Bayesian Hierarchical Modeling for Prediction of Extremes of Financial Indexes

Athanasios Kottas, University of California, Santa Cruz

April 2013

SIGFIRM-UCSC
Engineering-2, Room 403E
1156 High Street
Santa Cruz, CA 95064
831-459-2523
Sigfirm.ucsc.edu

The Sury Initiative for Global Finance and International Risk Management (SIGFIRM) at the University of California, Santa Cruz, addresses new challenges in global financial markets. Guiding practitioners and policymakers in a world of increased uncertainty, globalization and use of information technology, SIGFIRM supports research that offers new insights into the tradeoffs between risk and rewards through innovative techniques and analytical methods. Specific areas of interest include risk management models, models of human behavior in financial markets (the underpinnings of “behavioral finance”), integrated analyses of global systemic risks, and analysis of financial markets using Internet data mining.

Abel Rodríguez
Ziwei Wang
Athanasios Kottas

University of California, Santa Cruz, U.S.A.

Summary. We develop a novel Bayesian nonparametric model to assess the effect of systemic risks on multiple financial markets, and apply it to understand the behavior of the S&P500 sector indexes between January 1st, 2000 and December 31, 2011. More specifically, our model extends to multivariate extremes the Poisson process approach for extreme value analysis based on threshold exceedances. The intensity function 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. Furthermore, in order to capture changes in the risk structure over time, the intensity function associated with each of the underlying Poisson processes is modeled using a Dirichlet process mixture model that allows for clustering of extremes. Our application to the S&P500 data provides interesting insights about the nature of the 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 log returns appear to be driven mostly by systemic risks.

Keywords: Dirichlet Process Mixture Modelling; Multivariate Extreme Values; Non-homogenous Poisson Process; Nonparametric Bayes; Systemic Risk

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 crisis showed that novel techniques are required to understand how information spreads across financial markets during times of upheaval.

The statistical analysis of univariate extreme values is fairly well developed. In general, three distinct approaches are commonly used. One option is to model blockwise maxima (such as weekly or monthly maxima) using the generalized extreme value distribution (Fisher and Tippett, 1928; Gnedenko, 1943). The main shortcoming of this approach is that it ignores the information provided
by multiple extremes observed during the same block. Alternatively, the exceedances over a given threshold can be modelled using a generalized Pareto distribution (Pickands, 1975; Davison and Smith, 1990). Finally, the value of the exceedances and the time of their occurrence can be modeled jointly using a non-homogeneous Poisson process (Pickands, 1971; Coles and Tawn, 1996). For a detailed review of statistical models for univariate extremes see, for example, Kotz and Nadarajah (2000) and Coles (2001).

The three approaches mentioned above have been extended to model extremes of more general stochastic processes. For example, the blockwise maxima approach has been extended by assuming that observations are distributed conditionally independent according to a generalized extreme value distribution and introducing dependence in its parameters. Examples of this approach in the context of spatial and spatio-temporal modelling include Huerta and Sansó (2007), Cooley et al. (2007) and Sang and Gelfand (2009). The class of max-stable processes (Smith, 1990; Schlather, 2002) is another extension of the blockwise maxima approach that is particularly appealing from a theoretical perspective. However, likelihood-based inference for max-stable processes is difficult, since in most cases no closed-form expression for the likelihood is available. In addition, work on multivariate max-stable processes is still in its infancy. In a similar vein, extensions of the point process approach to time inhomogeneous processes have been considered (e.g., Smith, 1989). More recently, Kottas and Sansó (2007), Wang et al. (2011) and Kottas et al. (2012) considered a point process approach to modelling the extremes of time-inhomogeneous stochastic processes using nonparametric mixture models.

This paper focuses on extending the point process framework developed in Wang et al. (2011) 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). Figure 1 presents the most extreme negative log returns on four of those 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.

[Fig. 1 about here.]

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 and 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-homogenous Poisson process with unknown intensity function, which is assigned a flexible Dirichlet process mixture prior model. Hence, our model can be conceptualized as an example of a Cox process (Cox, 1955).

The rest of the paper is organized as follows. Section 2 develops our modelling approach and discusses its main properties. Section 3 presents a description of the computational approach to inference (with technical details included in the Appendix) along with a careful discussion on subjective prior elicitation. Section 4 illustrates the model using a detailed case study focusing on the S&P500 index. Finally, Section 5 concludes with discussion.

2. Modelling Approach

We focus on the negative log returns of a group of \( J \) related but distinct markets, which are defined as

\[
x_{i,j} = -100 \times \log \left( \frac{S_{i,j}}{S_{i-1,j}} \right),
\]

where \( S_{i,j} \) is the value of market \( j = 1, \ldots, J \) at time \( i = 1, \ldots, T \). Note that large positive values of \( x_{i,j} \) indicate a large drop in the price index associated with market \( j \), so for risk management purposes we are interested in large values of \( x_{i,j} \). Hence, for a given threshold \( u \), we focus our 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 \( \Lambda_0^* \) that is common to all markets,
while the second is associated with the idiosyncratic risk and has a cumulative intensity function $\Lambda^*_j$ 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 $\lambda^*_0$ and $\lambda^*_j$ are the Poisson process intensities associated with $\Lambda^*_0$ and $\Lambda^*_j$, respectively.

The modelling approach for the $\Lambda^*_j$ 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 $\gamma^*_j \equiv \Lambda^*_j(T) = \int_0^T \lambda^*_j(t)dt (< \infty)$ 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 $\Lambda_j$ 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, $\gamma_j$, 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, $\epsilon_j = \gamma^*_0/(\gamma^*_0 + \gamma^*_j)$, represents the proportion of exceedances in market $j$ that are associated with the systemic component. In addition, note that values of $\epsilon_j$ close to 1 (which are associated with $\gamma^*_0 \gg \gamma^*_j$) imply a stronger association in the pattern of extremes.

Because each $N_j(t)$ follows a Poisson process, the probability that at most $r$ exceedances will be observed in market $j$ during time period $[t_0, t_0 + \Delta]$ is

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

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 time points $t_0$ and $t_0 + \Delta$ is given by

$$\exp \{-[\Lambda_j(t_0 + \Delta) - \Lambda_j(t_0)]\} = \exp \{-[\Lambda^*_0(t_0 + \Delta) - \Lambda^*_0(t_0)]\} \times \exp \{-[\Lambda^*_j(t_0 + \Delta) - \Lambda^*_j(t_0)]\},$$

where the first term in the right-hand side 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.
2.1. Modelling 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 \( f^*_0 \) and \( f^*_1, \ldots, f^*_J \) associated with the systemic and idiosyncratic distribution functions \( F^*_0 \) and \( F^*_1, \ldots, F^*_J \), using Dirichlet process mixtures. In particular, we let

\[
f^*_j(t) = \int \psi(t | \mu, \tau) dG^*_j(\mu), \quad j = 0, 1, \ldots, J,
\]

where \( \psi(t | \mu, \tau) \) is a kernel density on \([0, T]\) indexed by the finite dimensional parameters \( \mu \) and \( \tau \). The random mixing distribution \( G^*_j \) is assigned a Dirichlet process prior (Ferguson, 1973) with precision parameter \( \alpha_j \) and baseline (centering) distribution \( H \), which is common to all \( G^*_j \). Thus, using the stick-breaking constructive definition of the Dirichlet process (Sethuraman, 1994),

\[
G^*_j(\cdot) = \sum_{l=1}^{\infty} \left\{ v_{j,l} \prod_{s<l} (1 - v_{j,s}) \right\} \delta_{\tilde{\mu}_{j,l}}(\cdot), \quad j = 0, 1, \ldots, J,
\]

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 random variables independent and identically distributed according to \( H \), and \( \{v_{j,1}, v_{j,2}, \ldots\} \) is another sequence of independent and identically distributed random variables according to a Beta\((1, \alpha_j)\) distribution.

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

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

where \( \mu \in [0, T] \) is a location parameter (the mean of the kernel distribution), \( \tau \in (0, \infty) \) can be interpreted as a scale parameter, and \( \mathbb{I}_\Omega(\cdot) \) denotes the indicator function on the set \( \Omega \). Because Dirichlet process mixtures allow for an infinite number of mixture components, the model is dense on the space of absolutely continuous distributions on \([0, T]\) as long as the baseline distribution \( H \) and the prior on the scale parameter \( \tau \) are selected to provide full support on the domain of the parameters of the rescaled beta kernel (see for example Diaconis and Ylvisaker, 1985). The precision parameter \( \alpha_j \) controls the relative weight of the components, with smaller values of \( \alpha_j \) favoring mixtures where a small number of components received very large weights. On the other hand, the baseline distribution \( H \) 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 \( \gamma^*_0, \gamma^*_1, \ldots, \gamma^*_J \). In the case of the rate associated with the common component, \( \gamma^*_0 \), 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 a more general,
zero-inflated gamma prior with density,
\[ p(\gamma^*_j \mid \pi) = (1 - \pi)\delta_0(\gamma^*_j) + \pi \text{Gam}(\gamma^*_j \mid a_{\gamma^*_j}, b_{\gamma^*_j}), \quad j = 1, \ldots, J. \]

Note that the case \( \gamma^*_j = 0 \) corresponds to \( \epsilon_j = 1 \), 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 \textit{idiosyncracy test}.

Finally, we note that the representation of the sector-specific distribution function \( F_j \) as a mixture of a systemic and an idiosyncratic component is reminiscent of the models for dependent random distributions discussed in Müller et al. (2004) and Kolossiatis et al. (2011). In spite of this connection, the motivation for our modelling approach is quite different. Indeed, while the motivation in the earlier work is to construct dependent random distributions, our model structure follows from the assumption that the point process associated with the presence of extreme values can be constructed as a superposition of Poisson processes.

\subsection{2.2. Hierarchical priors}

The model is completed by adding hyperpriors to model parameters. For the baseline distribution \( H \), we note that the parameter \( \mu \) in Equation (1) is constrained to the \([0, T]\) interval. Hence, a natural choice for \( H \) is another rescaled Beta distribution with density \( h(\mu) \propto (\mu/T)^{a-1} \{1 - (\mu/T)\}^{b-1}I_{[0,T]}(\mu) \). Although the full Bayesian model can be extended to incorporate a hyperprior for one or both of the parameters of \( H \), this was not necessary for the application to the S&P500 data. In fact, we applied the model using the default choice of a uniform distribution for \( H \); sensitivity of posterior inference results to this choice, as well as to the other hyperprior choices, is discussed in Section 4.2.

The remaining priors are selected for computational convenience. In particular, a conditionally conjugate gamma prior with shape parameter \( a_\alpha \) and rate parameter \( b_\alpha \) is assigned to the precision parameters \( \alpha_j \), \( j = 0, 1, \ldots, J \). Similarly, the conditionally conjugate prior for the mixing probability \( \pi \) is given by a \text{Beta}(a_\pi, b_\pi) distribution. Finally, the scale parameter \( \tau \) is assigned an inverse gamma distribution with shape \( a_\tau \) and scale \( b_\tau \). Details on informative hyperparameter elicitation are discussed in Section 3.2.

\subsection{2.3. Threshold Selection}

Application of our methodology requires specification of the value for the threshold \( u \). Popular approaches to threshold selection include graphical methods based on threshold stability and mean residual life plots, and heuristics based on the behavior of the so-called “tail fraction” (see Scarrott and MacDonald, 2012 for a detailed review). However, most of these tools are not directly applicable to our model. Alternatively, some authors (see, e.g., Tancredi et al.,
2006 and Behrens et al., 2004) have considered models that allow for threshold selection. However, these require an explicit model for the observations obtained below the threshold, a task that is beyond the scope of our application. Our modelling framework treats the choice of $u$ as a decision that should be made on the basis of subject-matter expertise and validated by investigating model fit. This approach is illustrated in Section 4 with the analysis of the S&P500.

3. Computation and prior elicitation

3.1. Markov chain Monte Carlo algorithm

The likelihood function associated with the non-homogeneous Poisson process giving rise to the exceedances in sector $j$ is $\gamma_j^n \exp\{-\gamma_j \prod_{k=1}^{n_j} f_j(t_{j,k})\}$. Hence, the joint posterior distribution for our model can be written as

$$p(\{\gamma_j^\ast\}, \{v_{j,l}\}, \{\bar{\mu}_{j,l}\}, \tau, \{\alpha_j\}, \pi \mid \text{Data}) \propto \prod_{j=1}^{J} (\gamma_0^\ast + \gamma_j^\ast)^{n_j} \exp\{- (\gamma_0^\ast + \gamma_j^\ast) \}
\times \prod_{j=1}^{J} \prod_{k=1}^{n_j} \left( \frac{\gamma_0^\ast}{\gamma_0^\ast + \gamma_j^\ast} \sum_{l=1}^{\infty} \left\{ v_{0,l} \prod_{s<l} (1 - v_{0,s}) \right\} \psi(t_{j,k} \mid \bar{\mu}_{0,l}, \tau) + \frac{\gamma_j^\ast}{\gamma_0^\ast + \gamma_j^\ast} \sum_{l=1}^{\infty} \left\{ v_{j,l} \prod_{s<l} (1 - v_{j,s}) \right\} \psi(t_{j,k} \mid \bar{\mu}_{j,l}, \tau) \right)$$

$$\times p(\pi)p(\tau)p(\gamma_0^\ast) \prod_{j=1}^{J} p(\gamma_j^\ast \mid \pi) \prod_{j=0}^{J} \prod_{l=1}^{\infty} h(\bar{\mu}_{j,l}) \text{Beta}(v_{j,l} \mid 1, \alpha_j) \prod_{j=0}^{J} p(\alpha_j)$$

(2)

where the priors $p(\pi)$, $p(\tau)$, $p(\gamma_0^\ast)$, $p(\gamma_j^\ast \mid \pi)$, for $j = 1, \ldots, J$, and $p(\alpha_j)$, for $j = 0, 1, \ldots, J$, are given in Section 2.

Since this posterior distribution is computationally intractable, we resort to a Markov chain Monte Carlo algorithm (Robert and Casella, 2005) for simulation-based inference. Given initial values for the different parameters, the algorithm proceeds by iteratively updating blocks of parameters by sampling from their posterior full conditional distribution. After an appropriate burn-in period, the algorithm produces a dependent sample that is approximately distributed according to (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 and James, 2001). Hence, for computational purposes, we replace the nonparametric
mixing distributions $G^*_0, G^*_1, \ldots, G^*_J$ with finite-dimensional approximations

$$G^*_j(\cdot) = \sum_{l=1}^{N} \left\{ u_{j,l} \prod_{s<l} (1 - u_{j,s}) \right\} \delta_{\mu^*_{j,l}}(\cdot), \quad j = 0, 1, \ldots, J.$$ 

where, as before, the $\mu^*_{j,l}$ are drawn independently from $H$ and the $u_{j,l}$ are independently $\text{Beta}(1, \alpha_j)$ distributed, 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 for each observation $t_{j,k}$ a pair of latent configuration variables, $(r_{j,k}, L_{j,k})$, where $r_{j,k} \in \{0, j\}$ is an indicator for the systemic or idiosyncratic component, whereas $L_{j,k} \in \{1, \ldots, N\}$ identifies the respective mixture component under the truncation approximation $G^*_0$ or $G^*_j$. More specifically, independently for $j = 1, \ldots, J$ and $k = 1, \ldots, n_j$, $\Pr(r_{j,k} = 0 \mid \gamma^*_0, \gamma^*_j) = 1 - \Pr(r_{j,k} = j \mid \gamma^*_0, \gamma^*_j) = \gamma^*_0/(\gamma^*_0 + \gamma^*_j)$, and, for $l = 1, \ldots, N$, $\Pr(L_{j,k} = l \mid r_{j,k} = 0, G^*_0) = u_{0,l} \prod_{s<l} (1 - u_{0,s})$ and $\Pr(L_{j,k} = l \mid r_{j,k} = j, G^*_j) = u_{j,l} \prod_{s<l} (1 - u_{j,s})$. 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 $\Pr(\xi_j = 0 \mid \pi) = 1 - \pi$. Inferences on $\Pr(\xi_j = 0 \mid \text{Data})$ provide an operational mechanism to implement the idiosyncrasy test discussed at the end of Section 2.1. Details of the algorithm are provided in the Appendix.

### 3.2. Hyperparameter elicitation

The hyperparameters associated with our nonparametric model can be elicited 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 $j \geq 1$ (which corresponds to $E\{\gamma^*_0 + \gamma^*_j\}$) by assuming that returns in the sector are normally distributed, so that $E\{\gamma^*_0 + \gamma^*_j\} \approx T \Phi(-u - \zeta_j)/\kappa_j$, where $\Phi$ denotes the cumulative distribution function for the standard normal distribution and $\zeta_j$ and $\kappa_j$ are rough estimates of the mean and standard deviation of returns for market $j$. The values of $\zeta_j$ and $\kappa_j$ can be obtained from historical data or expert knowledge. For simplicity, it can be assumed that $\zeta_j$ and $\kappa_j$ 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 densities \(f_0^*\) 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 parameters \(\alpha_0, \alpha_1, \ldots, \alpha_J\) (which, as mentioned in Section 2, control 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 \(\alpha_j\), see for example Escobar and West, 1995) can be elicited from a rough estimate of the frequency at which distress periods arise in market \(j\). Similarly, the value for the scale parameter \(\tau\) 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. 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.

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. These leaves us with sample sizes between 85 (for consumer staples) and 387 (for information technology) exceedances. A 2% drop in the market has been historically used as a threshold for trading curbs on program trades (which involve a basket of stocks from the S&P500 index where there are at least 15 stocks or where the value of the basket is at least $1 million). Although these particular trading curbs were abandoned in late 2007 because of their inefficacy in reducing market volatility, we believe that the 2% threshold is still a useful guideline to identify periods of distress in the market without excessively thinning the sample. Prices for the individual indexes were obtained from Bloomberg financial services; the corresponding tickers are
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 and 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 in 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 3.2, we chose the hyperparameters for $\gamma_0^*$ and $\gamma_j^*$ so that $a_{\gamma_0^*} = 7.32$, $b_{\gamma_0^*} = 0.06$, $a_{\gamma_j^*} = 1.32$, and $b_{\gamma_j^*} = 0.06$. 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 $\pi$, 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 $\pi$ close to zero. Finally, the prior for the scale parameter $\tau$ was selected so that $a_\tau = 5$ and $b_\tau = 2,400$ (which implies that $E\{\tau\} = 600$); for the precision parameter $\alpha$ 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 $\pi$) 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 2 (for further details on the sensitivity analysis see Section 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 2. The last two panels also provide summaries of the prior distribution over intensities induced by the prior choices discussed before. By comparing some of those estimates to the raw data presented in Figure 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 3 and 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 5 presents the posterior distribution for $\epsilon_1, \ldots, \epsilon_{10}$, 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 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 crisis. 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 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 economic conditions but whose consumption is otherwise relatively stable.

As we discussed in Section 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 6 shows the posterior distributions for the odds ratios 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 the other sectors (the estimated posterior probabilities are 1.000, 0.913 and 0.977 for
consumer staples, energy and information technology, respectively). Indeed, note that 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 caused 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.

[Fig. 6 about here.]

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 crossvalidation datasets being constructed by randomly selecting 20% of the observations from each sector to be used as held-out data. The remaining 80% of the data was used to fit our nonparametric model and generate nominal 90% highest posterior density (HPD) intervals for new exceedances. The true coverage of these HPD intervals was then evaluated on the held-out data. Figure 7 presents examples of crossvalidation samples and the corresponding intensities for two different sectors.

[Fig. 7 about here.]

We repeated the process described above for 10 different crossvalidation datasets, with the results being presented in Figure 8. As expected, there is variability in the coverage rates depending on the sector and the specific crossvalidation 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.

[Fig. 8 about here.]

In addition to the crossvalidation 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 time-rescaling theorem (see for example Daley and Vere-Jones, 2003), which ensures that, if \( \{t_{j,k} : k = 1, \ldots, n_j\} \) is a non-homogeneous Poisson process with intensity \( \Lambda_j(t) \), then the transformed point process \( \{\Lambda_j(t_{j,k}) : k = 1, \ldots, n_j\} \) is a homogeneous Poisson process with unit intensity. Hence, if the Poisson process assumption
is correct, we would expect the transformed inter-arrival times $z_1, \ldots, z_{n_j}$ defined as $z_k = 1 - \exp\left\{-\left[\Lambda_j(t_{j,k}; G_j, \tau) - \Lambda_j(t_{k-1,j}; G_j, \tau)\right]\right\}$ (with the convention $t_{0,j} = 0$) to be uniformly distributed in the unit interval. A variant of this check for extreme value models was originally proposed by Hill (1975) in the context of threshold selection.

Figure 9 presents quantile-quantile plots of the expected value of these transformed inter-arrival times, $E\{z_k \mid \text{Data}\}$ 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 systematically 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.

4.2. 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 scale parameter $\tau$, including a $\text{IGam}(5, 4000)$, a $\text{IGam}(10, 7200)$ and a $\text{IGam}(2, 500)$ 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 $\text{Gam}(4, 1/3)$, a $\text{Gam}(10, 2)$, a $\text{Gam}(2, 0.4)$, and a $\text{Gam}(3, 3)$. These hyperparameter choices imply prior expected values for $\alpha_j$ 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 $E\{\alpha_j\}$ naturally lead to somewhat smaller posterior means for the $\alpha_j$, but also to larger posterior values for $\tau$ and an increase in the posterior mean for some some of the $\epsilon_j$ (for example, for consumer staples we have $Pr(\epsilon_2 = 1 \mid \text{Data}) = 0.71$ under the $\text{Gam}(3, 3)$ prior). On the other hand, changes in the prior dispersion of $\alpha_j$ had no discernible effect on the individual posterior distributions of the hyperparameters, as long as the prior mean was kept constant.

To assess the effect of the baseline distribution $H$ 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 intensities $f_1^*, \ldots, f_{10}^*$. In addition, the posterior means for $\alpha$ and $\tau$ tend to be slightly lower under this baseline measure. However, the overall structure of the results is unchanged.
We also considered an alternative specification for the priors on $\gamma^*_0$ and $\gamma^*_1, \ldots, \gamma^*_0$ 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 $\Pr(0.2 < \gamma^*_0 (\gamma^*_0 + \gamma^*_j)^{-1} < 0.8) = 0.99$. 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 $\epsilon_1, \ldots, \epsilon_{10}$ 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 ($\Pr(\epsilon_2 = 1 \mid \text{Data}) \approx 1$), while for the other sectors we have very strong evidence for the presence of idiosyncratic components ($\Pr(\epsilon_j = 1 \mid \text{Data}) = 0$ for all $j = 1, 3, 4, \ldots, 10$).

Finally, we investigated the effect on posterior inferences of alternative prior distributions on $\pi$. In addition to the original $\text{Beta}(1/2, 2)$ prior (which favors the hypotheses that most the exceedances are generated by the systemic component of the model), we considered a uniform and a $\text{Beta}(1/2, 1/2)$ prior for $\pi$. While the $\text{Beta}(1/2, 1/2)$ 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 $\epsilon_j$s (for example, $\Pr(\epsilon_2 = 1 \mid \text{Data}) = 0.76$).

5. Discussion

We have discussed and illustrated a novel modelling approach for simultaneous risk assessment in multiple financial markets. Our approach models the process of exceedances as a superposition of two independent Poisson processes. The use of a superposition is specially appealing if markets share similar products. This is the situation, for example, when modelling 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.

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.

Acknowledgements

This research was supported in part by the National Science Foundation under awards SES 1024484 and DMS 0915272, and by funds granted by SIGFIRM at the University of California, Santa Cruz.
Appendix. Details of the Markov chain Monte Carlo algorithm

Here, we provide the details for posterior simulation from the proposed DP mixture model for the non-homogeneous Poisson process densities discussed in Section 3.1. 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:

(a) 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

\[
\Pr(r_{j,k} = 0, L_{j,k} = l | \cdots, \text{Data}) \propto \gamma_0^* \left[ u_{0,l} \prod_{s < l} \{1 - u_{0,s}\} \right] \psi(t_{j,k} | \mu_{0,l}, \tau),
\]

and

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

for \(l = 1, \ldots, N\).

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

\[
u_{j,l} | \cdots, \text{Data} \sim \text{Beta} \left(1 + M_{j,l}, \alpha_j + \sum_{r = l+1}^N M_{j,r}\right),
\]

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

(c) The full conditional distribution for the atoms \(\mu_{j,l}^*\) with \(j = 0, \ldots, J\) and \(l = 1, \ldots, N\) is given by

\[
p(\mu_{j,l}^* | \cdots, \text{Data}) \propto h(\mu_{j,l}^*) \prod_{(s,i) : r_{s,i} = j, L_{s,i} = l} \psi(t_{s,i} | \mu_{j,l}^*, \tau).
\]

Hence, if \(\{(s, i) : r_{s,i} = j, L_{s,i} = l,\} = \emptyset\), \(\mu_{j,l}^*\) 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 \(\mu_{j,l}^*\) using a random walk Metropolis-Hasting algorithm with symmetric logit-normal proposal distribution where new values \(\mu_{j,l}^{*(p)}\) are generated according to

\[
\text{logit} \left\{ \mu_{j,l}^{*(p)} \right\} | \mu_{j,l}^{*(c)} \sim \text{Normal} \left(\text{logit} \left\{ \mu_{j,l}^{*(c)} \right\}, \kappa_\mu^2 \right),
\]

where \(\mu_{j,l}^{*(c)}\) represents the current value of the chain and \(\kappa_\mu^2\) is a tuning parameter chosen so that the average acceptance rate is between 30% and 40%.
(d) Since the prior for $\gamma^*_0$ 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 $m_j = \sum_{k=1}^{n_j} I(r_{j,k}=0)$ denotes the number of observed exceedances in market $j$ that are associated with the systemic component.

(e) For $j = 1, \ldots, J$, the pairs $(\xi_j, \gamma^*_j)$ are conditionally independent from each other and can be updated by first updating $\xi_j$ so that $\xi_j = 1$ if $m_j < n_j$ or $\xi_j \mid \cdots, \text{Data} \sim \text{Bernoulli} \left( \pi \left\{ b_{\gamma^*_j} / (1+b_{\gamma^*_j}) \right\}^{\gamma^*_j} \right)$ if $m_j = n_j$. As before, $m_j = \sum_{k=1}^{n_j} I(r_{j,k}=0)$ denotes the number of observed exceedances in market $j$ that are associated with the systemic component. Once $\xi_j$ has been updated, $\gamma^*_j$ is updated by setting $\gamma^*_j = 0$ if $\xi_j = 0$ or by sampling $\gamma^*_j$ from a gamma distribution with shape parameter $a_{\gamma^*_j} + n_j - m_j$ and rate parameter $1 + b_{\gamma^*_j}$ if $\xi_j = 1$.

(f) The full conditional posterior for $\tau$ is

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

Since no direct sampler is available from this distribution, we update $\tau$ using a random walk Metropolis-Hasting algorithm with symmetric log-normal proposal, $\log(\tau(p)) \mid \tau(c) \sim \text{Normal}(\log(\tau(c)), \kappa^2_\tau)$, where $\kappa^2_\tau$ is again a tuning parameter.

(g) Because the prior on $\pi$ is conditionally conjugate, the full conditional for $\pi$ is given by a Beta $\left( J - \sum_{j=1}^{J} \xi_j + a_\pi, \sum_{j=1}^{J} \xi_j + b_\pi \right)$ distribution.

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

References


Fig. 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 seven events of significance to the markets: the bursting of the .com bubble (03/10/2000), the 09/11 terrorist attacks (09/11/2001), the stock market downturn of 2002 (09/12/2002), 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).
Fig. 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 of exceedances observed in each sector over the 12 year period under study. The last two figures in the bottom row present prior realizations for the intensity function (central panel) and the mean prior intensity function along with prior 95% pointwise credible intervals (right panel).
Fig. 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 intensity function, while the right panel shows the mean prior intensity function and prior 95% pointwise credible intervals.
Fig. 4. Posterior mean of the densities associated with the different components of the S&P500 index, along with posterior pointwise credible intervals.
Fig. 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.
Fig. 6. 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.
Fig. 7. Examples of crossvalidation datasets (plotted on the horizontal axis) and the density estimates associated with them.
Fig. 8. Results from the cross validation exercise to investigate the coverage rate of highest posterior density intervals associated with our nonparametric model.
Fig. 9. 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.