $\mu\{\partial(B_{\mathbf{s}} \cap A \cap A_{s_0})\} = 0$. Let

$$\zeta^{(1)}(t) = \max_{i \in \mathbb{N}} U_i^{(1)} V_i^{(1)}(t), \quad t \in T, \tag{7.13}$$

where $\sum_{i \in \mathbb{N}} \delta_{U_i^{(1)}}$ is a Poisson point process with intensity $(1-p)u^{-2}du$, and let $\mu^{(1)}$ be the exponent measure of the associated max-stable random vector $(\zeta^{(1)}(t_0), \ldots, \zeta^{(1)}(t_k))$. Then the choice $A = (1,\infty) \times [0,\infty)^k$ guarantees that $\mu(\cdot \cap A) = \mu^{(1)}(\cdot \cap A)$. Comparing the construction of $\zeta^{(1)}$ in (7.13) with the definition of the exponent measure, we see that $\mu^{(1)}$ is the intensity measure of the Poisson point process $\sum_{i \in \mathbb{N}} \delta_{(U_i^{(1)} V_i^{(1)}(t_0), \ldots, U_i^{(1)} V_i^{(1)}(t_k))}$ on E. Hence,

$$\mu(A) = \int_0^\infty (1-p)u^{-2} \int_{[u^{-1},\infty)} \mathbb{P}(V^{(1)}(t_0) \in dy) du$$
$$= (1-p) \int_0^\infty y \mathbb{P}(V^{(1)}(t_0) \in dy) = (1-p)\mathbb{E}V^{(1)}(t_0) = 1, \tag{7.14}$$

where the last equality follows from $\mathbb{E}V^{(1)}(t_0) = \mathbb{E}V(t_0)/(1-p)$. Furthermore, for $s_0 \geq 1$ and $\mathbf{s} \in [0,\infty)^k$ with $\mathbb{P}(\Delta\tilde{\mathbf{V}}^{(1)} \in \partial[\mathbf{0},\mathbf{s}]) = 0$,

$$\mu(B_{\mathbf{s}} \cap A \cap A_{s_0})/((1-p)\mu(A))$$
$$= \int_0^\infty u^{-2} \int_{[s_0 u^{-1}, \infty)} \mathbb{P}\Big(V^{(1)}(t_0) \in dy \Big| \Delta\mathbf{V}^{(1)} \leq \mathbf{s}\Big) \mathbb{P}(\Delta\mathbf{V}^{(1)} \leq \mathbf{s}) du$$
$$= \int_{[\mathbf{0},\mathbf{s}]} \int_{[0,\infty)} y s_0^{-1} \cdot \mathbb{P}\Big(V^{(1)}(t_0) \in dy \Big| \Delta\mathbf{V}^{(1)} = \mathbf{z}\Big) \mathbb{P}(\Delta\mathbf{V}^{(1)} \in d\mathbf{z})$$
$$= s_0^{-1} \int_{[\mathbf{0},\mathbf{s}]} \mathbb{E}\Big(V^{(1)}(t_0) \Big| \Delta\mathbf{V}^{(1)} = \mathbf{z}\Big) \mathbb{P}(\Delta\mathbf{V}^{(1)} \in d\mathbf{z}). \tag{7.15}$$

Equation (7.15) shows that the convergence in (7.12) holds for all continuity points $\mathbf{s} \in [0,\infty)^k$ of the distribution function of $\Delta\mathbf{V}^{(1)}$. $\qquad\square$

Remark 7.2.3. *1. If $V^{(1)}(t_0)$ is stochastically independent of the increments $\Delta\mathbf{V}^{(1)}$, we simply have $\mathbb{P}(\Delta\tilde{\mathbf{V}}^{(1)} \in d\mathbf{z}) = \mathbb{P}(\Delta\mathbf{V}^{(1)} \in d\mathbf{z})$. This is particularly the case if ζ admits a representation (7.3), which shows that (7.6) is indeed a special case of Theorem 7.2.1.*

2. If $p = \mathbb{P}(V(t_0) = 0) = 0$, the exponent measure μ of any finite-dimensional vector $\zeta = (\zeta(t_0), \ldots, \zeta(t_k))$, $t_0, \ldots, t_k \in T$, $k \in \mathbb{N}$, satisfies the condition $\mu\left(\{0\} \times [0,\infty)^k\right) = 0$, and following Proposition 7.2.7, the incremental representation of ζ according to (7.3) is given by $\zeta = \max_{i \in \mathbb{N}} U_i \cdot (1, \Delta\tilde{\mathbf{V}}_i)^\top$, where $\Delta\tilde{\mathbf{V}}_i$, $i \in \mathbb{N}$, are independent copies of $\Delta\tilde{\mathbf{V}} = \Delta\tilde{\mathbf{V}}^{(1)}$.

Remark 7.2.4. *In the above theorem, the thresholds $a(n)$ tend to ∞ to make $\{\eta(t_0) > a(n)\}$ a rare event. For statistical applications, $a(n)$ should also be chosen such that the number of exceedances $N(n) = \sum_{i=1}^n \mathbf{1}\{\eta_i(t_0) > a(n)\}$ converges to ∞ almost surely, where η_i, $i \in \mathbb{N}$, are independent copies of η. By the Poisson limit theorem, this is equivalent to the additional assumption that $\lim_{n \to \infty} a(n)/n = 0$, since then $n\mathbb{P}(\eta(t_0) > a(n)) = n/a(n) \to \infty$.*

Remark 7.2.5. *Engelke et al. (2012a) consider Hüsler-Reiss distributions (Hüsler & Reiss, 1989; Kabluchko, 2011) and obtain their limiting results by conditioning on certain extremal events $A \subset E$. They show that various choices of A are sensible in the Hüsler-Reiss case, leading to different limiting distributions of the increments of η. In case that ζ is a Brown-Resnick process and $A = (1,\infty) \times [0,\infty)^k$, the assertions of Theorem 7.2.1 and Engelke et al. (2012a, Thm. 3.3) coincide.*

Example 7.2.6 (Extremal Gaussian process (Schlather, 2002)). *A commonly used class of stationary yet non-ergodic max-stable processes on $\mathbb{R}^d$ is defined by*

$$\zeta(t) = \max_{i \in \mathbb{N}} U_i Y_i(t), \quad t \in \mathbb{R}^d, \tag{7.16}$$

where $\sum_{i \in \mathbb{N}} \delta_{U_i}$ is a Fréchet point process, $Y_i(t) = \max(0, \tilde{Y}_i(t))$, $i \in \mathbb{N}$, and the $\tilde{Y}_i$ are i.i.d. stationary, centered Gaussian processes with $\mathbb{E}(\max(0, \tilde{Y}_i(t))) = 1$ for all $t \in \mathbb{R}^d$ (Schlather, 2002; Blanchet & Davison, 2011). Note that in general, a $t_0 \in \mathbb{R}^d$ s.t. $Y_i(t_0) = 1$ a.s. does not exist, i.e., the process admits representation (7.7) but not representation (7.3). In particular, for the extremal Gaussian process we have $p = \mathbb{P}(V(t_0) = 0) = 1/2$ and the distribution of the increments in (7.10) becomes

$$\mathbb{P}(\Delta\tilde{\mathbf{V}}^{(1)} \in d\mathbf{z}) = \frac{1}{2}\mathbb{E}\Big[Y(t_0)\,\Big|\,(Y(t_1)/Y(t_0), \ldots, Y(t_k)/Y(t_0)) = \mathbf{z}, Y(t_0) > 0\Big]$$
$$\cdot \mathbb{P}\Big((Y(t_1)/Y(t_0), \ldots, Y(t_k)/Y(t_0)) \in d\mathbf{z}\,\Big|\,Y(t_0) > 0\Big).$$

While the Hüsler-Reiss distribution is already given by the incremental representation (7.3), cf. Kabluchko (2011), other distributions can be suitably rewritten, provided that the respective exponent measure μ is known.

Proposition 7.2.7. *Let $\zeta = (\zeta(t_0),\dots,\zeta(t_k))$ be a max-stable process on $T = \{t_0,\dots,t_k\}$ with standard Fréchet margins and suppose that its exponent measure μ is concentrated on $(0,\infty) \times [0,\infty)^k$. Define a random vector $\mathbf{W}$ via*

$$\mathbb{P}(\mathbf{W} \le \mathbf{s}) = \mu(B_{\mathbf{s}} \cap A), \quad \mathbf{s} \in [0,\infty)^k, \tag{7.17}$$

where $A = (1,\infty) \times [0,\infty)^k$ and $B_{\mathbf{s}} = \{\mathbf{x} \in [0,\infty)^{k+1} : (x^{(1)},\dots,x^{(k)}) \le x^{(0)}\mathbf{s}\}$. Then $\mathbf{W}$ is the incremental process of ζ in (7.3).

Proof. First, we note that (7.17) indeed defines a valid cumulative distribution function. To this end, consider the measurable transformation

$$T : (0,\infty) \times [0,\infty)^k \to (0,\infty) \times [0,\infty)^k, \ (x_0,\dots,x_k) \mapsto \left(x_0, \frac{x_1}{x_0},\dots,\frac{x_k}{x_0}\right).$$

Then, $T(B_{\mathbf{s}} \cap A) = (1,\infty) \times [\mathbf{0},\mathbf{s}]$ and the measure $\mu^T(\cdot) = \mu(T^{-1}((1,\infty) \times \cdot))$ is a probability measure on $[0,\infty)^k$. Since

$$\mu(B_{\mathbf{s}} \cap A) = \mu(T^{-1}((1,\infty) \times [\mathbf{0},\mathbf{s}])) = \mu^T([\mathbf{0},\mathbf{s}]),$$

the random vector $\mathbf{W}$ is well-defined and has law μ^T.

By definition of the exponent measure, we have $\zeta \stackrel{d}{=} \max_{i \in \mathbb{N}} \mathbf{X}_i$, where $\Pi = \sum_{i \in \mathbb{N}} \delta_{\mathbf{X}_i}$ is a PPP on E with intensity measure μ. Then, the transformed point process $T\Pi = \sum_{i \in \mathbb{N}} \delta_{(X_i^{(0)}, X_i^{(1)}/X_i^{(0)}, \dots, X_i^{(k)}/X_i^{(0)})}$ has intensity measure

$$\tilde{\mu}((c,\infty) \times [\mathbf{0},\mathbf{s}]) = \mu\left(T^{-1}\left((c,\infty) \times [\mathbf{0},\mathbf{s}]\right)\right)$$
$$= \mu(B_{\mathbf{s}} \cap ((c,\infty) \times [0,\infty)^k)) = c^{-1}\mu(B_{\mathbf{s}} \cap A)$$

for any $c > 0$, $\mathbf{s} \in [0,\infty)^k$, where we use the homogeneity property $c^{-1}\mu(d\mathbf{x}) = \mu(d(c\mathbf{x}))$. Thus, $T\Pi$ has the same intensity as $\sum_{i \in \mathbb{N}} \delta_{(U_i,\mathbf{W}_i)}$, where $\sum_{i \in \mathbb{N}} \delta_{U_i}$ is a Fréchet point process and $\mathbf{W}_i$, $i \in \mathbb{N}$, are i.i.d. vectors with law μ^T. Hence,

$$\zeta \stackrel{d}{=} \max_{i \in \mathbb{N}} T^{-1}\left((X_i^{(0)}, X_i^{(1)}/X_i^{(0)},\dots,X_i^{(k)}/X_i^{(0)})\right)$$
$$\stackrel{d}{=} \max_{i \in \mathbb{N}} T^{-1}\left((U_i,\mathbf{W}_i)\right) = \max_{i \in \mathbb{N}} U_i\mathbf{W}_i.$$

$\square$

Example 7.2.8 (Bivariate Pareto-based distribution, cf. Resnick (2008, Ex. 5.16) and the references therein)**.** *For $T = \{t_0,t_1\}$, the extreme value distribution*

$$\mathbb{P}(\zeta(t_0) \le x, \ \zeta(t_1) \le y) = \exp(-x^{-1} - y^{-1} + (x+y)^{-1}), \quad x,y > 0,$$

is the max-limit of a bivariate Pareto distribution. Using the density of the exponent measure,

$$\mu(dx,dy) = 2(x+y)^{-3}dxdy,$$

we get for $s \geq 0$

$$\mu(B_s \cap A) = \int_1^\infty \int_0^{sx} \mu(dx, dy)$$
$$= \int_1^\infty x^{-1} \left(1 - (1 + s)^{-2}\right) dx = 1 - (1 + s)^{-2}.$$

Thus, for W in the representation (7.3), *we have*

$$\mathbb{P}(W(t_1) \leq s) = 1 - (1 + s)^{-2}.$$

Example 7.2.9 (Symmetric logistic distribution, cf. Gumbel (1960))**.** *Choosing the index set $T = \{t_0, \ldots, t_k\}$, the symmetric logistic distribution is given by*

$$\mathbb{P}(\zeta(t_0) \leq x_0, \ldots, \zeta(t_k) \leq x_k) = \exp\left[-\left(x_0^{-q} + \ldots + x_k^{-q}\right)^{1/q}\right], \tag{7.18}$$

for $x_0, \ldots, x_k > 0$ and $q > 1$. Hence, the density of the exponent measure is

$$\mu(dx_0, \ldots, dx_k) = \left(\textstyle\sum_{i=0}^k x_i^{-q}\right)^{1/q - k - 1} \left(\textstyle\prod_{i=1}^k (iq - 1)\right) \textstyle\prod_{i=0}^k x_i^{-q-1} dx_0 \ldots dx_k.$$

Applying Proposition 7.2.7, the incremental process W in (7.3) *is given by*

$$\mathbb{P}(W(t_1) \leq s_1, \ldots W(t_k) \leq s_k) = \left(1 + \textstyle\sum_{i=1}^k s_i^{-q}\right)^{1/q - 1}.$$

7.2.1 Continuous sample paths

In this subsection, we provide an analog result to Theorem 7.2.1, replacing convergence in the sense of finite-dimensional distributions by weak convergence on function spaces. For a Borel set $U \subset \mathbb{R}^d$, we denote by $C(U)$ and $C^+(U)$ the space of non-negative and strictly positive continuous functions on U, respectively, equipped with the topology of uniform convergence on compact sets.

Theorem 7.2.10. *Let $K \subset \mathbb{R}^d$ be compact and $\{\eta(t) : t \in K\}$ a $C^+(K)$-valued process in the MDA of a max-stable process $\{\zeta(t) : t \in K\}$ as in* (7.3) *in the sense of weak convergence on $C(K)$. W.l.o.g., assume that $n^{-1} \max_{i=1}^n \eta_i(\cdot) \Rightarrow \zeta(\cdot)$ as $n \to \infty$. Let W be the incremental process from* (7.3) *and Z a Pareto random variable, independent of W. Then, for any sequence $a(n) \to \infty$, as $n \to \infty$, we have the weak convergence on $(0, \infty) \times C(K)$*

$$\left(\frac{\eta(t_0)}{a(n)}, \frac{\eta(\cdot)}{\eta(t_0)} \,\Big|\, \eta(t_0) > a(n)\right) \Rightarrow (Z, W(\cdot)).$$

Remark 7.2.11. *Analogously to Whitt (1970, Thm. 5), weak convergence of a sequence of probability measures P_n, $n \in \mathbb{N}$, to some probability measure P on $C(\mathbb{R}^d)$ is equivalent to weak convergence of $P_n r_j^{-1}$ to $P r_j^{-1}$ on $C([-j, j]^d)$ for all $j \geq 1$, where $r_j : C(\mathbb{R}^d) \to C([-j, j]^d)$ denotes the restriction to the cube $[-j, j]^d$. Hence, Theorem 7.2.10 remains valid if the compact set K is replaced by $\mathbb{R}^d$.*

Proof of Theorem 7.2.10. As the process ζ is max-stable and $\eta \in \text{MDA}(\zeta)$, similarly to the case of multivariate max-stable distributions (cf. Theorem 7.2.1),

$$\lim_{u \to \infty} u\mathbb{P}(\eta/u \in B) = \mu(B) \tag{7.19}$$

for any Borel set $B \subset C(K)$ bounded away from 0^K, i.e., $\inf\{\sup_{s \in K} f(s) : \ f \in B\} > 0$, and with $\mu(\partial B) = 0$ de Haan & Ferreira (2006, Cor. 9.3.2), where μ is the *exponent measure* of ζ, defined by

$$\mathbb{P}(\zeta(s) \le x_j, \ s \in K_j, \ j = 1, \dots, m)$$
$$= \exp\left[-\mu\left(\left\{f \in C(K) : \ \sup_{s \in K_j} f(s) > x_j \text{ for some } j \in \{1, \dots, m\}\right\}\right)\right]$$

for $x_j \ge 0$, $K_j \subset K$ compact. Thus, μ equals the intensity measure of the PPP $\sum_{i \in \mathbb{N}} \delta_{U_i W_i(\cdot)}$. For $z > 0$ and $D \subset C(K)$ Borel, we consider the sets

$$A_z = \{f \in C(K) : \ f(t_0) > z\}, \qquad B_D = \{f \in C(K) : f(\cdot)/f(t_0) \in D\}$$

and $A = A_1$. Again, $cB_D = B_D$ for any $c > 0$. Then, as $W(t_0) = 1$ a.s., we have $\mu(A_z) = \int_z^\infty u^{-2} du = z^{-1}$ and for $s_0 \ge 1$ and any Borel set $D \subset C(K)$ with $\mathbb{P}(W \in \partial D) = 0$, by (7.19), analogously to (7.12), we get

$$\mathbb{P}\left\{\eta(t_0)/a(n) > s_0, \ \eta(\cdot)/\eta(t_0) \in D \ \middle|\ \eta(t_0) > a(n)\right\}$$
$$\xrightarrow{n \to \infty} \frac{\mu(B_D \cap A_{s_0})}{\mu(A)} = \int_{s_0}^\infty u^{-2} \mathbb{P}\{uW(\cdot) \in B_D\} du = s_0^{-1}\mathbb{P}\{W(\cdot) \in D\},$$

which is the joint distribution of Z and $W(\cdot)$. $\qquad\square$

7.3 Incremental representation of mixed moving maxima processes

We show that under certain assumptions, processes of M3 type as in (7.5) also admit an incremental representations (7.3) or (7.7). We distinguish between M3 processes with strictly positive shape functions, for which an incremental representation (7.3) exists, and general non-negative shape functions, for which only the weaker representation (7.7) can be obtained.

Note that the marginal distributions of an M3 process M as in (7.5) are given by

$$\mathbb{P}(M(t_l) \le s_l, l = 0, \dots, k) = \exp\left(-\int_{C(\mathbb{R}^d)} \int_{\mathbb{R}^d} \max_{l=0}^k \left(\frac{f(t_l - t)}{s_l}\right) dt \, \mathbb{P}_F(df)\right), \tag{7.20}$$

for $t_0, \dots, t_k \in \mathbb{R}^d$, $s_0, \dots, s_k \ge 0$, $k \in \mathbb{N}$.

7.3.1 Mixed moving maxima processes with positive shape functions

Theorem 7.3.1. *Let M be an M3 process on $\mathbb{R}^d$ as in (7.5) with a shape function F with $F(t) > 0$ for all $t \in \mathbb{R}^d$. Let $\Pi_0 = \sum_{i \in \mathbb{N}} \delta_{(U_i, T_i, F_i)}$ be the corresponding PPP on*

$(0,\infty) \times \mathbb{R}^d \times C(\mathbb{R}^d)$ *with intensity* $u^{-2}du\, dt\, \mathbb{P}_F(df)$. *Then* M *admits a representation* (7.3) *with* $t_0 = 0$ *and incremental process* W *given by*

$$\mathbb{P}(W \in L) = \int_{C^+(\mathbb{R}^d)} \int_{\mathbb{R}^d} \mathbf{1}_{\{f(\cdot-t)/f(-t) \in L\}} f(-t)dt\, \mathbb{P}_F(df), \quad L \in \mathcal{B}(C^+(\mathbb{R}^d)). \qquad (7.21)$$

Proof. We consider the two Poisson point processes on $(0,\infty) \times C^+(\mathbb{R}^d)$

$$\Pi_1 = \sum_{i \in \mathbb{N}} \delta_{(U_i F_i(-T_i), F_i(\cdot - T_i)/F_i(-T_i))},$$

as a transformation of Π_0 and $\Pi_2 = \sum_{i \in \mathbb{N}} \delta_{(U_i', W_i(\cdot))}$, with W_i, $i \in \mathbb{N}$, being independent copies of W, and with $\sum_{i \in \mathbb{N}} \delta_{U_i'}$ being a Fréchet point process. Then the intensity measures of Π_1 and Π_2 satisfy

$$\begin{aligned}
&\mathbb{E}\Pi_1([z,\infty) \times L) \\
&= \int_{C^+(\mathbb{R}^d)} \int_{\mathbb{R}^d} \int_0^\infty u^{-2} \mathbf{1}_{\{uf(-t) \geq z\}} \mathbf{1}_{\{f(\cdot-t)/f(-t) \in L\}}\, du\, dt\, \mathbb{P}_F(df) \\
&= z^{-1} \int_{C^+(\mathbb{R}^d)} \int_{\mathbb{R}^d} \mathbf{1}_{\{f(\cdot-t)/f(-t) \in L\}} f(-t)\, dt\, \mathbb{P}_F(df) \\
&= z^{-1} \mathbb{P}(W \in L) = \mathbb{E}\Pi_2([z,\infty) \times L),
\end{aligned}$$

$L \in \mathcal{B}(C^+(\mathbb{R}^d))$, $z > 0$, and hence $\Pi_1 \overset{d}{=} \Pi_2$. The assertion follows from the fact that M is uniquely determined by Π_1 via the relation $M(\cdot) = \max_{(v,g) \in \Pi_1} vg(\cdot)$. $\qquad \square$

While the definition of W in (7.21) is rather implicit, in the following, we provide an explicit construction of the incremental process W, which can also be used for simulation. To this end, let $\sum_{i \in \mathbb{N}} \delta_{U_i''}$ be a Fréchet point process and let the distribution of $(S,G) \in C^+(\mathbb{R}^d) \times \mathbb{R}^d$ be given by

$$\begin{aligned}
\mathbb{P}((S, G) \in (B \times L)) &= \int_{C^+(\mathbb{R}^d)} \int_{\mathbb{R}^d} \mathbf{1}_{s \in B} \mathbf{1}_{f \in L} \frac{f(-s)}{\int f(r)dr}\, ds \left(\int f(r)dr \right) \mathbb{P}_F(df) \\
&= \int_{C^+(\mathbb{R}^d)} \int_{\mathbb{R}^d} \mathbf{1}_{s \in B} \mathbf{1}_{f \in L} f(-s)\, ds\, \mathbb{P}_F(df), \qquad (7.22)
\end{aligned}$$

$B \in \mathcal{B}^d$, $L \in \mathcal{B}(C^+(\mathbb{R}^d))$. In other words, $\mathbb{P}_G(df) = (\int f(r)dr)\, \mathbb{P}_F(df)$ and, conditional on $\{G = f\}$, the density function of the shift S is proportional to $f(-\cdot)$. Putting $W(\cdot) = G(\cdot - S)/G(-S)$, equation (7.21) is satisfied and with i.i.d. copies W_i, $i \in \mathbb{N}$, of W, we get that $\max_{i \in \mathbb{N}} U_i'' W_i(\cdot)$ is indeed an incremental representation (7.3) of the mixed moving maxima process M.

Remark 7.3.2 (M3 representation of Brown-Resnick processes)**.** *We consider the following two special cases of M3 processes:*

1. *Let* $\Sigma \in \mathbb{R}^{d \times d}$ *be a positive definite matrix and let the shape function be given by* $F(t) = (2\pi)^{-d/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}t^\top \Sigma^{-1} t\right\}$, $t \in \mathbb{R}^d$. *Then,* M *becomes the well-known*

Smith process. At the same time, by (7.22), $S \sim N(0,\Sigma)$ *and* $G \equiv F$. *Thus*

$$Y(t) = \exp\left\{-\tfrac{1}{2}(t-S)^\top \Sigma^{-1}(t-S) + \tfrac{1}{2}S^\top \Sigma^{-1}S\right\}$$
$$= \exp\left\{-\tfrac{1}{2}t^\top \Sigma^{-1}t + t^\top \Sigma^{-1}S\right\}.$$

Since $\mathbb{E}(t^\top \Sigma^{-1}S)^2 = t^\top \Sigma^{-1}t$, M *is equivalent to the Brown-Resnick process in* (7.24) *with variogram* $\gamma(h) = h^\top \Sigma^{-1}h$.

2. *For the one-dimensional Brown-Resnick process* ζ *in* (7.24) *with variogram* $\gamma(h) = |h|$, *i.e.,* Y *is the exponential of a standard Brownian motion with drift* $-|t|/2$, *Engelke* et al. (2011) *recently showed that the M3 representation is given by* $\{F(t) : t \in \mathbb{R}\} = \{Y(t) \mid Y(s) \leq 0 \ \forall s \in \mathbb{R} : t \in \mathbb{R}\}$, *i.e., the shape function is the exponential of a conditionally negative drifted Brownian motion. Having these two representations, it follows that the law of the conditional Brownian motion* F, *re-weighted by* $\int F(t)dt$ *and randomly shifted with density* $F(-\cdot)/\int F(t)dt$, *coincides with the law of* Y.

7.3.2 Mixed moving maxima processes with finitely supported shape functions

Let M be an M3 process on $\mathbb{R}^d$ as in (7.5). In contrast to Subsection 7.3.1, where the shape functions are required to take positive values, here, we allow for arbitrary shape functions with values in $[0, \infty)$.

Theorem 7.3.3. *The M3 process* M *as in* (7.5) *allows for an incremental representation of the form* (7.7), *with incremental processes* V_i *given by*

$$V_i(\cdot) = F_i(\cdot - R_i)/g(R_i).$$

Here R_i, $i \in \mathbb{N}$, *are i.i.d. copies of a random vector* R *with arbitrary density* g *satisfying* $g(t) > 0$ *for all* $t \in \mathbb{R}^d$, *and* F_i, $i \in \mathbb{N}$, *are i.i.d. copies of the random shape function* F.

Proof. With $\sum_{i \in \mathbb{N}} \delta_{U_i}$ being a Fréchet point process, we consider the process

$$\tilde{M}(t) = \max_{i \in \mathbb{N}} U_i F_i(t - R_i)/g(R_i), \quad t \in \mathbb{R}^d,$$

which clearly is of the form (7.7). Then,

$$\mathbb{P}(\tilde{M}(t_l) \leq s_l, l = 0, \ldots, k) = \exp\left(-\int_{C(\mathbb{R}^d)} \int_{\mathbb{R}^d} \max_{l=0}^{k} \frac{f(t_l - t)}{g(t)s_l} g(t)dt \, \mathbb{P}_F(df)\right).$$

The right-hand side coincides with the marginal distribution of M, which is given by (7.20). This concludes the proof. $\square$

Decomposing V as in (7.8) with $t_0 = 0$, we obtain the equality in distribution

$$V^{(1)}(\cdot) \stackrel{d}{=} \left(F(\cdot - R)/g(R) \mid -R \in \text{supp}(F)\right).$$

Applying Theorem 7.2.1 yields

$$
\begin{aligned}
\mathbb{P}(\Delta\tilde{\mathbf{V}}^{(1)} \in d\mathbf{z}) &= \mathbb{P}(F(-R)/g(R) > 0) \cdot \int_0^\infty y\mathbb{P}(V^{(1)}(0) \in dy,\, \Delta\mathbf{V}^{(1)} \in d\mathbf{z}) \\
&= \int_0^\infty y \int_{C(\mathbb{R}^d)} \int_{-\operatorname{supp}(f)} \mathbf{1}_{f(-t)/g(t)\in dy}\mathbf{1}_{(f(t_l-t)/f(-t))_{l=1}^k\in d\mathbf{z}}g(t)dt\, \mathbb{P}_F(df)dy \\
&= \int_{C(\mathbb{R}^d)} \int_{\operatorname{supp}(f)} f(t)\mathbf{1}_{(f(t_l+t)/f(t))_{l=1}^k\in d\mathbf{z}}dt\, \mathbb{P}_F(df).
\end{aligned}
\tag{7.23}
$$

The asymptotic conditional increments of $\eta \in \operatorname{MDA}(M)$ can be seen as a convolution of the shape function's increments with a random shift, whose density is given by the shape function itself. The distribution is particularly independent of the choice of the density g in Theorem 7.3.3.

Remark 7.3.4. *Section 7.3.1 considers the subclass of M3 processes with strictly positive shape functions and provides an incremental representation as in (7.3), which is nicely related to the conditional increments of η due to the property $W(0) = 1$. Section 7.3.2 applies to arbitrary M3 processes but only yields an incremental representation as in (7.7), for which the incremental process V does not directly represent the conditional increments of η.*

7.4 Outlook: Statistical applications

In univariate extreme value theory, a standard method for estimating the extreme value parameters fits all data exceeding a high threshold to a certain Poisson point process. This peaks-over-threshold approach has been generalized in Rootzén & Tajvidi (2006) to the multivariate setting. Conditioning on the event that at least one component of a random vector is large, the recent contribution Falk & Tichy (2012) analyzes the asymptotic distribution of exceedance counts of stationary sequences.

Here, we have suggested conditioning a stochastic process $\{\eta(t) : t \in T\}$ in the MDA of a max-stable process $\{\zeta(t) : t \in T\}$ such that it converges to the incremental processes W in (7.3). In this final section, some examples are provided of how these theoretical results can be used for statistical inference. The approach is a multivariate peaks-over-threshold method for max-stable processes, though the definition of extreme events differs from that in Rootzén & Tajvidi (2006); Falk & Tichy (2012).

In the sequel, suppose that $\eta_1, \ldots, \eta_n$, $n \in \mathbb{N}$, are independent observations of the random process η, already normalized to standard Pareto margins.

For a max-stable process ζ that admits an incremental representation (7.3), the statistical merit of the convergence results in Theorem 7.2.1 and Theorem 7.2.10 is the "deconvolution" of U and W, which allows to substitute estimation of ζ by estimation of the process W. As only the single *extreme* events converge to W, we define the index set of extremal observations as

$$
I_1(n) = \{i \in \{1, \ldots n\} : \ \eta_i(t_0) > a(n)\},
$$

for some fixed $t_0 \in T$. The set $\{\eta_i(\cdot)/\eta_i(t_0) : i \in I_1(n)\}$ then represents a collection of independent random variables that approximately follow the distribution of W. Thus, once the representation in (7.3) is known, both parametric and non-parametric estimation for the process W is feasible. For statistical inference it is necessary that the number of extremal observations $|I_1(n)|$ converges to ∞, as $n \to \infty$, cf. Remark 7.2.4.

Example 7.4.1 (Symmetric logistic distribution, cf. Example 7.2.9). *The dependence parameter $q \geq 1$ of the symmetric logistic distribution (7.18) can be estimated by perceiving the conditional increments of η in the MDA as realizations of W and maximizing the likelihood*

$$\mathbb{P}\big(W(t_1) \in ds_1, \ldots W(t_k) \in ds_k \,\big|\, q\big)$$
$$= \Big(1 + \textstyle\sum_{i=1}^k s_i^{-q}\Big)^{1/q-(k+1)} \Big(\textstyle\prod_{i=1}^k (iq-1)\Big) \textstyle\prod_{i=0}^k s_i^{-q-1} \; ds_1 \ldots ds_k.$$

Example 7.4.2 (Brown-Resnick processes, cf. Brown & Resnick (1977); Kabluchko *et al.* (2009)). *Let $\{Y(t) : t \in \mathbb{R}^d\}$ be a centered Gaussian process with stationary increments and $Y(t_0) = 0$ for some $t_0 \in \mathbb{R}^d$. Let $\gamma(t) = \mathbb{E}(Y(t) - Y(0))^2$, $t \in \mathbb{R}^d$, denote the variogram of Y. Then, with a Fréchet point process $\sum_{i \in \mathbb{N}} \delta_{U_i}$ and independent copies Y_i of Y, $i \in \mathbb{N}$, the* Brown-Resnick process for the variogram γ *is given by*

$$\zeta(t) = \max_{i \in \mathbb{N}} U_i \exp\big(Y_i(t) - \gamma(t - t_0)/2\big), \quad t \in \mathbb{R}^d. \tag{7.24}$$

Its distribution only depends on γ. Here, W from representation (7.3) is the log-Gaussian process. Hence, standard estimation procedures for Gaussian processes can be applied for statistical inference. Engelke et al. (2012a) explicitly construct several new estimators of the variogram γ based on the incremental representation, which also covers Hüsler-Reiss distributions.

Bibliography

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies* **18**(2):351–416.

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

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2009). Parametric and Nonparametric Volatility Measurement. In *Handbook of Financial Econometrics* (eds. Y. Ait-Sahalia & L. Hansen). Elsevier North-Holland, Amsterdam.

Andersen, T. G., Bollerslev, T. & Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics* **160**:220–234.

Andreou, E. & Ghysels, E. (2009). Structural Breaks in Financial Time Series. In *Handbook of Financial Time Series* (eds. T. Mikosch, J.-P. Kreiß, R. Davis & T. Andersen), chap. 37, pp. 839–870. Springer, Berlin, Heidelberg.

Artzner, P., Delbaen, F., Eber, J. M. & Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance* **9**(3).

Bandi, F. M. & Russell, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics* **79**:655–692.

Bandi, F. M. & Russell, J. R. (2008). Microstructure Noise, Realized Variance, and Optimal Sampling. *Review of Economic Studies* **75**:339–369.

Barndorff-Nielsen, O. E. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* **2**:1–37.

Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based Models and some of their Uses in Financial Economics. *Journal of the Royal Statistical Society, Series B* **63**(2):167–241.

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

Basrak, B. & Segers, J. (2009). Regularly varying multivariate time series. *Stochastic Process Appl* **119**(4):1055–1080.

Bauwens, L. & Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity.* Kluwer, Dordrecht.

Bauwens, L. & Hautsch, N. (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics* **4**(3):450–493.

Bauwens, L. & Hautsch, N. (2009). Modelling financial high frequency data using point processes. In *Handbook of Financial Time Series* (eds. T. G. Andersen, R. A. Davis, J.-P. Kreiß & T. Mikosch). Springer, Berlin.

Begon, M., Harper, J. L. & Townsend, C. R. (1990). *Ecology: individuals, populations and communities*, vol. 2nd. Blackwell Scientific Publications.

Berk, K. N. (1973). A Central Limit Theorem for m-Dependent Random Variables with Unbounded m. *Ann Probab* **1**(2):352–354.

Bialkowski, J., Darolles, S. & Le Fol, G. (2008). Improving VWAP strategies: A dynamic volume approach. *Journal of Banking & Finance* **32**(9):1709–1722.

Black, F. & Scholes, M. S. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3):637–54.

Blanchet, J. & Davison, A. C. (2011). Spatial modeling of extreme snow depth. *Ann Appl Stat* **5**:1699–1725.

Bollerslev, T. & Wright, J. H. (2001). High-frequency data, frequency domain inference, and volatility forecasting. *The Review of Economics and Statistics* **83**(4):596–602.

Bouchaud, J. & Potters, M. (2003). *Theory of financial risk and derivative pricing.* Cambridge Univ. Press, Cambridge, 2nd edn.

Bourgault, G. (1997). Spatial declustering weights. *Mathematical Geology* **29**:277–290.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics* **141**(2):876–912.

Brown, B. M. & Resnick, S. I. (1977). Extreme values of independent stochastic processes. *J Appl Probab* **14**:732–739.

Cartea, A. & Jaimungal, S. (2010). Modeling asset prices for algorithmic and high frequency trading. Available at SSRN: http://ssrn.com/abstract=1722202.

Centanni, S. & Minozzo, M. (2006). A Monte Carlo Approach to Filtering for a Class of Marked Doubly Stochastic Poisson Processes. *Journal of the American Statistical Association* **101**:1582–1597.

Chiles, J. P. & Delfiner, P. (1999). *Geostatistics: modeling spatial uncertainty.* John Wiley & Sons, New York.

Cont, R., Stoikov, S. & Talreja, R. (2010). A stochastic model for order book dynamics. *Oper Res* **58**(3):549–563.

Cosslett, S. R. & Lee, L.-F. (1985). Serial correlation in latent discrete variable models. *Journal of Econometrics* **27**(1):79–97.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* Wiley, New York.

Daley, D. J. & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods.* Probability and its Applications. Springer, New York, 2nd edn.

Daley, D. J. & Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure.* Probability and its Applications. Springer, New York, 2nd edn.

Davison, A. C. & Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proc R Soc Lond Ser A Math Phys Eng Sci* **468**:581–608.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann Probab* **12**(4):1194–1204.

de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory.* Springer, New York.

Deutsch, C. (1989). Declus: a fortran 77 program for determining optimum spatial declustering weights. *Computers & Geosciences* **15**(3):325–332.

Diggle, P. J., Menezes, R. & Su, T. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **59**(2):191–232.

Duval, C. & Hoffmann, M. (2011). Statistical inference across time scales. *Electronic Journal of Statistics* **5**:2004–2030.

Easley, D. & O'Hara, M. (1992). Time and the Process of Security Price Adjustment. *Journal of Finance* **47**:577–605.

Embrechts, P. (2000). *Extremes and Integrated Risk Management.* Risk Books.

Embrechts, P., Mikosch, T. & Klüppelberg, C. (1997). *Modelling Extremal Events: for Insurance and Finance (Stochastic Modelling and Applied Probability).* Springer, London.

Emery, X. & Ortiz, J. M. (2005). Histogram and variogram inference in the multigaussian model. *Stochastic Environmental Research and Risk Assessment* **19**:48–58.

Engelke, S., Kabluchko, Z. & Schlather, M. (2011). An equivalent representation of the Brown-Resnick process. *Statist Probab Lett* **81**:1150 – 1154.

Engelke, S., Malinowski, A., Kabluchko, Z. & Schlather, M. (2012*a*). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. (Available from `http://arxiv.org/abs/1207.6886`).

Engelke, S., Malinowski, A., Oesting, M. & Schlather, M. (2012*b*).   Representations of max-stable processes based on single extreme events.   (Available from `http://arxiv.org/abs/1209.2303`).

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* **50**:987–1007.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica* **68**:1–22.

Engle, R. F. & Lunde, A. (2003). Trades and quotes: A bivariate point process. *Journal of Financial Econometrics* **1**(2):159–188.

Engle, R. F. & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **66**:1127–1162.

Engle, R. F. & Russell, J. R. (2009). Analysis of High Frequency Financial Data. In *Handbook of Financial Econometrics* (eds. Y. Ait-Sahalia & L. Hansen). Elsevier North-Holland, Amsterdam.

Falk, M. & Tichy, D. (2012). Asymptotic conditional distribution of exceedance counts. *Adv Appl Probab* **44**(1):270–291.

Farrell, R. (1962). Representation of invariant measures. *Illinois J Math* **6**(3):447–467.

Fasen, V., Klüppelberg, C. & Schlather, M. (2010). High-level dependence in time series models. *Extremes* **13**(1):1–33.

Fernandes, M. & Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics* **130**(1):1–23.

Fisher, R. & Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* **24**(02).

Frey, R. (2000). Risk minimization with incomplete information in a model for high-frequency data. *Mathematical Finance* **10**(2):215–225.

Gentleman, R. (2008). *R Programming for Bioinformatics*. Chapman & Hall/CRC.

Ghysels, E. & Jasiak, J. (1998). GARCH for irregularly spaced financial data: The ACD-GARCH model. *Studies in Nonlinear Dynamics & Econometrics* **2**(4):133–149.

Ghysels, E., Harvey, A. & Renault, E. (1995). Stochastic volatility. Tech. rep., Toulouse - GREMAQ. URL `http://ideas.repec.org/p/fth/gremaq/95.400.html`.

Gilli, M. & Këllezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics* **27**(2):207–228.

Goldfeld, S. M. & Quandt, R. E. (1973). A markov model for switching regressions. *Journal of Econometrics* **1**(1):3–15.

Goodhart, C. A. E. & O'Hara, M. (1997). High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance* **4**:73–114.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation.* Oxford University Press.

Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ Inst Statist Univ Paris* **9**:171–173.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**(2):357–84.

Hamilton, J. D. (1993). *Time-series analysis.* Princeton Univerity Press, 1st edn.

Hamilton, J. D. (2008). Regime-switching models. In *The New Palgrave Dictionary of Economics* (eds. S. N. Durlauf & L. E. Blume). Palgrave Macmillan, Basingstoke.

Hansen, P. R. & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* **24**:127–161.

Hausman, J. A., Lo, A. W. & MacKinlay, A. C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* **31**:319–379.

Hautsch, N. (2004). *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models.* Lecture Notes in Economics and Mathematical Systems. Springer.

Hautsch, N. (2011). *Econometrics of Financial High-Frequency Data.* Springer.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**(1):83–90.

Heinrich, L. & Molchanov, I. S. (1999). Central limit theorem for a class of random measures associated with germ-grain models. *Advances in Applied Probability* **31**(2):283–314.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**:1163–1174.

Ho, L. P. & Stoyan, D. (2008). Modelling marked point patterns by intensity-marked Cox processes. *Statistics & Probability Letters* **78**:1194–1199.

Hüsler, J. & Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statist Probab Lett* **7**:283–286.

Isaaks, E. H. & Srivastava, R. M. (1989). *Applied geostatistics.* Oxford University Press.

Isham, V. (1985). Marked point processes and their correlations. In *Spatial Processes and Spatial Time Series Analysis* (ed. F. Droesbeke), pp. 63–75. Brussels: Publications des Facultés Universitaires Saint-Louis.

Journel, A. G. (1983). Nonparametric estimation of spatial distributions. *Mathematical Geology* **15**(3):445–468.

Juri, A. & Wüthrich, M. (2003). Tail dependence from a distributional point of view. *Extremes* **6**:213–246.

Kabluchko, Z. (2011). Extremes of independent Gaussian processes. *Extremes* **14**:285–310.

Kabluchko, Z., Schlather, M. & de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann Probab* **37**:2042–2065.

Kallenberg, O. (1986). *Random measures.* Academic Press, New York, 4th edn.

Karr, A. F. (1991). *Point Processes and Their Statistical Inference.* Marcel Dekker, Inc.

Lando, D. (1998). On Cox Processes and Credit Risky Securities. *Review of Derivatives Research* **2**(2/3):99–120.

Lange, T. & Rahbek, A. (2009). An introduction to regime switching time series models. In *Handbook of Financial Time Series* (eds. T. Mikosch, J.-P. Kreiß, R. Davis & T. Andersen), chap. 38, pp. pp. 871–887. Springer, Berlin, Heidelberg.

Langrock, R. (2011). Some applications of nonlinear and non-Gaussian state-space modelling by means of hidden Markov models. *Journal of Applied Statistics* **38**(12):2955–2970.

Laurini, F. & Tawn, J. A. (2008). Regular variation and extremal dependence of garch residuals with application to market risk measures. *Econometric Reviews* **28**(1-3):146–169.

Leadbetter, M. (1991). On a basis for 'peaks over threshold' modeling. *Stat Probab Lett* **12**(4):357 – 362.

Ledford, A. W. & Tawn, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* **65**(2):521–543.

Madhavan, A. (2002). VWAP strategies. In *Investment Guides, Transaction Performance: The Changing Face of Trading* (ed. B. Bruce), pp. 32–38. Institutional Investor Inc., New York.

Malinowski, A. (2009). Analysing interactions within high-frequency financial data using marked point processes and UHF-GARCH models. Diploma thesis.

Malinowski, A. & Schlather, M. (2011*a*). R Programming for Bioinformatics. R. Gentleman (2009). Chapman & Hall/CRC. ISBN 978-1-4200-6367-7. *Biometrical Journal* **53**(4):707–707.

Malinowski, A. & Schlather, M. (2011*b*). Refined analysis of interactions within high-frequency financial data through marked point process theory. In revision.

Malinowski, A., Schlather, M. & Zhang, Z. (2012*a*). Intrinsically weighted means of marked point processes. (Available from `http://arxiv.org/abs/1210.1335`).

Malinowski, A., Schlather, M. & Zhang, Z. (2012*b*). Marked point process adjusted tail dependence analysis for high-frequency financial data. Submitted.

Marshall, A. W. & Olkin, I. (1983). Domains of attraction of multivariate extreme value distributions. *The Annals of Probability* **11**(1):168–177.

Martens, M. (2002). Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets* **22**(6):497–518.

Mase, S. (1996). The threshold method for estimating total rainfall. *Annals of the Institute of Statistical Mathematics* **48**(2):201–213.

McCulloch, R. E. & Tsay, R. S. (2001). Nonlinearity in high-frequency financial data and hierarchical models. *Studies in Nonlinear Dynamics & Econometrics* **5**(1).

Meddahi, N. & Renault, E. (2004). Temporal aggregation of volatility models. *Journal of Econometrics* **119**(2):355–379.

Meddahi, N., Renault, E. & Werker, B. (2006). GARCH and irregularly spaced data. *Economics Letters* **90**(2):200–204.

Meinguet, T. & Segers, J. (2010). Regularly varying time series in Banach spaces. (Available from `http://arxiv.org/abs/1001.3262`).

Møller, J. & Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC.

Møller, J., Syversveen, A. R. & Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**:451–482.

Myllymäki, M. (2009). Statistical models and inference for spatial point patterns with intensity-dependent marks. Tech. rep., University of Jyväskylä.

Myllymäki, M. & Penttinen, A. (2009). Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. *Statistica Neerlandica* **63**(4):450–473.

Nelson, D. & Cao, C. (1992). Inequality Constraints in the Univariate GARCH Model. *Journal of Business & Economic Statistics* **10**:229–35.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**(3):318–334.

Ober, U., Malinowski, A., Schlather, M. & Simianer, H. (2012). The expected linkage disequilibrium in finite populations revisited. Unpublished manuscript.

Olea, R. (2007). Declustering of clustered preferential sampling for histogram and semivariogram inference. *Mathematical Geology* **39**:453–467.

Pacurar, M. (2008). Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys* **22**(4):711–751.

Padoan, S. A., Ribatet, M. & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *J Amer Statist Assoc* **105**:263–277.

Piger, J. (2009). Econometrics: Models of regime changes. In *Encyclopedia of Complexity and Systems Science* (ed. R. A. Meyers), pp. pp. 2744–2757. Springer.

Politis, D. & Sherman, M. (2001). Moment estimation for statistics from marked point processes. *Journal Of The Royal Statistical Society Series B* **63**(2):261–275.

Politis, D., Romano, J. & Wolf, M. (1999). *Subsampling.* Springer, New York.

Racicot, F., Théoret, R. & Coën, A. (2008). Forecasting Irregularly Spaced UHF Financial Data: Realized Volatility vs UHF-GARCH Models. *International Advances in Economic Research* **14**:112–124.

Resnick, S. I. (2008). *Extreme Values, Regular Variation and Point Processes.* Springer, New York.

Rootzén, H. & Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12**:917–930.

Russell, J. R. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data. Working Paper, University of Chicago.

Rydberg, T. H. & Shephard, N. (2000). A Modelling Framework for the Prices and Times of Trades Made on the New York Stock Exchange. In *Nonlinear and Nonstationary Signal Processing* (eds. W. J. Fitzgerald, R. L. Smith, A. T. Walden & P. C. Young), vol. 1, pp. 217–246. Cambridge University Press.

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

Schlather, M. (2001). On the second-order characteristics of marked point processes. *Bernoulli* **7**:99–117.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes* **5**:33–44.

Schlather, M. & Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90**:139–156.

Schlather, M., Ribeiro, Jr, P. J. & Diggle, P. J. (2004). Detecting Dependence between Marks and Locations of Marked Point Processes. *Journal of the Royal Statistical Society, Series B* **66**(1):79–93.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models: In Econometrics, Finance and Other Fields* (eds. D. R. Cox, D. V. Hinkley & O. E. Barndorff-Nielsen). Chapman & Hill, New York.

Smith, R. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.

Stoyan, D. (1984). On correlations of marked point processes. *Math Nachr* **116**:197–207.

Stoyan, D. & Stoyan, H. (1994). *Fractals, random shapes and point fields.* John Wiley & Sons Ltd, Chichester.

Stoyan, D. & Stoyan, H. (2000). Improving ratio estimators of second order point process characteristics. *Scandinavian Journal of Statistics* **27**:641–656.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and its Applications.* John Wiley & Sons Ltd, Chichester, 2nd edn.

Strokorb, K. & Schlather, M. (2012). Characterizing extremal coefficient functions and extremal correlation functions. (Available from `http://arxiv.org/abs/1205.1315`).

Sun, Z. & Engle, R. F. (2007). When is noise not noise – a microstructure estimate of realized volatility. NYU Working Paper No. FIN-07-047.

Taylor, S. J. (1986). *Modelling Financial Time Series.* Wiley, Chichester.

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications.* Springer.

Wang, Y. & Stoev, S. A. (2010). On the structure and representations of max-stable processes. *Adv Appl Probab* **42**:855–877.

Whitt, W. (1970). Weak Convergence of Probability Measures on the Function Space $C[0,\infty)$. *Ann Math Statist* **41**(3):939–944.

Zhang, M. Y., Russell, J. R. & Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics* **104**:179–207.

Zucchini, W. & MacDonald, I. (2009). *Hidden Markov Models for Time Series: An Introduction Using R.* Chapman & Hall, London.