UNIVERSITY OF CALIFORNIA
SANTA CRUZ

A BAYESIAN NONPARAMETRIC MODELING FRAMEWORK FOR EXTREME VALUE ANALYSIS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS AND STOCHASTIC MODELING

by

Ziwei Wang

December 2012

The Dissertation of Ziwei Wang
is approved:

Professor Abel Rodríguez, Co-chair

Professor Athanasios Kottas , Co-chair

Professor Bruno Sansó

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

A Bayesian nonparametric modeling framework for extreme value analysis

by

Ziwei Wang

Extreme value theory studies the tail behavior of a stochastic process, and plays a key role in a wide range of applications. Understanding and quantifying the behavior of rare events and the associated uncertainties is practically important for risk assessment, since such unexpected events can result in massive losses of wealth and high cost in human life. In this dissertation, we present a Bayesian nonparametric mixture modeling framework for the analysis of extremes with applications in financial industry and environmental sciences. In particular, the modeling is built from the point process approach to analysis of extremes, under which the pairwise observations, comprising the time of excesses and the exceedances over a high threshold, are assumed to arise from a non-homogeneous Poisson process. To relax the time homogeneity restriction, implicit in traditional parametric methods, a nonparametric Dirichlet process mixture model is presented to provide flexibility in estimation of the joint intensity of extremes, the marginal intensity over time, and different types of return level curves for one financial market. This class of models is then expanded to assess the effect of systemic risk in multiple financial markets. In this case, the process generating the extremes is modeled as a superposition of two Poisson process. This approach provides a decomposition of the risk associated with each individual market into two components: a systemic risk component and an idiosyncratic risk component. Finally, we extend the point process framework to model spatio-temporal extremes

from environmental processes observed at multiple spatial locations over a certain time interval. Specifically, a spatially varying mixing distribution, assigned a spatial Dirichlet process prior, is incorporated into the model to develop inference for spatial interpolation of risk assessment quantities for high-level exceedances. The modeling approaches are illustrated with a number of simulated and real data examples.

To my family,

who offered me endless love, support, and encouragement

throughout my whole life.

Acknowledgments

It has been a great fortunate to me to write this dissertation as part of my doctoral journey in department of Applied Mathematics and Statistics at UCSC have been guided and assisted by a wonderful group of faculties and members throughout my education.

My first debt of gratitude must go to my co-advisors, Professor Athanasios Kottas and Professor Abel Rodríguez, for their enormous guidance, motivations, and endless support throughout my Ph.D. study. I will always remember their inspiration, patience, and great efforts in supervising me. I want to thank for their perseverance, commitment to research work and all the encouragement when I faced computational challenges and difficult time during the research. Specifically, to Thanasis for his sense of humor and enthusiastic attitude towards research and life that guides me during these years (and always corrects my Greek pronunciation). To Abel for his insightful discussions about the research and the sharing of his knowledgeable programing tricks.

In addition to my advisors, Professor Bruno Sansó deserves the special thanks for reading the whole dissertation and contributing valuable advices. Further, he has done outstanding jobs of overseeing the remarkably collegial department by serving as chair during my time in the department.

I want to thank other professors in the department who taught courses and shared their knowledge so that I grow as a Bayesian. I also have met many colleagues within the department whom I wish to thank.

Last but not least, I would like to thank my husband with whose endless love and

support that allows me to finish this journey. I would also like to express my gratitude to my parents, for giving birth to me at the first place and supporting me to study abroad. Their precious love, understanding and continuous encouragement bring huge supports into my life and help me to achieve who I am in the end.

Chapter 1

Introduction

1.1 Motivation

In the past 100 years, China has seen several severe floods caused by abnormal weather phenomena, such as heavy rainfall and snowstorms. These natural catastrophes have brought not only large economic losses, but also a high cost in human life. Between 1.45 to 3.7 million people died and 28.5 million people were affected by the Yangtze River floods of 1931, which has been considered the deadliest natural disaster since 1900. More recently, when the Yangtze suffered its worst floods in 50 years during the 1998 summer, more than 4000 people lost their lives, around 14 million people were left homeless and 240 million people were affected directly by the floods. These floods also resulted in serious economic losses; 5 million houses were destroyed, 12 million houses were damaged and submerged, and 25 million acres of land flooded; the total estimated loss was over $20 billion ($US) (NCDC, 1998).

Besides environmental science, stock markets have shown numerous large downward

price movements in the last ten years. This phenomenon can be the result of major catastrophic events, financial crises or speculative stock market bubbles. The seven most significant events that lead to a stock market crash from 2000 to 2011 are: the bursting of the dot-com bubble (03/10/2000), the 09/11 terrorist attacks (09/11/2001), the stock market downturn of 2002 (09/12/2001), the bursting of the Chinese bubble (02/27/2007), the bankruptcy of Lehman Brothers (09/16/2008), Dubai's debt standstill (11/27/2009), and the beginning of the European sovereign debt crisis (08/27/2010). Such extreme price movements in stock market often result in largely unexpected, but widespread losses of wealth. For instance, the stock market decline in 2008 caused the Great Recession and massive economic losses to the investigators, market participants, and the entire financial industry. Therefore, monitoring such rare events and quantifying huge losses is very important for our society.

It is well known that the actual distributions of stock market returns is typically heavy-tailed relative to the "bell-shaped" normal distribution. Hence, an appropriately heavy-tailed distribution would provide more accurate estimates of Value-at-Risk (VAR), a benchmark measure of portfolio risk (Smith, 2003). However, since the number of extreme observations is usually small relative to the entire data set, we have little information of the tails. On the other hand, the presence of the extreme values in standard statistical analysis, determined to be outliers, can lead to a large distortion effect on sample means and variances. For instance, estimated regression models using the least square approach are highly sensitive to extreme observations. Thus, when statisticians apply traditional statistical methods to describe the main part of the distribution, they may use most of the data, but remove the extreme values. However, since the extreme events may cause enormous loss of human life and property, it is important to under-

stand and capture the behavior of such rare events as well as predict their future occurrance.

One way to address the above two problems is to consider statistical approaches that focus exclusively on modeling the tail of the distributions. Extreme value analysis, a method that focuses on investigating the behavior of rare events and the associated uncertainties, is an important research field in risk assessment. It has been widely applied in environmental sciences and financial industry. For instance, extreme value analysis provides a framework to characterize the extremal behavior of the daily returns by quantifying possible losses during the period of financial turbulences in stock market.

1.2 Outline and Contributions

Chapter 2 reviews relevant background and concepts applied throughout this dissertation. In the first part of Chapter 2, we provide a brief review of univariate extreme value theory with emphasis on the point process approach, the one primarily used in this dissertation. In particular, under the point process approach, the pairwise observations comprising the time of excesses and the exceedances over a high threshold, are assumed to arise from a non-homogeneous Poisson process (NHPP) with a non-negative and locally integrable intensity function. The three approaches discussed in Section 2.1 are parametric methods that rely on strong assumptions about the distribution of observations. Specifically, it is assumed that all the samples are independent and identically distributed from a distribution F_0. However, many datasets do not satisfy these conditions. For instance, large values of observations in most financial and environmental situations often occur in clusters. In addition, some series of data points

are often seasonal or dependent on the price level, such as the stock index or rainfall data. In particular, a serious limitation of the limiting form for the intensity function of extreme values, under the traditional parametric point process approach, is that it is homogeneous over time. This restriction will likely lead to unreasonable inferences when the occurrence of the extremal events varies across space or time.

The main contribution of this dissertation is a novel class of models that provide flexible inference methods under the point process setting for extreme value analysis. In particular, we focus on modeling the extremal events generated from general stochastic processes evolving over time and possibly also over space. We develop a Bayesian nonparametric modeling framework to model the occurrence of extremes with the aim of relaxing the restrictive aspects of standard parametric methods. The Bayesian paradigm offers clear advantages for development of such modeling, since it allows exploration of flexible hierarchical model formulations and proper incorporation of full predictive uncertainty. In addition, nonparametric mixture prior models provide a rich inference framework for functions, densities or distributions. Background on the main nonparametric mixture modeling approach utilized throughout this dissertation is introduced in Section 2.2. Specifically, we discuss relevant definitions and properties for the Dirichlet process prior, Dirichlet process mixture models, including related computational strategies for posterior simulation.

In Chapter 3, we provide a flexible modeling framework for the analysis of extremes of temporal stochastic processes using a point process approach. To achieve flexible shapes and temporal heterogeneity for the intensity of extremes, we utilize a Bayesian nonparametric Dirichlet process mixture model. Particular emphasis is placed on the choice of the mixture

kernel to ensure desirable results for the implied tail behavior of the marginal extreme value distribution. The mixture nature of the nonparametric model enables more general inferences than traditional parametric methods. To our knowledge, this approach provides the first attempt to fully nonparametric modeling for extremes from a single time series, with flexible resulting inference for the joint intensity of extremes, the marginal intensity over time, and for different types of return level curves. The methodology is illustrated with a simulated data example and a real data example involving returns of the Dow Jones index over a five year period.

In Chapter 4, we extend the point process framework developed in Chapter 3 to assess the effect of systemic risks on multiple financial markets. The point process of extremes on each market is modeled as a superposition of two Poisson processes, that can be interpreted as the systemic and idiosyncratic market risk components. Dirichlet process mixture models are used to provide flexible inference for the corresponding temporal intensities through data-driven clustering of extreme values. We apply the model to understand the behavior of the S&P500 sector indexes between January 1st, 2000 and December, 31, 2011. As a result, our application provides interesting insights about the relevant risks associated with different economic sectors. For example, our model suggests that there are few idiosyncratic risks associated with the consumer staples sector, whose extreme negative returns seem to be driven mostly by systemic risks.

Chapter 5 presents an approach to modeling and risk assessment for extremes of environmental processes evolving over time and recorded at a number of spatial locations. We extend the earlier modeling framework discussed in Chapter 3 to capture the temporal heterogeneity for the intensity of extremes at any spatial location. In particular, we utilize a logit-normal

mixture model for the corresponding Poisson process density. A spatial Dirichlet process prior for the mixing distributions completes the nonparametric spatio-temporal model formulation. We develop inference for spatial interpolation of risk assessment quantities for high-level exceedances of the environmental process. The methodology is tested with a synthetic data example and is further illustrated with analysis of rainfall exceedances recorded over a period of 50 years from a region in South Africa.

Finally, Chapter 6 concludes with a summary of the main contributions of this dissertation.

Chapter 2

Background for Extreme Value Analysis and Bayesian Nonparametric Models

2.1 Univariate Extreme Value Theory

Extreme value theory (EVT), which focuses on the study of the tail behavior of a stochastic process, plays a key role in a number of fields, such as environmental sciences and finance. The literature on extreme value analysis for independent and identically distributed observations is well developed. One popular approach is to model blockwise maxima using the generalized extreme value distribution (Fisher & Tippett, 1928; Gnedenko, 1943). Alternative approaches include the well known *peaks over threshold* (POT) models (Pickands, 1975; Davison & Smith, 1990), and the point process approach (Pickands, 1971). Under the former, the exceedances over a given threshold are modeled using a generalized Pareto distribution. The latter focuses on jointly modeling the exceedances and the time of their occurrence using a

non-homogeneous Poisson process (Pickands, 1971; Smith, 1989; Coles & Tawn, 1996).

In this section we introduce the three approaches mentioned above. Our purpose is to review the relevant background materials that will underlie the work presented in this dissertation. For a detailed review of statistical models for univariate extremes see, for example, Embrechts et al. (1997), Kotz & Nadarajah (2000a) and Coles (2001).

2.1.1 Block Maxima Models

As a classical approach in extreme value theory, the *block maxima* models focus on modeling the blockwise maximum. Let $X_1, \ldots, X_r$ be a sequence of independent and identically distributed random variables with common distribution function F_0, and let $M_r = \max(X_1, \ldots, X_r)$. The distribution of M_r is related to $F_0(x)$ through $\Pr(M_r \leq x) = \{F_0(x)\}^r$. However, since the exact distribution function F_0 is usually unknown, we seek approximate families of models for $F_0(x)^r$. Note that, if $F_0(x) < 1$, the limit of $F_0(x)^r$ converges to 0 as $r \to \infty$, which implies that F_0 is a degenerating function. Therefore, the limit of M_r will consequently degenerate to the upper end point of F_0. Such degeneracy can be avoided by considering a linear renormalization $(M_r - b_r)/a_r$ with sequences of normalizing constants $\{a_r > 0\}$ and $\{b_r\}$.

The extreme types theorem of Fisher & Tippett (1928) and Gnedenko (1943) indicates that if there exist sequences of constants $a_r > 0$ and $b_r \in \mathbb{R}$, such that

$$\Pr\{(M_r - b_r)/a_r \leq x\} \to K(x) \qquad \text{as } r \to \infty, \tag{2.1}$$

where K is a non-degenerate distribution function, then K belongs to one of the three families

of the distribution:

$$\text{Gumbel}: \quad K(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, \qquad -\infty < x < \infty;$$

$$\text{Fréchet}: \quad K(x) = \begin{cases} 0, & x \le b, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & x > b; \end{cases}$$

$$\text{Weibull}: \quad K(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-b}{a}\right)\right]^{\alpha}\right\}, & x < b, \\ 1, & x \ge b, \end{cases}$$

for some $a > 0$, $b \in \mathbb{R}$, and $\alpha > 0$. These three types of distributions have distinct characterizations with regard to their tail behaviors. The Gumbel distribution is obtained when F_0 has exponential-type tails such as the normal or exponential distributions. The Fréchet distribution arises from distributions with heavier polynomial-type tails such as the student-t or Pareto distributions. The Weibull distribution arises when F_0 has finite support as is the case with the uniform or beta distributions. The theorem implies that if the sequences of a_r and b_r exist, then the limit distribution of the linear transformation $(M_r - b_r)/a_r$ must belong to one of the three possible extreme value distributions, that is the three distributions provide the only possible limits for $(M_r - b_r)/a_r$ no matter what the underlying distribution F_0 is.

A compact way to express K is to reformulate it by encompassing all three types of the extreme value families into a single family of distribution function (Von Mises, 1954; Jekinson, 1955), known as the generalized extreme value (GEV) distribution,

$$K(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{2.2}$$

where $z_+ = \max(z, 0)$, and μ, σ, and ξ are location, scale, and shape parameters, respectively. The scale parameter σ is restricted to be positive, and μ, ξ can be arbitrary real numbers. The

shape parameter ξ is determined by the tail behavior of F. In particular, the case $\xi > 0$ corresponds to the Fréchet family, $\xi < 0$ corresponds to the Weibull distribution, and the limit $\xi \to 0$ refers to the Gumbel distribution. Hence, each of the three types of distributions can be treated as a special case of the GEV family.

Traditionally, one of the goals of extreme value theory is to estimate the so-called return level function; see, for example, Coles (2001). The m-th year *return level* q_m corresponds to the level which one expects the annual maximum to be exceeded only once in every m-year periods, or equivalently, the level with probability $1/m$ of being exceeded by the annual maximum in a given year. Such quantity is defined as the $1-1/m$ quantile of the annual maximum distribution, which under the GEV model reduces to,

$$q_m = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-1/m)\}^{-\xi}\right], & \xi \neq 0, \\ \mu - \sigma \log\{-\log(1-1/m)\}, & \xi = 0. \end{cases} \tag{2.3}$$

The generalized extreme value distribution is usually fitted to blockwise maxima (for example, daily maxima of hourly data, or monthly maxima of daily data). However, modeling the block maximum series directly may be wasteful in practice, because it uses only the value of blockwise maxima and ignores most of the data in each block. For instance, if we model the annual maxima of daily rainfall, we discard the other 364 observations. Moreover, multiple extreme events may occur within a single block rather than other blocks. Discarding these values leads to loss of valuable information.

2.1.2 Peaks Over Threshold Models

An alternative to the GEV approach referred as *peaks over threshold* (POT) approach (Smith, 1984; Davison & Smith, 1990), consists of modeling the distribution of exceedances over a given threshold u rather than the block maximum. In the case when observations are i.i.d. from F_0, the conditional distribution of the exceedances over a threshold u is related to F_0,

$$\Pr\{X \leq u+y \mid X > u\} = \frac{F_0(u+y)}{1-F_0(u)}, \qquad y > u. \tag{2.4}$$

Balkema & de Haan (1974) and Pickands (1975) show that for large enough threshold u, the limit of the above conditional distribution converges in distribution to the generalized Pareto distribution,

$$\Pr\{X \leq u+y \mid X > u\} = 1 - \left\{1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right\}_+^{-1/\xi}, \tag{2.5}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$. Moreover, Pickands (1975) shows that the threshold exceedances have limiting distribution within the GPD family if and only if the underlying distribution F_0 for block maxima are in the domain of attraction of GEV family, with the same shape parameter ξ. Therefore, as in the GEV family, $\xi < 0$ corresponds to a bounded distribution; $\xi > 0$ corresponds to a heavy-tailed distribution; and $\xi \to 0$ corresponds to a light-tailed distribution.

Choice of the threshold is a classical bias-variance trade-off problem. Choosing a sufficiently high threshold will cause small sample of extreme data points. Conversely, setting the threshold too low will cause the asymptotic result to be invalid. The threshold selection is usually based on an extended explorative statistical analysis, see Coles (2001) and Embrechts et al. (1997).

2.1.3 Point Process Characterizations

The point process approach introduced by Pickands (1971) and Smith (1989) represents a third alternative to model extremes. Before providing details of the point process approach, we take a brief detour to define the non-homogeneous Poisson process. The non-homogeneous Poisson process, over a space $\mathcal{A}$, with intensity function $\lambda(\cdot)$, which is a non-negative and local integrable function, is defined as a counting process with independent increments that satisfies two properties. First, for any subset $A \subset \mathcal{A}$, the number of points in set A, $N(A)$, satisfies

$$N(A) \sim \mathsf{Poi}(\Lambda(A))$$

where $\Lambda(A) = \int_A \lambda(t)\mathrm{d}t$ is the expected number of points in A. Second, for any pair non-overlapping sets $A, B \subset \mathcal{A}$, $N(A)$ and $N(B)$ are independent random variables. Then, the likelihood function resulting from any NHPP generated point pattern $\{t_i : i = 1, \ldots, n\} \subset \mathcal{A}$, where $0 \leq t_1 < t_2 < \ldots < t_n \leq T$, can be expressed as

$$L(\lambda(\cdot); \text{Data}) \propto \exp\left\{-\int_{\mathcal{A}} \lambda(t)\mathrm{d}t\right\} \prod_{i=1}^{n} \lambda(t_i). \tag{2.6}$$

Since the intensity is locally integrable, that is, $\int_A \lambda(t)\mathrm{d}t < \infty$, for all bounded $A \subset \mathcal{A}$, a NHPP intensity function can be represented through a density function and a parameter that defines the total intensity over the observation window. Specifically, $\lambda(\cdot) = \gamma f(\cdot)$, where $\gamma \equiv \Lambda(\mathcal{A}) = \int_{\mathcal{A}} \lambda(t)\mathrm{d}t$ is the total integrated intensity, and $f(\cdot)$ is a density function on $\mathcal{A}$ that fully controls the shape of the intensity function.

Consider again a sequence $X_1, \ldots, X_r$ of independent random distribution with common distribution F_0. To jointly model the time and value of exceedances as a point process

(e.g., Joe et al., 1992), that is, for regularly spaced observations, we consider the ordered pairs $\{(j, X_j) : j = 1, \ldots, r\}$, where the first entry denotes the period over which each observation is collected. If we restrict attention to those observations that fall above a given threshold u, then our sample is thinned to the pairs $\{(Z_i, Y_i) : i = 1, \ldots, n\}$, where $n \leq r$, Y_i is the value of the i-th exceedance and Z_i is the time at which the i-th exceedance occurred. The pairs $\{(Z_i, Y_i) : i = 1, \ldots, n\}$ can be regarded as arising from a two-dimensional point process $\{N(A) : A \subset \mathcal{A} = \{1, \ldots, r\} \times [u, \infty)\}$. Pickands (1971) showed that the limiting form of this point process as $u \to \infty$ is a bivariate non-homogeneous Poisson process with intensity function

$$\frac{1}{\sigma}\left\{1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right\}_+^{-1/\xi-1} \tag{2.7}$$

where the parameters $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$. As before the value of the shape parameter ξ controls the tail behavior of F_0. Figure 2.1 shows a two-dimensional point process including the time and the value of exceedances over a high threshold u, for the set $A_y = [t, t + \Delta t] \times [y, \infty)$, $y > u$.

The shape parameter ξ in Equation (2.7) shares the same properties as in the GEV and GPD family. Note that this common property implies that there are connections between these three approaches of extreme value theory (see, Coles, 2001). For example, the GEV and GPD distribution can both be derived from the Poisson process approach. In particular, the probability that the blockwise maximum is less than a specified level y is equivalent to the probability that there are no events occurring in the region $A_y = [t, t + \Delta t] \times [y, \infty)$. Hence, the distribution of the blockwise maximum is given by

$$P(y) = \Pr(\max\{X_1, \ldots, X_r\} \leq y) = \exp\left\{-\Delta t\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{2.8}$$

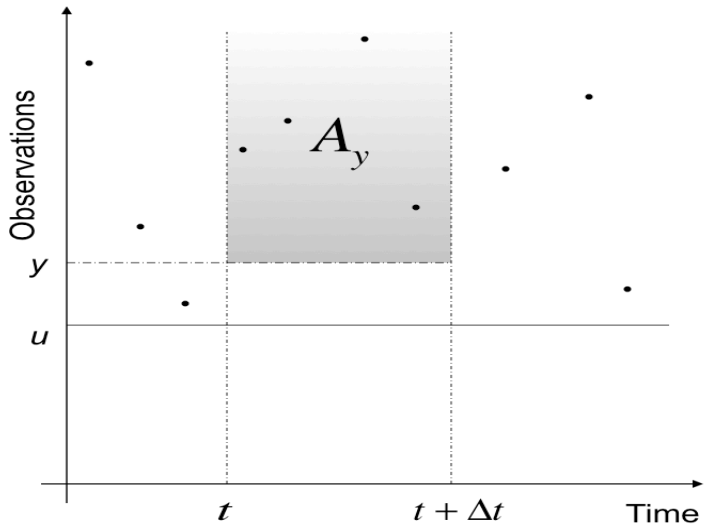


Figure 2.1: Illustration of two-dimensional point process including times of excesses and exceedances over a specified threshold u.

which corresponds to the generalized extreme value distribution in (2.2). Similarly, we can compute the conditional distribution of the exceedances over the threshold u as the generalized Pareto distribution. Recall that $\Lambda(A) = \int_A \lambda(t,y)\mathrm{d}t\mathrm{d}y$, thus, for the region $A_y = [t, t+\Delta t] \times [y, \infty)$, $\Lambda(A_y) = \Delta t[1+\xi(y-\mu)/\sigma]_+^{-1/\xi}$. The conditional probability of $X > u+y$ given $X > u$ is

$$\frac{\Lambda(A_{u+y})}{\Lambda(A_u)} = \frac{\Delta t[1+\xi(\frac{u+y-\mu}{\sigma})]_+^{-1/\xi}}{\Delta t[1+\xi(\frac{u-\mu}{\sigma})]_+^{-1/\xi}} = \left\{1+\frac{\xi y}{\sigma+\xi(u-\mu)}\right\}_+^{-1/\xi},$$

which leads to Equation (2.5).

Statistical inference under the Poisson process approach is straightforward. We could obtain the MLE of parameters (μ, σ, ξ) of the intensity function $\lambda(\cdot)$ by maximizing the likelihood function in Equation (2.6). Alternatively, Bayesian approaches can be used by combining the likelihood function with appropriate priors for $\{\mu, \sigma, \xi\}$. An advantage of Bayesian methods is that they allow us to include additional sources of information, such as expert knowledge.

This is practically important in extreme value analysis because data is sparse. Another reason is that regularity assumptions are not required in the Bayesian setting as heavily as in MLE estimation. Indeed, Smith (1985) shows that the asymptotic theory of maximum likelihood estimation is non-regular when $\xi < -0.5$. Bayesian inference for this class of models, including elicitation of informative priors from experts, is discussed in Coles & Tawn (1996).

2.2 Bayesian Nonparametric Methods

In Bayesian statistics, a nonparametric model refers to a probability model in which the indexing parameter is infinite-dimensional. In contrast to the traditional parametric methods with finite number of parameters and some distributional assumptions, Bayesian nonparametric models lead to a flexible framework for modeling complex structured data. For a detailed review of Bayesian nonparametric modeling see, for example, Müller & Quintana (2004).

Finite mixture models (e.g. the finite Gaussian mixtures) are widely used for the inference in complex datasets. However, one major limitation in this approach is that one needs to specify the number of mixture components a *priori*. The Bayesian nonparametric mixture models provide an alternative to avoid this restriction by placing a rich class of priors on the random mixing distribution. In particular, the Dirichlet process (DP) prior (Ferguson, 1973) is the most popular prior used in this setting. This chapter briefly reviews the Bayesian nonparametric models implemented throughout this dissertation, with an emphasis on the Dirichlet process.

2.2.1 The Dirichlet Process

The Dirichlet process (DP), originally developed by Ferguson (1973), has been widely applied as a prior for random distributions in Bayesian nonparametric models. Formally, the DP is defined as a stochastic process whose realizations are random probability measures on some probability space Θ. Hence, it is a distribution on the spaces of all possible distributions.

A Dirichlet process is characterized by two parameters: a baseline measure G_0, and a positive precision parameter α. Let G and G_0 be probability measures over Θ. The random probability measure G is distributed according to the Dirichlet process, denoted by $G \sim \mathsf{DP}(\alpha, G_0)$ if, for any finite and disjoint partition $A_1, A_2, \ldots, A_k$ of Θ, the vector $(G(A_1), G(A_2), \ldots, G(A_k))$ is Dirichlet distributed, that is,

$$(G(A_1), \ldots, G(A_k)) \sim \mathsf{Dir}(\alpha G_0(A_1), \ldots, \alpha G_0(A_k)) \tag{2.9}$$

Hence, analogously to the Gaussian process (whose marginal distributions are Gaussian), any finite measurable partition of the Dirichlet process follows a Dirichlet distribution. The base distribution G_0 specifies the center of the Dirichlet process, in the sense that for any measurable set $A \subseteq \Theta$,

$$\mathsf{E}(G(A)) = G_0(A).$$

Moreover, the precision parameter α controls how close the realization G is to G_0, as

$$\mathsf{Var}(G(A)) = \frac{G_0(A)(1 - G_0(A))}{\alpha + 1},$$

larger value of α results in small variability in Dirichlet process realizations. Figure 2.2 shows the simulated sample path from a Dirichlet process with a standard normal base distribution, and different values of α.

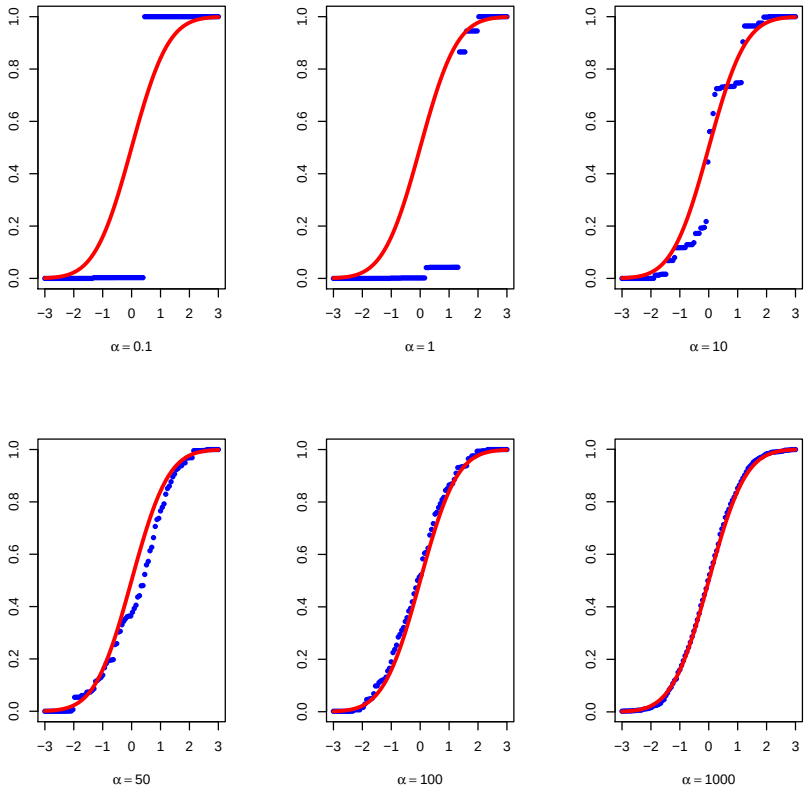


Figure 2.2: Cdf sample path from a DP(α, $G_0 = \mathrm{N}(0,1)$) prior, for different values of α. The solid red line denotes the cdf of G_0.

Another distinctive feature of the Dirichlet process is the conjugacy. Assuming $\theta_1, \ldots, \theta_n$ is a sequence of independent and identically distributed random samples from G, and $G \sim \mathrm{DP}(\alpha, G_0)$. Then,

$$G|\theta_1, \ldots, \theta_n \sim \mathrm{DP}\left(\alpha + n, \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^{n} \delta_{\theta_i}\right),$$

where δ_θ is the Dirac point mass centered at θ. Smaller value of α results in placing less weight on the prior base distribution, whereas larger value of α leads to placing less weight on the empirical distribution. On the other hand, as the sample size $n \to \infty$, the posterior process con-

verges to the empirical distribution of the sample, which suggests that the posterior associated with a DP prior is consistent.

Now we seek for the expression of the predictive distribution of a new observation θ_{n+1} given the observations $\theta_1, \ldots, \theta_n$. Consider again $\theta_1, \ldots, \theta_n$ is a sequence of i.i.d. samples from G, and $G \sim \mathsf{DP}(\alpha, G_0)$. As in standard Bayesian methods, the predictive distribution of $\theta_{n+1} \mid \theta_1, \ldots, \theta_n$ is obtained by integrating out the random distribution G, that is,

$$\theta_{n+1} \mid \theta_1, \ldots, \theta_n \sim \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{\theta_i} \tag{2.10}$$

Pólya Urn Scheme

The resulting Equation (2.10) can be interpreted via a popular metaphor, known as Pólya urn scheme, described by Blackwell & MacQueen (1973). The scheme can be visualized as the following process. Consider an urn with α blue balls initially. Then, we randomly draw a ball from the urn at each time. If the ball drawn is blue, we return the ball to the urn and put a new ball with a new color to the urn; if a non-blue ball is drawn, then we return the ball and add a new ball with the same color to the urn. Conceptually, whenever a blue ball is drawn from the urn, it means that we sample a new cluster from the base distribution $G_0(\cdot)$ with the probability proportional to α. The Pólya urn scheme represents the procedure to sample parameters from G, though G is not observable. On the other hand, the scheme also implies a clustering property of the Dirichlet process, which implicitly partition n objects into n^* clusters. Let $\theta_1^*, \ldots, \theta_{n^*}^*$ be the distinct values of $\theta_i, i = 1, \ldots, n$, and n_k be the size of k-th cluster, i.e.,

$n_k = |\{i : \theta_i = \theta_k^*\}|, k = 1, \ldots n^*$. Then, the Equation (2.10) can be rewritten as

$$\theta_{n+1} \mid \theta_1, \ldots, \theta_n \sim \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{k=1}^{n^*} n_k \delta_{\theta_k^*}$$

Note that the probability of drawing a non-blue color ball is proportional to the number of the existing certain color balls in the urn. As a result, the more often a non-blue certain color is drawn, the higher probability of being drawn again in the future.

Chinese Restaurant Process

As the n observations can be specified by n^* distinct values, the Dirichlet process implicitly partitions the data. The induced distribution of partitions is described by a different metaphor, called the Chinese restaurant process. Imagine we have a Chinese restaurant with infinite set of tables, with each table serving only one dish. Customers enter the restaurant sequentially and choose a table to sit. The first customer arrives and sit at the first table. The second customer arrives and decides either to join an occupied table k with probability proportional to the number of customers at that table n_k, or sit at a new table with probability proportional to α. Here, the infinite number of tables correspond to clusters, and customers correspond to observations. More specifically, each θ_i is analogous to a customer and the occupied table at which the customer sits is equivalent to the distinct value θ_k^*. When the customer sits at a new table, he/she samples a dish from G_0, otherwise, he/she sits at an occupied table and shares the dish with other diners sitting there. Importantly, the Chinese restaurant process defines not only a distribution over partitions, but also an exchangeable distribution over permutations. Indeed, the probability distribution is invariant to the order in which customers are assigned to tables.

Stick-breaking Construction

The stick-breaking construction, first proposed by Sethuraman (1994), provided an explicit constructive definition of the Dirichlet process, and leads to an effective MCMC algorithms for the Dirichlet process mixture models (see section 2.2.2). A key aspect of the stick-breaking construction is that the random measure G can be shown as an infinite weighted sum of point masses. Specifically, a realization G from $\mathsf{DP}(\alpha, G_0)$ has an almost sure representation of the form,

$$G = \sum_{l=1}^{\infty} \omega_l \delta_{\theta_l} \tag{2.11}$$

where $\{\theta_1, \theta_2, \ldots\}$ are independent and identically distributed samples from the base distribution G_0.

The associate weights $\{\omega_1, \omega_2, \ldots\}$ are generated from the stick-breaking process, which is $\omega_1 = \nu_1$, and for $l > 2$, $\omega_l = \nu_l \prod_{s<l}(1-\nu_s)$ with $\{\nu_1, \nu_2, \ldots\}$ another independent and identically distributed samples from a $\mathsf{Beta}(1, \alpha)$ distribution. Figure 2.3 shows the analogy of the stick-breaking construction. Starting with a stick of a unit length, we sequentially partition the stick into infinite pieces with associated weights. We first break the stick into a portion of the stick with length ν_1, and the remaining of the stick with length $1-\nu_1$. Then, the length of the second break is determined by a random weight ν_2 of the remaining length $(1-\nu_1)$. We recursively break the stick using the same mechanism until the stick is broken into infinite partitions.

Equation (2.11) shows that samples drawn from the Dirichlet process are discrete with probability one. Moreover, the stick-breaking construction provides the basis for the definitions of dependent random measures, i.e. dependent Dirichlet process (MacEachern, 1999, 2000).

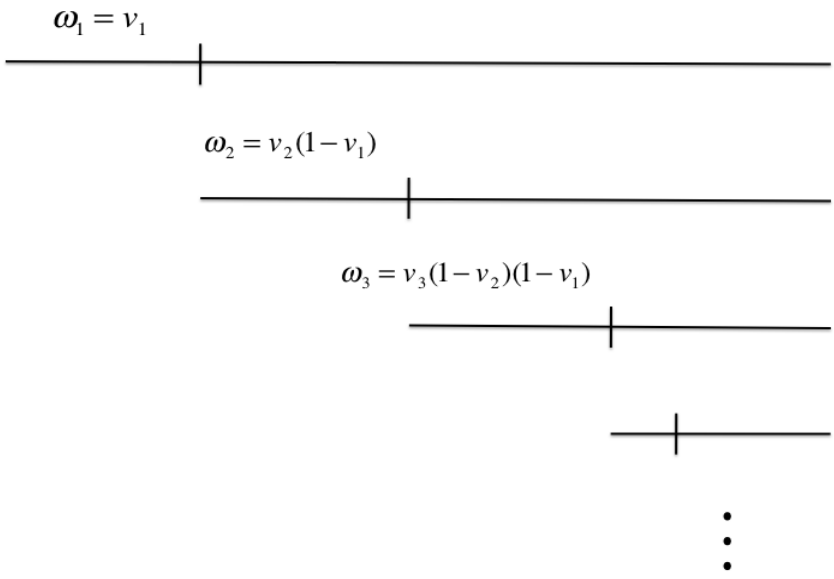


Figure 2.3: Stick-breaking process

2.2.2 Dirichlet Process Mixture Models

Since the random measure G is composed as a countable mixture of weighted sums, multiple observations from G are likely to take identical values. Therefore, the Dirichlet process is not appropriate to directly model the continuous observations. A natural alternative usage of the Dirichlet process is as a prior over the random distribution G of mixture components, resulting in a Dirichlet process mixture model (Antoniak, 1974; Lo, 1984; Escobar, 1994; Escobar & West, 1995). In particular, the Dirichlet process mixture density (or probability mass) function can be expressed as $f(\cdot;G) = \int k(\cdot|\theta)\mathrm{d}G(\theta)$, where $k(\cdot|\theta)$ is a parametric density function indexed by parameter vector θ, and G is a random mixing distribution. Due to the the discreteness of G, the model partitions the observations into independent clusters with observations assigned

to each cluster, assumed to be distributed according to k. In such case, the DP mixture model allows not only clustering when the number of clusters is unknown, but also generating flexible shapes for the mixture density function $f(\cdot;G)$. Therefore, the DP mixture model can be used in the context of nonparametric mixing for applications where clustering of the observations is practically relevant as in, e.g., density estimation, classification, and regression. Recall that for the DP(α, G_0) prior, the precision parameter α controls how close the realizations G is to G_0. In the DP mixture model, α controls the number of distinct mixture components $n^*(n^* < n)$ (Antoniak, 1974; Escobar & West, 1995). For instance, for density estimation problems with moderately large sample sizes, n, a useful approximation to the prior expectation for the number of clusters is given by $\alpha\log\{(\alpha+n)/\alpha\}$.

There are multiple computational strategies developed for model inference in the Dirichlet process mixture models. One commonly used approach exploits the Pólya urn representation to avoid dealing with the infinite dimensional G (Escobar & West, 1995; Bush & MacEachern, 1996; MacEachern & Müller, 1998; Neal, 2000). In this approach, the mixing component is updated one at a time, depending on the most recently sampled values of other components. A drawback of this sampler is that it is difficult when the base measure G_0 is not conjugate to the kernel density $k(\cdot|\theta)$, as the approach involves integrating over the mixing components. Alternatively, we use the *block Gibbs sampler*, based on the truncated stick-breaking representation of the DP (Ishwaran & Zarepour, 2000; Ishwaran & James, 2001). Ishwaran & James (2001) shows that the truncation method closely approximates the infinite-dimensional Dirichlet process by a finite mixture model with a finite truncation level N, when N is chosen large enough. Typically, N is chosen using standard DP properties, such that $(\alpha/(\alpha+1))^N = \varepsilon$,

for small ε (Ishwaran & Zarepour, 2000). The truncated approximation of G is given by,

$$G^N(\cdot) = \sum_{l=1}^{N} p_l \delta_{\vartheta_l}(\cdot) \tag{2.12}$$

where ϑ_l are independent realizations from the base distribution G_0, and $\mathbf{p} = \{p_l, l = 1, \ldots, N\}$ are the associated weights defined using the stick-breaking construction subject to the constraint $p_N = 1 - \sum_{l=1}^{N-1} p_l$. The joint distribution for the random weights $\mathbf{p}$, given α, is defined as

$$f(\mathbf{p}|\alpha) = \alpha^{N-1} p_N^{\alpha-1} (1-p_1)^{-1} (1-(p_1+p_2))^{-1} \times \ldots \times (1 - \sum_{l=1}^{N-2} p_l)^{-1} \tag{2.13}$$

which is a special case of the generalized Dirichlet distribution (Connor & Mosimann, 1969).

Chapter 3

Nonparametric Mixture Modeling for Extreme Value Analysis

3.1 Introduction

As we discussed in Section 2.1, the literature on univariate extreme values associated with independent and identically distributed observations is extensive. In contrast, the literature on modeling extremes generated from more general stochastic processes evolving over time and/or space is less well developed. In this work, we focus on modeling the extremes of inhomogeneous temporal processes. In particular, we consider an extension of the point process approach discussed in Section 2.1.3 that allows for a general structure for the intensity function of the underlying Poisson point process. More specifically, we use nonparametric mixtures of bivariate kernels to model the intensity function associated with the times and values of the exceedances over a given threshold. A related approach was discussed in Kottas & Sansó (2007)

where mixtures of bivariate beta kernels were used to model the intensity function of the point process. Here, we provide a more scientifically relevant modeling framework for extremes by considering alternative types of mixtures for the Poisson process intensity which ensure that the marginal distributions of the underlying process belong to the Fréchet domain of attraction. Moreover, we develop inferences for important extreme value analysis functionals, including different types of return level functions. Our approach shares some similarities with that of Coles et al. (1994), who extends the point process approach to allow for temporal dependence across locations. Nonparametric mixture models have also been applied to create more flexible models for the extremes of independent and identically distributed observations by using mixtures of Pareto distributions to model the exceedances over a given threshold (Tressou, 2008).

The rest of this chapter is organized as follows. In Section 3.2, we present the proposed nonparametric mixture modeling approach, including discussion of details of the model formulation, theoretical results, and definition of return level functions. Section 3.3 provides details on model implementation, including methods for prior specification, posterior simulation and inference. In Section 3.4, we illustrate the methodology using a simulated data example and a real data set of daily returns of the Dow Jones index over a five year period. Finally, Section 5.5 concludes with a summary and discussion of possible extensions.

3.2 Nonparametric Point Process Modeling for Analysis of Extremes

3.2.1 Background and Motivation

The work presented in this chapter is concerned with extending the point process approach described in Section 2.1.3 for modeling the tails of a general stochastic process $\{X_t : t \in [0,T]\}$. We denote by F_t the marginal distribution function for X_t, which is related to the conditional distribution of the exceedances over a threshold u at time t through $\Pr(X_t \leq y \mid X_t > u) = \{F_t(y) - F_t(u)\}/\{1 - F_t(u)\}$, for $y \geq u$. Moreover, recall that the pairwise observations $\{(j, X_j) : j = 1, \ldots, r\}$ defined in Section 2.1.3, where the first entry denotes the period over which each observation is collected. If we restrict attention to those observations that fall above a given threshold u, then our sample is thinned to the pairs $\{(Z_i, Y_i) : i = 1, \ldots, n\}$, where Y_i is the value of the i-th exceedance and Z_i is the time at which the i-th exceedance occurred. In our approach, the pairs $\{(Z_i, Y_i) : i = 1, \ldots, n\}$ are again treated as a realization from a non-homogeneous Poisson process, here, on $\mathcal{A} = [0,T] \times [u, \infty)$, with intensity function $\lambda(t,y)$, so that $N(A) \sim \text{Poi}\{\int_A \lambda(t,y)\mathrm{d}t\mathrm{d}y\}$ for any measurable set $A \subset \mathcal{A}$.

We focus on modeling the intensity function of extremes to provide more flexible inference than the limiting parametric intensity in Equation (2.7), which, for instance, is restricted by time homogeneity. To this end, we formulate a mixture model for $\lambda(t,y)$ by exploiting the connection of the Poisson process intensity with a density function. We build on a modeling approach originally developed in Kottas & Sansó (2007), which has also been applied to analysis of immunological studies (Ji et al., 2009) and neuronal data analysis (Kottas & Behseta, 2010; Kottas et al., 2012). Our objective is to develop a flexible inferential framework for ex-

treme value analysis. Hence, in contrast to this earlier work, we seek more structured modeling for the kernel in the mixture representation for $\lambda(t,y)$ to achieve a balance between desirable theoretical properties for the tail behavior of the marginal distribution F_t, and general inference for key extreme value analysis functionals.

3.2.2 The Modeling Approach

To generate a flexible model for extreme value analysis under the point process approach, we aim at estimating nonparametrically the intensity function, $\lambda(t,y)$, over time and exceedance values. The key observation underlying our modeling approach is that the Poisson process intensity function can be decomposed as $\lambda(\cdot) = \gamma f(\cdot)$, where $\gamma \equiv \Lambda(\mathcal{A}) = \int_{\mathcal{A}} \lambda(t,y)\mathrm{d}t\mathrm{d}y$ is the total intensity of exceedances, and $f(\cdot) = \lambda(\cdot)/\Lambda(\mathcal{A})$ is a density function on $\mathcal{A}$, which fully controls the shape of the intensity function. The implicit assumption is that $\Lambda(\mathcal{A}) < \infty$, which can be justified by noting that the Poisson process definition implies that $\exp\{-\Lambda(\mathcal{A})\} = \Pr(\{X_t < u : t \in [0,T]\})$. Hence, provided the threshold u and the underlying stochastic process are such that $\Pr(\{X_t < u : t \in [0,T]\}) > 0$, the previous identity implies that $\Lambda(\mathcal{A}) < \infty$. For example, this condition is satisfied if $\{X_t : t \in [0,T]\}$ is a Brownian motion; in general, continuous sample paths would likely be needed if one seeks more specific conditions on the underlying stochastic process such that $\Pr(\{X_t < u : t \in [0,T]\}) > 0$ holds true.

Under this formulation for the intensity of extremes, we can express the Poisson process likelihood function as

$$L(\gamma, f(\cdot); \{(t_i, y_i) : i = 1, \ldots, n\}) \propto \exp(-\gamma)\gamma^n \prod_{i=1}^{n} f(t_i, y_i). \tag{3.1}$$

Hence, the problem of estimating the intensity function for the point process of exceedances can be broken down into two independent problems, namely, estimating the total intensity of the Poisson process, and estimating the probability density associated with the distribution of exceedances over the region $\mathcal{A}$.

To generate a rich prior for the Poisson process density, we consider a nonparametric mixture model,

$$f(t,y) \equiv f(t,y;G) = \int k(t,y \mid \theta) \mathrm{d}G(\theta), \tag{3.2}$$

where $k(t,y \mid \theta)$ is a parametric density on $\mathcal{A}$ indexed by parameter vector θ, and G is a random mixing distribution. As reviewed in Section 2.2, placing a Dirichlet process prior $\mathsf{DP}(\alpha, G_0)$ on G results in a Dirichlet process mixture model for $f(t,y;G)$. To study model properties as well as for posterior simulation, we make use of the Dirichlet process stick-breaking definition introduced in Section 2.2.1. This specification leads to the following mixture model for the intensity of extremes,

$$\lambda(t,y) \equiv \lambda(t,y;G,\gamma) = \gamma\, f(t,y;G) = \gamma \int k(t,y \mid \theta) \mathrm{d}G(\theta), \qquad G \mid \alpha, \psi \sim \mathsf{DP}(\alpha, G_0), \tag{3.3}$$

where ψ collects the parameters of the centering distribution G_0; as discussed in Section 3.3, the full Bayesian model involves priors for hyperparameters α and ψ.

Since the purpose of studying extreme values is often to extrapolate the tail behavior of the distribution beyond the observed range of exceedances, and accurate extrapolation in this setting heavily depends on properties of the tail of the density $f(t,y)$, the choice of the mixture kernel $k(t,y \mid \theta)$ is a critical aspect of the model formulation. Indeed, note that, unlike other applications of Poisson processes to spatial modeling, in this problem the nature of the

argument of the dimensions associated with $\mathcal{A} = [0,T] \times [u,\infty)$ is very different. Hence, in specifying the mixture kernel density for model (3.3), we consider a product form

$$k(t,y \mid \theta) \equiv k(t,y \mid \theta_1,\theta_2) = k_1(t \mid \theta_1)k_2(y \mid \theta_2), \tag{3.4}$$

that is, kernel components k_1 and k_2 are independent before mixing. However, after mixing over the random G, dependence is induced to the resulting mixture model $f(t,y;G)$ for the bivariate Poisson process density.

A key objective of our modeling approach is to remove the restriction of time homogeneity implied by (2.7), and thus for the intensity in the time direction, we seek as general a specification as possible. Because of its flexibility, a (rescaled) beta distribution emerges as a natural choice for the kernel component over time,

$$k_1(t \mid \theta_1) = \frac{1}{T}\frac{\Gamma(\tau)}{\Gamma(\kappa\tau/T)\Gamma(\{1-\kappa/T\}\tau)}\left(\frac{t}{T}\right)^{\kappa\tau/T-1}\left(1-\frac{t}{T}\right)^{\{1-\kappa/T\}\tau-1}, \quad t \in (0,T) \tag{3.5}$$

where $\theta_1 = (\kappa,\tau)$, $\kappa \in (0,T)$ is the mean of the beta distribution, and $\tau > 0$ is a scale parameter.

The choice of kernel component $k_2(y \mid \theta_2)$ is more delicate. The previous attempt in Kottas & Sansó (2007) to nonparametric mixture modeling for extremes under the point process approach utilized kernels with bounded support, defined by a bivariate beta distribution, and thus the underlying F_t was implicitly assumed to have compact support. This is restrictive for many applications. Moreover, kernel k_2 is used to capture through mixing the tail behavior of the underlying distribution where we do not expect, for instance, multimodalities. Hence, applying the mixture model with a beta density for k_2 may lead to overfitting, which is especially damaging for extrapolation. In contrast, the asymptotic theory for extremes suggests what the

tail behavior is, and we can make use of that information to improve inference under the mixture model.

Hence, we build the intensity function in the exceedances direction from a kernel defined through a special case of the generalized Pareto distribution,

$$k_2(y \mid \theta_2) = \frac{1}{\sigma}\left\{1+\frac{\xi(y-u)}{\sigma}\right\}^{-1/\xi-1}, \quad y \geq u \tag{3.6}$$

such that $\theta_2 = (\sigma, \xi)$ with $\sigma > 0$ and $\xi > 0$. The location parameter is set to the specified threshold value u to ensure that the resulting mixture kernel $k(t, y \mid \theta_1, \theta_2)$ has support on $\mathcal{A}$. Moreover, we focus on the $\xi > 0$ range for the shape parameter, which ensures that the corresponding marginal distributions belong to the Fréchet maximum domain of attraction, that is, we are modeling an underlying stochastic process with heavy tailed behavior; see specific theoretical result in Section 3.2.3. Specifically, the distribution for X_t is in the Fréchet domain of attraction if, for sufficiently large x, $\Pr(X_t > x) \approx Cx^{-\rho}L(x)$, where C and ρ are non-negative quantities, which are constants in x, and $L(x)$ is a slowly varying function, that is, $L(x)$ satisfies $\lim_{x\to\infty} L(sx)/L(x) = 1$, for all $s > 0$ (Embrechts et al., 1997). The tail index parameter ρ has a useful interpretation as a risk indicator – larger values of ρ^{-1} lead to larger probability of exceeding the specified level x – and its estimation has been considered extensively in the literature; see, for instance, the related discussion and references in Tressou (2008).

An appealing feature of the mixture model formulation in (3.3) – (3.6) is that the model for the intensity of extremes can be interpreted as accommodating time inhomogeneities through local adaptive fitting of generalized Pareto distributions, where the mode of the beta kernel associated with each distinct mixture component serves to localize the effect of the gen-

eralized Pareto kernel in time. This feature provides flexibility with respect to capturing non-standard underlying intensity shapes as they are suggested by the data. In addition, note that if $\alpha \to 0^+$, which results in a single mixture component, and if the beta kernel component is reduced to a uniform, we recover as a special case the parametric model for the intensity function in (2.7) with $\mu = u$.

Traditionally, one of the key goals of extreme value theory is to estimate the return level function of the process, which is strongly connected to the intensity function. In the case of non-homogeneous processes, we can define two different types of return level functions. For a given point $t_0 \in (0,T)$ and a small $\varepsilon > 0$, we define the ε-conditional return level curve as given by the solution to the Equation $\Pr(\{X_t > x_m : t \in [t_0 - \varepsilon, t_0 + \varepsilon]\}) = m^{-1}$, for different values of m. Under the nonparametric mixture model, for any $x > u$,

$$\Pr(\{X_t > x : t \in [t_0 - \varepsilon, t_0 + \varepsilon]\}; G, \gamma) =$$
$$1 - \exp\left[-\gamma \int \{1 - K_2(x \mid \sigma, \xi)\} \{K_1(t_0 + \varepsilon \mid \kappa, \tau) - K_1(t_0 - \varepsilon \mid \kappa, \tau)\} \, \mathrm{d}G(\kappa, \tau, \sigma, \xi)\right], \quad (3.7)$$

where K_1 and K_2 denote the distribution functions for the beta and generalized Pareto kernel components, respectively. Hence, the ε-conditional return level x_m at time t_0 corresponds approximately to a realization of the process that would be exceeded only once in every m periods if additional (imaginary) draws were to be taken according to the underlying F_{t_0}.

We can also define a marginal return level curve through the average intensity function $\tilde{\Lambda}([x,\infty)) = T^{-1} \int_0^T \int_x^\infty \lambda(t,y) \mathrm{d}t \mathrm{d}y$, for $x > u$. Proceeding as before, we define the marginal return level curve as the solution to $\Pr(\tilde{X} > x_m) = m^{-1}$, where

$$\Pr(\tilde{X} > x; G, \gamma) = 1 - \exp\left[-T^{-1}\gamma \int \{1 - K_2(x \mid \sigma, \xi)\} \, \mathrm{d}G(\sigma, \xi)\right], \quad (3.8)$$

and $\tilde{X}$ corresponds to the outcome associated with an "average" period. Hence, unlike the conditional return level curve, which provides information about the likelihood of extremes at a specific time point t_0, the marginal return level curve provides an average over all $t \in [0, T]$.

To understand the relationship between conditional and marginal return level curves, it is useful to compare equations (3.7) and (3.8) with the tail probability obtained by modeling $\lambda(t, y)$ using (2.7). Since in that case the intensity function is time homogeneous, the marginal and conditional tail probabilities agree for the traditional parametric model, and correspond to the one generated from (2.3). For non-homogeneous processes the marginal and conditional return level curves provide important and distinct insights into the behavior of the underlying stochastic process. While marginal return level curves can be used to assess what extremes look like on a "normal" period, the conditional return level curves can be used to examine specific past dates, providing insights about the behavior of the underlying process on a particularly "good" or "bad" period.

3.2.3 Theoretical Properties

As stated in Section 3.2.2, we focus on the case of the shape parameter ξ in (3.6) is positive, which indicates that the corresponding marginal distributions belong to the Fréchet maximum domain of attraction. The specific theoretical result under our modeling approach is formulated below as Theorem 1. Key to the proof of this theorem uses a result of independent interest that relates a representation of the tail probability for the underlying process marginals at any specific time point to the conditional Poisson process density at that time point. We therefore state this result first as the following lemma.

LEMMA 1. *Consider a stochastic process* $\{X_t : t \in [0,T]\}$ *with right-continuous sample paths, and the point process whose points comprise the time, t, and value, y, of exceedances of process* $\{X_t : t \in [0,T]\}$ *above a given threshold u. Assume a non-homogeneous Poisson model for the point process with intensity function* $\lambda(t,y) = \gamma f(t,y)$*, for* $(t,y) \in [0,T] \times [u,\infty)$*, where* $\gamma = \int_{\mathcal{A}} \lambda(t,y)dtdy$*. Then, for any specified time point* t_0*,*

$$Pr(X_{t_0} > x \mid X_{t_0} > u) = \int_x^\infty f(y \mid t_0)dy = \int_x^\infty f(t_0,y)dy/\{f(t_0)\}, \quad x > u.$$

Proof. Consider a generic $x > u$, where u is the given threshold, and fix a time point t_0. Using continuity of probability measure and the right-continuity of the sample paths of the underlying stochastic process $\{X_t : t \in [0,T]\}$, we have $\text{pr}(X_{t_0} > x) = \lim_{\Delta t \to 0} \text{pr}\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\}\right)$.

The same argument applies to $\text{pr}(X_{t_0} > u) = \lim_{\Delta t \to 0} \text{pr}\left(\{X_t > u : t \in [t_0, t_0 + \Delta t)\}\right)$, resulting in

$$\Pr(X_{t_0} > u) = \lim_{\Delta t \to 0} \left\{ 1 - \exp\left(- \int_u^\infty \int_{t_0}^{t_0 + \Delta t} \lambda(t,y) \mathrm{d}t \, \mathrm{d}y \right) \right\} \tag{3.9}$$

based on the Poisson assumption for the point process of exceedances.

Next, define M_0 as the number of exceedances in time interval $[t_0, t_0 + \Delta t)$. Then, we can write

$$\Pr\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\}\right) = \sum_{m \geq 1} \Pr\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\} \cap \{M_0 = m\}\right).$$

Using the Poisson process assumption, $\lim_{\Delta t \to 0} (\Delta t)^{-1} \Pr(M_0 = m) = 0$, for $m \geq 2$, resulting in $\Pr(X_{t_0} > x) = \lim_{\Delta t \to 0} \Pr\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\} \cap \{M_0 = 1\}\right)$. Based again on the Poisson

process structure,

$$\Pr\left(\{X_t > x : t \in [t_0, t_0+\Delta t)\} \cap \{M_0 = 1\}\right)$$

$$= \Pr(1 \text{ event in } [t_0, t_0+\Delta t) \times [x, \infty)) \times \Pr(0 \text{ events in } [t_0, t_0+\Delta t) \times [u, x))$$

$$= \left\{ \exp\left(-\int_x^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \left(\int_x^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \right\} \times \left\{ \exp\left(-\int_u^{x} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \right\},$$

and thus,

$$\Pr(X_{t_0} > x) = \lim_{\Delta t \to 0} \left(\int_x^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \exp\left(-\int_x^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right)$$
$$\exp\left(-\int_u^{x} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) = \lim_{\Delta t \to 0} \int_x^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y. \quad (3.10)$$

Combining Equation (3.9) and (3.10), we obtain for any $x > u$,

$$\Pr(X_{t_0} > x \mid X_{t_0} > u) = \lim_{\Delta t \to 0} \frac{\mathrm{pr}\left(\{X_t > x : t \in [t_0, t_0+\Delta t)\}\right)}{\mathrm{pr}\left(\{X_t > u : t \in [t_0, t_0+\Delta t)\}\right)}$$
$$= \frac{\lim_{\Delta t \to 0} (\Delta t)^{-1} \int_x^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y}{\lim_{\Delta t \to 0} (\Delta t)^{-1} \left\{ 1 - \exp\left(-\int_u^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \right\}}.$$

The limit in the numerator yields $\int_x^{\infty} \lambda(t_0, y) \mathrm{d}y$. For the denominator, we use a first-order Maclaurin series expansion of the function $g(\Delta t) = \exp\left(-\int_u^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right)$ to obtain $g(\Delta t) = g(0) + (\Delta t) g'(0) + R_2(\Delta t) = 1 - (\Delta t) \int_u^{\infty} \lambda(t_0, y) \mathrm{d}y + R_2(\Delta t)$, where $\lim_{\Delta t \to 0} (\Delta t)^{-1} R_2(\Delta t) = 0$. Hence,

$$\lim_{\Delta t \to 0} (\Delta t)^{-1} \left\{ 1 - \exp\left(-\int_u^{\infty} \int_{t_0}^{t_0+\Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \right\} = \int_u^{\infty} \lambda(t_0, y) \mathrm{d}y = \gamma f(t_0),$$

and thus finally,

$$\Pr(X_{t_0} > x \mid X_{t_0} > u) = \frac{\int_x^{\infty} \lambda(t_0, y) \mathrm{d}y}{\gamma f(t_0)} = \frac{\gamma \int_x^{\infty} f(t_0, y) \mathrm{d}y}{\gamma f(t_0)} = \int_x^{\infty} f(y \mid t_0) \mathrm{d}y.$$

□

THEOREM 1. *Assume a non-homogeneous Poisson process model on $\mathcal{A} = [0,T] \times [u,\infty)$ for the times and values of the exceedances, given a fixed threshold u, of a stochastic process $\{X_t : t \in [0,T]\}$ with right-continuous sample paths. Consider the mixture model defined by (3.3) – (3.6) for the Poisson process intensity function. Then, the marginal distributions of the process, $Pr(X_t > x)$, belong to the Fréchet maximum domain of attraction.*

Proof. Consider a sufficiently large, generic $x > u$, where u is the given threshold. We seek to prove that the marginal distributions of the underlying process satisfy $\Pr(X_t > x) \approx Cx^{-\rho}L(x)$, where ρ is the tail index parameter and $L(x)$ is a slowly varying function.

First, note that, since $1 - F_t(u)$ is a positive constant in x, $\Pr(X_t > x)$ belongs to the Fréchet maximum domain of attraction if and only if $\Pr(X_t > x \mid X_t > u) = \{1 - F_t(x)\}/\{1 - F_t(u)\}$ does, and importantly, both distributions have the same tail index parameter. Hence, it suffices to work with conditional distribution $\Pr(X_t > x \mid X_t > u)$ at a specific time point t.

To complete the proof, we employ the truncation approximation G^N to the Dirichlet process representation for mixing distribution G, which as discussed in Section 2.2.2, provides the version of the mixture model applied to the data. Then based on Lemma 1, we obtain

$$\begin{aligned}
\Pr(X_t > x \mid X_t > u) &= \int_x^\infty f(y \mid t)\mathrm{d}y \\
&= \frac{\sum_{l=1}^N p_l k_1(t \mid \kappa_l, \tau_l) \left\{1 + \sigma_l^{-1}\xi_l(x-u)\right\}^{-1/\xi_l}}{\sum_{l=1}^N p_l k_1(t \mid \kappa_l, \tau_l)} \\
&= \sum_{l=1}^N \omega_l (A_l + B_l x)^{-1/\xi_l}
\end{aligned}$$

where $\omega_l = p_l k_1(t|\kappa_l,\tau_l)/\{\sum_{l=1}^N p_l k_1(t|\kappa_l,\tau_l)\}$, $A_l = 1 - \sigma_l^{-1}\xi_l u$, and $B_l = \sigma_l^{-1}\xi_l$. Note that the weights ω_l depend on the specified time point t, but not on level x. Next, letting $\rho =$

$\min\{\xi_l^{-1} : l = 1,...,N\}$ and $l^* = \arg\min\{\xi_l^{-1} : l = 1,...,N\}$, we can write

$$\begin{aligned}\Pr(X_t > x \mid X_t > u) &= \sum_{l=1}^{N} \omega_l (A_l + B_l x)^{-1/\xi_l} \\ &= x^{-\rho} \sum_{l=1}^{N} \omega_l \left(\frac{A_l}{x} + B_l\right)^{-\frac{1}{\xi_l}} x^{-\frac{1}{\xi_l}+\rho} \\ &= x^{-\rho} L(x).\end{aligned}$$

Now,

$$\lim_{x\to\infty} \left(\frac{A_l}{x} + B_l\right)^{-\frac{1}{\xi_l}} x^{-\frac{1}{\xi_l}+\rho} = \begin{cases} B_l^{-1/\xi_l} & \text{if } l = l^* \\ 0 & \text{otherwise} \end{cases}$$

and therefore $\lim_{x\to\infty} L(sx)/L(x) = 1$, for any $s > 0$, which completes the argument. □

The practical utility of the lemma is that it enables time-dependent inference for tail probabilities of the marginal distributions of the underlying process – which is observed only through its exceedances above the given threshold – based on the nonparametric mixture model for the point process density. Note that the result of Lemma 1 is not specific to the particular modeling approach as its proof utilizes only the $\lambda(t,y) = \gamma f(t,y)$ formulation for the Poisson process intensity. However, the rest of the proof for Theorem 1 uses the mixture representation for the Poisson process density and the specific mixture kernel built from (3.5) and (3.6). In fact, the argument relies on a truncation approximation to the Dirichlet process representation, which is also used in the posterior simulation approach; see Section 3.3.1. The assumption of right-continuous sample paths for the underlying process is needed for the proof of the lemma. From a theoretical point of view, this assumption is not restrictive given the availability of results on existence of versions of stochastic processes with right-continuous sample paths.

3.3 Implementation Details

3.3.1 MCMC Posterior Simulation

Based on the form of the Poisson process likelihood in (3.1), the marginal posterior distribution for γ is analytically available as a Gamma distribution under a Gamma prior or the marginal reference prior, which is given by $p(\gamma) \propto \gamma^{-1} 1(\gamma > 0)$ (Kottas & Behseta, 2010). In particular, under the latter prior, the joint posterior distribution is proper, and $p(\gamma \mid \text{Data})$ is simply a $\mathsf{Gamma}(n, 1)$ distribution.

Inference for the Poisson process density requires the computation of the posterior distribution for the random mixing distribution G and the Dirichlet process prior hyperparameters. Full posterior inference under Dirichlet process mixture models can be obtained by using the truncated version of G,

$$G^N(\cdot) = \sum_{l=1}^{N} p_l \delta_{\zeta_l}(\cdot),$$

where the ζ_l are independent draws from the base distribution G_0 and $p_1, \ldots, p_N$ are the associated weights defined using a stick-breaking construction under the constraint $p_N = 1 - \sum_{l=1}^{N-1} p_l$. Introducing configuration variables $\mathbf{L} = (L_1, \cdots, L_n)$, where $L_i = l$ if and only if the mixing parameter corresponding to observation (t_i, y_i) is given by ζ_l, the hierarchical model for the data

is written as:

$$(t_i, y_i) \mid \kappa_{L_i}, \tau_{L_i}, \sigma_{L_i}, \xi_{L_i} \overset{ind.}{\sim} k_1(t_i \mid \kappa_{L_i}, \tau_{L_i}) k_2(y_i \mid \sigma_{L_i}, \xi_{L_i}), \quad i = 1, \ldots, n$$

$$L_i \mid \mathbf{p} \overset{ind.}{\sim} \sum_{l=1}^{N} p_l \delta_l(L_i), \qquad i = 1, \ldots, n$$

$$\mathbf{p} \mid \alpha \sim f(\mathbf{p} \mid \alpha)$$

$$\zeta_l = (\kappa_l, \tau_l, \sigma_l, \xi_l) \overset{ind.}{\sim} G_0(\zeta_l \mid \psi), \qquad l = 1, \ldots, N$$

where the induced prior $f(\mathbf{p} \mid \alpha)$ for the vector of weights $\mathbf{p} = (p_1, \ldots, p_N)$, given α, is given by a generalized Dirichlet distribution as introduced in Section 2.2.2. The structure of the centering distribution, G_0, and its hyperparameters, ψ, is discussed in Section 3.3.3, where we also discuss the prior choices for α and ψ to complete the model.

We employ a blocked Gibbs sampler to obtain samples from the full posterior distribution $p(\sigma, \xi, \kappa, \tau, \mathbf{L}, \mathbf{p}, \alpha, \psi \mid \text{Data})$. Details of the posterior simulation algorithm are provided next.

MCMC Details

Here, we provide the details for the simulation from the full posterior distribution of the hierarchical model for the Poisson process density. The Markov chain Monte Carlo algorithm applied to iteratively update model parameters according to the following steps:

- Updating L_i, $i = 1, \ldots, n$. Each draw of L_i is from a discrete distribution, $\sum_{l=1}^{N} \tilde{p}_{l,i} \delta_l(L_i)$, where the vector of revised weights, $\tilde{p}_l \propto (p_1 k(t_i, y_i | \zeta_1), \ldots, p_N k(t_i, y_i | \zeta_N))$.

- Updating $\mathbf{p}$. The full conditional of $\mathbf{p}$ is proportional to a generalized Dirichlet distribution with parameters $(M_1 + 1, \ldots M_{N-1} + 1)$ and $(\alpha + \sum_{l=2}^{N} M_l, \ldots \alpha + M_N)$ (Ishwaran

& James, 2001), where M_l is the size of each component, defined as $M_l = |\{i : L_i = l\}|, l = 1, \ldots N$. Therefore, $\mathbf{p}$ can be sampled using the following two steps. First, sample the latent parameter $V_l^* \sim \text{beta}(1 + M_l, \alpha + \sum_{r=l+1}^{N} M_r)$ and then let $p_1 = V_1^*; p_l = V_l^* \prod_{r=1}^{l-1}(1 - V_r^*), l = 1, 2, \ldots, N-1; p_N = 1 - \sum_{l=1}^{N-1}(1 - V_r^*)$.

- Updating α. The updates for α are generic for any choice of kernel in the Dirichlet process mixture model; details are given in Ishwaran & Zarepour (2000).

- Updating the centering distribution parameters. Based on the conditionally conjugate priors used for b_σ, b_ξ, and b_τ (see prior specification details in Section 3.3.3), each of the corresponding posterior full conditionals is available in closed form. In particular, for b_σ this is a Gamma distribution with shape parameter $1 + Na_\sigma$ and scale parameter $\left(d_\sigma^{-1} + \sum_{l=1}^{N} \sigma_l^{-1}\right)^{-1}$. Moreover, b_ξ has an inverse Gamma posterior full conditional with shape parameter $2 + Na_\xi$ and scale parameter $d_\xi + \sum_{l=1}^{N} \xi_l$. Finally, the full conditional for b_τ is a Gamma distribution with shape parameter $1 + Na_\tau$ and scale parameter $\left(d_\tau^{-1} + \sum_{l=1}^{N} \tau_l^{-1}\right)^{-1}$.

- Updating $(\kappa_l, \tau_l, \sigma_l, \xi_l)$, for $l = 1, ..., N$. Let n^* be the number of distinct clusters in vector $\mathbf{L}$, and $L_1^*, \cdots L_{n^*}^*$ the distinct values in vector $\mathbf{L}$. Then if $l \notin \{L_j^* : j = 1, \ldots, n^*\}$, $(\kappa_l, \tau_l, \sigma_l, \xi_l)$ is drawn from the centering distribution G_0. If $l \in \{L_j^* : j = 1, \ldots, n^*\}$, the posterior full conditional for $(\kappa_l, \tau_l, \sigma_l, \xi_l)$ is given by:

$$p(\kappa_l | \cdots, \text{Data}) \propto \text{d}G_0^\kappa(\kappa_l) \prod_{\{i:L_i=l\}} k_1(t_i \mid \kappa_l, \tau_l)$$

$$p(\tau_l | \cdots, \text{Data}) \propto \text{d}G_0^\tau(\tau_l) \prod_{\{i:L_i=l\}} k_1(t_i \mid \kappa_l, \tau_l)$$

$$p(\sigma_l|\cdots,\text{Data}) \propto \text{d}G_0^{\sigma}(\sigma_l) \prod_{\{i:L_i=l\}} k_2(y_i \mid \sigma_l,\xi_l)$$

$$p(\xi_l|\cdots,\text{Data}) \propto \text{d}G_0^{\xi}(\xi_l) \prod_{\{i:L_i=l\}} k_2(y_i \mid \sigma_l,\xi_l)$$

where $k_1(t \mid \kappa_l,\tau_l)$ and $k_2(y \mid \sigma_l,\xi_l)$ is given by (3.5) and (3.6), respectively, and G_0^{κ}, G_0^{τ}, G_0^{σ}, G_0^{ξ} are centering distributions defined in details in Section 3.3.3. Since no direct sampler is available for these distributions, we employ separate Gaussian random walk Metropolis steps on appropriately transformed versions of the parameters, that is, logarithmic transformations for τ_l, σ_l and ξ_l, and a logit transformation for κ_l. In all cases, the variances of the Gaussian proposals were tuned to obtain acceptance rates of around 20% to 30%.

3.3.2 Posterior Inference

Using the posterior samples for $G^N \equiv \{(p_l,\kappa_l,\tau_l,\sigma_l,\xi_l) : l = 1,...,N\}$, we can obtain full inference for the joint intensity of extremes,

$$\lambda(t,y;G,\gamma) = \gamma \sum_{l=1}^{N} p_l \; k_1(t \mid \kappa_l,\tau_l) k_2(y \mid \sigma_l,\xi_l).$$

Besides the intensity function, we can also obtain the marginal density of exceedance times that describes the information about when the extreme events will be more likely to happen,

$$f(t;G) = \sum_{l=1}^{N} p_l \; k_1(t \mid \kappa_l,\tau_l),$$

and for tail probabilities of the underlying process based on Lemma 1.

Similarly, approximate inferences for ε-conditional return level curves can be ob-

tained by replacing G in Equation (3.7) with its truncation approximation G^N,

$$\Pr(X_t > x : t \in [t_0 - \varepsilon, t_0 + \varepsilon]; G, \gamma) =$$
$$1 - \exp\left[-\gamma \sum_{l=1}^{N} p_l \left\{1 - K_2(x \mid \sigma_l, \xi_l)\right\} \left\{K_1\left(t_0 + \varepsilon \mid \kappa_l, \tau_l\right) - K_1\left(t_0 - \varepsilon \mid \kappa_l, \tau_l\right)\right\}\right],$$

and marginal return level curve in Equation (3.8) can be expressed as,

$$\Pr(X_t > x) = 1 - \exp\left[-T^{-1}\gamma \sum_{l=1}^{N} p_l \left(1 - K_2(x \mid \sigma_l, \xi_l)\right)\right].$$

3.3.3 Prior Specification

We assume that the different components of the centering distribution G_0 are independent, that is,

$$G_0(\kappa, \tau, \sigma, \xi) = G_0^{\kappa}(\kappa) G_0^{\tau}(\tau) G_0^{\sigma}(\sigma) G_0^{\xi}(\xi).$$

For G_0^{σ}, we use an inverse Gamma distribution with fixed shape parameter $a_\sigma = 2$, which implies infinite prior variance, and random mean parameter b_σ, which is assigned an exponential prior with mean d_σ. Hence, the specific choice of d_σ allows us to control the prior mean value for σ while being relatively non-informative about this choice. For G_0^{ξ}, we take an exponential distribution with mean b_ξ; an inverse Gamma prior with shape parameter 2 and mean d_ξ is placed on b_ξ. We suggest that the values of d_σ and d_ξ are selected to reflect the scale of the data under a single component of the mixture model. In particular, with a prior guess at the mean and variance for the exceedance values, we can numerically solve for σ and ξ from the Equation of the mean and variance of the generalized Pareto distribution. Then, we set the solutions to d_σ and d_ξ, which are the prior means for σ and ξ, respectively.

Regarding the parameters of the beta kernel component, for G_0^{κ}, we work with a beta distribution, with fixed parameters, for the scaled mean κ/T. For G_0^{τ}, we take an inverse Gamma distribution with shape parameter equal to 2 and mean parameter b_τ to which we place an exponential prior with mean d_τ. To specify d_τ and the parameters for G_0^{κ}, we study the implied prior for the marginal density of exceedance times. Based on the connection between the parametric and nonparametric formulations for the point process model, a non-informative specification may be built from a uniform prior mean for this marginal density. On the other hand, for some applications we may wish to encourage priors that favor clustering of extreme values, and this can also be achieved through appropriate specification of G_0^{κ} and G_0^{τ}. We provide illustrations of both scenarios with the data examples of Section 3.4.

Finally, we use a Gamma prior for α, and the reference prior, discussed in Section 3.3.1, for γ.

3.4 Illustrations

3.4.1 Simulation Study

To illustrate our modeling approach, we first consider a simulated data set where observations were generated according to a non-linear regression model, $X_t = \mu(t) + We_t$, with mean function

$$\mu(t) = -0.5 + 1.6(t/T) + 0.5\sin(-5.4 + 10.8(t/T)) + 1.1\{1 + 4(2(t/T) - 1)^2\}^{-1}.$$

Here, the noise terms e_t are independently distributed according to a Student t distribution with 3 degrees of freedom for every t, and $W = 0.32$. This choice implies that $\text{var}\{We_t\} = 0.3$ for

all t. The raw dataset contains $T = 10,000$ observations equally spaced in the interval $[0,T]$; to assemble the final data set, we retain observations that are larger than the threshold $u = 2.1$, which results in $n = 525$ extreme observations.

We assign a Gamma prior to the precision parameter α with mean 5 and variance 2.5. Also, following the approach discussed in Section 3.3.3, we set $d_\sigma = 0.296$ and $d_\xi = 0.257$. Moreover, we set $d_\tau = 300$, and consider two prior choices for G_0^κ. The first is a beta distribution for κ/T with mean 0.5 and variance $1/28$, whereas the second prior is based on a uniform distribution for κ/T. The effect of these two choices on the implied prior for the marginal density of exceedances over time is illustrated in the first two rows of Figure 3.1. The first row plots 10 prior realizations for this marginal density, while the second row shows the prior mean along with 95% pointwise credible intervals. Both prior specifications induce a large degree of variability for the marginal density of exceedance times, with individual realizations being highly multimodal. However, the first prior choice tends to favor exceedances located in the middle of the time interval, while the second prior implies a more uniform distribution of exceedances.

The algorithm discussed in Section 3.3.1 was used to fit our model. A total of 4,000 posterior samples were used for all inferences. These samples were obtained after thinning a sample of 200,000 from which 40,000 iterations were discarded as burn-in period. Posterior mean estimates of the joint intensity function $\lambda(t,y)$ are presented in the third row of Figure 3.1, while the last row shows the posterior mean and 95% credible intervals for the marginal density of exceedance times. Note that posterior inference is quite robust to the specific prior choices.

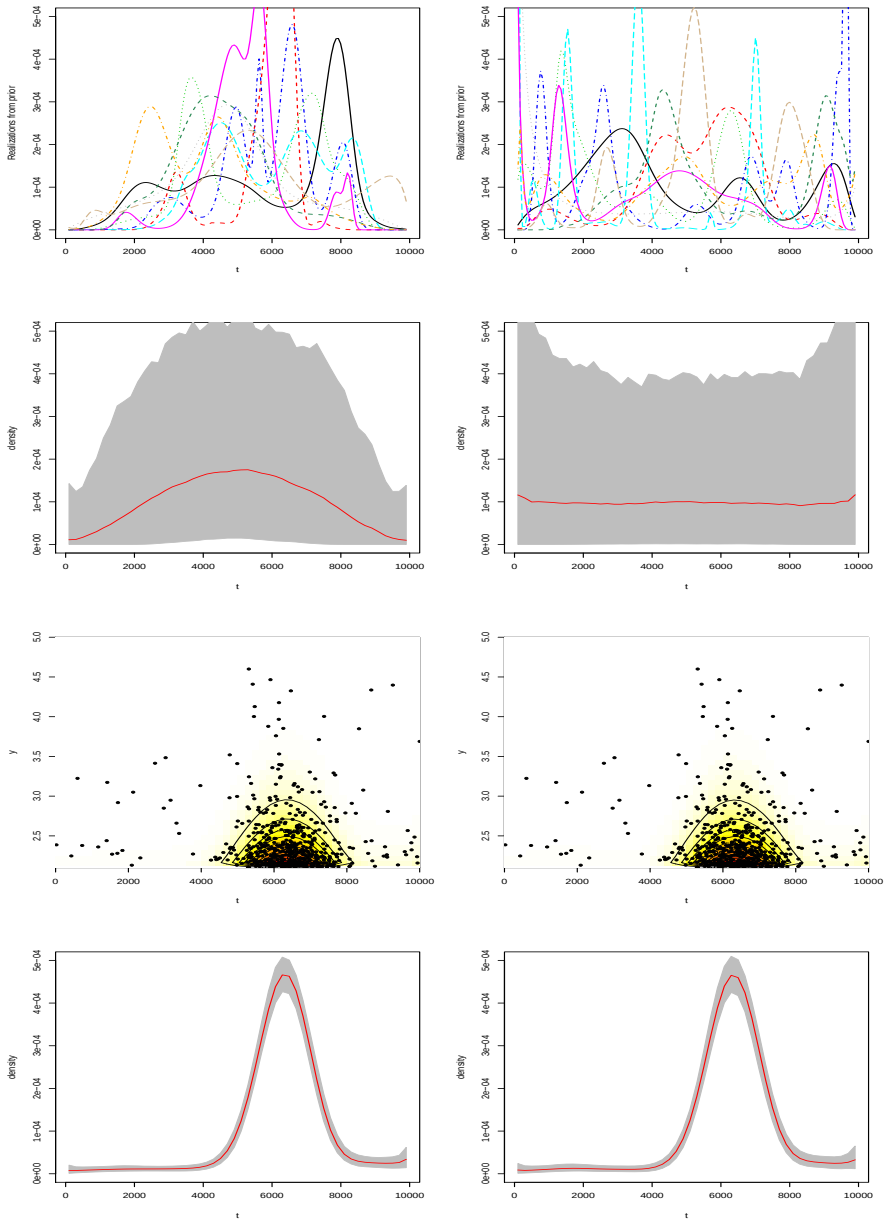
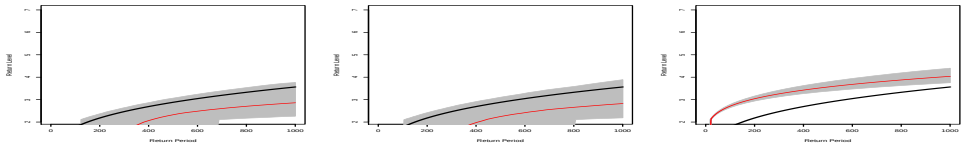


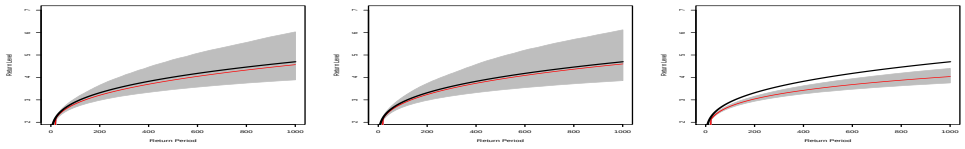
Figure 3.1: Simulation study. From top to bottom, 10 prior realizations of the marginal density of exceedance times; prior mean (red line) and 95% intervals (grey bands) of the marginal density of exceedance times; posterior mean for the bivariate intensity function; and posterior mean and 95% intervals for the marginal density of exceedance times. The left column corresponds to the prior choice involving a beta distribution for κ/T with mean 0.5 and variance $1/28$, and the right column to the prior based on a uniform distribution for κ/T.

In addition to providing estimates of the intensity function, we are interested in investigating the ability of the model to estimate the tails of the stochastic process, and in comparing its performance with the parametric model discussed in Section 2.1.3. For this purpose, we present in Figure 3.2 the true conditional return level function corresponding to four time points, along with posterior estimates generated under the parametric model for the intensity function given in (2.7), as well as under the nonparametric model, using the two prior choices discussed above. For the parametric model, we utilize a normal prior on the location parameter μ with mean 3.23 and variance 10; an inverse Gamma prior on the scale parameter σ with shape parameter 2 and mean 0.43; and an exponential prior on the shape parameter ξ with mean 0.048. The parametric model is fitted using a Gaussian random walk Metropolis algorithm that samples jointly the three parameters on an appropriately transformed scale for each parameter.

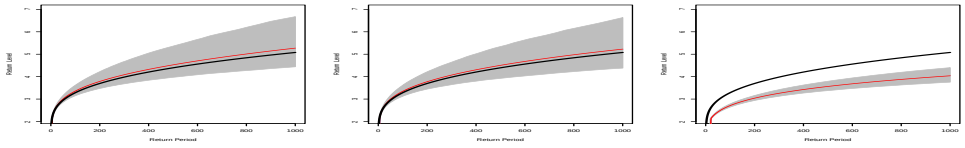
The right column of Figure 3.2 shows that the parametric model performs poorly at capturing the true return level curve at all four time points. Moreover, the credible intervals are very narrow, which suggests that the parametric model dramatically underestimates the uncertainty in situations where the process is not homogeneous in time. On the other hand, the estimates generated by the nonparametric model are quite accurate, particularly for time points $t_0 = 5100$ and $t_0 = 6500$ around which a relatively large number of exceedances are concentrated. Again, posterior inference is robust to the two prior choices. Finally, note that the nonparametric mixture model estimates for the marginal return level curve are almost identical to those generated by the parametric model, involving only a minimal increase in posterior uncertainty.



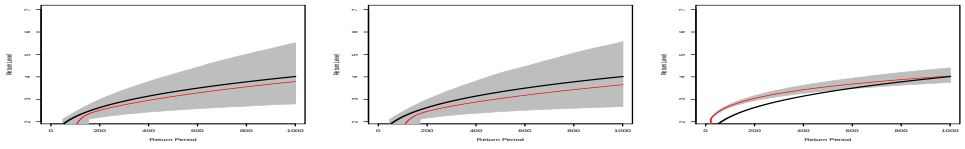
(a) $t_0 = 500$



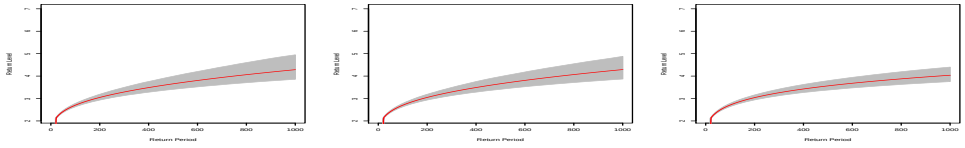
(b) $t_0 = 5100$



(c) $t_0 = 6500$



(d) $t_0 = 9500$



(e) Average Return Levels

Figure 3.2: Simulation study. The top four rows plot the posterior mean (red line) and 95% intervals (grey bands), and the true 1000-observation return level (black line) at four time points. The bottom row includes the posterior mean and 95% interval estimates of the marginal return levels. Results are shown for the nonparametric model under the beta distribution for κ/T with mean 0.5 and variance $1/28$ (left column), the nonparametric model under the uniform distribution for κ/T (middle column), and the parametric model (right column).

3.4.2 Dow Jones Data

Here, we discuss the analysis of extremes of the daily returns for the Dow Jones index between September 11, 1995 and September 7, 2000; a previous analysis of this data set is presented in Coles (2001). Modeling the lower tail of the distribution for returns of financial assets is critical to compute risk measures such as the Value at Risk or the Expected Shortfall.

In the sequel we work with the negative log returns of the index,

$$y_t = -100 \times \log\left(\frac{x_t}{x_{t-1}}\right),$$

where x_t is the closing price at day t. Note that, in this case, drops in the index correspond to positive values of y_t, while increases correspond to negative values. Figure 3.3 shows the $n = 82$ values of y_t above the threshold $u = 1.5$. The vertical dashed lines mark the dates at which three financial crises started: the mini-crash on October 27, 1997, the Russian financial crisis on August 17, 1998, and the bursting of the dot-com bubble on March 10, 2000. We see that the three biggest drops in the index align well with these three financial crises, and that a large number of exceedances concentrate around those dates.

For the analysis of this data set we used priors that are similar to the second prior specification discussed in Section 3.4.1. Figure 3.4 shows 10 prior realizations for the marginal density of exceedance times, as well as prior mean and 95% interval estimates for this density. Given that the time period under study includes multiple crises around which extreme values may cluster, a prior choice that favors multimodal exceedance time densities is arguably justified. At the same time, the corresponding prior mean is fairly close to a uniform density with wide uncertainty bands. Hence, even though our prior favors the clustering of extreme values,

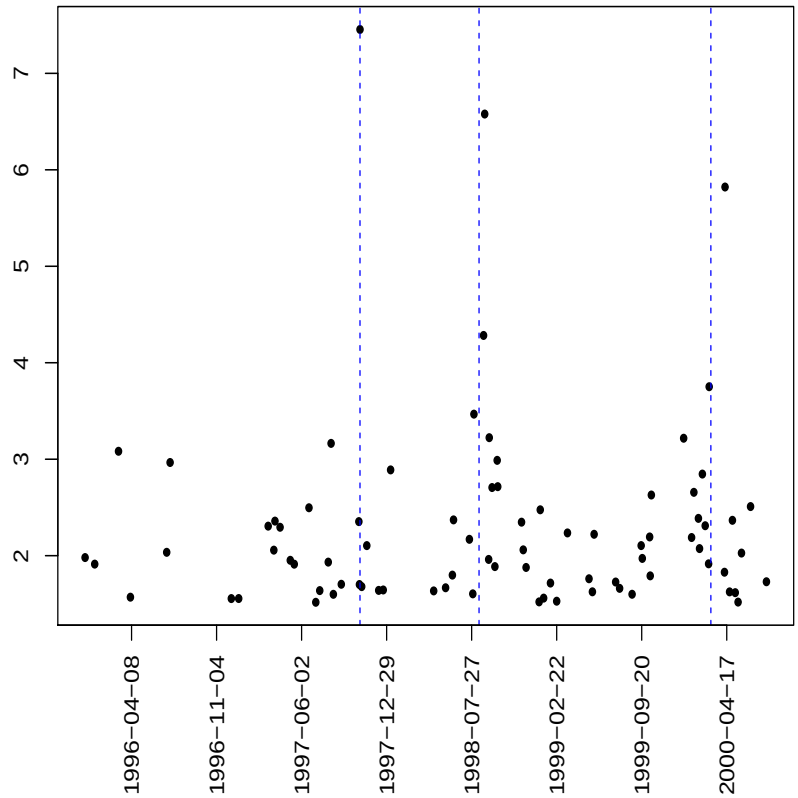


Figure 3.3: Dow Jones data. Plot of the extreme negative log returns for the Dow Jones index above threshold $u = 1.5$, from September 11, 1995 to September 7, 2000. The blue dashed lines indicate the time points of three financial crises: the mini-crash on October 27, 1997, the Russian financial crisis on August 17, 1998, and the bursting of the dot-com bubble on March 10, 2000, respectively.

we make no prior assumption about the location of such clusters.

Figure 3.4 includes posterior inference results for the bivariate intensity function, and for the marginal density of exceedance times. Note that the nonparametric model captures reasonably well the localized characteristics of the raw data and cyclical nature of the business cycle. Interestingly, the estimates also suggest an increasing risk of extreme loses over the time period under study. Both of these features are captured by the model even though it does not

contain any explicit term to account for them.

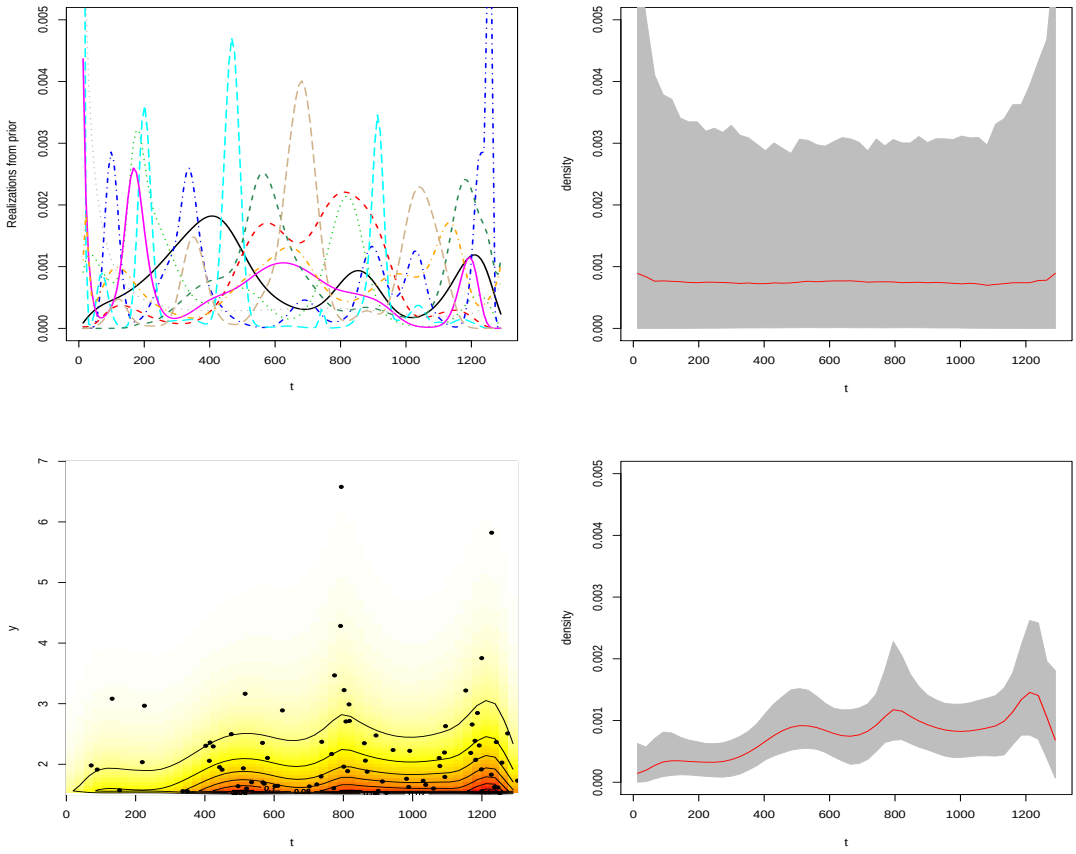


Figure 3.4: Dow Jones data. Clockwise from top left, 10 prior realizations of the marginal density of exceedance times; prior mean (red line) and 95% intervals (grey bands) of the marginal density of exceedance times; posterior mean and 95% interval estimates for the marginal density of exceedance times; and posterior mean of the bivariate intensity function.

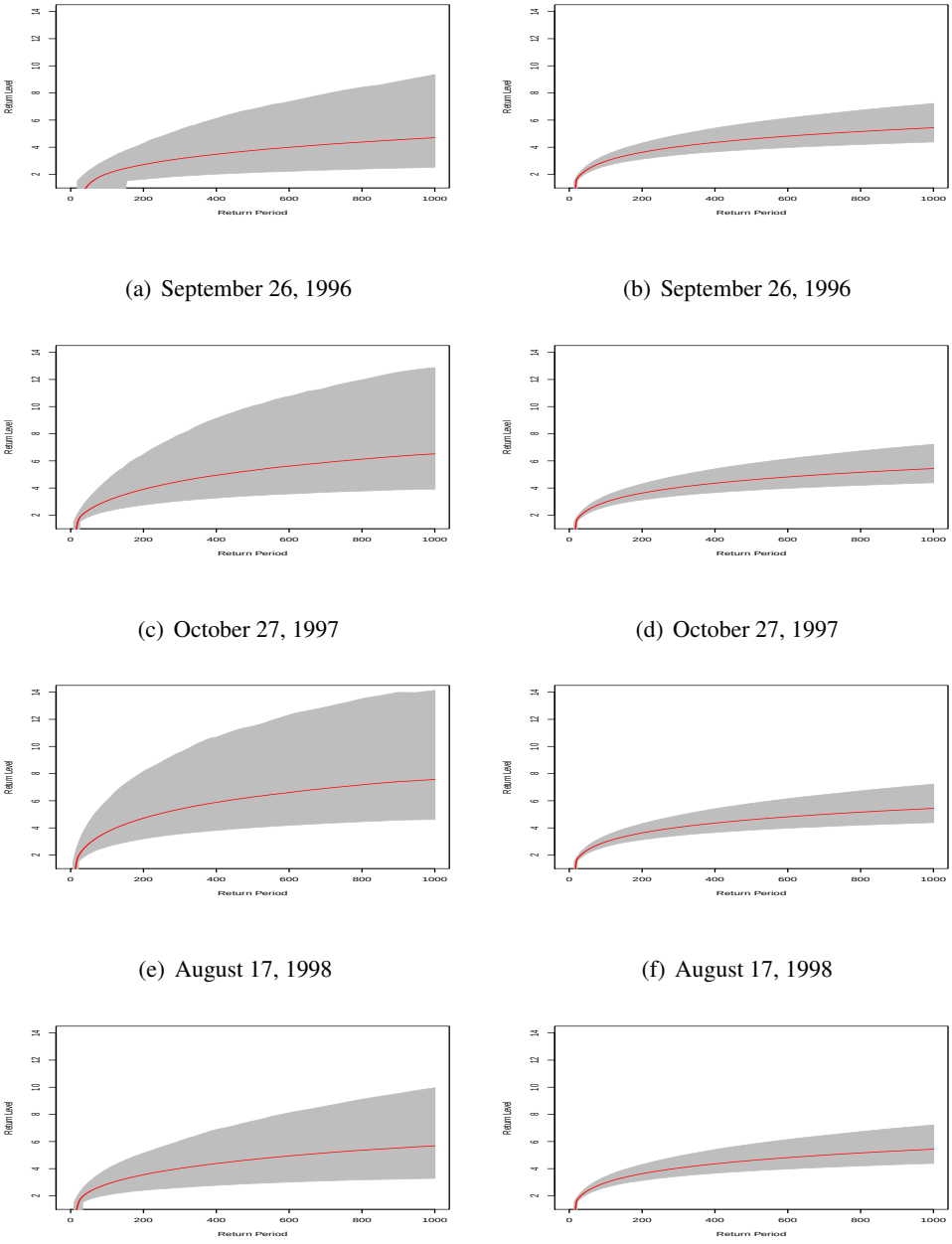
Next, we report on inference for ε-conditional return level curves, obtained using Equation (3.7) with $\varepsilon = 0.5$, that is, daily conditional return level curves. Based on the nonparametric mixture model, the left column of Figure 3.5 shows posterior mean and 95% interval estimates for the conditional return level curve at four specific dates, September 26, 1996, October 27, 1997, August 17, 1998, and July 20, 1999. Note that the posterior mean estimates of

the return level curve at October 27, 1997 and August 17, 1998 are uniformly above the ones at September 26, 1996 and July 20, 1999. This is consistent with the fact that the former two dates fall within periods of financial distress, while the latter do not. In addition, we compare the nonparametric model estimates against those generated by the parametric model; see the right column of Figure 3.5. Since the parametric model is unable to capture the time inhomogeneity in the data, it produces the same point estimate at all dates with much narrower posterior uncertainty bands. Finally, Figure 3.6 shows posterior point and interval estimates for the marginal return level curve under both the parametric and nonparametric models. We note that the point estimates are similar, but the uncertainty levels associated with the nonparametric model are higher.

Model Checking

To check the Poisson process model assumptions, in-sample goodness of fit was investigated using quantile-quantile plots. Since the marginal process of the exceedance times is also a Poisson process, the time-rescaling theorem (see for example Daley & Vere-Jones, 2003) is applied for checking the Poisson process assumption in the time direction. More specifically, if $\{t_i : i = 1,\ldots,n\}$ is a realization from the non-homogeneous Poisson process with cumulative intensity $\Lambda(t)$, then the transformed point process $\{\Lambda(t_i) : i = 1,\ldots n\}$ is a homogeneous Poisson process with unit intensity. Hence, if the Poisson process assumption is correct, the transformed inter-arrival times $z_1,\ldots,z_n$, defined as $z_i = 1-\exp\{-[\Lambda(t_i;G^N)-\Lambda(t_{i-1};G^N)]\}$ (with the convention $\Lambda(0;G^N)=0$) are independent uniform random variables on $[0,1]$.

On the other hand, to check the Poisson process assumption in both directions, we



(a) September 26, 1996

(b) September 26, 1996

(c) October 27, 1997

(d) October 27, 1997

(e) August 17, 1998

(f) August 17, 1998

(g) July 20, 1999

(h) July 20, 1999

Figure 3.5: Dow Jones data. The posterior mean (red line) and 95% interval estimates (grey bands) of the 1000-day conditional return level curves at four different dates under the nonparametric model (left column) and the parametric model (right column).

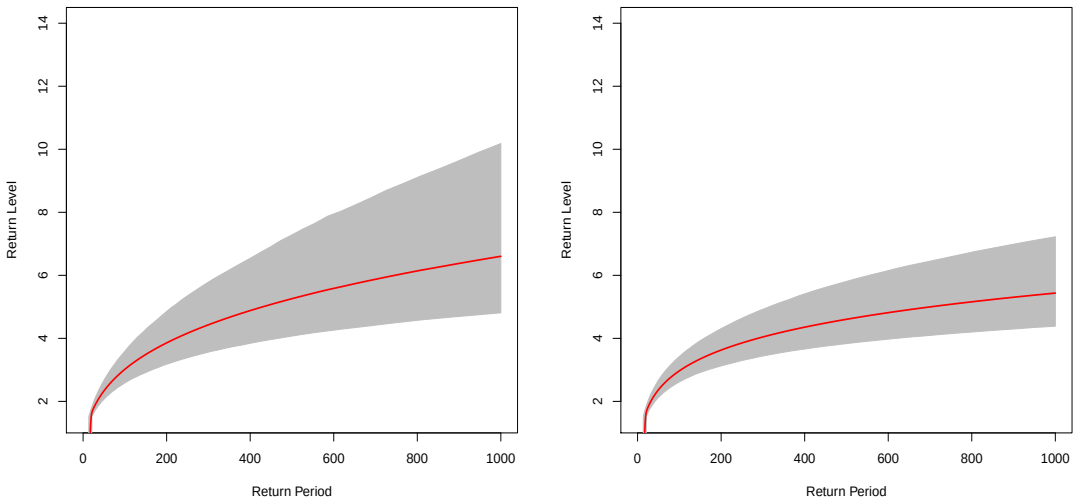


Figure 3.6: Dow Jones data. The posterior mean (red line) and 95% interval estimates (grey bands) of the marginal 1000-day return level curve under the nonparametric model (left panel) and the parametric model (right panel).

studied the conditional distribution for exceedances given time. Specifically, the conditional cumulative distribution function $F(y|t;G^N) = \int_u^y f(s \mid t;G^N)\mathrm{d}s$ is available at any specific point (t,y). Hence, if the Poisson process assumption holds, the random variables $u_i = F(y_i \mid t_i;G^N)$, for $i = 1,\ldots,n$, are uniformly distributed in the unit interval (Taddy & Kottas, 2012).

The left panel of Figure 3.7 shows the quantile-quantile plot for the transformed inter-arrival times against the quantiles of a uniform distribution, and the right panel of Figure 3.7 shows the conditional distribution quantile-quantile plot. The quantile-quantile plots are estimated from each MCMC posterior sample, and the posterior mean is shown in red and 95% credible intervals are shown in blue. These results do not indicate any drastic deviations from the Poisson process assumptions

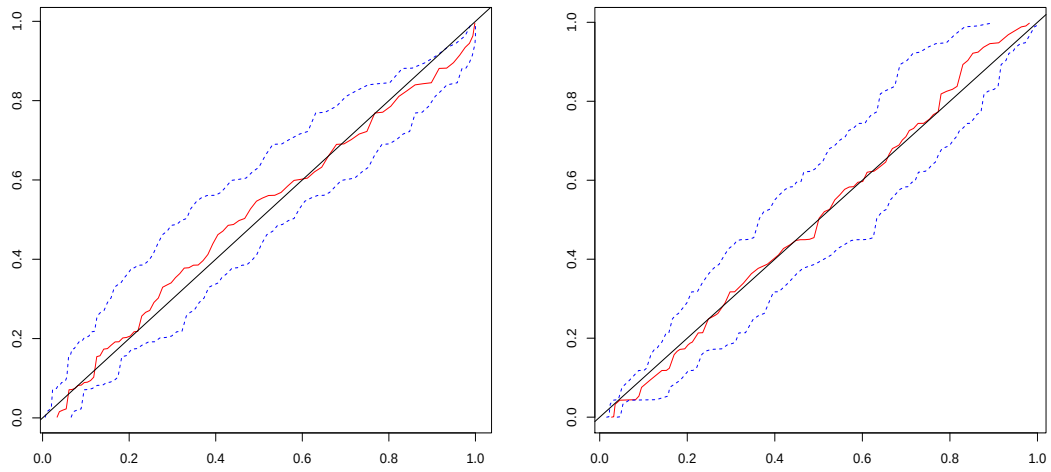


Figure 3.7: Dow Jones data. Posterior quantile-quantile plots (posterior mean and 95% credible intervals) for $1-\exp\{-[\Lambda(t_i;G^N)-\Lambda(t_{i-1};G^N)]\}$ (left panel), and for $\int_u^{y_i} f(s \mid t_i;G^N)\mathrm{d}s$ (right panel), respectively.

Sensitivity Analysis

To assess prior sensitivity for the nonparametric model, Figures 3.8 and 3.9 provide inference results under three different prior choices for the parameters of the beta distribution for G_0^{κ}, for the exponential prior, with mean d_τ, for the mean parameter b_τ of the inverse-Gamma distribution for G_0^{τ}, and for the parameters of the gamma prior for α. Specifically, the left column of Figures 3.8 and 3.9 corresponds to the first prior specification with a uniform distribution for κ/T, $d_\tau = 40$, and $\mathsf{E}(\alpha) = 5$, $\mathrm{var}(\alpha) = 2.5$; the middle column to the second prior choice using a beta distribution for κ/T with $\mathsf{E}(\kappa/T) = 0.5$ and $\mathrm{var}(\kappa/T) = 1/28$, $d_\tau = 300$, and $\mathsf{E}(\alpha) = 5$, $\mathrm{var}(\alpha) = 2.5$; and the right column to the third prior choice involving a uniform distribution for κ/T, $d_\tau = 2$, and $\mathsf{E}(\alpha) = 1$, $\mathrm{var}(\alpha) = 1$. As in earlier examples, Figure 3.8 shows prior realizations for the marginal density of exceedance times, and prior mean and

interval estimates for this density. The first prior favors unimodal densities and results in a slightly U-shaped prior mean density. Although the second prior, which is the same as the first prior considered in Section 3.4.1, encourages multimodal density realizations, it yields a unimodal prior mean density. Finally, the third prior is chosen to strongly favor U-shaped density realizations, including a U-shaped prior mean density with a relatively low level of uncertainty associated with it; clearly, this is a prior choice that would not be recommended for this particular problem.

Figure 3.8 plots posterior estimates for the intensity function and for the density of exceedance times. The estimates under the first two priors show features that are similar to those we obtained under the original prior. The estimates under the third prior, which strongly supports the absence of localized features, capture the increasing trend in risk, but not the clustering of extremes. Figure 3.9 presents inference for the daily conditional return level curves at the same time points we considered above, along with estimates of the marginal return level curve. Again, posterior inference under the first two priors is similar to that obtained under the original prior. However, results under the third prior differ, particularly in terms of the uncertainties that the model attaches to the posterior mean estimates.

3.5 Discussion

We have presented a Bayesian nonparametric model for the analysis of extremes under a generalization of the point process approach. Our model is built to relax the time homogeneity restriction through flexible mixture modeling for the intensity function of extremes.

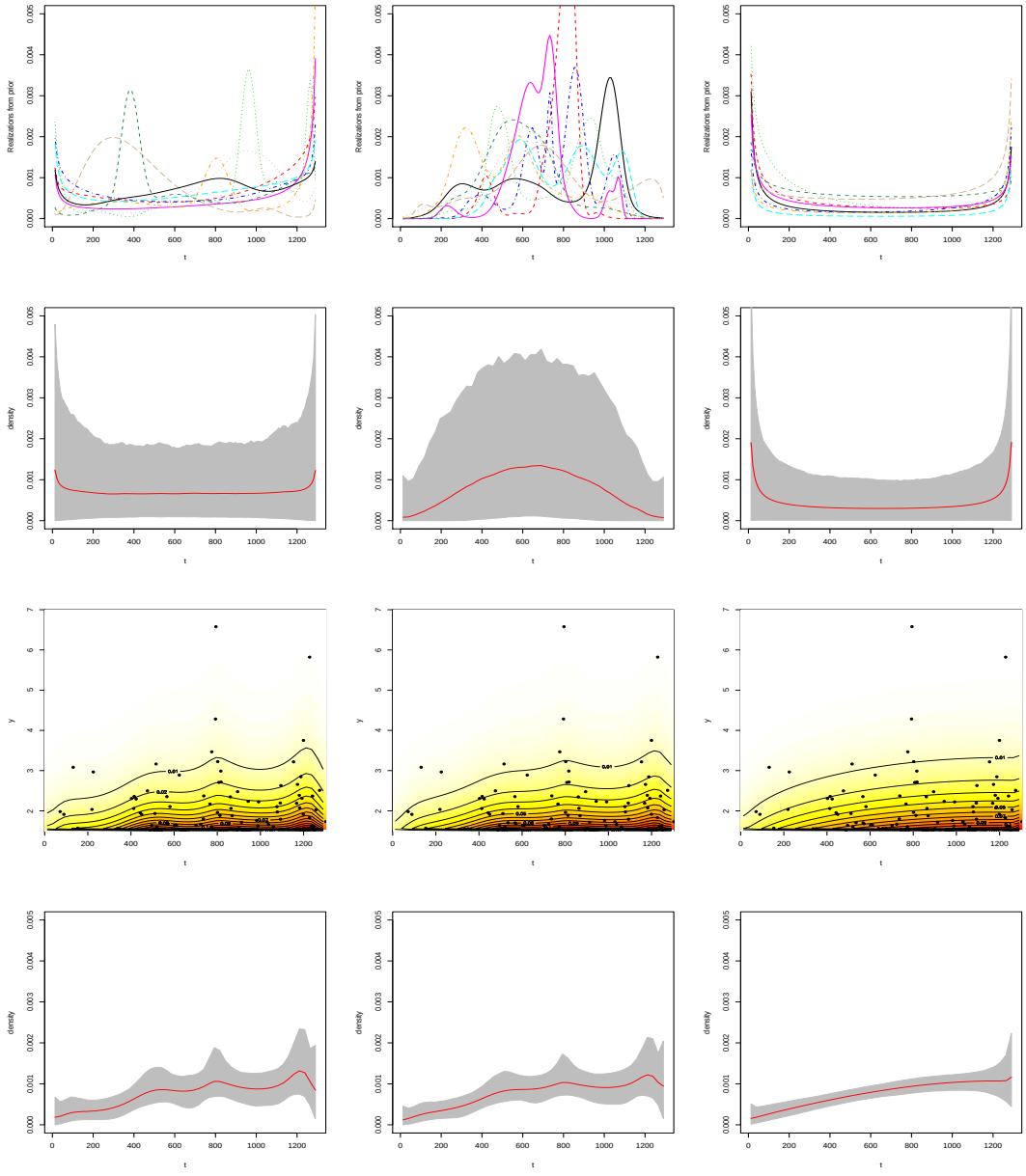
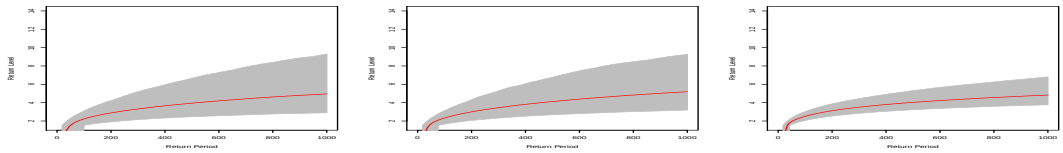
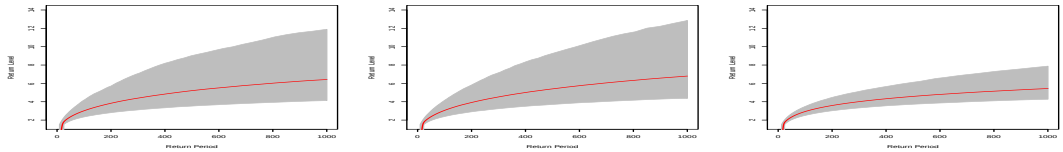


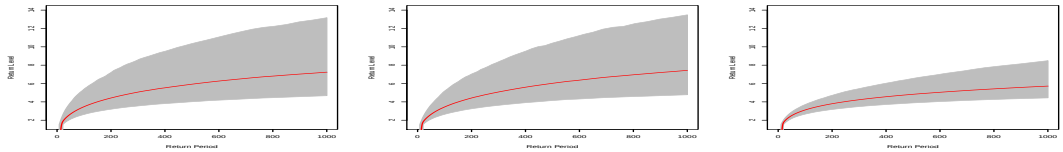
Figure 3.8: Prior sensitivity analysis results for the Dow Jones data; see Section 3.4.2 for details about the three priors corresponding to the columns of the figure. From top to bottom, 10 prior realizations for the marginal density of exceedance times; prior mean (red line) and 95% interval estimates (grey bands) of the marginal density of exceedance times; posterior mean of the bivariate intensity function; and posterior mean and 95% interval estimates of the marginal density of exceedance times.



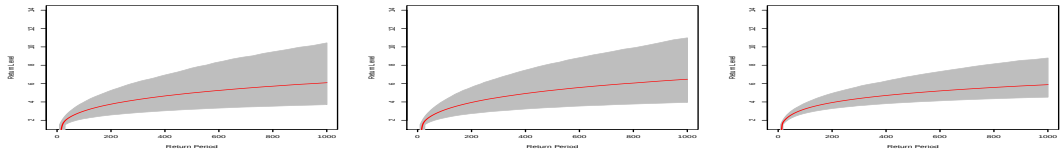
(a) September 26, 1996



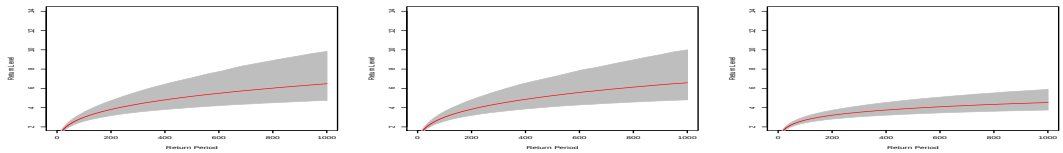
(b) October 27, 1997



(c) August 17, 1998



(d) July 20, 1999



(e) Marginal Return Levels

Figure 3.9: Prior sensitivity analysis results for the Dow Jones data; see Section 3.4.2 for details about the three priors corresponding to the columns of the figure. The top four rows include the posterior mean (red line) and 95% interval estimates (grey bands) of the 1000-day conditional return level curves at four different dates. The bottom row plots the corresponding estimates for the 1000-day marginal return level curve.

Particular emphasis has been placed on the mixture model formulation for this intensity function to obtain desirable properties for the tail behavior of the underlying process whose extremes are recorded. Our empirical results suggest that the model is quite robust to the choice of priors when sample sizes are moderately large, as in our simulation study, but might be affected by strongly informative priors when sample sizes are relatively small, as in the Dow Jones data application. As a general strategy we suggest the use of priors similar to the ones discussed in Section 3.4.2, which allow for clustering of extreme values when such a feature is suspected in the data.

The starting point of the mixture model formulation is the product kernel specification in (3.4), motivated by the different nature of the arguments that comprise the support of the bivariate Poisson process. A consequence of this specification is that the implied tail index parameter does not depend on time; see the proof of Theorem 1. From a practical point of view, this is arguably not a serious limitation, since the typically small number of exceedances will likely not suffice to inform temporally dependent tail index indicators. Nevertheless, this methodological extension can be developed through choice of an appropriate bivariate mixture kernel $k(t,y \mid \theta)$. For instance, a possible modification of the form in (3.4) involves the same beta kernel component for $k_1(t)$ with a conditional Pareto distribution for $k_2(y \mid t)$ defined by extending the shape parameter ξ in (3.6) to a parametric function $\xi(t)$. Then, the same argument as in the proof of Theorem 1 yields a temporally dependent tail index indicator. In particular, the choice $\xi(t) = \exp(\beta_0 + \beta_1 t)$, with Dirichlet process mixing on the real-valued parameters β_0 and β_1, leads to a mixture model that includes the model of Section 3.2.2 as a special case.

Chapter 4

Nonparametric Modeling for Systemic Risk Assessment in Correlated Financial Markets

4.1 Introduction

One important lesson learned from the financial crisis of 2007 and 2008 is that the behavior of markets during periods of distress can dramatically deviate from their behavior during periods of calm. In particular, the financial crises showed that novel techniques are required to understand how information spreads across financial markets during times of upheaval.

This chapter focuses on extending the point process framework developed in Chapter 3 to simultaneously model risks in multiple financial markets. The model we discuss provides a decomposition of the risk associated with each of the individual markets into two components: a systemic risk component, whose features are shared by all markets, and an idiosyncratic risk component, which is specific to each sector. To motivate this type of decomposition, consider

the returns associated with the ten sectors making up the S&P500 index (see also Section 4.4). Figure 4.1 presents the most extreme *negative* log returns on all of the ten sectors between January 1, 2000 and December 31, 2011. It is clear from the figure that all sectors present an increased frequency of extreme values around periods of distress, such as the so-called "dot com" bubble burst in March of 2000 and the climax of the financial crises in September 2008. However, it is also clear that certain features are particular to specific sectors, such as the increased number of extreme returns in the energy sector in 2004 and 2005.

The decomposition achieved by our model is similar in spirit to the ideas underlying the traditional Capital Asset Pricing (CAP) model (Treynor, 1961, 1962; French, 2003). In the CAP model, linear regression is used to relate the returns of an individual security to those of the market, also allowing for a decomposition of risks into a systemic and an idiosyncratic component. However, and unlike the CAP model, our model focuses on patterns associated with extreme values of the index, and uses a nonparametric methodology that does not assume (implicitly or explicitly) that returns arise from a Gaussian distribution. Furthermore, our model is dynamic in nature, allowing for the structure of the different risks to evolve over time.

Variations of the CAP model that focus on extreme returns have been discussed in Barnes & Hughes (2002), Allen et al. (2009) and Chang et al. (2011), among others. These papers use quantile regression instead of ordinary linear regression to relate the returns of individual securities to those of the market. The approach we pursue here is completely different as it does not involve regression analysis. Instead, we model the returns that exceed a given threshold as a non-homogeneous Poisson process with unknown intensity function, which is assigned a flexible prior based on Dirichlet process mixture models. Hence, our model can be

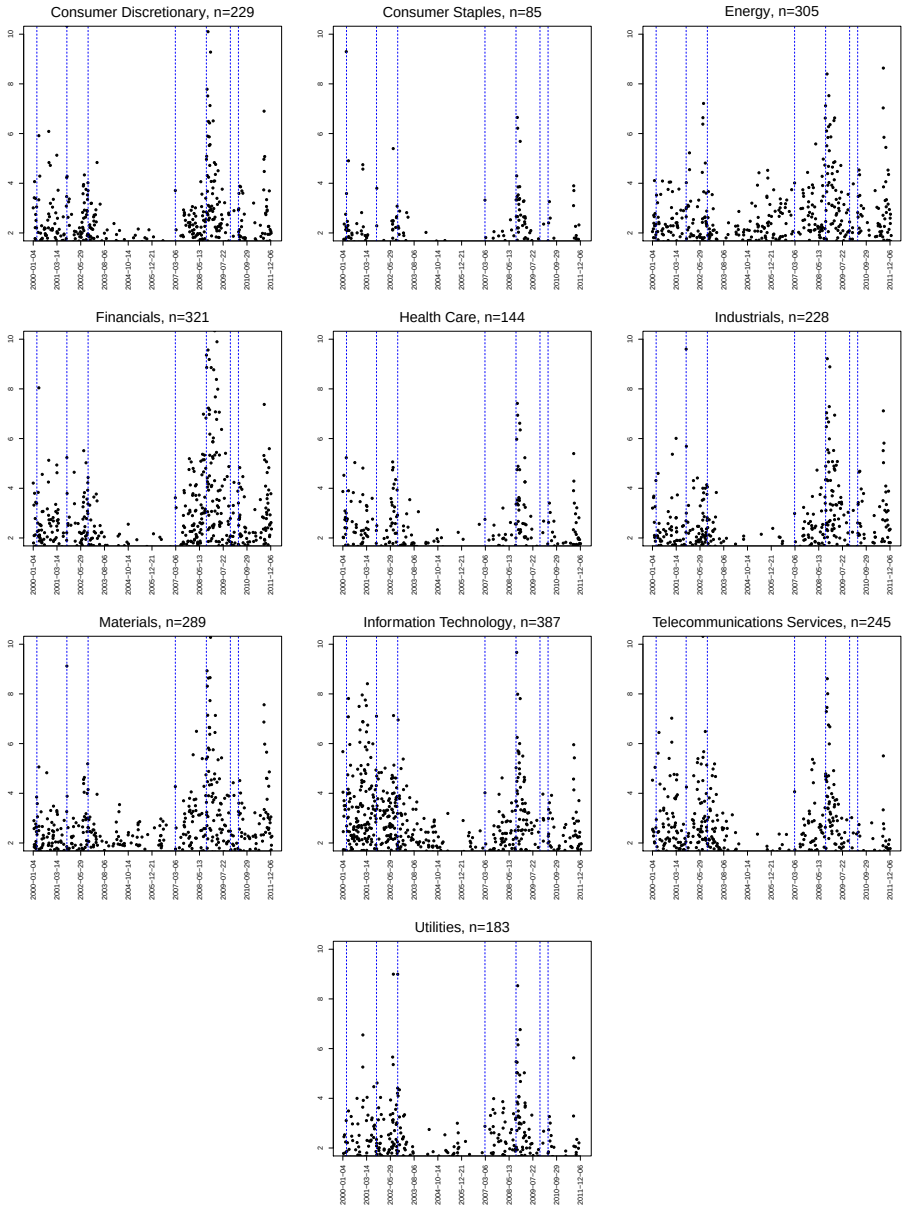


Figure 4.1: Negative log returns above 2% for four sectors of the S&P500 index (consumer staples, energy, financials and information technology). Vertical dotted lines identify events of significance to the markets, such as the bursting of the .com bubble (03/10/2000), the 09/11 terrorist attacks (09/11/2001), the stock market downturn of 2002 (09/12/2001), the bursting of the Chinese bubble (02/27/2007), the bankruptcy of Lehman Brothers (09/16/2008), Dubai's debt standstill (11/27/2009) and the beginning of the European sovereign debt crisis (08/27/2010).

conceptualized as an example of a Cox process (Cox, 1955).

The rest of this chapter is organized as follows: Section 4.2 provides a detailed description of our modeling approach and discusses its main properties. Section 4.3 presents a description of our proposed computational approach along with a careful discussion on subjective prior elicitation. Section 4.4 illustrates the model using a detailed case study focusing on the S&P500 index. Finally, Section 4.5 discusses some limitations of the model and the possibilities for further extensions and applications.

4.2 Modeling Approach

As in Section 3.4.2, we focus on the negative log returns of a group of J related but distinct markets, which are defined as

$$y_{i,j} = -100 \times \log\left(\frac{x_{i,j}}{x_{i-1,j}}\right),$$

where $x_{i,j}$ is the value of market $j = 1,\ldots,J$ at time $i = 1,\ldots,T$. Note that large positive values of $y_{i,j}$ indicate a large drop in the price index associated with market j, so for risk management purpose we are interested in *large* values of $y_{i,j}$. Hence, for a given threshold u, focus attention on the collections of times $\{t_{j,k} : k = 1,\ldots,n_j, j = 1,\ldots,J\}$, where $t_{j,k}$ is the date associated with the appearance of the k-th negative log return in sector j that is larger than u.

Our methodology relies on the point process approach to extreme value analysis developed in Pickands (1971) and Smith (1990). More specifically, for each market j, we regard the collection of times $\{t_{j,k} : k = 1,\ldots,n_j\}$ at which exceedances occur as a realization from a point process $N_j(t)$ defined on $[0,T]$. In turn, each $N_j(t)$ is constructed as the superposition

of two independent, non-homogeneous Poisson processes. The first such process accounts for systemic risk and has a cumulative intensity function Λ_0^* that is common to all markets, while the second is associated with the idiosyncratic risk and has a cumulative intensity function Λ_j^* that is specific to each market. Because of properties of superpositions of Poisson processes, this assumption implies that each $N_j(t)$ is also a non-homogeneous Poisson process with cumulative intensity $\Lambda_j(t) = \Lambda_0^*(t) + \Lambda_j^*(t)$, and intensity function $\lambda_j(t) = \lambda_0^*(t) + \lambda_j^*(t)$, where λ_0^* and λ_j^* are the Poisson process intensities associated with Λ_0^* and Λ_j^*, respectively.

The modelling approach for the Λ_j^* builds from the direct connection of a non-homogeneous Poisson process cumulative intensity/intensity function with a distribution/density function. Specifically, for $j = 0, 1, ..., J$, we can write $\Lambda_j^*(t) = \gamma_j^* F_j^*(t)$, where $\gamma_j^* \equiv \Lambda_j^*(T) = \int_0^T \lambda_j^*(t)\mathrm{d}t$ ($< \infty$) is the rate parameter controlling the total number of exceedances and $F_j^*(t) = \Lambda_j^*(t)/\Lambda_j^*(T)$ is a distribution function on $[0, T]$ that controls how the exceedances are distributed over time. Hence, the sector-specific cumulative intensity function Λ_j can be written as

$$\Lambda_j(t) = \gamma_j F_j(t) = \{\gamma_0^* + \gamma_j^*\} \left\{ \frac{\gamma_0^*}{\gamma_0^* + \gamma_j^*} F_0^*(t) + \frac{\gamma_j^*}{\gamma_0^* + \gamma_j^*} F_j^*(t) \right\}.$$

Our construction implies that the market-specific exceedance rate, γ_j, is simply the sum of the systemic and idiosyncratic rates, while the market-specific distribution function, F_j, can be written as a mixture of the systemic and idiosyncratic distribution functions. The corresponding weight, $\varepsilon_j = \gamma_0^*/(\gamma_0^* + \gamma_j^*)$, represents the proportion of exceedances in market j that are associated with the systemic component. In addition, note that values of ε_j close to 1 (which are associated with $\gamma_0^* \gg \gamma_j^*$) imply a stronger association in the pattern of extremes.

Because the processes generating the exceedances were assumed to be Poisson pro-

cesses, computing the probability that at most r exceedances will be observed in market j on a period $[t_0, t_0+\Delta]$ is simply

$$\sum_{i=0}^{r} \frac{\left\{\Upsilon_j(t_0,\Delta)\right\}^i \exp\left\{-\Upsilon_j(t_0,\Delta)\right\}}{i!},$$

where $\Upsilon_j(t_0,\Delta) = \Lambda_j(t_0+\Delta) - \Lambda_j(t_0)$. These exceedance probabilities are easier to interpret than the intensity functions through which the model is defined. For example, the probability that no exceedances are observed in market j between t_0 and $t_0+\Delta$ is simply

$$\exp\left\{-\left[\Lambda_j(t_0+\Delta) - \Lambda_j(t_0)\right]\right\} =$$

$$\exp\left\{-\left[\Lambda_0^*(t_0+\Delta) - \Lambda_0^*(t_0)\right]\right\} \times \exp\left\{-\left[\Lambda_j^*(t_0+\Delta) - \Lambda_j^*(t_0)\right]\right\},$$

where the first term in the right-hand expression corresponds to the probability of no exceedance due to the systemic component and the second term corresponds to the probability of no exceedance due to the idiosyncratic component. Hence, our model implies a multiplicative risk structure.

4.2.1 Modeling the Intensity Functions

In order to generate a flexible model that can capture changes in the pattern of extreme events over time, we model the densities $f_0^*, f_1^*, \ldots, f_J^*$ associated with the systemic and idiosyncratic distribution $F_0^*, F_1^*, \ldots, F_J^*$ using Dirichlet process mixture models. In particular, we let

$$f_j^*(t) \equiv f_j^*(t; G_j^*, \tau) = \int k(y \mid \mu_j, \tau) \mathrm{d}G_j^*(\mu_j), \qquad j = 0, 1, \ldots, J,$$

where the function $k(y \mid \mu, \tau)$ is a kernel indexed by the finite dimensional parameters μ and τ, while G_j^* is a random mixing distribution that follows a Dirichlet process prior,

$$G_j^*(\cdot) = \sum_{l=1}^{\infty} \omega_{j,l} \delta_{\tilde{\mu}_{j,l}}(\cdot), \qquad j = 0, 1, \ldots, J,$$

where $\delta_a(\cdot)$ denotes the degenerate measure at a, the atoms $\tilde{\mu}_{j,1}, \tilde{\mu}_{j,2}, \ldots$ form a sequence of independent random variables identically distributed according to the baseline measure H, and $\omega_{j,l}$ denotes the corresponding weights arise through a stick-breaking construction.

Because in our application the support for the point process is a compact set, a natural choice for the kernel $k(t \mid \mu, \tau)$ is the rescaled beta density,

$$\frac{1}{T} \frac{\Gamma(\tau)}{\Gamma(\mu\tau/T)\Gamma(\{1-\mu/T\}\tau)} \left(\frac{t}{T}\right)^{\mu\tau/T-1} \left(1 - \frac{t}{T}\right)^{\{1-\mu/T\}\tau-1} \mathbb{I}_{[0,T]}(t) \tag{4.1}$$

where $\mu \in [0, T]$ is a location parameter and $\tau \in (0, \infty)$ can be interpreted as a bandwidth parameter. Because the Dirichlet process mixtures allow for an infinite number of mixture components, the model has full support on the space of absolutely continuous distribution on $[0, T]$ as long as the baseline measure H and the prior on the bandwidth are selected to provide full support on the domain of the parameters of the rescaled beta kernel. The precision parameter α controls the relative weight of the components, with smaller values of α favoring mixtures where a small number of components received very large weights. On the other hand, the baseline measure H controls the location of the mixture components.

Besides a prior on the densities $f_0^*, f_1^*, \ldots, f_J^*$, full prior specification for the intensity functions $\Lambda_1(t), \ldots, \Lambda_J(t)$ requires priors for the rate parameters $\gamma_0^*, \gamma_1^*, \ldots, \gamma_1^*$. In the case of the rate associated with the systemic component, γ_0^*, a natural choice is a Gamma distribution with shape parameter $a_{\gamma_0^*}$ and rate parameter $b_{\gamma_0^*}$. For the idiosyncratic component, we use slightly

more general zero-inflated gamma prior with density,

$$p(\gamma_j^* \mid \pi) = (1-\pi)\delta_0(\gamma_j^*) + \pi \mathsf{Gamma}(\gamma_j^* \mid a_{\gamma_j^*}, b_{\gamma_j^*}) \qquad j \geq 1.$$

Note that the case $\gamma_j^* = 0$ corresponds to $\varepsilon_j = 1$, i.e., all exceedances in market j are driven by systemic risks. Hence, this zero inflated prior allows us to formally test for the presence of idiosyncratic risks. In the sequel we refer to this test as the *idiocyncracy test*.

It is worthwhile to mention that the representation of the sector-specific F_j as a mixture of a systemic and an idiosyncratic component is reminiscent of the models for dependent random measures discussed in Müller et al. (2004) and Kolossiatis et al. (2011). However, in spite of this connection, the motivation for our modeling approach is quite different. Indeed, while the original motivation in Müller et al. (2004) and Kolossiatis et al. (2011) is to construct dependent random measure, our construction follows from the assumption that the point process associated with the presence of extreme values can be constructed as a superposition of Poisson processes.

4.2.2 Hierarchical Priors

The model is completed by eliciting hyperpriors for all model parameters. For the baseline measure H, we note that the parameter μ in Equation (4.1) is constrained to the $[0,T]$ interval. Hence, a natural choice for H is another rescaled Beta distribution with density

$$h(\mu) = \frac{1}{T}\frac{\Gamma(a_\mu + b_\mu)}{\Gamma(a_\mu)\Gamma(b_\mu)}\left(\frac{\mu}{T}\right)^{a_\mu - 1}\left(1 - \frac{\mu}{T}\right)^{b_\mu - 1}$$

The remaining priors are selected for computational convenience. In particular, a gamma prior with shape parameter a_α and rate parameter b_α is assigned to the precision param-

eter α, leading to a conditionally conjugate model. Similarly, the mixing probability π is given a $\mathsf{Beta}(a_\pi, b_\pi)$ prior, again leading to conditionally conjugate model. Finally, the bandwidth parameter τ is assigned inverse Gamma distributions with shape a_τ and scale b_τ. Details on informative hyperparameter elicitation are discussed in Section 4.3.2.

4.3 Computation and Prior Elicitation

4.3.1 Markov Chain Monte Carlo Algorithms

The likelihood function associated with the non-homogeneous Poisson process giving rise to the exceedances in sector j is simply

$$\exp\left\{-\gamma_j\right\}\gamma_j^{n_j}\prod_{k=1}^{n_j} f_j(t_{j,k})$$

Hence, the joint posterior distribution for our model reduces to

$$\begin{aligned} p(\{\gamma_j^*\},\{v_{j,l}\},\{\tilde{\mu}_{j,l}\},\tau,\{\alpha_j\},\pi \mid \text{Data}) &\propto \prod_{j=1}^{J}\exp\left\{-(\gamma_0^*+\gamma_j^*)\right\}(\gamma_0^*+\gamma_j^*)^{n_j} \\ &\times \prod_{j=1}^{J}\prod_{k=1}^{n_j}\left(\frac{\gamma_0^*}{\gamma_0^*+\gamma_j^*}\sum_{l=1}^{\infty}\omega_{0,l}\psi\left(t_{j,k}\mid\tilde{\mu}_{0,l},\tau\right)+\frac{\gamma_j^*}{\gamma_0^*+\gamma_j^*}\sum_{l=1}^{\infty}\omega_{j,l}\psi\left(t_{j,k}\mid\tilde{\mu}_{j,l},\tau\right)\right) \\ &\times p(\pi)p(\tau)p(\gamma_0^*)\prod_{j=1}^{J}p(\gamma_j^*\mid\pi)\prod_{j=0}^{J}\prod_{l=1}^{\infty}p(\tilde{\mu}_{j,l})p(v_{j,l}\mid\alpha_j)\prod_{j=0}^{J}p(\alpha_j) \quad (4.2) \end{aligned}$$

Since this posterior distribution is computational intractable, we resort to a Markov chain Monte Carlo algorithm (Robert & Casella, 2005) for approximate, simulation-based inference. Given initial values for the different parameters, the algorithm proceeds by iteratively updating blocks of parameters by sampling from their full conditional posterior distribution. After an appropriate burn-in period, the algorithm produces a dependent sample that is approx-

imately distributed according to (4.2). The ergodic theorem ensures that posterior summaries of interest such as posterior means or posterior quantiles can be estimated using these posterior samples.

To sample from the posterior distribution associated with the nonparametric component of the model, we resort to a blocked Gibbs sampler (Ishwaran & James, 2001). Hence, for computational purposes, we replace the nonparametric mixing distributions $G_0^*, G_1^*, \ldots, G_J^*$ with finite-dimensional approximations $G_0^N, G_1^N, \ldots, G_J^N$ where

$$G_j^N(\cdot) = \sum_{l=1}^{N} p_{j,l} \delta_{\mu_{j,l}^\star}(\cdot)$$

where, as before, $\mu_{j,l}^\star$ are i.i.d. samples from baseline measure H for $j = 0, \ldots, J$ and the weights $p_{j,l} = u_{j,l} \prod_{s<l}(1-u_{j,s})$, where $u_{j,l} \sim \mathsf{Beta}(1, \alpha_j)$ for $l = 1, \ldots, N-1$, but $u_{j,N} = 1$ to ensure that the weights sum to 1.

Furthermore, we expand the model by introducing indicator variables $L_{j,k} \in \{1, 2, \ldots, N\}$ and $r_{j,k} \in \{0, j\}$ such that $\Pr(L_{j,k} = l \mid \{u_{j,l}\}) = u_{j,l} \prod_{s<l}\{1-u_{j,s}\}$ and $\Pr(r_{j,k} = 0 \mid \gamma_0^*, \gamma_j^*) = 1 - \Pr(r_{j,k} = j \mid \gamma_0^*, \gamma_j^*) = \gamma_0^*/(\gamma_0^* + \gamma_j^*)$ independently for every $j = 1, \ldots, J$ and $k = 1, \ldots, n_j$. In addition to the aforementioned indicator variables associated with the mixture representation of the intensity functions, we also introduce a set of binary indicators $\xi_1, \ldots, \xi_J$ such that $\Pr(\xi_j = 0 \mid \pi) = 1 - \pi$. Note that inferences on $\Pr(\xi_j = 0 \mid \text{Data})$ provide an operational mechanism to implement the idiosyncrasy test we discussed at the end of Section 4.2.1.

After these indicator variables have been introduced, the MCMC algorithm iterates through the following general steps:

(a) Jointly update each pair $(r_{j,k}, L_{j,k})$ independently by sampling from a multinomial distri-

bution.

(b) Update each stick-breaking weight $u_{j,l}$ by sampling from an appropriate Beta distribution.

(c) Update each $\mu^{\star}_{j,l}$ using a random walk Metropolis-Hasting algorithm with symmetric logit-normal proposal distributions.

(d) Update γ^{*}_{0} from a Gamma distribution.

(e) Jointly update each pair (ξ_j, γ^{*}_{j}) by first sampling ξ_j from a Bernoulli distribution and then either setting $\gamma^{*}_{j} = 0$ if $\xi_j = 0$ or sampling γ^{*}_{j} from a Gamma distribution if $\xi_j = 1$.

(f) Update τ using a random walk Metropolis-Hasting algorithm with symmetric log-normal proposal distributions.

(g) Update π by sampling from Beta distribution.

MCMC Details

Here, we provide the details for posterior simulation from the proposed DP mixture model for the non-homogeneous Poisson process densitie. Simulation is based on the blocked Gibbs sampler, including Metropolis-Hastings steps for some of the parameters. In particular, the MCMC algorithm iteratively updates model parameters through the following steps:

- Each pair $(r_{j,k}, L_{j,k})$ for $j = 1, \ldots, J$ and $k = 1, \ldots, n_j$ is conditionally independent a posteriori and can be updated jointly from a multinomial distribution with

$$\Pr(r_{j,k} = 0, L_{j,k} = l \mid \cdots, \text{Data}) \propto \gamma^{*}_{0} \left[u_{0,l} \prod_{s<l} \{1 - u_{0,s}\} \right] \psi(t_{j,k} \mid \mu^{\star}_{0,l}, \tau),$$

and

$$\Pr(r_{j,k} = j, L_{j,k} = l \mid \cdots, \text{Data}) \propto \gamma_j^* \left[u_{j,l} \prod_{s<l} \{1 - u_{j,s}\} \right] \psi(t_{j,k} \mid \mu_{j,l}^\star, \tau),$$

for $l = 1, \ldots, N$.

- The update for the stick-breaking ratios $\{v_{j,l}\}$ follows the standard recipe described in Ishwaran & James (2001). In particular, these parameters are conditionally independent and their posterior full conditional distribution reduces to

$$v_{j,l} \mid \cdots, \text{Data} \sim \text{Beta}\left(1 + M_{j,l}, \alpha_j + \sum_{r=l+1}^{N} M_{j,r}\right),$$

where $M_{j,l} = \sum_{k=1}^{J} \sum_{i=1}^{n_k} \mathbb{I}_{(L_{k,i}=l, r_{k,i}=j)}$.

- The full conditional distribution for the atoms $\mu_{r,l}^\star$ with $r = 0, \ldots, J$ and $l = 1, \ldots, N$ is given by

$$p(\mu_{r,l}^\star \mid \cdots, \text{Data}) \propto h(\mu_{r,l}^\star) \prod_{\{(j,k): L_{j,k}=l, r_{j,k}=r\}} \psi(t_{j,k} \mid \mu_{r,l}^\star, \tau).$$

Hence, if $\{(j,k) : L_{j,k} = l, r_{j,k} = r\} = \emptyset$, $\mu_{r,l}^\star$ can be updated by simply sampling from the baseline measure H. Otherwise, since this full conditional does not correspond to any known distribution, we update $\mu_{r,l}^\star$ using a random walk Metropolis-Hasting algorithm with symmetric logit-normal proposal distribution where new values $\mu_{r,l}^{\star(p)}$ are generated according to

$$\text{logit}\left\{\mu_{r,l}^{\star(p)}\right\} \mid \mu_{r,l}^{\star(c)} \sim \text{N}\left(\text{logit}\left\{\mu_{r,l}^{\star(c)}\right\}, \kappa_\mu^2\right),$$

where $\text{logit}\left\{\mu_{r,l}^{\star(c)}\right\}$ is the current value of the chain and κ_μ^2 is a tuning parameter chosen so that the average acceptance rate is between 30% and 40%.

- Since the prior for γ_0^* is conditionally conjugate, we update this parameter by sampling from

$$\gamma_0^* \mid \cdots, \text{Data} \sim \text{Gamma}\left(a_{\gamma_0^*} + \sum_{j=1}^{J} m_j, b_{\gamma_0^*} + J\right)$$

where $m_j = \sum_{k=1}^{n_j} \mathbb{I}_{(r_{j,k}=0)}$ denotes the number of observed exceedances in market j that are associated with the systemic component.

- For $j = 1, \ldots, J$, the pairs (ξ_j, γ_j^*) are conditional independent from each other and can be updated by first updating ξ_j so that $\xi_j = 1$ if $m_j < n_j$ or

$$\xi_j \mid \cdots, \text{Data} \sim \begin{cases} \delta_1 & m_j < n_j \\ \text{Ber}\left(\frac{\pi\left\{b_{\gamma_j^*}/(1+b_{\gamma_j^*})\right\}^{a_{\gamma_j^*}}}{1-\pi+\pi\left\{b_{\gamma_j^*}/(1+b_{\gamma_j^*})\right\}^{a_{\gamma_j^*}}}\right) & m_j = n_j \end{cases}$$

As before, $m_j = \sum_{k=1}^{n_j} \mathbb{I}_{(r_{j,k}=0)}$ denotes the number of observed exceedances in market j that are associated with the systemic component. Once ξ_j has been updated, γ_j^* is updated by setting $\gamma_j^* = 0$ if $\xi_j = 0$ or by sampling γ_j^* from a Gamma distribution with shape parameter $a_{\gamma_j^*} + n_j - m_j$ and rate parameter $1 + b_{\gamma_j^*}$ if $\xi_j = 1$.

- The full conditional posterior for τ is

$$p(\tau \mid \cdots, \text{Data}) \propto p(\tau) \prod_{j=1}^{J} \prod_{k=1}^{n_j} \psi(t_{j,k} \mid \mu^{\star}_{r_{j,k}, L_{j,k}}, \tau).$$

Since no direct sampler is available from this distribution, we update τ using a random walk Metropolis-Hasting algorithm with symmetric log-normal proposals,

$$\log\left\{\tau^{(p)}\right\} \mid \tau^{(c)} \sim \text{N}\left(\log\left\{\tau^{(c)}\right\}, \kappa_\tau^2\right)$$

where κ_τ^2 is again a tuning parameter.

- Because the prior on π is conditionally conjugate, the posterior full conditional for π is a beta distribution,

$$\pi \mid \cdots, \text{Data} \sim \text{Beta}\left(J - \sum_{j=1}^{J} \xi_j + a_\pi, \sum_{j=1}^{J} \xi_j + b_\pi\right).$$

- The precision parameters $\alpha_0, \alpha_1, \ldots, \alpha_J$ can be updated independently using the algorithm described in Escobar & West (1995).

4.3.2 Hyperparameter Elicitation

The hyperparameters associated with our nonparametric model can be selected using historical and/or expert information that is typically available for most liquid financial markets. We recommend that this elicitation process be complemented with a careful sensitivity analysis over a reasonable range of prior beliefs.

Consider first the parameters $\gamma_0^*, \gamma_1^*, \ldots, \gamma_{10}^*$, which control the total number of exceedances observed in each market and the relative distribution of these exceedances between the systemic and idiosyncratic component of the model. Because of their role in the model, we can elicit a value for the expected number of extremes in a given sector j (which corresponds to $\mathsf{E}\{\gamma_0^* + \gamma_j^*\}$) by assuming that returns in the sector are normally distributed, so that $\mathsf{E}\{\gamma_0^* + \gamma_j^*\} \approx T\Phi(\{u - \zeta_j\}/\kappa_j)$, where Φ denotes the cumulative distribution function for the standard normal distribution and ζ_j and κ_j are rough estimates of the mean and standard deviation of returns for market j. The values of ζ_j and κ_j can be obtained from historical data or expert knowledge. For simplicity, it can be assumed that ζ_j and κ_j are the same for every market, leading to a model where sectors are exchangeable, but this is not required. Similarly,

we can exploit the interpretation of $\gamma_0^*/(\gamma_0^*+\gamma_j^*)$ as the proportion of exceedances arising from the systemic component to elicit expert information about the most likely value of such rate, as well as a high probability range for its value. This same information can be used to provide informative priors for $1-\pi$, the prior probability that the risk in a given market is entirely driven by the systemic component.

Consider now eliciting the hyperparameters associated with the density function f_0^* and $f_1^*,\ldots,f_J^*$. A common feature of extreme returns in financial time series is that they tend to cluster over time (e.g., Mandelbrot, 1963). Hence, the prior for the precision parameter $\alpha_0,\alpha_1,\ldots,\alpha_J$ (which, as mentioned in Section 4.2, controls the number of components in the mixture) should favor a multimodal distribution. A rough value for the number of components (which can be used to select the prior mean of α_j) can be elicited from a rough estimate of the frequency at which distress periods arise in market j. Similarly, the value for the bandwidth parameter τ can be elicited from prior information about the length of distress periods. Finally, in the absence of prior information about the time at which distress periods occur, we recommend that H be selected so that the prior mean for f_0^* and $f_1^*,\ldots,f_J^*$ is close to uniform.

4.4 An Application to the Returns of the S&P500 Sectors

The Standard & Poor's 500, or S&P500 index, is a commonly watched stock market index in the U.S. The S&P500 is constructed as a market-value weighted average of the prices of the common stock of 500 publicly traded companies. Standard & Poor's, which publishes the index, selects the companies included in the S&P500 index to be representative of the industries

in the U.S. economy. These companies are commonly grouped into ten economic sectors (consumer discretionary, consumer staples, energy, financials, health care, industrials, materials, information technology, telecommunication services and utilities), with the largest (consumer discretionary) including 81 companies and the smallest (telecommunication services) including only 8. In addition to the overall S&P500 index, Standard & Poor's publishes separate indexes for each of these sectors. The behavior of these sector-specific indexes is of independent interest; for example, the performance of the industrial component of the S&P500 is sometimes used by analysts as a leading indicator of future economic growth. More generally, it is clear that different sectors react differently to the same economic shocks.

The data analyzed in this section corresponds to the negative log returns above 2% on each of the ten sectors that make up the S&P500 index between January 1, 2000 and December 31, 2011. Prices for the individual indexes were obtained from Bloomberg financial services; the corresponding tickers are S5COND, S5CONS, S5ENRS, S5FINL, S5HLTH, S5INDU, S5MATR, S5INFT, S5TELS, S5UTIL. All inferences reported in this section are based on 3,000 quasi-independent samples obtained after a burn in period of 20,000 iterations and thinning of the original chain every 50 iterations. Convergence of the MCMC algorithm was monitored using trace plots as well as the R statistic discussed in Gelman & Rubin (1992). In particular, we ran four independent chains started from overdispersed initial values and compared between and within chain variability in the value of the log likelihood function and some of the hyperparameters in the model. No lack of convergence was evident from these diagnostics. The algorithm was implemented in `C/C++` and total execution time was approximately 16 hours on a MacBook laptop with a 2 GHz Intel Core 2 Duo processor and 2GB of memory.

Following the approach described in Section 4.3.2, we chose the hyperparameters for γ_0^* and γ_j^* so that $a_{\gamma_0^*} = 7.32$, $b_{\gamma_0^*} = 0.06$, $a_{\gamma_0^*} = 1.32$, and $b_{\gamma_0^*} = 0.06$. These values were chosen on the basis of a 0 mean return with 18% annualized volatility for the S&P500, along with the prior beliefs that on average 85% of the observed exceedances, and with .99 probability at least 50% of them, arise from the systemic component of the model. These numbers reflect our prior belief that the systemic component of the risk will explain a majority of the exceedances observed in the data. This is further emphasized by our choice for the hyperparameters for π, which are selected as $a_\pi = 0.5$ and $b_\pi = 2$. This choice implies that $E\{\pi\} = 0.2$ and places a high probability on values of π close to zero. Finally, the prior for the bandwidth parameter τ was selected so that $a_\tau = 5$ and $b_\tau = 2{,}400$ (which implies that $\mathsf{E}\{\tau\} = 400$). For the precision parameter α we choose $a_\alpha = 4$ and $b_\alpha = 1/3$ (so that $E\{\alpha\} = 12$, leading to highly multimodal intensity functions a priori), and the parameters for the baseline measure $a_\mu = 1$ and $b_\mu = 1$. In all cases the posterior distributions for all hyperparameters appear to be concentrated relative to the corresponding prior distributions. In addition, we note that moderate changes in these assumptions (e.g., assuming that on average only 50% of the exceedances come from the systemic component, or a 2% positive annualized return for the S&P500 with a 25% annualized volatility, or a uniform prior on π) lead to essentially equivalent posterior inferences for the intensity functions. The only inferences that are somewhat affected by moderate changes in the hyperparameters are those associated with the idiosyncrasy test discussed in Section 4.2 (for further details on the sensitivity analysis see Section 4.4.2).

Estimates of the overall intensities $\lambda_1(t), \ldots, \lambda_{10}(t)$ associated with each of the ten components of the S&P500 index can be seen in Figure 4.2. The last two panels also provide

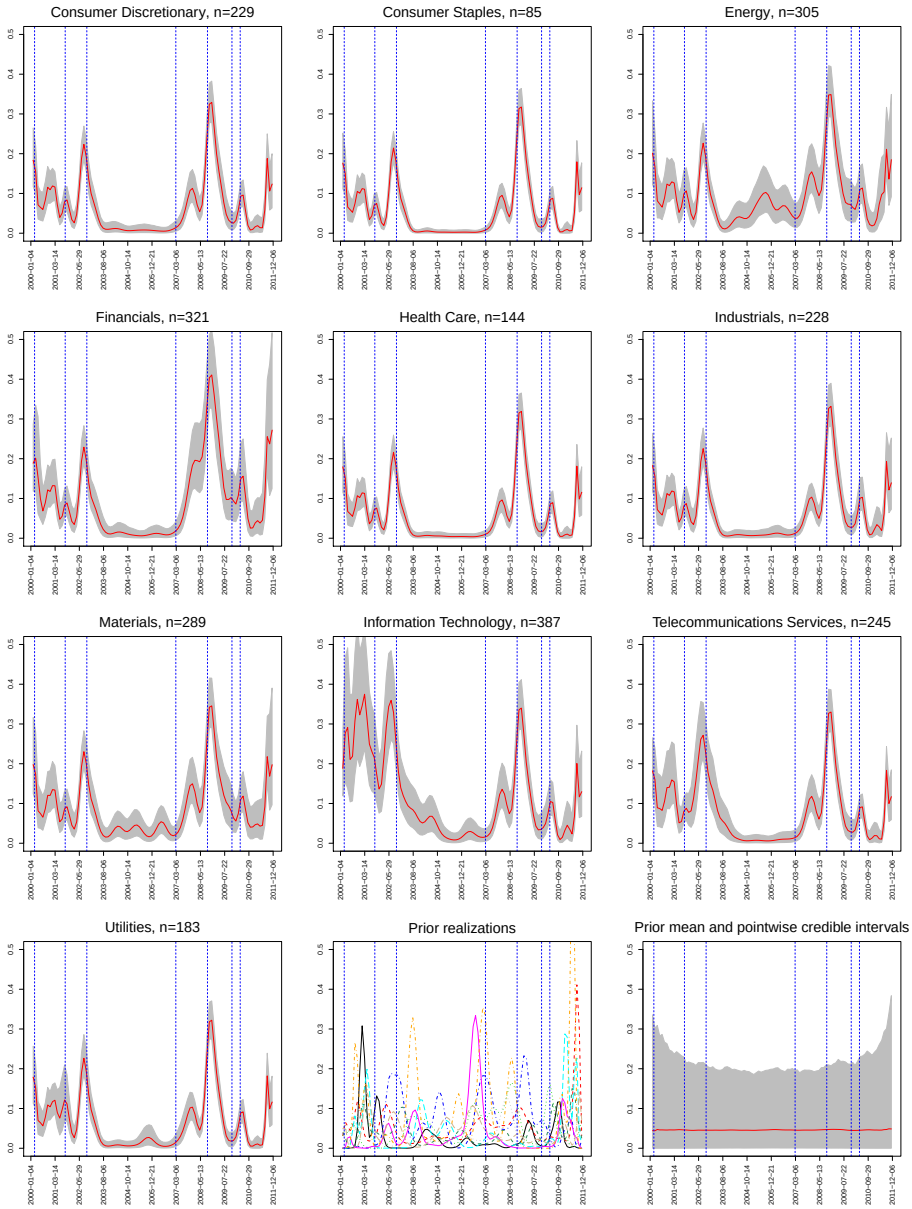


Figure 4.2: Posterior mean of the overall intensity associated with the different components of the S&P500 index, along with posterior pointwise credible intervals. The headers on each panel include the number exceedances observed in each sector over the 12 year period under study. The last two figures in the bottom line present realizations from the prior distribution on the intensity function (central panel) along with the mean prior intensity function and prior 95% pointwise credible intervals (right panel).

summaries of the prior distribution over intensities that are induced by the prior choices discussed before. By comparing some of those estimates to the raw data presented in Figure 4.1, it becomes clear that the model faithfully reproduces the main features of the data. Furthermore, the uncertainty associated with these estimates is relatively low.

Next, Figures 4.3 and 4.4 show estimates of the densities f_0^* and $f_1^*, \ldots, f_{10}^*$ associated with the systemic and idiosyncratic risk intensities. In addition, Figure 4.5 presents the posterior distribution for $\varepsilon_1, \ldots, \varepsilon_{10}$, the proportion of the risk attributable to the systemic component in each of the ten sectors. Note that in half the sectors (consumer discretionary, consumer staples, health care, industrials and utilities) the proportion of extremes associated with the systemic component is at least 80%, while for the rest (energy, financials, information technology, telecommunications and materials) the proportion is between 40% and 60%. In addition, note that the density for the systemic risk shows peaks that coincide, or shortly follow, important stock market events. On the other hand, the behavior of the idiosyncratic risk varies drastically with the economic sector and, in most cases, can be explained by factors that are clearly sector-specific. For example, the energy and utilities sectors present big increases in idiosyncratic risk during 2005, a period that corresponded to sharp increases in oil prices but that was otherwise relatively calm. On the other hand, the idiosyncratic risk associated with the financial sector increases dramatically after the summer of 2007. An oversized idiosyncratic risk for this sector after 2007 is clearly reasonable as financials were the main driver of the recent crises. Similarly, the idiosyncratic risks associated with the information technology and telecommunication services sectors are particularly elevated between 2000 and 2002, a period that included the bursting of the so-called dot-com bubble. Finally, note that the idiosyncratic risk associated

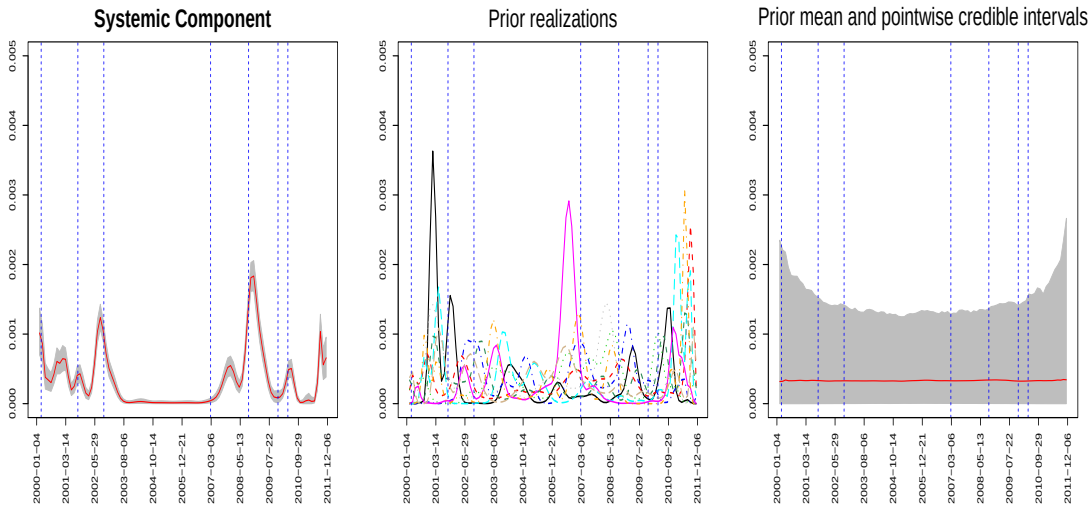


Figure 4.3: The left panel shows the posterior mean of the density associated with systemic risk component of the S&P500 index, including posterior pointwise credible intervals. The central panel shows realizations from the prior density function, while the right panel shows the mean prior density function and prior 95% pointwise credible intervals.

with consumer staples is almost negligible over the whole period under study, with our idiosyncrasy test suggesting that there is moderate evidence for the absence of idiosyncratic risk in this sector of the S&P500 index. This is reasonable, as the consumer staples sector includes companies that produce and trade basic necessities whose consumption might be affected by general market trends but have otherwise little sensitivity to intrinsic risk factors. Figure 4.6 shows the histograms of the posterior distribution of the hyperparameters α_0, α_j, and τ with the prior densities provided in red. We observed moderate to sufficient learning for τ and precision parameter in sectors having a larger proportion of extremes associated with the idiosyncratic component. As expected, there was less learning for precision parameter in other sectors with relatively small number of extremes associated with the idiosyncratic component. Posterior densities for these parameters is shown to concentrate around the corresponding prior densities.

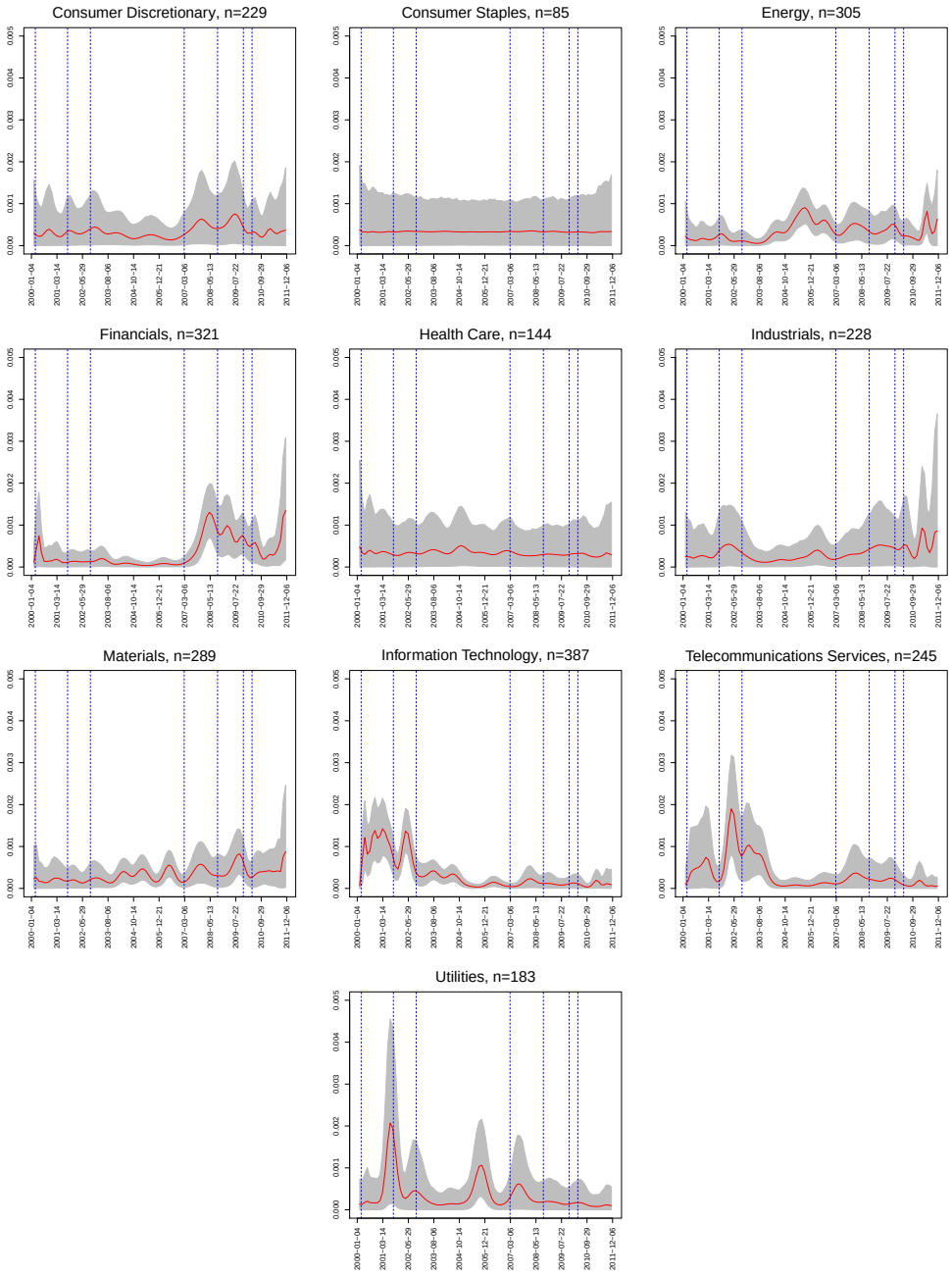


Figure 4.4: Posterior mean of the densities associated with the different components of the S&P500 index, along with posterior pointwise credible intervals.

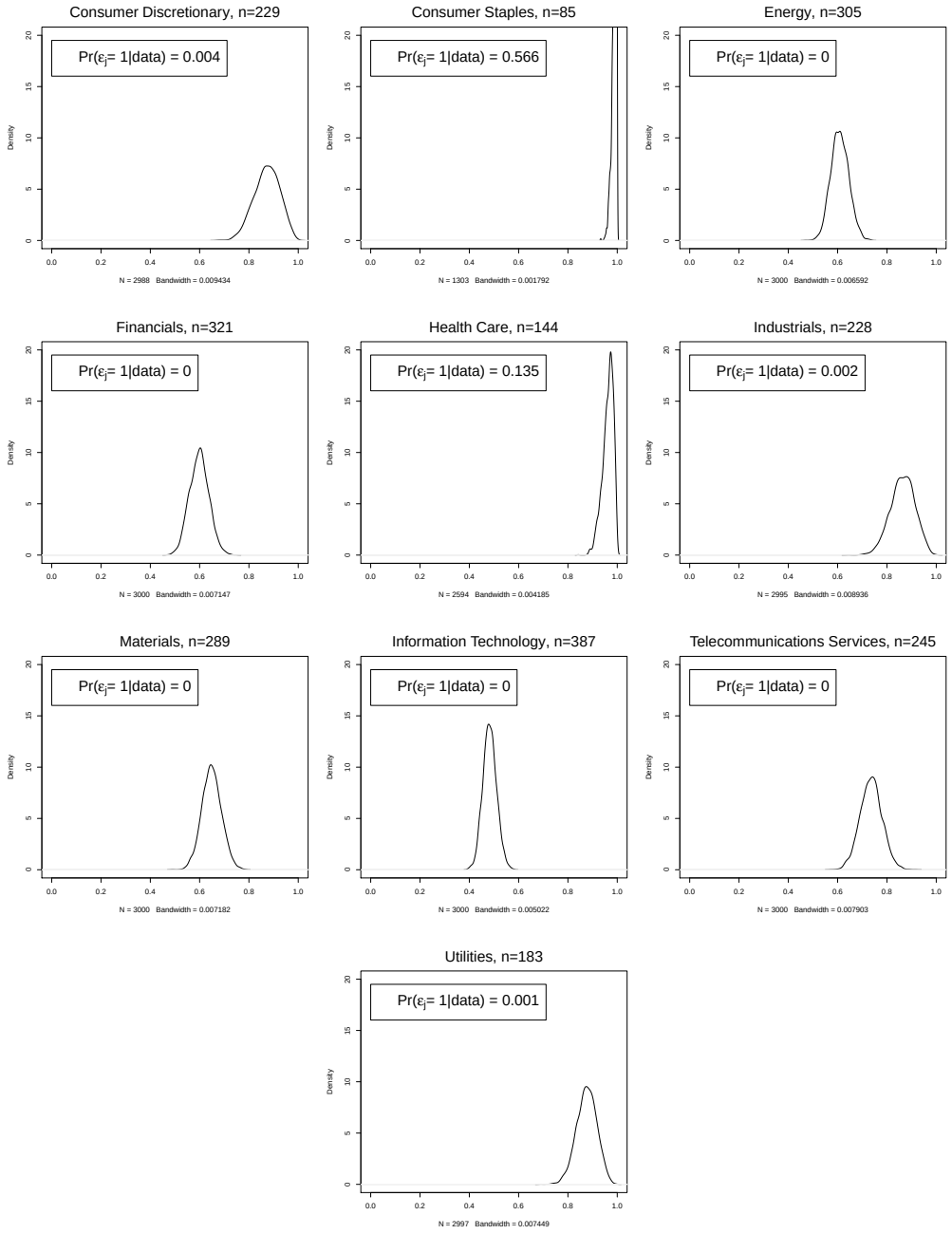


Figure 4.5: Posterior distribution for the overall proportion of risk attributable to the systemic component on each of the ten components of the S&P500 index.

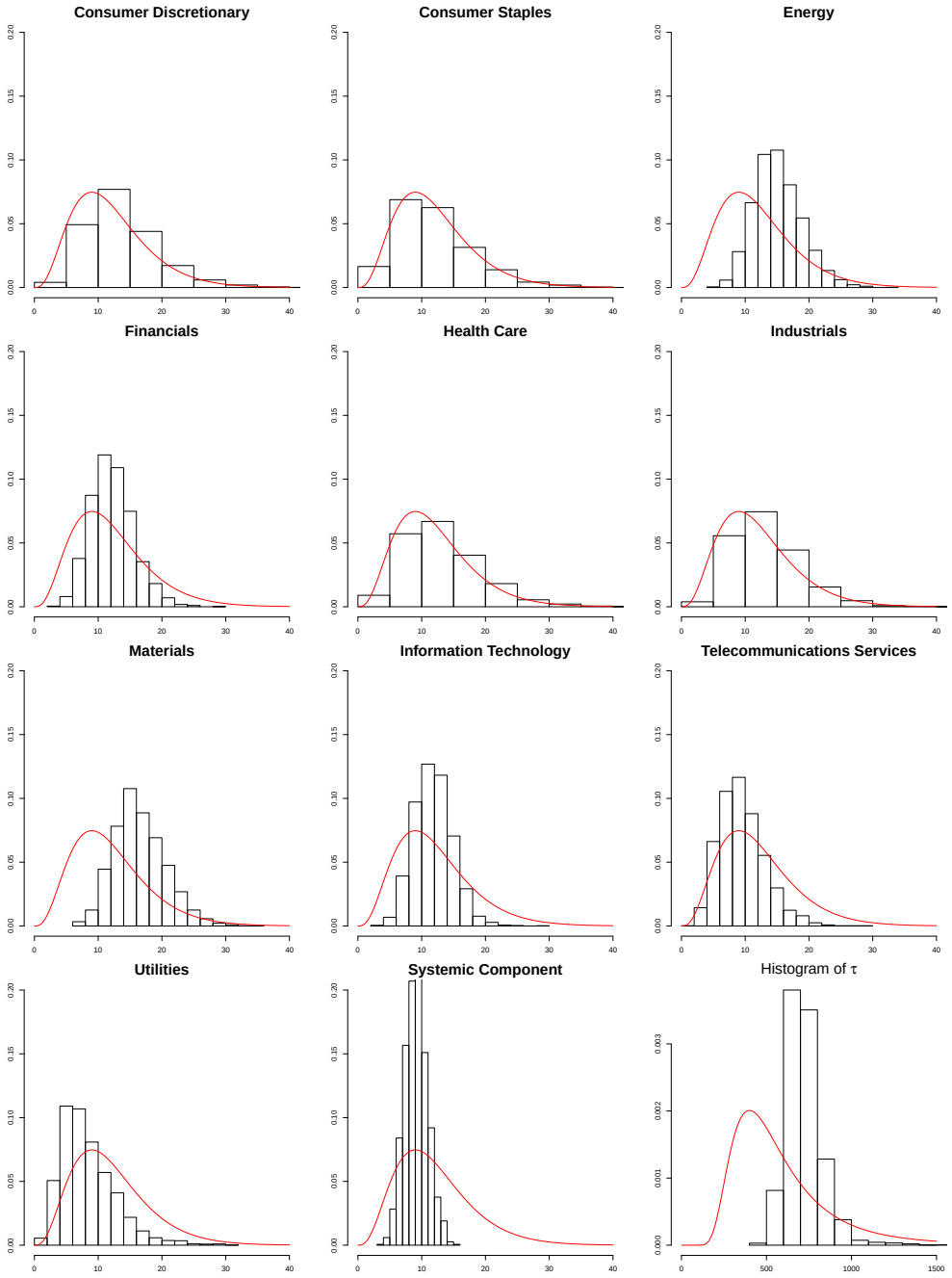


Figure 4.6: Posterior distribution of the precision parameter α_j in market j, α_0, and bandwidth parameter τ. Prior densities are provided in red line.

As we discussed in Section 4.2, we can alternatively quantify the level of risk through the probability of observing at least one exceedance during a given period of time. Figure 4.7 shows the posterior distributions for the odds ratio of the probability of at least one exceedance in the month starting two weeks after the bankruptcy of Lehman Brothers versus the probability of at least one exceedance in the month ending two weeks before the bankruptcy for four different sectors. Note that all sectors show an increase in risk after Lehman Brothers bankruptcy. However the increase in risk is lower for financials than it is for all the sectors (the estimated posterior probabilities are 1.000, 0.913 and 0.977 for consumer staples, energy and information technology, respectively). Indeed, note that the systemic risk increases after the bankruptcy of Lehman Brothers but the idiosyncratic risk associated with financials actually decreases (as does the one for energy, although to a lesser degree), while the idiosyncratic risks associated with information technology and consumer staples increased. The increase in risk in the information technology and consumer staples sectors can be explained by the fact that one of the main effects of Lehman's bankruptcy was a collapse in the short term corporate debt market. Hence, although the bankruptcy of Lehman Brothers actually reduced the uncertainty in the financial sector of the economy, it cause real damage to companies in other sectors that are extremely dependent on short term debt. Note that companies that are part of the S&P500 energy sector are typically not reliant in short term funding, hence the limited impact of Lehman's bankruptcy in their idiosyncratic risk.

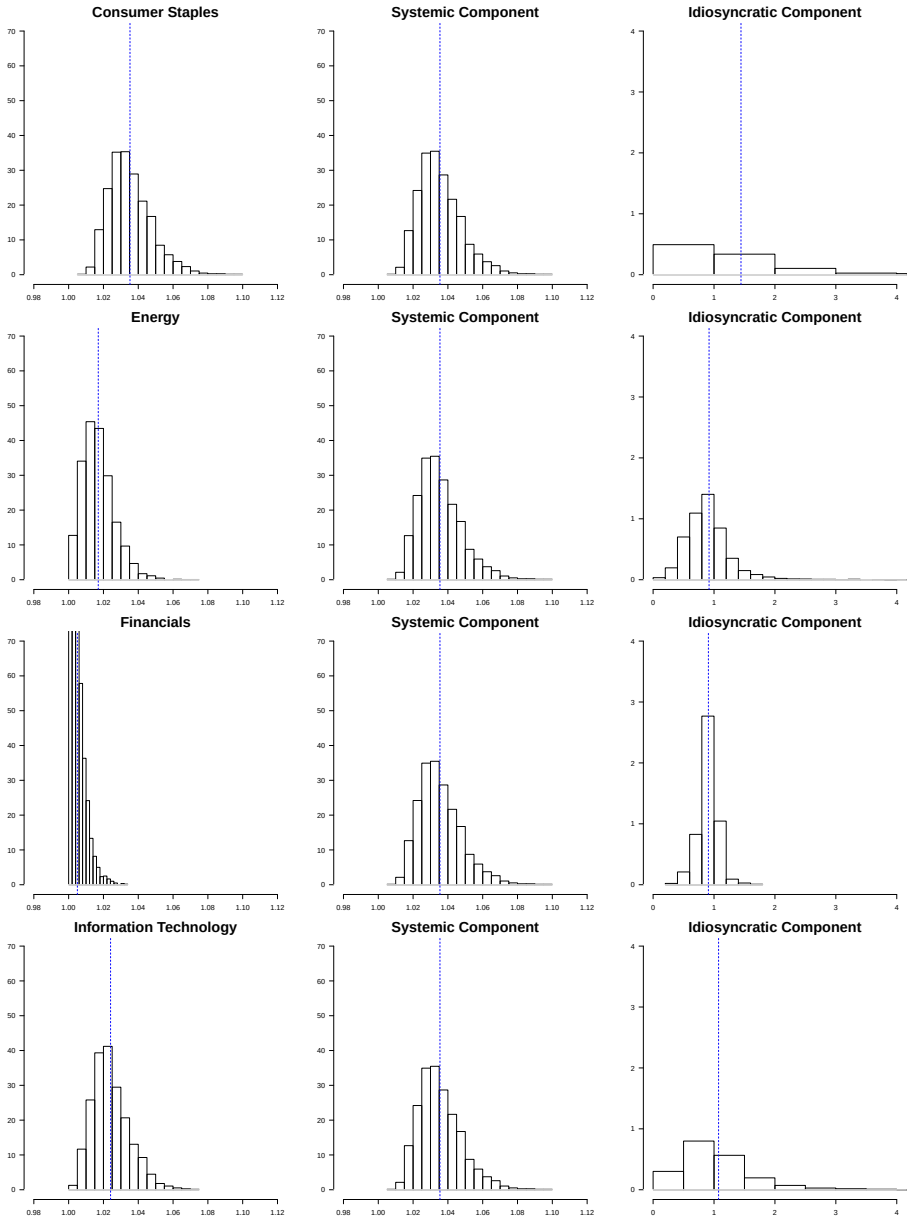


Figure 4.7: Posterior distributions for the odds ratio of the probability of at least one exceedance in the month starting two weeks after the bankruptcy of Lehman Brothers against the probability of at least one exceedance in the month ending two weeks before the bankruptcy for four different sectors. The vertical line corresponds to the mean of the posterior distribution.

4.4.1 Model Validation

The model was validated in this dataset using two different approaches. First, an out-of-sample cross validation exercise was conducted, with cross-validation datasets being constructed by randomly selecting 20% of the observations from each sector to be used as hold-out data. The remaining 80% of the data was used to fit our nonparametric model and generate nominal 90% highest posterior densities (HPD) intervals for new exceedances. The true coverage of these HPD intervals is then evaluated on the held-out data. Figure 4.8 presents examples of cross-validation samples and the corresponding intensities for ten sectors.

We repeated the process described above for 10 different cross-validation datasets, with the results being presented in Figure 4.9. As expected, there is variability in the coverage rates depending on the sector and the specific cross-validation dataset. However, the results suggest that for the most part the real coverage rates are in line with the nominal coverage, which suggest that the model does not under or overfit.

In addition to the cross-validation exercise described above, in-sample goodness of fit was investigated using quantile-quantile plots for the posterior distribution of inter-event times. More specifically, we use the similar strategy discussed in Section 3.4.2 to check the Poisson process assumption in time direction. Figure 4.10 presents quantile-quantile plots of expected value of these transformed inter-arrival times, $\mathsf{E}\{z_k \mid \text{Data}\}$ for $k = 1, \ldots, n_j$, against the quantiles of a uniform distribution for each of the ten S&P500 sectors. For the most part the residuals follow a straight diagonal line. However, there is some evidence of poor fit for a couple of sectors. In particular, note that for consumer staples our model tends to systemati-

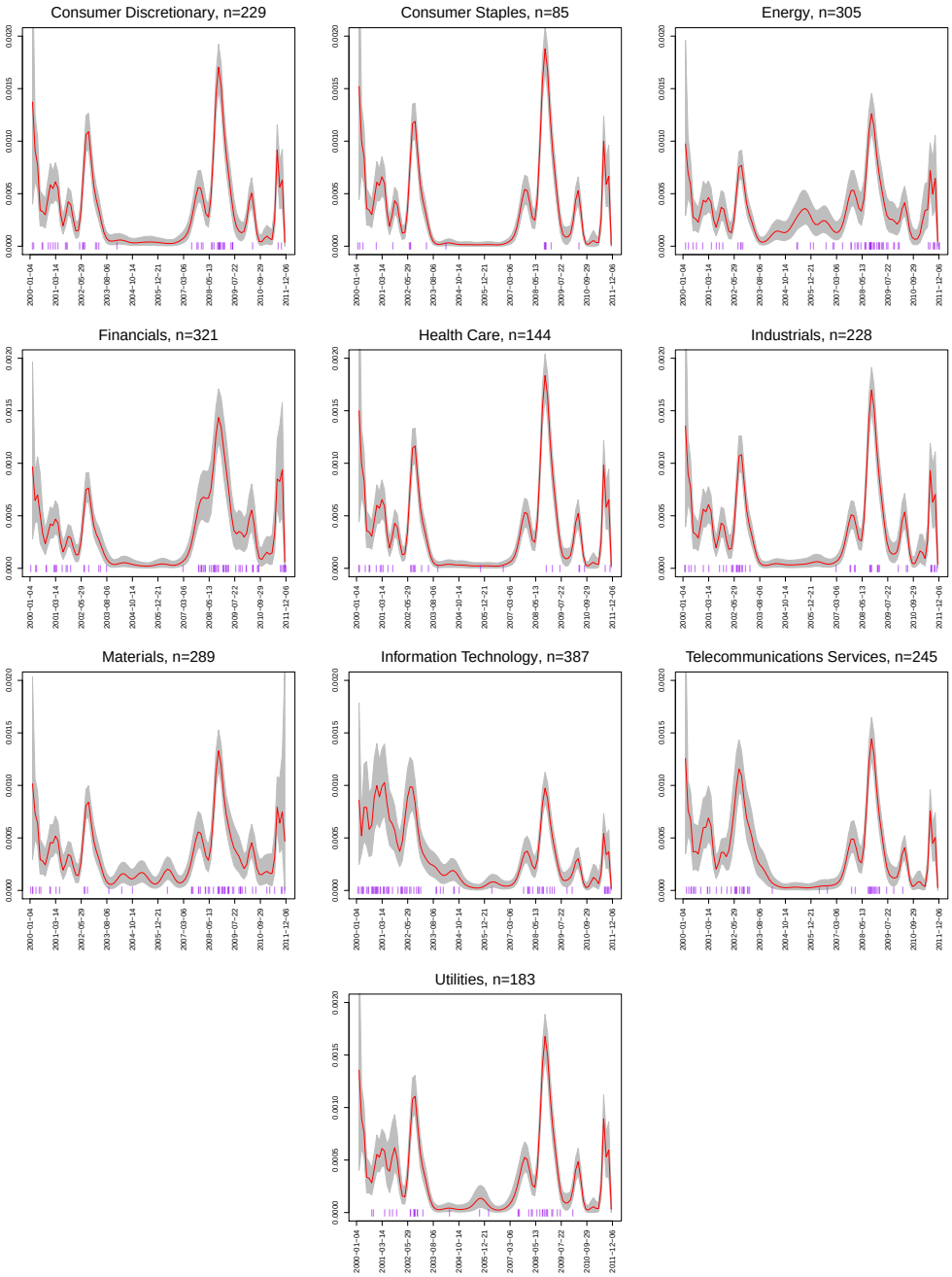


Figure 4.8: Examples of cross-validation datasets and the density estimates associated with them. The dots in the horizontal axis correspond to the held-out samples.

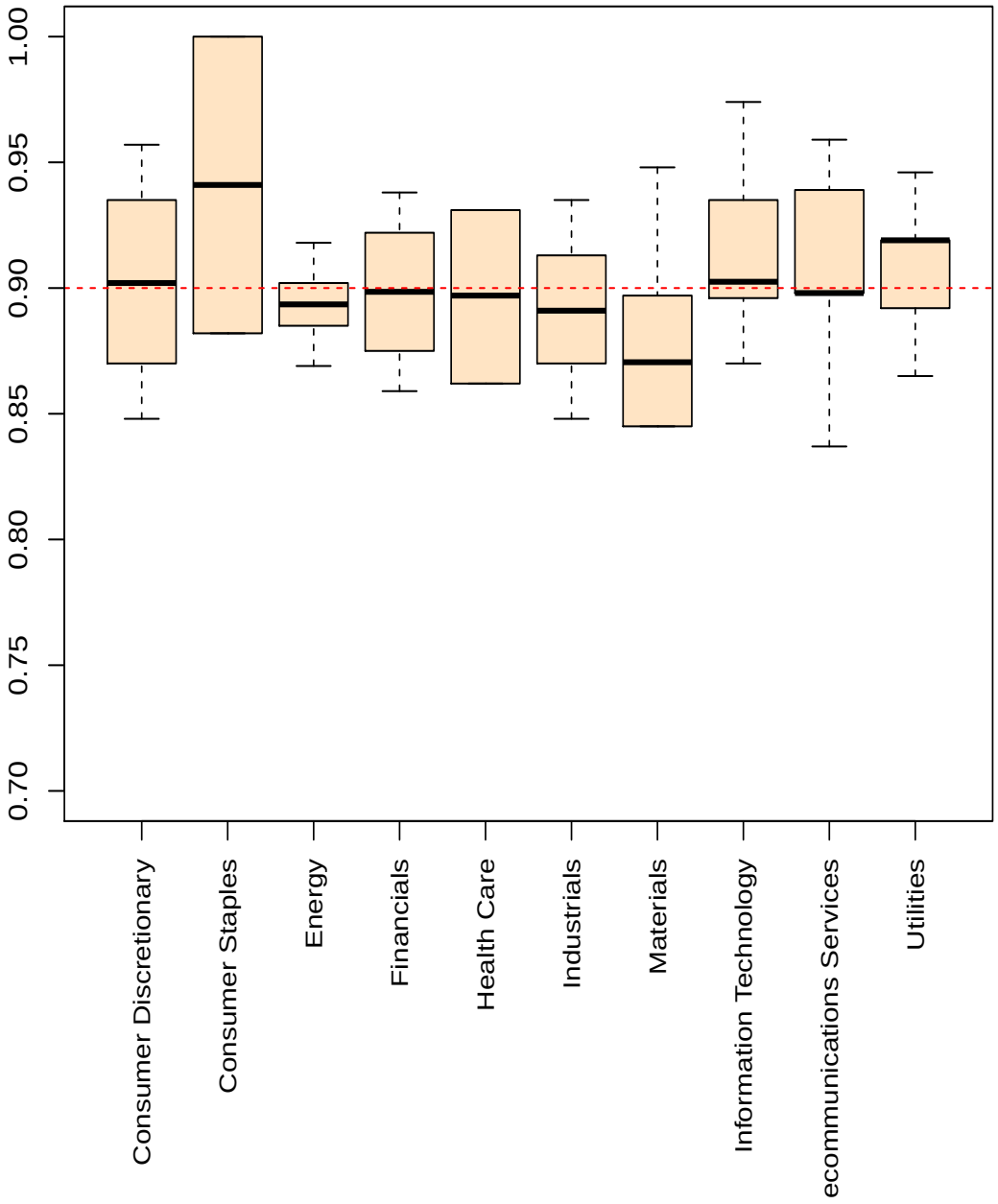


Figure 4.9: Results from the cross validation exercise to investigate the coverage rate of highest posterior density intervals associated with our nonparametric model.

cally predict somewhat shorter inter-arrival periods than those that would be expected under the Poisson model. Similar, but less dramatic biases can also be seen for information technology and financials.

4.4.2 Sensitivity Analysis

We carried out a comprehensive sensitivity analysis to assess the effect of prior distributions on posterior inferences. First, we considered three alternative sets of hyperparameters for the bandwidth parameter τ, including a $\mathsf{InvGamma}(5,4000)$, a $\mathsf{InvGamma}(10,7200)$ and a $\mathsf{InvGamma}(2,500)$ priors. The hyperparameters were selected to represent a range of situations where the prior mean is both larger and smaller than the one used for our previous analysis, as well as different levels of concentration. Posterior inferences were mostly unaffected under any of these scenarios.

Next, we considered four alternative prior specifications for the concentration parameters $\alpha_0,\ldots,\alpha_{10}$, including a $\mathsf{Gamma}(4,1/3)$, a $\mathsf{Gamma}(10,2)$, a $\mathsf{Gamma}(2,2/5)$, and a $\mathsf{Gamma}(3,3)$. These hyperparameter choices imply prior expected values for α_j of 12, 5, 5 and 1, respectively. Inferences for the intensity function were mostly unchanged under these prior distributions. However, inferences for individual hyperparameters were somewhat affected. In particular, smaller values for $\mathsf{E}\{\alpha_j\}$ naturally lead to somewhat smaller posterior means for the α_js, but also to larger posterior values for τ and an increase in the posterior mean for some of the ε_js (for example, for consumer staples we have $\Pr(\varepsilon_2 = 1 \mid \text{Data}) = 0.71$ under the $\mathsf{Gamma}(3,3)$ prior). On the other hand, changes in the prior dispersion of α_j had no discernible effect on the individual posterior distributions of the hyperparameters, as long as the prior mean was the

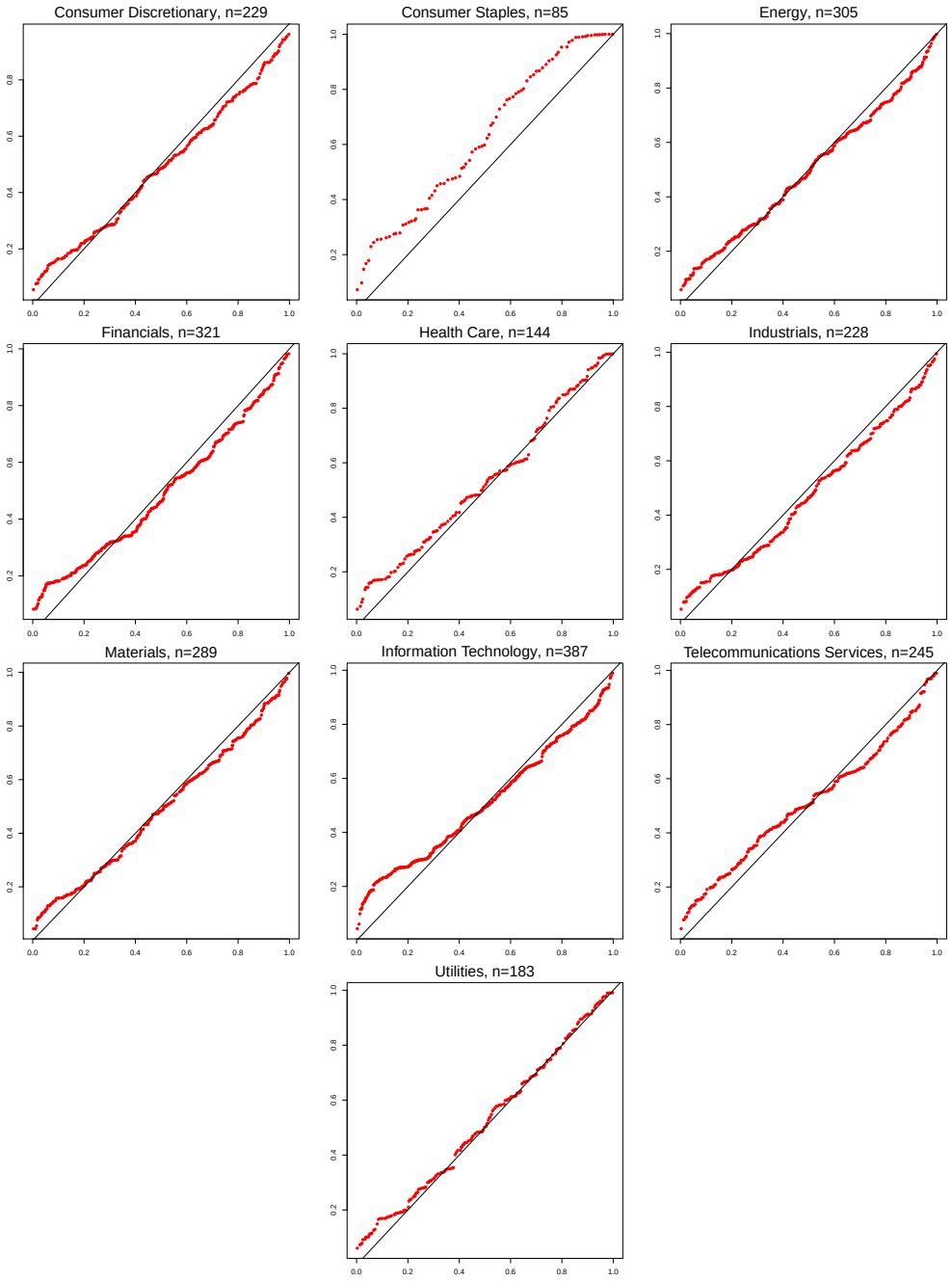


Figure 4.10: Quantile-quantile plot of expected value of transformed inter-arrival times against the quantiles of a uniform distribution for each of the ten S&P500 sectors. The red dots corresponds to the posterior mean of the expected value of transformed inter-arrival times.

same.

To asses the effect of the baseline measure H on posterior inferences, we considered an alternative rescaled beta distribution with $a_\mu = b_\mu = 3$. This prior tends to favor the localization of distress periods towards the middle of the time series. This alternative baseline measure leads to somewhat smoother estimates for the density functions $f_1, \ldots, f_{10}$, and to more unimodal estimates for the idiosyncratic densities $f_1^*, \ldots, f_{10}^*$. In addition, the posterior means for α and τ tend to be slightly lower under this baseline measure.

We also considered an alternative specification for the priors on γ_0^* and $\gamma_1^*, \ldots, \gamma_{10}^*$ where we assume that the number of extreme events in each sector is consistent with a 2% positive annualized return and a 25% annualized volatility for the S&P500, while only 50% of the exceedances come from the systemic component and $\Pr(0.2 < \gamma_0^*/(\gamma_0^* + \gamma_j^*) < 0.8) = 0.99$. This leads to $a_{\gamma_0^*} = 7.65$, $b_{\gamma_0^*} = 0.65$, $a_{\gamma_j^*} = 7.65$ and $b_{\gamma_j^*} = 0.65$ for $j \geq 1$. As before, these new priors have very little impact on the inference of the intensity functions. However, the inferences on the weights $\varepsilon_1, \ldots, \varepsilon_{10}$ are significantly affected. In particular, the idiosyncrasy test now provides very strong evidence that shocks in consumer staples are driven exclusively by the systemic component ($\Pr(\varepsilon_2 = 1 \mid \text{Data}) \approx 1$), while for the other sectors we have very strong evidence for the presence of idiosyncratic components ($\Pr(\varepsilon_j = 1 \mid \text{Data}) = 0$ for all $j = 1, 3, 4, \ldots, 10$).

Finally, we investigated the effect on posterior inferences of alternative prior distributions on π. In addition to the original $\mathsf{Beta}(1/2, 2)$ prior (which favors the hypotheses that most the exceedances are generated by the systemic component of the model), we considered a uniform and a $\mathsf{Beta}(1/2, 1/2)$ prior for π. While the $\mathsf{Beta}(1/2, 1/2)$ had a negligible effect on

posterior inferences, the use of a uniform prior lead again to an increase in the posterior mean for some of the ε_js (for example, $\Pr(\varepsilon_2 = 1 \mid \text{Data}) = 0.76$).

Posterior probabilities for all hyperparameters appear to be more concentrated relative to the prior distribution.

4.5 Discussion

We have discussed and illustrated a novel class of model for simultaneous risk assessment in multiple financial markets. Our approaches models the process of exceedances as a superposition of two stochastic processes. The use of a superposition is specially appealing if markets share similar products. This is the situation, for example, when modeling the returns from multiple mutual funds or exchange-traded funds. In that case, the interpretation of a systemic component is straightforward, as it simply reflects the fact that these different products invest in some of the same securities. However, the idea is justifiable in a much more general setting. Indeed, the systemic component can be interpreted as representing global shocks that affect markets simultaneously.

Another interesting application of the model described here in the world of finance is to reduced-form credit risk models, which also use Cox processes to model credit default events (for example, see Lando, 1998). Extending our model to jointly estimate default probabilities over multiple sectors would be relatively straightforward.

The approach discussed in this paper focuses on modeling the frequency of extreme values and pays no attention to the size of losses (except in that they are greater than the thresh-

old u). Introducing the size of the losses by modeling a bivariate Poisson process (along the lines discussed in Chapter 3) is conceptually a relatively straightforward extension that will also be considered elsewhere.

Chapter 5

Nonparametric Spatial Modeling for Extreme Values from Environmental Processes

5.1 Introduction

Extreme value analysis plays a key role for a number of problems in the environmental sciences. Extreme natural phenomena, such as severe droughts, unusually low temperatures or torrential rains, are rare but catastrophic events, which can result in large economic losses and high cost in human life. Therefore, risk analysis to quantify the uncertainty associated with such extreme events is both scientifically relevant and practically important for effective environmental policy making.

In extreme value analysis for environmental problems interest lies in very large or very small values of variables associated with a physical process, which is typically recorded over both time and space. Statistical inference and prediction for rare events is complicated

by the fact that observations corresponding to the center of the distribution, which are the most abundant, carry little information about the tails. This implies particular challenges for the study of dynamical variations of the process under study. For instance, rainfall records could show a steady average behavior over time, while the amount of rainfall of the largest storms may be increasing. Capturing also spatial dependence for processes observed at a number of monitoring stations adds to the challenge for modeling and inference.

The literature on extreme value analysis for independent and identically distributed observations is fairly well developed; see details in Section 2.1. However, there is a relatively smaller collection of modeling methods for extremes from stochastic processes evolving over time and space, although this is an active research area in the more recent literature. The Bayesian paradigm offers clear advantages in this setting, since it allows exploration of flexible hierarchical model formulations and proper incorporation of full predictive uncertainty. The main theme of Bayesian modeling approaches has been to extend in a hierarchical fashion the parametric distributions used in extreme value analysis. In particular, the observed block maxima or threshold exceedances are typically assumed to arise conditionally independent from the generalized extreme value or generalized Pareto distribution, respectively, with temporally and spatially dependent parameters. Common approaches to introduce the spatio-temporal dependence to the parameters include dynamic linear models (Huerta & Sansó, 2007) and Gaussian processes (Cooley et al., 2007; Sang & Gelfand, 2009). The approach in Sang & Gelfand (2010) fits within the same framework, but relaxes the conditional independence assumption in the first stage of the hierarchical model. Alternatively, one may consider max-stable processes (Smith, 1990; Schlather, 2002; Reich & Shaby, 2012), which provide another

natural generalization of the generalized extreme value distribution. Although this approach is appealing from a theoretical perspective, likelihood-based inference for max-stable processes is difficult, since in most cases no closed-form expression for the likelihood is available when more than two or three locations are involved. Hence, most inference procedures for max-stable processes employ composite likelihoods (Padoan et al., 2010; Genton et al., 2011), which are unappealing from a Bayesian perspective. In addition, there are other more recent contributions include copula-based semiparametric methods (Fuentes et al., 2012).

In this chapter, we build on the Bayesian nonparametric modeling framework developed in Chapter 3, where a mixture model for the non-homogeneous Poisson process intensity has been developed to overcome the restrictive aspects of the standard parametric form, most notably, the homogeneity for the intensity of exceedance times. Here, we focus on the time dimension under the bivariate NHPP approach and study the practically important methodological extension on spatial modeling for the exceedance time intensities. Our objective is to retain inferential flexibility for the temporal intensity while incorporating nonparametric spatial dependence into the modeling. To this end, we represent the NHPP density at each site through a mixture of logit-normal kernels, and use a spatial Dirichlet process for the mixing distributions to drive the nonparametric (non-Gaussian and non-stationary) spatial dependence. A prior probability model for the spatial surface of total exceedance intensities completes the model specification. We develop methods for Markov chain Monte Carlo posterior simulation, and for spatial interpolation of risk assessment quantities for high-level exceedances.

Our illustrative data analysis involves rainfall exceedances, using data from the Cape Floristic Region located in the southwestern coastline of South Africa. The times of exceedances

are based on daily precipitation records from monitoring stations across South Africa between year 1950 to 1999. However, the proposed methodology is generally applicable to modeling and spatial prediction of threshold exceedances from different types of environmental variables.

The outline of this chapter is as follows. In Section 5.2, we develop the spatial nonparametric modeling approach, including the detailed nonparametric spatial model formulation for extremes from environmental time series. In Section 5.3, we discuss the implementation details regarding posterior simulation and inference. Section 5.4 illustrates the methodology using a simulated data example and the rainfall data. Finally, Section 5.5 concludes with a summary.

5.2 Modeling Approach

The nonparametric mixture modeling framework outlined in Chapter 3.2.2 combines the appealing features of the point process approach to extreme value analysis with the inferential power of Bayesian nonparametric prior models. To our knowledge, the approach proposed in Chapter 3 provides the first attempt to fully nonparametric modeling for extremes from a single time series, with flexible resulting inference for the joint intensity of extremes, the marginal intensity over time, and for different types of return level curves.

Here, we study more general spatio-temporal data structures involving threshold exceedances from environmental processes observed at multiple spatial locations over a certain time interval (which, again, without loss of generality is transformed to $[0,1]$). More specifically, let $\mathcal{S} \subset \mathbb{R}^2$ be the geographic region under study, and $\mathbf{s}_{\text{obs}} = (\mathbf{s}_1, \ldots, \mathbf{s}_m)$ the m distinct locations in $\mathcal{S}$ where the process is observed. Hence, the full data set comprises $\{(t_i(\mathbf{s}_j), Y_i(\mathbf{s}_j)) :$

$i = 1, ..., n_j$; $j = 1, ..., m\}$, where $n_j \equiv n_{\mathbf{s}_j}$ is the number of threshold exceedances at location $\mathbf{s}_j$, $t_i(\mathbf{s}_j)$ is the time at which the i-th exceedance occurred at location $\mathbf{s}_j$, and $Y_i(\mathbf{s}_j)$ is the value of that exceedance. For such problems, it is of interest to explore spatial modeling extensions for the NHPP intensity of extremes while retaining the flexibility of a fully nonparametric inference framework. However, this is a non-trivial extension and, in this chapter, we take the first step in this direction by focusing on the time dimension under the point process approach.

We therefore consider only the times of threshold exceedances $\{t_i(\mathbf{s}_j) : i = 1, ..., n_j\}$ from each observed spatial location $\mathbf{s}_j \in \mathcal{S}$, for $j = 1, ..., m$. Following the definition of the bivariate NHPP assumed under the general approach, for any generic location $\mathbf{s} \in \mathcal{S}$, the point pattern $\{t_i(\mathbf{s}) : i = 1, ..., n_\mathbf{s}\}$ is a realization from a temporal NHPP on $[0, 1]$. The corresponding temporal intensity function at location $\mathbf{s}$ is denoted by $\lambda_\mathbf{s}(t)$, where this is the appropriate marginal of the bivariate NHPP intensity $\lambda_\mathbf{s}(t, y)$. Here, we seek to develop a nonparametric prior model for $\{\lambda_\mathbf{s}(t) : t \in [0, 1]; \mathbf{s} \in \mathcal{S}\}$, that is, for a collection of temporal NHPP intensities evolving over (continuous) space. The key inferential objectives are twofold: to allow general time-inhomogeneous shapes for the intensity of threshold exceedances at each specific spatial location; and to enable flexible inference for these spatially varying temporal intensities and for implied risk assessment functionals. The implicit assumption is that of a smooth evolution of the intensities across space, although the proposed model for the spatial dependence is nonparametric relaxing both of the customary assumptions of Gaussianity and stationarity.

Regarding the choice of threshold u, we view its specification as a component of scientific or policy making considerations for the particular problem at hand. Hence, for any substantive application of the methodology, the threshold would be chosen in consultation with

the domain experts. The threshold value can be site-specific, and this is how we envision the model to be applied in general settings. Since the examples of Section 5.4 involve a small geographic region, we work with a constant threshold value across space for our illustrative data analyses, and without loss of generality, retain the non-spatially varying notation for the threshold.

5.2.1 Mixture Modeling for the Temporal Intensity of Threshold Exceedances

To build the prior model for $\{\lambda_{\mathbf{s}}(t) : t \in [0,1]; \mathbf{s} \in \mathcal{S}\}$, we follow the similar strategy discussed in Section 3.2.2. In particular, for any spatial location $\mathbf{s} \in \mathcal{S}$, we utilize the decomposition of the intensity function into the total intensity $\gamma_{\mathbf{s}}$ and the NHPP density function $f_{\mathbf{s}}(t)$ on $[0,1]$, such that $\lambda_{\mathbf{s}}(t) = \gamma_{\mathbf{s}}\ f_{\mathbf{s}}(t)$. Here, $\gamma_{\mathbf{s}} = \int_0^1 \lambda_{\mathbf{s}}(t)\mathrm{d}t$, where $\gamma_{\mathbf{s}} < \infty$ based on the NHPP definition that imposes local integrability for the intensity function. Now, for any observed point pattern $\{t_i(\mathbf{s}) : i = 1, ..., n_{\mathbf{s}}\}$ of exceedance times at location $\mathbf{s}$, the NHPP likelihood can be written as

$$L(\lambda_{\mathbf{s}}(\cdot)) \equiv L(\gamma_{\mathbf{s}}, f_{\mathbf{s}}(\cdot)) \propto \exp(-\gamma_{\mathbf{s}})\gamma_{\mathbf{s}}^{n_{\mathbf{s}}} \prod_{i=1}^{n_{\mathbf{s}}} f_{\mathbf{s}}(t_i(\mathbf{s})). \tag{5.1}$$

The full likelihood requires an extension of (5.1) to include the data from all locations, but this expression highlights the practical utility of the $(\gamma_{\mathbf{s}}, f_{\mathbf{s}}(\cdot))$ representation for the NHPP intensity. Namely, it allows us to build the model for the spatially varying intensities through a nonparametric prior model for spatially dependent densities. A prior model for the spatial surface $\{\gamma_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ will also be needed, but owing to the factorization in (5.1), the estimation of its parameters proceeds independently of the model for the NHPP densities.

We propose a mixture model formulation for the spatially varying NHPP densities,

$f_{\mathbf{s}}(t) \equiv f(t;G_{\mathbf{s}}) = \int k(t \mid \theta)\mathrm{d}G_{\mathbf{s}}(\theta)$, for $t \in [0,1]$ and $\mathbf{s} \in \mathcal{S}$. Here, $k(t \mid \theta)$ is the parametric kernel density supported by the unit interval, and $G_{\mathbf{s}}$ is the random mixing distribution indexed by spatial location $\mathbf{s}$. To meet our inferential goals, we need an appropriate kernel density that enables general, possibly multimodal shapes for the mixture density at any location, as well as a nonparametric prior model for the (uncountable) collection of mixing distributions $G_{\mathcal{S}} = \{G_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ that allows flexible inference for spatial interpolation of the intensity of extremes.

Regarding the mixture kernel, the Beta distribution used in previous chapters is a natural choice given the range of shapes the Beta density achieves, and the fact that it is directly bounded to $[0,1]$. However, the lack of a conditionally conjugate distribution for the parameters of the Beta density makes implementation of posterior simulation challenging even when modeling a single density with a nonparametric mixture of Beta densities. This challenge is exacerbated in terms of both modeling and implementation of inference in our context which involves a collection of spatially related densities. Hence, we work with a more convenient modeling platform based on a logit-normal kernel,

$$k(t \mid \theta, \tau^2) = (2\pi\tau^2)^{-1/2} t^{-1}(1-t)^{-1} \exp\left\{ -\left[\log\left(\frac{t}{1-t}\right) - \theta\right]^2 / 2\tau^2 \right\}, \quad t \in [0,1]. \tag{5.2}$$

Note that this density arises through the logistic transformation, $t = \exp(z)/(1+\exp(z))$, of a $\mathrm{N}(\theta, \tau^2)$ density for z. As discussed below, this provides a significant advantage in the formulation of the nonparametric prior model for $G_{\mathcal{S}}$ and in MCMC posterior simulation, since we can work with a (spatially dependent) mixture of normals for the logit-transformed exceedance times. The potential drawback of the logit-normal kernel is that it is susceptible to boundary effects due to the logit transformation, $\mathrm{logit}(t) = \log(t/(1-t))$, and the normal distribution tails.

However, in practice, the nonparametric mixture structure allows robust inference under both kernel choices provided the data do not maintain high intensity at the edges of the observation window; see, e.g., the empirical comparison in Taddy & Kottas (2012).

Now, for any spatial location $\mathbf{s} \in \mathcal{S}$, the proposed mixture model for the density of threshold exceedance times is expressed as

$$f_{\mathbf{s}}(t) \equiv f(t; G_{\mathbf{s}}, \tau^2) = \int k(t \mid \theta, \tau^2) \mathrm{d}G_{\mathbf{s}}(\theta), \quad t \in [0,1] \tag{5.3}$$

where $k(t \mid \theta, \tau^2)$ is given by (5.2). Therefore, the NHPP density is modeled with a semi-parametric mixture based on nonparametric mixing with respect to only the location parameter of the logit-normal kernel. This mixture model formulation strikes a good balance between model flexibility and computational feasibility. Location mixtures of logit-normals can capture non-standard density shapes, including skewness or multimodality; however, this may come at the expense of a larger number of mixture components than what would be needed under the model that includes mixing also with respect to the scale parameter of the logit-normal kernel. Although it is possible to extend the spatial nonparametric model to include location-scale mixing, this more general representation requires a more complex prior for $G_{\mathcal{S}}$ and more complicated methods for posterior simulation. The scale parameter τ^2 of the kernel can be viewed as a bandwidth parameter, which is estimated from the data based on an inverse Gamma prior.

5.2.2 The Spatial Nonparametric Prior Model

To build the spatial dependence in the prior model for the threshold exceedance time densities, $\{f(t; G_{\mathbf{s}}, \tau^2) : \mathbf{s} \in \mathcal{S}\}$, we use the spatial DP prior (Gelfand et al., 2005) for the col-

lection of corresponding mixing distributions $G_{\mathcal{S}} = \{G_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$. The spatial DP defines a nonparametric prior for the distribution of random fields, and it can thus be used to develop semiparametric models for spatial or spatio-temporal data by replacing customary Gaussian process (GP) specifications for spatial random effects distributions. Central to its development is the DP prior constructive definition. According to this definition, if $G \sim \mathsf{DP}(\alpha, G_0)$, G admits an almost sure representation of the form $\sum_{l=1}^{\infty} w_l \delta_{\vartheta_l}$; see details in Section 2.2.1. Under the standard model setting with DP priors, the ϑ_l are either scalar or vector valued, and thus G_0 is supported by a possibly multivariate, albeit finite dimensional, Euclidean space.

To model nonparametrically the distribution of a random field over region $\mathcal{S} \subset \mathbb{R}^2$, G_0 is extended to a parametric stochastic process $G_{0,\mathcal{S}}$ over the region of interest, a natural choice for which is a GP (possibly after transformation of the spatial random effects parameters). Hence, the (almost sure) representation for spatial DP prior realizations becomes

$$G_{\mathcal{S}} = \sum_{l=1}^{\infty} w_l \delta_{\vartheta_{l,\mathcal{S}}}$$

where the $\vartheta_{l,\mathcal{S}} = \{\vartheta_l(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ are independent realizations from $G_{0,\mathcal{S}}$. We take a GP for $G_{0,\mathcal{S}}$ with constant mean function ζ, constant variance η^2, and isotropic exponential correlation function, that is, $\mathrm{Corr}(\vartheta_l(\mathbf{s}), \vartheta_l(\mathbf{s}') \mid \rho) = \exp(-\rho||\mathbf{s}-\mathbf{s}'||)$, where $\rho > 0$ is the range parameter. As discussed in Section 5.3.1, the full Bayesian model is completed with priors for the precision parameter α and for the GP hyperparameters, $\phi = (\zeta, \eta^2, \rho)$.

Therefore, the spatial DP prior model involves a countable mixture of GP realizations with weights defined through stick-breaking as in the standard DP prior. Consequently, for any finite set of spatial locations $(\mathbf{s}_1, ..., \mathbf{s}_r)$, the spatial DP prior induces a DP prior for the finite

collection of mixing distributions $(G_{\mathbf{s}_1}, ..., G_{\mathbf{s}_r})$; the centering distribution of this DP prior is the r-dimensional normal induced by the GP used for $G_{0,\mathcal{S}}$. This is a key property of the spatial DP prior with respect to both simulation-based model fitting and predictive inference for spatial interpolation. Spatial DPs provide an illustration of dependent DPs (MacEachern, 2000) in that they yield a stochastic process of random distributions, one at each location in $\mathcal{S}$. These distributions are dependent but such that, at each index value, the distribution is a univariate DP.

Hence, for any location $\mathbf{s} \in \mathcal{S}$, the spatial DP prior yields a location DP mixture of logit-normals following the formulation in (5.3). The DP mixture model interpretation is also valid for any finite collection of locations, with the additional structure of spatial dependence induced to the threshold exceedance time densities by the spatial dependence in the mixing distributions. It is important to note that the spatial DP generates non-stationary spatial surface realizations with non-Gaussian finite dimensional distributions, even when the centering GP is isotropic. Moreover, if $G_{\mathbf{s}}$ and $G_{\mathbf{s}'}$ denote the marginal distributions at generic locations $\mathbf{s}$ and $\mathbf{s}'$, then the continuity of the $\vartheta_{l,\mathcal{S}}$ (implied by the exponential correlation function of $G_{0,\mathcal{S}}$) yields that, as the distance between $\mathbf{s}$ and $\mathbf{s}'$ gets smaller, the difference between $G_{\mathbf{s}}$ and $G_{\mathbf{s}'}$ gets smaller. Formally, for any $\varepsilon > 0$, $\lim_{||\mathbf{s}-\mathbf{s}'||\to 0} \Pr(\mathcal{L}(G_{\mathbf{s}}, G_{\mathbf{s}'}) < \varepsilon) = 1$, where $\mathcal{L}$ is the Lévy distance (MacEachern, 2000; Gelfand et al., 2005); see Guindani & Gelfand (2006) for a detailed study of smoothness properties for spatial DP realizations. Hence, the level of dependence between $G_{\mathbf{s}}$ and $G_{\mathbf{s}'}$, and thus between $f(t; G_{\mathbf{s}}, \tau^2)$ and $f(t; G_{\mathbf{s}'}, \tau^2)$, is driven by the distance between the spatial locations. The practical implication is that in predictive inference for spatial interpolation, we learn more from locations $\mathbf{s}'$ nearby $\mathbf{s}$ than from more distant locations, a desirable property for densities that are expected to evolve relatively smoothly across space.

5.2.3 The Spatial Model for the Total Intensity of Threshold Exceedances

To complete the model specification for the spatially dependent intensity functions, we turn to a spatial probability model for the total intensity surface. Here, we work with a GP-based hierarchical specification, using a GP prior for the log-intensity surface $\{\beta(\mathbf{s}) = \log(\gamma_{\mathbf{s}}) : \mathbf{s} \in \mathcal{S}\}$. The first stage arises from the NHPP assumption, namely, the observed exceedance counts are assumed Poisson distributed,

$$n_j \mid \beta_j \overset{ind.}{\sim} \text{Poisson}(\exp(\beta_j)) \qquad\qquad j = 1, \ldots, m,$$

where $\beta_j = \log(\gamma_{\mathbf{s}_j})$. The GP prior for $\{\beta(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ is assumed to have constant mean λ, constant variance κ^2, and power exponential correlation function $\exp(-\psi||\mathbf{s} - \mathbf{s}'||^a)$ for fixed $a \in [1,2]$. Finally, we place a normal prior $\text{N}(m_\lambda, S_\lambda^2)$ on λ, an inverse gamma prior $\text{IG}(a_{\kappa^2}, b_{\kappa^2})$ (with mean $b_{\kappa^2}/(a_{\kappa^2} - 1)$ provided $a_{\kappa^2} > 1$) on κ^2, and a uniform prior $\text{Unif}(0, b_\psi)$ on ψ. MCMC posterior simulation, as well as specification of the hyperpriors for λ, κ^2 and ψ, is discussed in Section 5.3.2.

We note that the exceedance counts will typically take small to moderate values. If the particular application involves also a small number of spatial locations (as for the data sets considered in Section 5.4), the relatively simple model specification discussed above is arguably a suitable choice. For problems involving data from a large number of locations and/or where physical information is available, more structured GP mean functions or non-stationary covariance functions can be entertained. The fact that the prior model for the total intensities is specified independently of that for the NHPP densities is an asset in this respect.

5.3 Posterior Simulation and Inference

5.3.1 Hierarchical Model Formulation

Inference for the intensity of extremes across space requires posterior simulation for: the random mixing distribution $G_{\mathbf{s}}$ over a number of spatial locations $\mathbf{s}$, including interpolation at new locations; the spatial DP prior hyperparameters; and the parameters of the GP-based spatial model for the total intensity of exceedances.

In general, nonparametric Bayesian inference for related distributions (indexed by time, space, or covariate values) requires some form of replication, although imbalance in the replicate responses can be handled. In the absence of replication, posterior simulation can be overly sensitive to the prior specification and predictive inference at new index points will inevitably fall back exclusively to the prior. In our context, the replication is provided by the set of threshold exceedance times, $\{t_i(\mathbf{s}_j) : i = 1,...,n_j\}$, at each observed site $\mathbf{s}_j$, $j = 1,...,m$.

However, in contrast to earlier applications of dependent nonparametric prior models to geostatistics problems, our observations are random times arising as the events of the underlying NHPP at each site. This aspect of the data structure creates a challenge in matching the observed times to form response vectors across sites. An option is to discretize the time interval under study into time units specified such that at most one exceedance time is included in each time unit from each location; the default choice would be the unit at which the data is recorded, e.g., a day for daily rainfall records. We can then construct time-ordered response vectors, of dimension between 1 and m, that include an entry for all sites for which there was an exceedance at the specific time unit. For the hierarchical data model, the k-th response vector is

assigned a vector of mixing parameters, $\theta_k(\mathbf{s}_{\text{obs}}) = (\theta_k(\mathbf{s}_1), ..., \theta_k(\mathbf{s}_m))$, where the $\theta_k(\mathbf{s}_{\text{obs}})$ arise conditionally independent from the DP prior induced by the spatial DP at the vector of observed locations $\mathbf{s}_{\text{obs}}$. This hierarchical model formulation is along the lines in Gelfand et al. (2005) and Kottas et al. (2008) for independent and temporally dependent replicates, respectively.

Our data examples involve small exceedance counts over the time period of interest; with daily records over 50 years, the realized n_j range roughly between 15 to 250. Hence, constructing the replicates as discussed above results in response vectors with a very small number of entries relative to m. We are thus working with data structures where replication is present, but there is essentially only a single observation at any location. Here, we apply the spatial DP model under this scenario, where the i-th observation at the j-th location, $t_i(\mathbf{s}_j)$, is assigned a vector of mixing parameters, $\theta_{ij}(\mathbf{s}_{\text{obs}}) = (\theta_{ij}(\mathbf{s}_1), ..., \theta_{ij}(\mathbf{s}_m))$, from which only $\theta_{ij}(\mathbf{s}_j)$ is used in the hierarchical model representation. A similar approach to implementing a dependent DP prior model for time series problems can be found in Rodriguez & ter Horst (2008).

More specifically, let $z_{ij} = \text{logit}(t_i(\mathbf{s}_j))$, for $i = 1, ..., n_j;\ j = 1, ..., m$, be the logit-transformed observations. Then, the hierarchical model for the data can be expressed as

$$\begin{aligned} z_{ij} \mid \theta_{ij}(\mathbf{s}_{\text{obs}}), \tau^2 &\overset{ind.}{\sim} \text{N}(\theta_{ij}(\mathbf{s}_j), \tau^2), \quad && i = 1, ..., n_j;\ j = 1, ..., m \\ \theta_{ij}(\mathbf{s}_{\text{obs}}) \mid G_{\mathbf{s}_{\text{obs}}} &\overset{i.i.d.}{\sim} G_{\mathbf{s}_{\text{obs}}}, && i = 1, ..., n_j;\ j = 1, ..., m \end{aligned}$$

where $G_{\mathbf{s}_{\text{obs}}} \mid \alpha, \phi \sim \text{DP}(\alpha, G_{0,\mathbf{s}_{\text{obs}}})$, that is, the DP prior induced by the spatial DP. Therefore, $G_{0,\mathbf{s}_{\text{obs}}}$ is an m-variate normal distribution with mean vector $\zeta \mathbf{1}_m$ and covariance matrix $\Sigma = \eta^2 R(\rho)$, with $R_{jj'}(\rho) = \exp(-\rho ||\mathbf{s}_j - \mathbf{s}_{j'}||)$, for $j, j' = 1, ..., m$. Here, $\mathbf{1}_m$ denotes an m-dimensional vector with all its elements equal to 1.

We use blocked Gibbs sampling for MCMC posterior simulation. The approach is based on a truncation approximation to the DP prior for $G_{\mathbf{s}_{\text{obs}}}$ defined through $G^N_{\mathbf{s}_{\text{obs}}} = \sum_{l=1}^N p_l \delta_{\vartheta_l(\mathbf{s}_{\text{obs}})}$, where the $\vartheta_l(\mathbf{s}_{\text{obs}}) = (\vartheta_l(\mathbf{s}_1), ..., \vartheta_l(\mathbf{s}_m))$ are independent realizations from $G_{0,\mathbf{s}_{\text{obs}}}$, and the weights $\mathbf{p} = \{p_l : l = 1, ..., N\}$ are defined using the DP stick-breaking construction subject to the constraint $p_N = 1 - \sum_{l=1}^{N-1} p_l$. The truncation level N can be chosen to any desired level of accuracy, using standard DP properties (e.g., Ishwaran & Zarepour, 2000). Then, the model can be fit to the data without the need to impute the mixing parameter vectors $\theta_{ij}(\mathbf{s}_{\text{obs}})$. To this end, we introduce configuration variables $\mathbf{L} = \{L_{ij} : i = 1, ..., n_j;\ j = 1, ..., m\}$, where $L_{ij} = l$, for $l = 1, ..., N$, if and only if $\theta_{ij}(\mathbf{s}_{\text{obs}}) = \vartheta_l(\mathbf{s}_{\text{obs}})$. Hence, the hierarchical model for the data becomes

$$\begin{aligned} z_{ij} \mid \{\vartheta_l(\mathbf{s}_{\text{obs}})\}, L_{ij}, \tau^2 &\overset{ind.}{\sim} \text{N}(\vartheta_{L_{ij}}(\mathbf{s}_j), \tau^2), \quad i = 1, ..., n_j;\ j = 1, ..., m \\ L_{ij} \mid \mathbf{p} &\overset{i.i.d.}{\sim} \textstyle\sum_{l=1}^N p_l \delta_l(L_{ij}) \quad\quad i = 1, ..., n_j;\ j = 1, ..., m \end{aligned} \tag{5.4}$$

where $\vartheta_l(\mathbf{s}_{\text{obs}}) \mid \phi \overset{i.i.d.}{\sim} G_{0,\mathbf{s}_{\text{obs}}}$, for $l = 1, ..., N$, and the prior for $\mathbf{p}$, given α, is given by a generalized Dirichlet distribution (Ishwaran & James, 2001). The hierarchical model is completed with hyperpriors for the spatial DP parameters: a gamma prior for α, a normal prior $\text{N}(m_\zeta, S^2_\zeta)$ for ζ, an inverse gamma prior $\text{IG}(a_{\eta^2}, b_{\eta^2})$ for η^2, and a uniform prior $\text{Unif}(0, b_\rho)$ for ρ. The prior specification of these parameters follows the approach in Gelfand et al. (2005), in particular, b_ρ is chosen using the *range* interpretation of the GP correlation parameter ρ; see details in Section 5.3.4. We also provide details of MCMC posterior simulation for model (5.4) in the next section.

5.3.2 Markov Chain Monte Carlo Posterior Simulation

In this section, we provide the details for MCMC posterior simulation from the spatial DP model for the NHPP densities as well as the GP model for the total NHPP intensity surface.

Regarding the former, simulation from the posterior distribution of model (5.4) is based on the blocked Gibbs sampler, including Metropolis-Hastings steps. In particular, the MCMC algorithm iteratively updates model parameters according to the following steps.

- Updating L_{ij}, $i = 1,\ldots,n_j$; $j = 1,\ldots,m$. Each L_{ij} is drawn from a discrete distribution on $\{1,...,N\}$ with probabilities proportional to $p_l k(z_{ij}|\vartheta_l(\mathbf{s}_j),\tau^2)$, for $l = 1,\ldots,N$.

- Updating α and $\mathbf{p}$. The draws for these parameters are generic for any choice of kernel in the DP mixture model; details are given in Ishwaran & Zarepour (2000).

- Updating $\vartheta_l(\mathbf{s}_{\text{obs}})$, $l = 1,\ldots,N$. Let n^* be the number of distinct components in vector $\mathbf{L}$, and $\mathbf{L}^* = \{L_k^* : k = 1,\ldots,n^*\}$ the set of distinct elements. If $l \notin \mathbf{L}^*$, then $\vartheta_l(\mathbf{s}_{\text{obs}})$ is drawn from the normal centering distribution $G_{0,\mathbf{s}_{\text{obs}}}$. If $l \in \mathbf{L}^*$, the posterior full conditional for $\vartheta_l(\mathbf{s}_{\text{obs}})$ is proportional to $\mathrm{N}_m(\vartheta_l(\mathbf{s}_{\text{obs}}) \mid \zeta\mathbf{1}_m, \Sigma)\prod_{\{(i,j):L_{ij}=l\}}\mathrm{N}(z_{ij} \mid \vartheta_l(\mathbf{s}_j),\tau^2)$, a form which results in an m-variate normal distribution.

- Updating the centering GP parameters. The full conditional for ζ can be derived as a normal distribution with mean $\left(n^*\mathbf{1}_m'\Sigma^{-1}\mathbf{1}_m + S_\zeta^{-2}\right)^{-1}\left(\mathbf{1}_m'\Sigma^{-1}\sum_{k=1}^{n^*}\vartheta_{L_k^*}(\mathbf{s}_{\text{obs}}) + m_\zeta S_\zeta^{-2}\right)$ and variance $\left(n^*\mathbf{1}_m'\Sigma^{-1}\mathbf{1}_m + S_\zeta^{-2}\right)^{-1}$. Given their high posterior correlation, we update η^2 and ρ as a block with a joint random walk Metropolis-Hasting step based on a bivariate normal proposal distribution (on the log scale for η^2 and the logit scale for ρ/b_ρ). To

achieve good mixing, we estimate the proposal covariance matrix from the output of an initial chain based on separate updates for η^2 and ρ, using a Metropolis-Hasting step for ρ and sampling η^2 from its full conditional which is available as an inverse Gamma distribution.

- Updating τ^2. The posterior full conditional for τ^2 is an inverse Gamma distribution with shape parameter $a_{\tau^2} + 0.5\sum_{j=1}^{m} n_j$ and scale parameter $b_{\tau^2} + 0.5\sum_{j=1}^{m}\sum_{i=1}^{n_j}(z_{ij} - \vartheta_{L_{ij}}(\mathbf{s}_j))^2$.

Turning to the model of Section 5.2.3 for the total intensity surface, the MCMC posterior sampling steps are as follows.

- Updating β_j, $j = 1,\ldots,m$. The posterior full conditional for each β_j is proportional to $\exp(n_j\beta_j - \exp(\beta_j))p(\beta_j \mid \{\beta_r : r \neq j\})$, where $p(\beta_j \mid \{\beta_r : r \neq j\})$ denotes the normal distribution for β_j, conditional on $\{\beta_r : r \neq j\}$, implied by the GP prior for $\{\beta(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$. Hence, β_j can updated using slice sampling (as in Example 4 of Damien et al., 1999).

- Updating the GP prior parameters. The GP mean parameter λ is sampled from its normal posterior full conditional distribution, whereas (κ^2, ψ) are updated jointly with a Metropolis-Hasting step designed similarly to the one for (η^2, ρ) discussed above.

Convergence of the MCMC algorithms was assessed by visually inspecting the trace plots associated with various parameters of interest, as well as by computing standard diagnostic criteria. For instance, for the spatial DP model hyperparameters, the *R* statistic values (Gelman & Rubin, 1992) were below 1.1 after 40,000 iterations. All inferences reported in Section 5.4 are based on 3,000 posterior samples obtained after discarding the first 50,000 iterations and

thinning the remaining 150,000 every 50 observations. Both MCMC algorithms were implemented in the C programming language. The code for the spatial DP mixture model executed at a rate of 1,500 iterations per minute on a MacBook laptop with a 2 GHz Intel Core 2 Duo and 2 GB of memory.

5.3.3 Posterior Inference for Risk Assessment

The resulting posterior samples from MCMC algorithms can be used to extend the inference to spatial interpolation based on a set of M new locations, $\mathbf{s}_{\text{new}} = (\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_M)$. Spatial interpolation for the total intensity surface $\{\gamma_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ proceeds through standard GP predictive computing based on the implied conditional normal distribution for $(\beta(\tilde{\mathbf{s}}_1), ..., \beta(\tilde{\mathbf{s}}_M))$ given $(\beta(\mathbf{s}_1), ..., \beta(\mathbf{s}_m))$. In Section 5.4, we illustrate with posterior mean estimates for the $\{\gamma_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ surface.

GP predictive calculations are also central for spatial interpolation of the NHPP densities, in conjunction with the spatial DP structure for the set of mixing distributions that includes the new sites. Specifically, under the DP truncation approximation,

$$G^N_{(\mathbf{s}_{\text{obs}}, \mathbf{s}_{\text{new}})} = \sum_{l=1}^{N} p_l \delta_{(\vartheta_l(\mathbf{s}_{\text{obs}}), \vartheta_l(\mathbf{s}_{\text{new}}))},$$

where now the $(\vartheta_l(\mathbf{s}_{\text{obs}}), \vartheta_l(\mathbf{s}_{\text{new}}))$ arise independent from an $(m+M)$-variate normal distribution with mean vector $\zeta \mathbf{1}_{m+M}$ and covariance matrix with structure that extends the one in $G_{0,\mathbf{s}_{\text{obs}}}$. Hence, having obtained posterior samples for the $\vartheta_l(\mathbf{s}_{\text{obs}})$ (and $\mathbf{p}$), the additional sampling needed to complete the posterior realizations for $G^N_{(\mathbf{s}_{\text{obs}}, \mathbf{s}_{\text{new}})}$ is from M-variate conditional normal distributions to impute $\vartheta_l(\mathbf{s}_{\text{new}})$ given $\vartheta_l(\mathbf{s}_{\text{obs}})$, for $l = 1, ..., N$.

With posterior samples for the mixing distribution available at any desired set of sites, we can report different types of risk assessment inference. For any (observed or new) site $\mathbf{s}$, point estimates for the density or intensity of exceedance times can be obtained along with corresponding uncertainty bands. This inference is immediate from the definition of the mixture model for the NHPP density function,

$$f(t; G_{\mathbf{s}}^N, \tau^2) = \sum_{l=1}^{N} p_l k\left(t \mid \vartheta_l(\mathbf{s}), \tau^2\right),$$

or intensity function,

$$\lambda(t; G_{\mathbf{s}}^N, \tau^2) = \gamma_{\mathbf{s}} \sum_{l=1}^{N} p_l k\left(t \mid \vartheta_l(\mathbf{s}), \tau^2\right).$$

Using the NHPP definition, we can compute risk surface estimates defined through the probability of a specific number of threshold exceedances within any time interval of interest. For both data examples of Section 5.4, we illustrate with the probability of at least one exceedance in a given month across a number of years. Letting (t_1, t_2) denote the time interval of interest (e.g., a specific month), the risk surface probability of at least one exceedance is given by

$$1 - \exp\left(-\gamma_{\mathbf{s}} \int_{t_1}^{t_2} f(t; G_{\mathbf{s}}^N, \tau^2)\mathrm{d}t\right) = 1 - \exp\left\{-\gamma_{\mathbf{s}} \sum_{l=1}^{N} p_l \left(\int_{t_1}^{t_2} k(t \mid \vartheta_l(\mathbf{s}), \tau^2)\mathrm{d}t\right)\right\},$$

with each integral term readily computed through a difference of two normal cdf values. Our illustration represents an admittedly narrow example of risk assessment, since inference for the risk surface probability is not accompanied by impact analysis and/or vulnerability evaluation. The results presented here are merely meant to demonstrate the capacity of the nonparametric modeling approach for flexible inference which can potentially be placed in the context of a

broader analysis for problems that involve additional information for more general risk assessment.

5.3.4 Prior Specification

We follow an approach along the lines in Gelfand et al. (2005) to specify the priors for the hyperparameters of the spatial DP model. In general, we center the normal prior for ζ at 0, and set the shape parameter of inverse gamma priors to small values that yield large (possibly infinite) prior variance. Then, working with a single component of the spatial DP mixture model, the marginal variance for the response on the logit scale can be decomposed into a sum of three terms involving the prior mean of τ^2, the prior mean of η^2, and the prior variance of ζ. Hence, with a rough guess at the range of the logit-transformed exceedance times, we can complete the prior specification for ζ, η^2 and τ^2. To specify the Unif$(0, b_\rho)$ prior for ρ, we use the *range of dependence* interpretation of this parameter for the centering GP of the spatial DP prior. In particular, under the exponential correlation function, $3/\rho$ is the distance between sites that yields correlation 0.05. The range of dependence is usually assumed to be a fraction of the maximum interpoint distance (say, $d_{\max}$) over the geographic region under study. Hence, since $3/b_\rho < 3/\rho$, we specify b_ρ such that $3/b_\rho = cd_{\max}$, for $c \leq 1$; $c = 1$ was used as a conservative choice for the data examples of Section 5.4. Finally, the role α plays in controlling the number of distinct mixture components (as discussed in Section 2.2.1) can be used to guide the choice of its exponential prior.

A similar prior choice strategy can be used for the GP-based model for the total intensity surface. The approach is the same for the correlation parameter ψ. Here, the marginal

mean and variance for the site-specific exceedance counts can be expressed in terms of the λ and κ^2 parameters, using the first two moments of the lognormal distribution (induced for $\gamma_{\mathbf{s}}$ by the GP prior model). Hence, the prior means for these parameters are specified through proxies for the center and range of the number of exceedances across the region.

5.4 Data Illustrations

5.4.1 Synthetic Data Example

Our first illustration involves simulated data based on the same region and 25 observed sites (Figure 5.1), time interval (years 1950 – 1999), and time unit (days) as the real data discussed in Section 5.4.2. The times of exceedances at each site were generated using a two-state, time-inhomogeneous Markov chain, with spatial structure introduced by making the transition probabilities spatially dependent. Specifically, let $\nu_t(\mathbf{s})$ be an indicator variable such that $\nu_t(\mathbf{s}) = 1$ if an exceedance occurs at time t and location $\mathbf{s}$. Then, the data is simulated according to

$$\Pr(\nu_t(\mathbf{s}) = 1 \mid \nu_{t-1}(\mathbf{s}) = k) = \Phi(\mu_{k,t}(\mathbf{s})), \quad k \in \{0,1\},$$

for $t = 1,\ldots,T = 18,262$, and a given $\nu_0(\mathbf{s})$, where

$$\begin{pmatrix} \mu_{0,t}(\mathbf{s}) \\ \\ \mu_{1,t}(\mathbf{s}) \end{pmatrix} = \begin{pmatrix} 0.25\sin(4\pi t T^{-1}) \\ \\ 0.25\cos(4\pi t T^{-1}) \end{pmatrix} + \begin{pmatrix} \varepsilon_0(\mathbf{s}) \\ \\ \varepsilon_1(\mathbf{s}) \end{pmatrix}.$$

Here, $\{\varepsilon_k(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$, for $k = 0,1$, are independent realizations from an isotropic GP with mean -2.7, variance 1, and correlation function $\exp\{-0.2||\mathbf{s}-\mathbf{s}'||\}$. The number of realized exceedances across the 25 sites ranges between 14 and 224. The true probability of no ex-

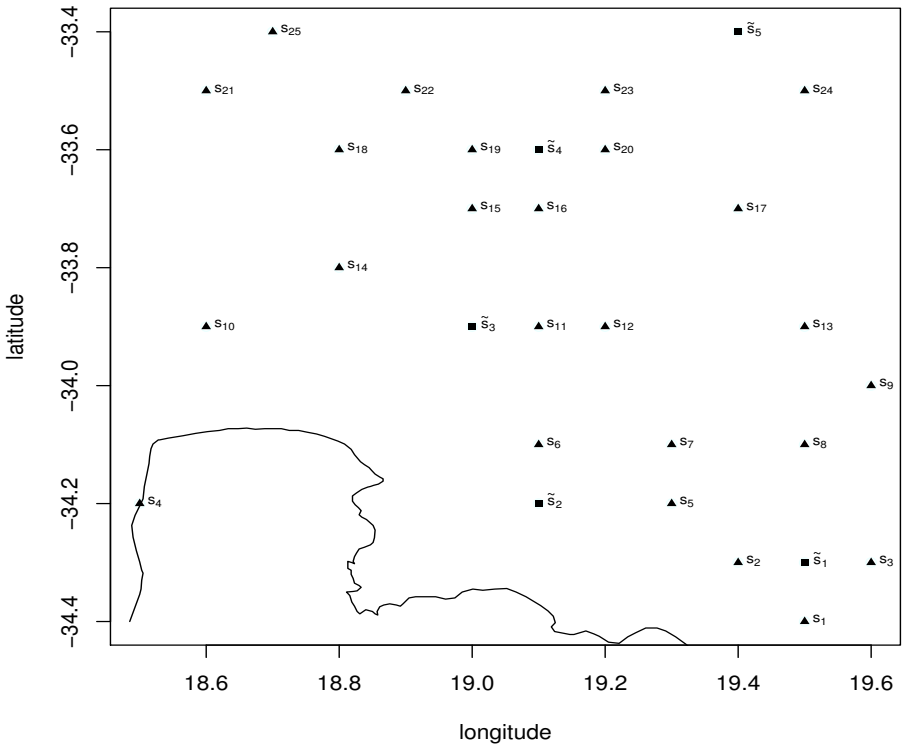


Figure 5.1: Geographic map of the southwest coastline area subregion of the Cape Floristic Region in South Africa. The map shows the 25 spatial locations $(\mathbf{s}_1,...,\mathbf{s}_{25})$ which comprise the observed set of sites for the data examples of Section 5.4, and the 5 new sites $(\tilde{\mathbf{s}}_1,...,\tilde{\mathbf{s}}_5)$ used for prediction of the exceedance times density.

ceedances at site $\mathbf{s}$ during time period $(t_0+1,\ldots,t_0+R)$ can be expressed conditional on the state at time t_0, in particular, when $\nu_{t_0}(\mathbf{s})=0$, it is given by $\prod_{t=t_0+1}^{t_0+R}\{1-\Phi(\mu_{0,t}(\mathbf{s}))\}$, whereas if $\nu_{t_0}(\mathbf{s})=1$, it is obtained as $\{1-\Phi(\mu_{1,t}(\mathbf{s})\}\prod_{t=t_0+2}^{t_0+R}\{1-\Phi(\mu_{0,t}(\mathbf{s}))\}$.

Note that the synthetic data generating mechanism is completely unrelated to our statistical model. Hence, this simulation example is used to illustrate the flexibility of the nonparametric mixture model to reconstruct risk surfaces from general stochastic processes. In addition,

the simulation is intended to demonstrate that, despite its flexibility, the nonparametric mixture model does not overfit the data. Indeed, although our statistical model can potentially capture non-stationary and non-separable behavior, the underlying data generating process is stationary and separable.

We follow the strategy discussed in the Section 5.3.4 to specify the model hyperpriors. Regarding the spatial DP parameters, we place a normal prior on ζ with mean 0 and variance 10, an $\mathsf{InvGamma}(3,12)$ prior on η^2, and a $\mathsf{Unif}(0,2.34)$ prior on ρ. An exponential prior with mean 3 is assigned to the spatial DP precision parameter α. For the GP-based model for the total intensity surface, we assign a normal prior to λ with mean 3.74 and variance 10, an $\mathsf{InvGamma}(2,0.52)$ prior to κ^2, and a $\mathsf{Unif}(0,2.34)$ prior to ψ. Finally, the scale parameter τ^2 of the logit-normal kernel is assigned an $\mathsf{InvGamma}(3,3)$ prior. We observed significant prior-to-posterior learning for all the spatial DP hyperparameters, and for τ^2. As expected, given the nature of the observables for the total intensity surface model and the small number of spatial locations, there was less learning for parameters λ, κ^2 and ψ; nevertheless, posterior densities for these parameters were noticeably concentrated relative to the corresponding prior densities.

Figure 5.2 shows an image plot of the true surface for the number of exceedances, computed from simulated data over a grid of sites, along with the model-based point estimate (posterior mean) for the total intensity surface. Note that, even though the data generating process does not imply that the exceedance counts at a given location follow a Poisson distribution, our model provides reasonable estimates capturing the underlying spatial heterogeneity.

Figure 5.3 presents posterior mean estimates for the risk surface probability of at least one exceedance in the month of June in three given years (year 1954, 1964, and 1974). These

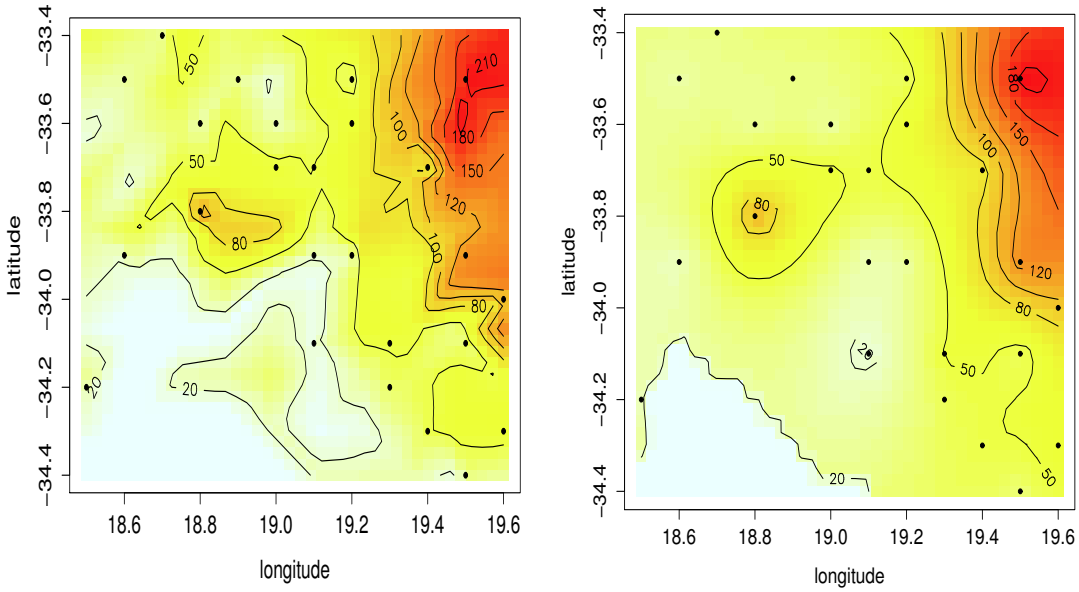
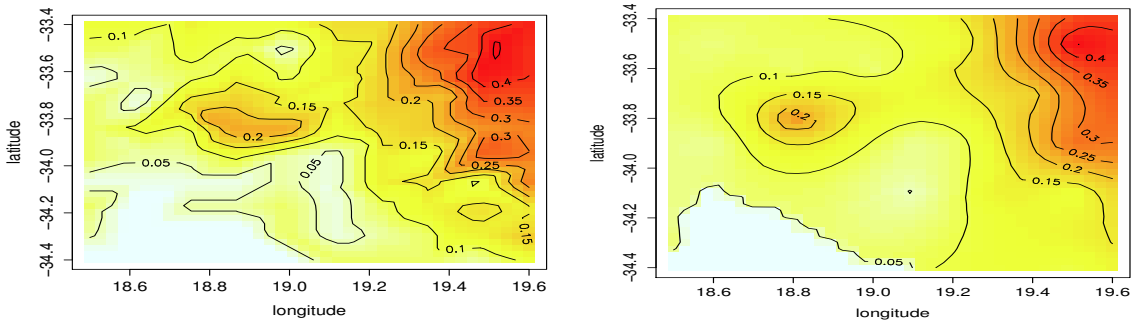


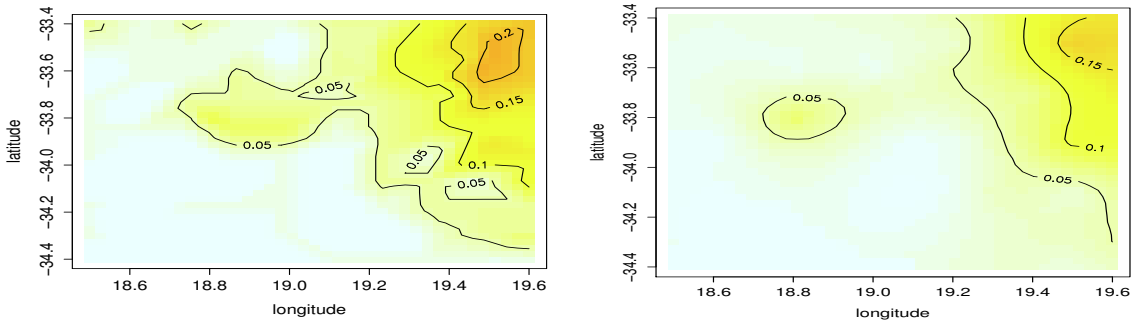
Figure 5.2: Synthetic data example. Image plot of the true surface for the number of exceedances (left panel) and the posterior mean of $\{\gamma_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ (right panel).

maps illustrate the ability of the model to capture both temporal or spatial heterogeneity. The point estimates generated by the model tend to be smoother than the true surfaces, but capture very well the patterns implied by the underlying stochastic process.

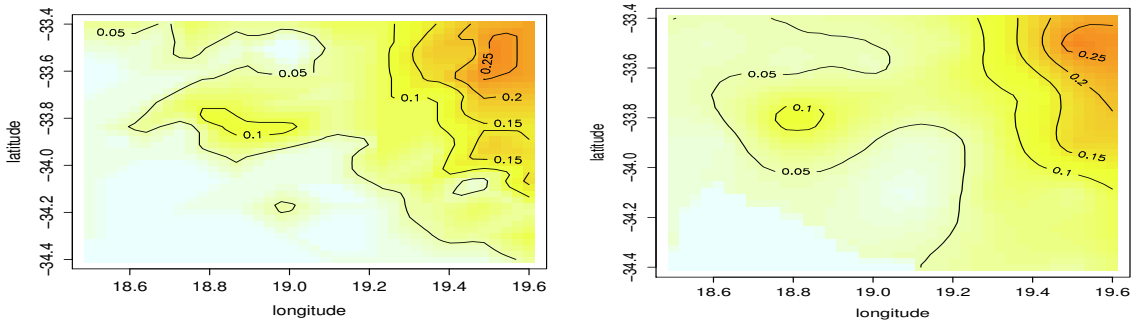
To supplement the graphical comparison results with a quantitative measure of model assessment, we report on the coverage of 95% (equal-tail) credible intervals. Based on a grid of 601 spatial locations (including the 25 observed sites), the proportion of 95% credible intervals for the site-specific probability of at least one exceedance that contain the corresponding true value is: 96.3% for June 1954, 93.3% for June 1964, and 94.3% for June 1974. A potential concern for complex Bayesian nonparametric models is that they may overfit the data with undesirable implications in prediction. In this respect, the results above are encouraging,



(a) June 1954



(b) June 1964



(c) June 1974

Figure 5.3: Synthetic data example. True surface (left panels) and posterior mean estimate (right panels) for the probability of at least one exceedance in the month of June for year 1954, 1964, and 1974 (from top to bottom).

since the data arise from a stochastic mechanism with simpler structure than what the spatial nonparametric mixture model can accommodate.

5.4.2 Rainfall Precipitation Data

Here, we present an illustration with rainfall exceedances from data collected in the Cape Floristic Region in South Africa. The Cape Floristic Region is located in the southwestern coastline of South Africa covering roughly 90,000 km^2. Although it is the smallest of the six recognized floral kingdoms in the world, it has the highest diversity, density and endemism of the flora species. The Cape Floristic Region has a semi-mediterranean climate pattern. In the west of the region, around Cape town and Paarl, the climate is characterized by hot dry summers and cool wet winters. Moving to the east, rainfall tends to be uniformly distributed over the year. Because the entire region lies between the southwestern ocean and the northeastern L-shaped mountain system, known as Cape Fold Mountains, the precipitation varies significantly. Specifically, rainfall ranges from 300 – 500 millimeters in the lowlands and 1,000 – 3,300 millimeters in the mountain areas. A previous analysis of annual rainfall maxima at 1,078 grid cells over the entire region is presented in Sang & Gelfand (2009). The raw data consists of the daily grid-aggregated precipitation, obtained via certain interpolation techniques (Hewitson & Crane, 2005), based on records at monitoring stations across South Africa between 1950 and 1999.

For an illustrative data example, and considering the topography of the region and the climate pattern discussed above, we work with a subregion of the Cape Floristic Region. In particular, we select 25 sites ($\mathbf{s}_1, ..., \mathbf{s}_{25}$ in Figure 5.1) from the southwest coastline area

including the city of Cape Town and vicinity; the longitude and latitude of the specific subregion range from $(18.5, 19.6)$ and $(-33.4, -34.4)$, respectively. Moreover, in the interest of cross-validation for spatial prediction, we consider 5 additional sites where data is available, but not used in fitting the model; these sites are denoted by $\tilde{\mathbf{s}}_1, ..., \tilde{\mathbf{s}}_5$ in Figure 5.1. To assemble the final data set with the times of exceedances at each site, we set the threshold to $u = 350$ millimeters. The range for the number of exceedances across the 25 observed sites is from 14 to 241.

Given that the region and time interval are the same with the simulated data set, there are similarities in the hyperpriors of the spatial DP model for the NHPP density and of the GP model for the total intensity surface. In particular, we place a normal prior on ζ with mean 0 and variance 10, an $\mathsf{InvGamma}(3, 12)$ prior on η^2, a $\mathsf{Unif}(0, 2.34)$ prior on ρ, and an exponential prior with mean 3 on α. Moreover, we assign a normal prior to λ with mean 3.95 and variance 10, an $\mathsf{InvGamma}(2, 0.6)$ prior to κ^2, and a $\mathrm{Unif}(0, 2.34)$ prior to ψ. Finally, τ^2 is assigned an $\mathsf{InvGamma}(3, 3)$ prior. Regarding prior-to-posterior learning for the model hyperparameters, results were consistent with the ones for the synthetic data discussed in Section 5.4.1.

The posterior mean and 95% uncertainty bands for the exceedance times density at the 25 monitoring sites are plotted in the top 5 rows of Figure 5.4, while the bottom row shows the predicted density at the five new sites shown in Figure 5.1. (Note that the bottom row panels include the histograms of the observed exceedance times, although data at these 5 sites were not used in the model fitting.) In general, the nonparametric mixture model captures well the heterogeneity of the rainfall exceedance times across space. For the observed sites, the estimates become more accurate with larger number of realized exceedances. Nonparametric spatial interpolation is illustrated with the estimates at the new sites, where predictive inference

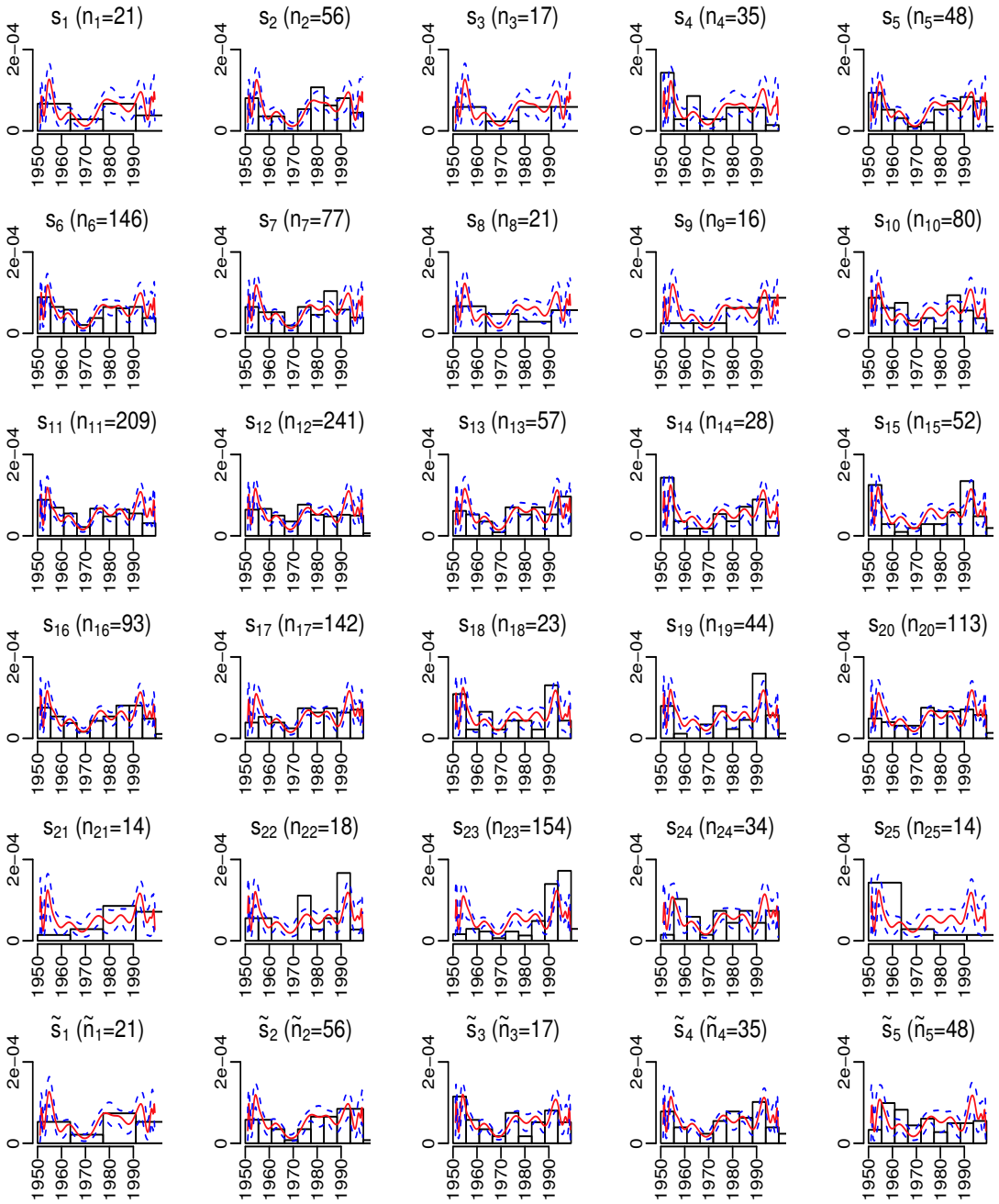


Figure 5.4: Precipitation data. Posterior mean (red solid line) and 95% interval estimates (blue dashed lines) of the exceedance time density functions at the 25 observed sites (top 5 rows) and at 5 new sites (bottom row). Each panel indicates the corresponding exceedance count and shows a histogram of the observed exceedance times.

is more accurate when interpolating at locations that have a number of monitoring sites nearby; for example, contrast the estimates at sites $\tilde{\mathbf{s}}_4$ and $\tilde{\mathbf{s}}_5$.

The posterior mean estimate for the total rainfall exceedance intensity across the region is shown in Figure 5.5 (right panel); as a point of reference, the left panel of Figure 5.5 includes the image plot of the realized number of exceedances at the 25 monitoring sites. The model estimates a larger intensity of extremes in the central part of the region relative to the northwestern and southeastern parts.

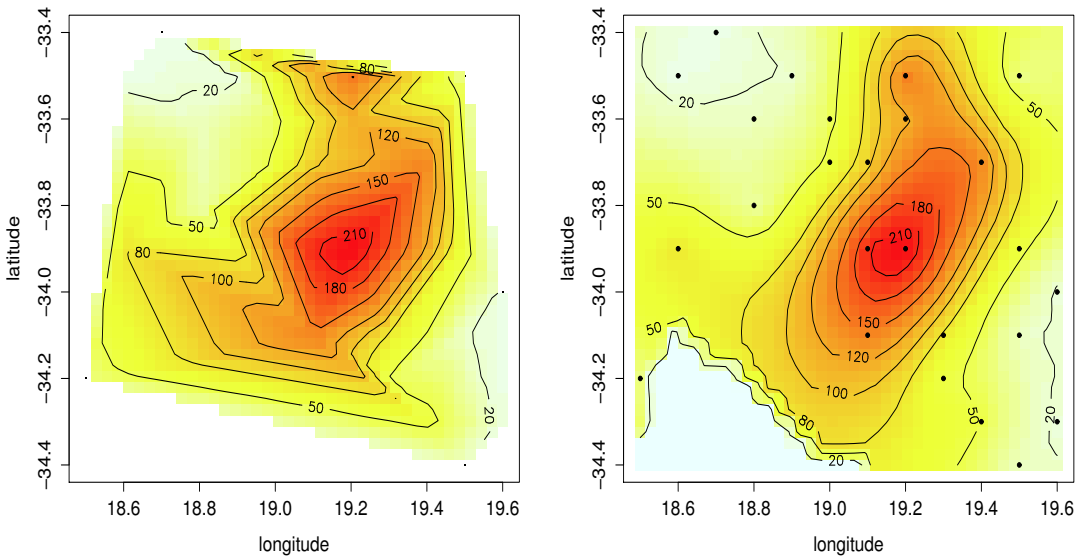


Figure 5.5: Precipitation data. Image plot of the observed number of exceedances at the 25 monitoring sites (left panel) and the posterior mean estimate of $\{\gamma_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ (right panel).

Finally, we report inference for the risk surface probability of at least one exceedance in a particular month at different years. Recall that the climate pattern in the studied region is mediterranean with cool wet winters. Hence, we focus on months when large rainfall is to be expected, and in particular, we choose the month of June. Figure 5.6, 5.7, and 5.8 plot the

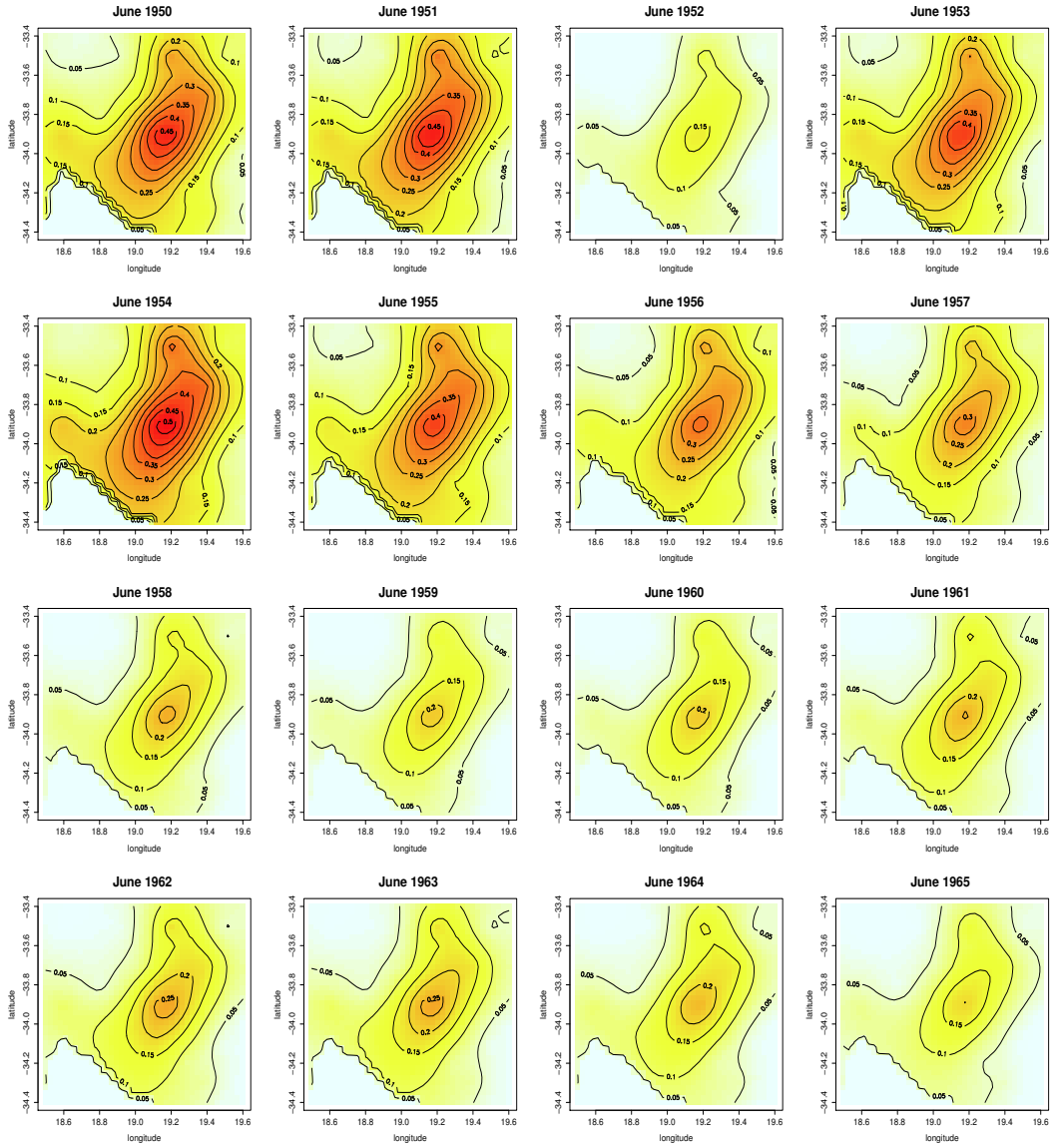


Figure 5.6: Precipitation data. Posterior mean estimate for the risk surface probability of at least one exceedance in the month of June for the year 1950 – 1965.

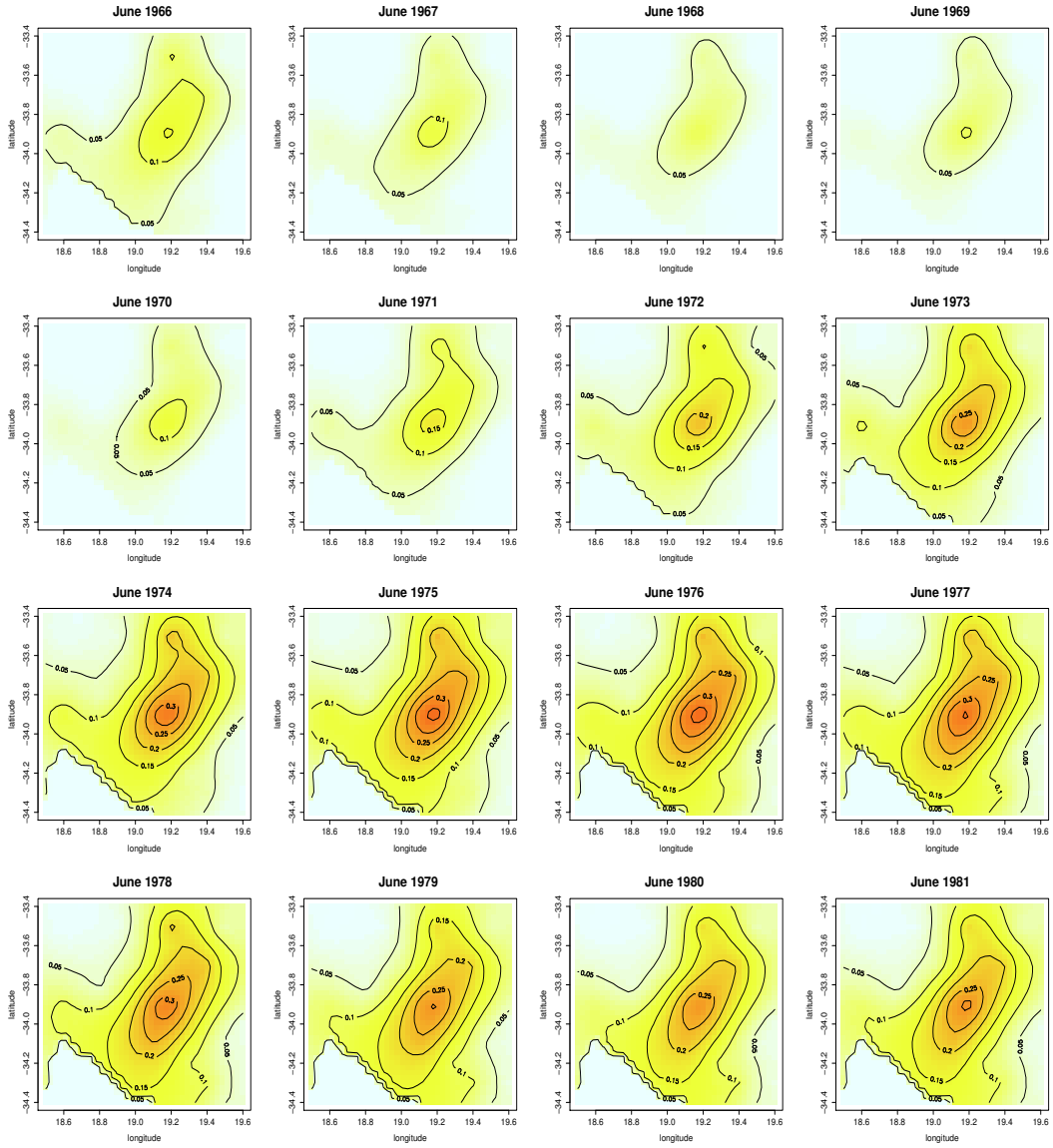


Figure 5.7: Precipitation data. Posterior mean estimate for the risk surface probability of at least one exceedance in the month of June for the year 1966 – 1981.

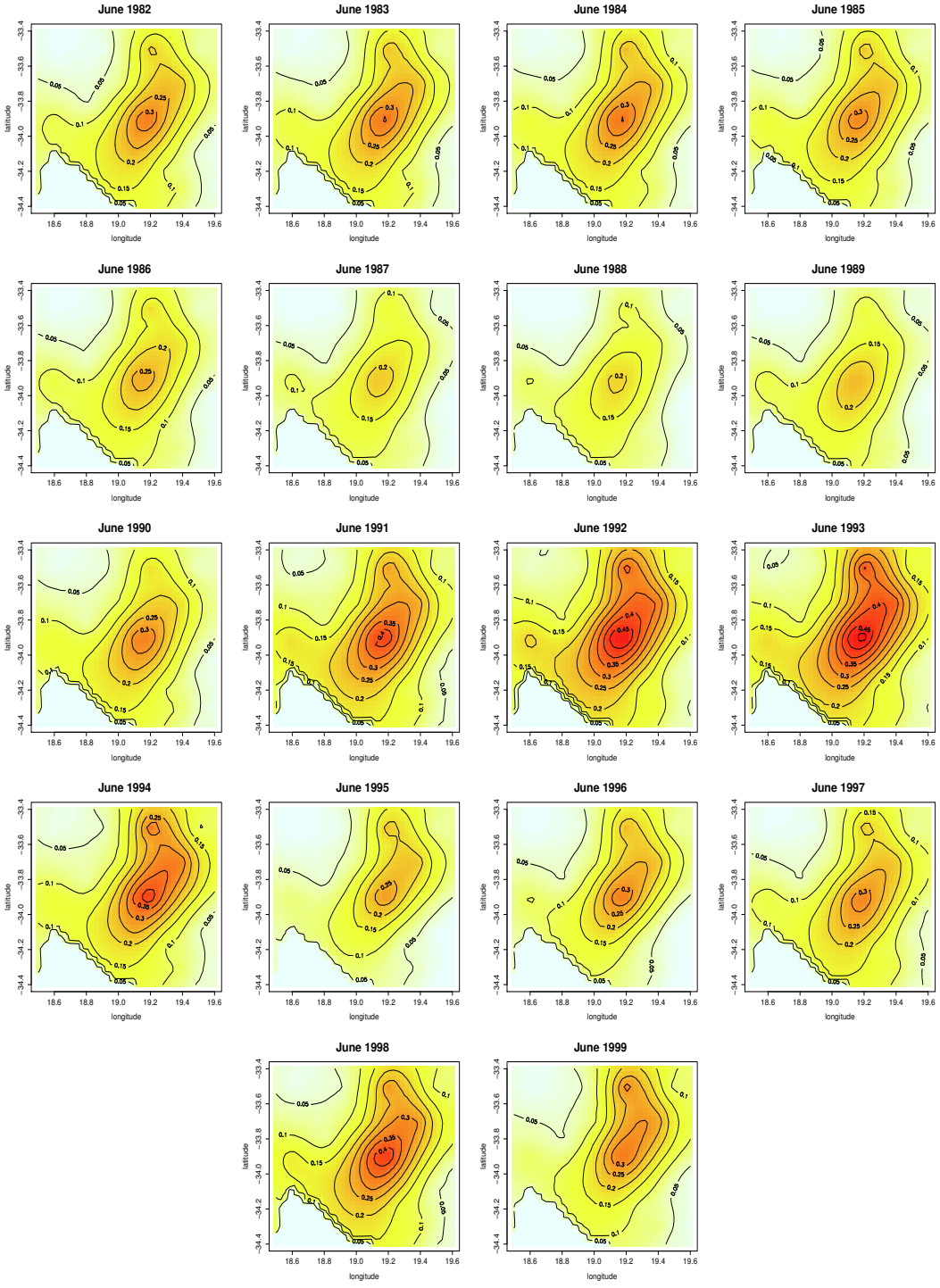


Figure 5.8: Precipitation data. Posterior mean estimate for the risk surface probability of at least one exceedance in the month of June for the year 1982 – 1999.

posterior mean estimates at all fifty years. The nonparametric mixture model estimates spatially varying risk surfaces with both intensity and shapes changing across years. The overall pattern reveals higher probabilities of at least one exceedance over June in the center of the studied region, with idiosyncratic features in certain years, for instance, the second mode more clearly seen in June of the years from 1954 to 1956 and again in June of the years from 1992 to 1994.

5.5 Summary

We have developed a Bayesian nonparametric model for the analysis of extremes from environmental variables observed over time and across a number of monitoring sites. The methodology builds on the point process approach to extreme value analysis through a nonparametric mixture model for the spatially varying intensities. The modeling approach allows general time-inhomogeneous shapes for the intensity of threshold exceedances at each specific site, as well as nonparametric spatial interpolation for practically important risk assessment functionals. A posterior simulation algorithm to implement such inference has been designed. The model has been tested with a simulated data set and applied to rainfall exceedances recorded over a time period of 50 years from a subregion of the Cape Floristic Region in South Africa.

Our data examples included a small number of sites because of the relatively small size of the geographic region under study. Moreover, the intention was to demonstrate the capacity of the spatial nonparametric mixture model to provide useful inference results under moderate sample sizes. For extreme value analysis applications, the number of observations from each site will typically be small to moderate. However, one can envision practically im-

portant scenarios that involve a large number of observed sites (at least, in the thousands). For such cases, standard posterior simulation methods are not practical for implementation of the spatial DP mixture model. Alternative cost-effective MCMC algorithms for large data sets (e.g., Guha, 2010) may provide a platform for expanding the practical utility of the proposed methodology.

Chapter 6

Conclusion

This dissertation concludes with a summary of the three and half years' research work on Bayesian nonparametric modeling for extreme value analysis, which plays an important role in financial industry and environmental sciences. This work is motivated by the need to better characterize the extremal behavior of observations generated from general stochastic processes. Our main contribution is the development of a novel class of modeling approaches that provides flexible inference methods for analysis of extremes under a generalization of the point process approach.

We have presented the methodological contributions of the dissertation through model specifications and a number of data examples in Chapter 3, 4, and 5. The first aim of this dissertation is to show the flexibility of the proposed nonparametric mixture model in modeling the extremal events. In particular, the model formulation combines the point process approach of extreme value analysis with the inferential power on Bayesian nonparametric prior models. Adopting a Bayesian nonparametric mixture modeling approach provides a natural means for

propagating uncertainty into clustering, arguably resulting in richer inferences about extremes. Hence, compared to the traditional parametric point process approach under which the intensity of extremes is homogeneous in time, the key feature of the proposed Bayesian nonparametric mixture model allows us to capture the changes in the shape of the underlying intensity function over time. Additionally, we extend the model to characterize spatially varying intensities of extremes by incorporating spatial dependence into the modeling. In such case, it allows for capturing the flexible shapes and temporal heterogeneity for the intensity of extremes at any particular location. Hence, our modeling approaches provide an appealing feature in modeling extremes from general stochastic processes evolving over time and/or space.

From a practical point of view, modeling extremal events is vital to many of the risk management issues related to finance and insurance. Unlike using methods based on normality assumptions which are likely to underestimate tail risk, the presented work in this dissertation focuses on utilizing extreme value analysis to assess the tail-related risk of financial series. In particular, the different type of return levels developed in Chapter 3 provide an important and distinct insight of the behavior of extreme price movements, especially what extremes look like on a "normal" period, or on a particular "good" or "bad" period. Furthermore, our modeling approach presented in Chapter 4 allows for modeling simultaneous risk in multiple correlated financial markets. In that case, the risk associated with each individual market can be decomposed into two components: the systemic risk and the idiosyncratic risk. This feature is specially appealing if different markets share similar products. Thus, this work provides an interesting statistical modeling tool to help financial analysts and risk managers quantify extreme market risks, and to give portfolio managers guidance for their purchases and sales of various financial

instruments.

In addition to finance, climate change in environmental process is another active field in which extreme value analysis is widely applied. The modeling approach developed in Chapter 5 incorporated spatial dependence to model the intensity of rare natural events across multiple locations. The practical implication is that it enables researchers to interpolate the intensity of extremes at unobserved locations by borrowing strength from neighboring regions. More importantly, it enables estimation of risk assessment functionals, which may help decision-makers in government or industry build effective environmental policy.

In summary, the methods developed in this dissertation provide ample evidence of our research contributions both theoretically and practically. Extreme value analysis is a wide-open research area. As indicated before, there are numerous well-developed literatures on modeling extreme values, which implies this area is still growing. While we applied the modeling approaches herein with applications in finance and environmental sciences, it is believed that the methodologies would be widely applicable to other data analysis applications in the near future.

Bibliography

ALLEN, D., GERRANS, P., SINGH, A. & POWELL, R. (2009). Quantile regression: its application in investment analysis. *JASSA* **1**, 1–12.

ANTONIAK, C. (1974). Mixture of dirchlet processes with applications to bayesian nonparametric problems. *Annals of Statistics* **2**, 1152–1174.

BALKEMA, A. A. & DE HAAN, L. (1974). Residual lifetime at great age. *Ann. Probab.* **2**, 792–804.

BARNES, M. & HUGHES, W. (2002). A quantile regression analysis of the cross section of stock market returns, working papers. *Federal Reserve Bank of Boston* **02-2**.

BLACKWELL, D. & MACQUEEN, J. (1973). Ferguson distributions via pólya urn schemes. *The Annals of Statistics* **1**, 353–355.

BUSH, C. & MACEACHERN, S. (1996). A semiparametric bayesian model for randomised block designs. *Biometrika* **83**, 275–285.

CASSON, E. & COLES, S. (1999). Spatial regression models for extremes. *Extremes* **1**, 449–468. DOI: 10.1023/A:1009931222386.

CHANG, M., HUNG, J. & NIEH, C. (2011). Reexamination of capital asset pricing model (capm): An application of quantile regression. *African Journal of Business Management* **5**, 12684–12690.

COLES, S., TAWN, J. & SMITH, R. L. (1994). A seasonal Markov model for extremely low temperatures. *Environmetrics* **5**, 221–239.

COLES, S. G. (2001). *An introduction to statistical modeling of extreme values*. Springer-Verlag, New York.

COLES, S. G. & POWELL, E. A. (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review* **64**, 119–136.

COLES, S. G. & TAWN, J. (1996). A Bayesian analysis of extreme rainfall data. *Applied Statistics* **45**, 463–478.

CONNOR, R. J. & MOSIMANN, J. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association* **64**, 194–206.

COOLEY, D., NYCHKA, D. & NAVEAU, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**, 824–840.

COOLEY, D. & SAIN, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 381–402.

COX, D. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society* **17**, 129–164.

DALEY, D. & VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes (Second ed.)*. Springer.

DAMIEN, P., WAKEFIELD, J. & WALKER, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society, Series B* **61**, 331–344.

DAVISON, A. & SMITH, R. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B* **52**, 393–442.

DE IORIO, M., MÜLLER, P., ROSNER, R. & MACEACHERN, S. N. (2004). An anova model for dependent random measures. *Journal of the American Statistical Association* **99**, 205–215.

EMBRECHTS, P., KLÜPPELBERG, C. & MIKOSCH, T. (1997). *Modeling extremal events for insurance and finance*. Berlin: Springer-Verlag.

ESCOBAR, M. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89**, 268–277.

ESCOBAR, M. & WEST, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics* **2**, 615–629.

FISHER, R. A. & TIPPETT, L. H. C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Mathematical proceedings of the Cambridge Philosophical Society* **24**, 180–190.

FRENCH, C. (2003). The treynor capital asset pricing model. *Journal of Investment Management* .

FUENTES, M., HENRY, J. & REICH, B. J. (2012). Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes* **15**. Forthcoming.

GELFAND, A. E., KOTTAS, A. E. & MACEACHERN, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035.

GELMAN, A. & RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457–511. DOI: doi:10.1214/ss/1177011136.

GENTON, M. G., MA, Y. & SANG, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika* **98**, 481–488.

GNEDENKO, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423–453.

GUHA, S. (2010). Posterior simulation in countable mixture models for large datasets. *Journal of the American Statistical Association* **105**, 775–786. DOI: 10.1198/jasa.2010.tm09340.

GUINDANI, M. & GELFAND, A. E. (2006). Smoothness properties and gradient analysis under spatial Dirichlet process models. *Methodology and Computing in Applied Probability* **8**, 159–189.

DE HAAN, L. (1985). Extremes in higher dimensions: the model and some statistics. In *Proc. 45th Sess. Int. Statist. Inst*. The Hague: International Statistical Institute.

HEWITSON, B. C. & CRANE, R. G. (2005). Gridded area-averaged daily precipitation via conditional interpolation. *Journal of Climate* **18**, 41–57.

HUERTA, G. & SANSÓ, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics* **14**, 285–299.

ISHWARAN, H. & JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173.

ISHWARAN, H. & ZAREPOUR, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and Beta two-parameter process hierarchical models. *Biometrika* **87**, 371–390.

JEKINSON, A. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Quarterly Journal of the Royal Meteorological Society* **81**, 158–172.

JI, C., MERL, D., KEPLER, T. B. & WEST, M. (2009). Spatial mixture modelling for unobserved point processes: Examples in immunofluorescence histology. *Bayesian Analysis* **4**, 297–316.

JOE, H., SMITH, R. L. & WEISSMAN, I. (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society, Series B* **54**, 171–183.

KOLOSSIATIS, M., GRIFFIN, J. & STEEL, M. (2011). On bayesian nonparametric modeling mf two correlated distributions. *Statistics and Computing* .

KOTTAS, A. (2006). Dirichlet process mixtures of Beta distributions, with applications to density and intensity estimation. In *Proceedings of the Workshop on Learning with Nonparametric Bayesian Methods, 23rd International Conference on Machine Learning*. Carnegie Mellon University, Pittsburgh, Pennsylvania.

KOTTAS, A. & BEHSETA, S. (2010). Bayesian nonparametric modeling for comparison of single-neuron firing intensities. *Biometrics* **66**, 277–286.

KOTTAS, A., BEHSETA, S., MOORMAN, D. E., POYNOR, V. & OLSON, C. R. (2012). Bayesian nonparametric analysis of neuronal intensity rates. *Journal of Neuroscience Methods* **203**, 241–253.

KOTTAS, A., DUAN, J. A. & GELFAND, A. E. (2008). Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures. *Biometrical Journal* **50**, 29–42.

KOTTAS, A. & SANSÓ, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Jounal of Statistical Planning and Inference* **137**, 3151–3163.

KOTZ, S. & NADARAJAH, S. (2000a). *Extreme Value Distributions - Theory and Applications*. Imperial College Press.

KOTZ, S. & NADARAJAH, S. (2000b). *Extreme Value Distributions - Theory and Applications*. Imperial College Press.

LANDO, D. (1998). On cox processes and credit risk securities. *Review of Derivatives Research* **2**, 99–120.

LO, A. (1984). On a class of Bayesian nonparametric estimates: I. density estimates". *The Annals of Statistics* **12**, 351–357.

MACEACHERN, S. & MÜLLER, P. (1998). Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics* **7**.

MACEACHERN, S. N. (1999). Dependent nonparametric process. *In ASA Proceedings of the Section on Bayesian Statistical Science* .

MACEACHERN, S. N. (2000). Dependent dirichlet process. Technical report, Ohio State University, Department of Statistics.

MANDELBROT, B. (1963). The variation of certain speculative prices. *Journal of Business* **36**, 394–419.

MÜLLER, P. & QUINTANA, F. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**.

MÜLLER, P., QUINTANA, F. & ROSNER, G. (2004). Hierarchical meta-analysis over related non-parametric bayesian models. *Journal of the Royal Statistical Society, Series B* **66**, 735–749.

NCDC (1998). Flooding in China, summer 1998. http://lwf.ncdc.noaa.gov/oa/reports/chinaflooding/chinaflooding.html.

NEAL, R. (1997). Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. Technical Report 9702, Universit of Toronto, Department of Statistics.

NEAL, R. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265.

PADOAN, S. A., RIBATET, M. & SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**, 263–277.

PICKANDS, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability* **8**, 745–756.

PICKANDS, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 119–131.

REICH, B. J. & SHABY, B. A. (2012). A hierarchical Bayesian analysis of max-stable spatial processes. *Annals of Applied Statistics* Forthcoming.

RIBATET, M., COOLEY, D. & DAVISON, A. (2010). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica* **Submitted**.

ROBERT, C. & CASELLA, G. (2005). *Monte Carlo Statistical Methods (Second Edition ed.)*. Springer.

RODRIGUEZ, A. & DUNSON, D. B. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis* **6**, 145–178.

RODRIGUEZ, A., DUNSON, D. B. & GELFAND, A. E. (2008). The nested dirichlet process. *Journal of the American Statistical Association* **103**, 1131–1154.

RODRIGUEZ, A. & TER HORST, E. (2008). Bayesian dynamic density estimation. *Bayesian Analysis* **3**, 339–366.

SANG, H. & GELFAND, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16**, 407–426.

SANG, H. & GELFAND, A. E. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 49–65.

SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5**, 33–44.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

SMITH, R. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science* **4**, 367–393.

SMITH, R. L. (1984). *Threshold methods for sample extremes, In Tiago de Oliveira, J.*, pp. 621–638. Statistical Extremes and Applications, Reidel, Dordrecht.

SMITH, R. L. (1985). Maximum likelihood estimation in a class os non-regular cases. *Biometrika* **72**, 67–90.

SMITH, R. L. (1990). Max-stable processes and spatial extremes. Technical report, University of North Carolina at Chapel Hill.

SMITH, R. L. (2003). *Statistics of extremes, with applications in environment, insurance and finance. Extreme Values in Finance, Telecommunications and the Environment*, chapter 1. pp.1-78. Chapman & Hall/CRC Press, London.

STEPHENSON, A. & TAWN, J. A. (2004). Bayesian inference for extremes: accounting for the three extremal types. *Extremes* **7**, 291–307. DOI: 10.1007/s10687-004-3479-6.

TADDY, M. A. & KOTTAS, A. (2012). Mixture modeling for marked Poisson processes. *Bayesian Analysis* **7**, 335–362. DOI: 10.1214/12-BA711.

TEH, Y., JORDAN, M., BEAL, M. & BLEI, D. (2006). Sharing clusters among related groups: Hierarchical dirichlet process. *Journal of the American Statistical Association* **101**, 1566–1581.

TRESSOU, J. (2008). Bayesian nonparametrics for heavy tailed distribution. Application to food risk assessment. *Bayesian Analysis* **3**, 367–392.

TREYNOR, J. (1961). Market value, time, and risk. Unpublished manuscript.

TREYNOR, J. (1962). Toward a theory of market value of risky assets. Unpublished manuscript.

VON MISES, R. (1954). *La distribution de la plus grande de n valeurs, In: Selected Papers*, volume II, 271-294. American Mathematical Society, Providence, RI.

WEST, M. & HARRISON, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer - Verlag, New York, second edition edition.