Bayesian analysis of extreme values by mixture modeling

Leonardo Bottolo

University of Pavia, Italy and Athens University of Economics, Greece

lbottolo@eco.unipv.it

Guido Consonni *

University of Pavia, Italy

guido.consonni@unipv.it

Petros Dellaportas

Athens University of Economics, Greece

petros@aueb.gr

Antonio Lijoi

University of Pavia, Italy

lijoi@unipv.it

Abstract

Modeling of extreme values in the presence of heterogeneity is still a relatively unexplored area. We consider losses pertaining to several related categories. For each category, we view exceedances over a given threshold as generated by a Poisson process whose intensity is regulated by a specific location, shape and scale parameter. Using a Bayesian approach, we develop a hierarchical mixture prior, with an unknown number of components, for each of the above parameters. Computations are performed using Reversible Jump MCMC. Our model accounts for possible clustering effects and takes advantage of the similarity across categories, both for estimation and prediction purposes. Detailed guidance on the specification of the prior distribution is provided, together with an assessment of inferential robustness. The method is illustrated throughout using a data set on large claims against a well-known insurance company over a 15-year period.

Keywords: Heterogeneity; Insurance claim and loss; Partition; Predictive distribution; Reversible Jump MCMC.

*Address for correspondence: Dipartimento di Economia Politica e Metodi Quantitativi, Via S. Felice 5, 27100 Pavia, Italy.

1 Introduction and summary

Modeling of extreme data is now a well established branch of modern statistics, see e.g. Smith (1989), Embrechts, Klüppelberg, and Mikosch (1997) and Coles (2001).

On the other hand, the Bayesian analysis and modeling of extreme values is, to our knowledge, relatively scarce, notable exceptions being Coles and Powell (1996), and references therein, Coles and Tawn (1996), Smith (1999), Smith and Goodman (2000), Coles and Pericchi (2001). This is unfortunate for several reasons, among which we like to emphasize:

- extreme values are, by definition, rare and thus incorporating prior structure and information into the analysis is often essential to obtain meaningful answers;
- prediction of (extreme) losses is a crucial enterprise, especially in insurance and finance, and the Bayesian method provides a simple and clear cut answer that duly incorporates estimation uncertainty;
- heterogeneity of data is increasingly recognized as an important issue in applications. Again the Bayesian approach allows a variety of clustering techniques for modeling extreme data.

Motivated by the analysis of a data set on seven types of insurance claims carried out in Smith and Goodman (2000), we propose a Bayesian hierarchical model that embodies clustering of type-effects. While our treatment is general, we present our method for ease and concreteness of exposition with reference to the insurance claim data. Major features of our model are:

- the number of clusters is not fixed in advance; instead it is left to the data and the model to determine a posterior probability distribution on this quantity;
- a *distinct* clustering structure is allowed for each of the main parameters of the model: in this way claim-type heterogeneity can be more flexibly accounted for;
- predictive distributions of extreme losses can be routinely computed within our framework. Furthermore the predictions of losses for claim-types on which very

few observations are available can be performed more reliably through a borrowing-strength procedure;

- Markov chain Monte Carlo (MCMC) methods are used to compute estimates and predictions; in particular Reversible Jump MCMC (RJMCMC) technology, see Green (1995), is exploited for model determination.

The structure of the paper is as follows. Section 2 is devoted to the statistical modeling of extreme data characterized by heterogeneity. In particular, subsection 2.1 describes a data set on large claims in the insurance industry together with the standard modeling approach that motivated our search for a new model; subsection 2.2 briefly reviews some Bayesian modeling strategies, in particular partition and mixture models to deal with heterogeneity and clustering, while subsection 2.3 presents a detailed description of the structural features of the proposed model, together with a graphical-model representation. Section 3 offers some guidance to model specification so that our method can be readily implemented in an applied context. Section 4 examines the data set on insurance claims, presents the basic findings, compares them with previous results, and carries out some sensitivity analysis. Finally, Section 5 contains some concluding remarks. To facilitate the exposition, the computational details of the RJMCMC algorithm are relegated to the Appendix.

2 Modeling heterogeneity in extreme data

Heterogeneity is a pervasive concept in statistical analysis. In this section we treat this topic with explicit reference to the modeling of extreme values. Our analysis is general; however, for a more concrete motivation and illustration, we discuss the issue in the context of insurance claims.

2.1 Claim-type effects in the insurance industry

Assessment of the risk of an insurance company depends crucially on the prediction of the size and frequency of large claims. Reliable predictions, in turn, require accurate statistical models. Smith and Goodman (2000), henceforth S&G, analyze data on large claims against a well-known international company over a 15-year period. To protect confidentiality the data were coded so that neither the actual size, nor the time frame, are disclosed. The data comprise a collection of 425 claims classified into one of seven types (e.g. fire, flood, liability, etc.). It is common to aggregate claims relative to a single source into a single "loss", in order to prevent dependency among the data. When this is done, the number of data, which we still name "claims", reduces to 393. A further data reduction occurs when only large losses exceeding a specified threshold are considered. In particular only six claim-types need to be examined for such losses. Typical questions that have to be addressed refer to the choice of a suitable probabilistic model for very large claims, the estimation of the parameters and the prediction of overall losses. S&G, besides providing an informative description of the data set, deal extensively with these topics. In particular they discuss issues such as: a point-process representation of the exceedances, the choice of suitable "thresholds", outlier detection, model estimation and diagnostic fits, trend- and claim-type effects. With reference to the latter issues, they conclude that there seems to be no convincing evidence of any long-term trend. On the other hand, claim-type effects are significant and ignoring them can lead to serious model misspecification and inappropriate parameter estimates. We like to emphasize at this stage that S&G analyze claim-type effects using a hierarchical Bayesian model assuming exchangeability of the parameters. The main disadvantage of the latter model is that types are treated symmetrically and correspondingly parameter estimates are shrunk towards a common point.

2.2 Clustering *via* mixture distributions

Suppose there exist I claim-types, so that the data can be arranged as $y_i = (y_{il}, i = 1, \ldots, I, l = 1, \ldots, n_i)$. The standard Bayesian approach to model claim-type heterogeneity is to assume a specific parameter η_i, say, for the distribution of y_i and to model hierarchically $\eta = (\eta_1, \ldots, \eta_I)$. Several options are available at this stage depending on the amount of structural prior information. If there exist claim-type specific covariates and these are deemed to be relevant, then a regression model on some suitable transformation of η_i may be attempted. A special case of this set-up is the *partial exchangeability* model. Here a set of categorical factors is used to classify types into groups, or clusters, each group corresponding to a specific combination of the factor-levels , see e.g. Bernardo and Smith (1994, Section 4.2.2 and 4.6). Types belonging to the same group are deemed to be "similar" to each other: accordingly the η_is within each group are modeled as exchangeable. Trivially, when no classification factor is employed, one recovers the standard model of full exchangeability. On the other hand, one can go beyond the standard assumption of full exchangeability by assuming a prior distribution on several cluster-structures and the corresponding posterior may be regarded as a Bayesian cluster analysis. These concepts are at the heart of partition models, see Hartigan (1990), Malec and Sedransk (1992) and Consonni and Veronese (1995).

Another approach that bears a resemblance to partition models, but seems more amenable to computations, is based on mixture distributions. Conditionally on the number of mixture-components k, the weights w_js and the hyperparameters δ_js , the η_is are assumed to be iid according to a finite-mixture distribution

$$\eta_i \,|k, w, \delta \overset{iid}{\sim} \sum_{j=1}^{k} w_j p_j(\cdot|\delta_j), \tag{1}$$

where $p_j(\cdot|\cdot)$ denotes the jth component distribution, w stands, here and in the sequel, for the vector $(w_1, \ldots, w_k)$, $\sum_{j=1}^{k} w_j = 1$, and similarly for $\delta = (\delta_1, \ldots, \delta_k)$. Priors are now assigned, possibly using a further hierarchy, to k, w and δ.

It is convenient to rewrite slightly (1) using, independently for each η_i, a latent variable z_i such that $\Pr(z_i = j|k, w) = w_j$. The variable z_i identifies the mixture component from which η_i is drawn. Formula (1) can thus be replaced with

$$\begin{aligned} \eta_i \,|\delta, z_i &\overset{ind}{\sim} p_{z_i}(\cdot\,|\delta_{z_i}) \\ z_i \,|k, w &\overset{iid}{\sim} \Pr(z_i = j\,|k, w) = w_j. \end{aligned}$$

The employment of the allocation variables z_is turns out to be a useful tool especially from the computational viewpoint.

Bayesian mixture models with an unknown number of components have become a powerful tool for statistical analysis following the work of Richardson and Green (1997), which makes use of the RJMCMC algorithm introduced by Green (1995). Green and Richardson (2001) further elaborate on the issue and make comparisons between parametric and nonparametric mixture models. For our purposes the paper by Nobile and Green (2000), henceforth N&G, is especially useful, since it deals with mixture models for parameters, as we do, and not for data, as in Richardson and Green (1997).

2.3 A hierarchical mixture model

We follow Smith (1989) and we view exceedances as generated by a Poisson process of intensity

$$\Lambda_y = \left(1 + \xi \frac{y - \mu}{\psi}\right)_+^{-1/\xi}, \tag{2}$$

where $x_+ = \max(0, x)$. In equation (2) μ represents a location parameter, ξ a shape parameter, while ψ accounts for scale. One of the advantages of the above representation is that it can be immediately extended to cover more complex situations, such as those with time-dependent parameters, see e.g. S&G.

Consider now data $y = (y_{il})$, $\quad i = 1, \ldots, I; \quad l = 1, \ldots, n_i(u_i)$, where y_{il} represents the lth exceedance over threshold u_i for claim-type i.

Conditionally on μ_i, ψ_i and ξ_i we assume that, independently, the point process of exceedances of type i is generated according to the Poisson process (2) with parameters μ_i, ψ_i and ξ_i. This is the only step in the model we share with the paper by S&G.

Following N&G we specify a hierarchical mixture prior distribution for the collection of the μ_i, ψ_i and ξ_i parameters. When no specific indication is given, conditional independence is tacitly assumed.

Let

$$\begin{aligned} \mu_i \left| k^\mu, w^\mu, \theta^\mu, \sigma^\mu \right. &\overset{iid}{\sim} \sum_{j=1}^{k^\mu} w_j^\mu N(\theta_j^\mu, (\sigma_j^\mu)^2), \\ \log \psi_i \left| k^\psi, w^\psi, \theta^\psi, \sigma^\psi \right. &\overset{iid}{\sim} \sum_{j=1}^{k^\psi} w_j^\psi N(\theta_j^\psi, (\sigma_j^\psi)^2), \\ \xi_i \left| k^\xi, w^\xi, \theta^\xi, \sigma^\xi \right. &\overset{iid}{\sim} \sum_{j=1}^{k^\xi} w_j^\xi N(\theta_j^\xi, (\sigma_j^\xi)^2). \end{aligned} \tag{3}$$

Notice that we postulate, for each of the three parameters of the Poisson process, a distinct mixture distribution, with specific number of components, weights, means and variances.

To complete the model specification, the distributions of the hyperparameters appearing in (3) must be detailed. For simplicity we only present those referring to μ. Similar expressions hold with respect to $\log\psi$ and ξ. We assume that

$$
\begin{array}{rcl}
k^{\mu} & \stackrel{iid}{\sim} & Unif\{1,\ldots,I\}, \\
w^{\mu}\,|k^{\mu} & \sim & Dir(d_1^{\mu},\ldots,d_{k^{\mu}}^{\mu}), \\
\theta_j^{\mu}\,|\tau^{\mu} & \stackrel{ind}{\sim} & N(m_j^{\mu},\dfrac{1}{\tau^{\mu}}), \\
(\sigma_j^{\mu})^{-2} & \stackrel{ind}{\sim} & Ga(a_j^{\mu},b_j^{\mu}), \\
\tau^{\mu} & \sim & Ga(a^{\tau^{\mu}},b^{\tau^{\mu}}),
\end{array}
\tag{4}
$$

where $Unif\{1,\ldots,I\}$ denotes the uniform distribution over the integers $\{1,\ldots,I\}$, Dir stands for the Dirichlet distribution, $Ga(a,b)$ denotes the gamma distribution with expectation a/b.

The joint distribution of all random quantities involved is now fully specified and the corresponding Directed Acyclic Graph (DAG) is depicted in Figure 1.

Figure 1 about here

3 Specification of prior distributions

Consider first the prior distribution on w^{μ}. A useful way to think about the choice of the d_j^{μ}s is in terms of the induced probability distributions on the *partitions* of the I types for each fixed k^{μ}. This is the approach taken in N&G which we have found especially illuminating. In our example, set, for instance, $k^{\mu}=3$. Then the claim-types parameters μ_i can be regarded as random draws from one of three distinct Normal populations with weights $w_j^{\mu}, \quad j=1,2,3$. The latter in turn are random and governed by the coefficients d_j^{μ}. For each fixed value of $d^{\mu}=(d_j^{\mu})$ one can compute the probability of the various partitions of six objects of degree less than or equal to three and check the reasonableness of the assignments against such distribution. For example, if one wishes to favor aggregation of claim-types, then one would want to place the highest probability on the partition of degree (i.e. number of groups) one, where types belong to the same population, followed by those of degree two and finally by that of degree three. In practice some default methods are called for and we recommend using $d_j^{\mu}=1$ for all j. This choice can be shown to favor rather strongly the assumption of a single cluster and is thus also useful for comparison purposes with works based on the assumption of full exchangeability. More generally, the above choice of prior encourages aggregation of types into a few number of clusters. Similar considerations apply to the choice of d_j^{ψ} and d_j^{ξ}.

Table 1 reports the prior probabilities induced on the various partitions of $I = 6$ objects from a prior $w \sim Dir(1, \ldots, 1)$ (we omit superscript since our considerations are general). More precisely the first column of Table 1 reports selected values of k ranging from one to three. Column 2 describes the structure of the partitions of the index set $\{1, \ldots, 6\}$ in term of the cardinality of its clusters. For example for $k = 2$, $(4, 2)$ indicates that a partition has two clusters, one having four elements while the other two. In the third column, to the right of $(4, 2)$, a full description of all partitions is provided. For example if $\pi = (1, 1, 2, 1, 1, 2)$ this means that elements one, two, four and five belong to the same cluster (indicated with 1), while elements three and six belong to cluster two. Notice that $\pi = (2, 2, 1, 2, 2, 1)$ would have provided an equivalent description of the partition and is accordingly omitted. Finally the last column reports the probability of each partition appearing to its left. Clearly the probability of the overall cluster structure is obtained by multiplying the previous probability by the number of partitions in the set, e.g. the probability of the cluster structure $(5, 1)$ is $6 \times 0.0223 = 0.1338$.

Table 1 about here

Consider now the distribution of the θ_j^{μ}s. Notice that the expectations m_j^{μ}s are fixed, while the variances are all equal but random $(\frac{1}{\tau^{\mu}})$. In general (hierarchical) mixture models are more sensitive to variance, rather than mean, specifications. In particular $\sqrt{\frac{1}{\tau^{\mu}}}$ represents variability *between* components of the mixture while the σ_j^{μ}s are measures of *within* components variability. We claim that the former should be appreciably larger than the latter: in this way we are fairly stringent about the variability within each component but are largely uninformative relative to the location of the components (see below for details). Clearly special care is needed in assigning these prior distributions, since the structure of the mixture as well as the parameter estimates are highly dependent upon them.

As for the values of the coefficients m_j^{μ}s, the default choice we suggest is to take them all equal to m^{μ}, say. Using an Empirical Bayes approach, we found that a fruitful procedure is to compute the MLE for each μ_i, $\hat{\mu_i}$ and then set $m^\mu = \frac{\min\{\hat{\mu_i}\}+\max\{\hat{\mu_i}\}}{2}$.

We now consider the assignment of the coefficients a_j^μ and b_j^μ governing the distribution of $(\sigma_j^\mu)^{-2}$. Since we are willing to believe that two μ_is come from the same mixture component if they are "close", the range of values of each σ_j^μ should be fairly "small". The meaning of "small" is of course context-dependent, and it is not immediate for the parameters μ_i, ψ_i and ξ_i in the Poisson process model. As a consequence the procedure of N&G (Subsection 3.2) cannot be straightforwardly applied in our setting. We therefore propose a modification that again relies on some empirically based estimates. First of all we fix $a_j^\mu = 3$ for all j: this guarantees that $(\sigma_j^\mu)^2$ will have finite second-order moment. Next we evaluate the control parameter Δ^μ, which represents with (high) probability p_0 the maximal discrepancy between two distinct μ_is coming from the same mixture component. In other words $\Pr(|\mu_i - \mu_{i'}| \leq \Delta^\mu) = p_0$. Clearly small values of Δ^μ will favor a greater number of mixture components, and conversely for large values of Δ^μ. This is the procedure we propose: we compute all pairs $\hat{\Delta}_{ii'} = |\hat{\mu}_i - \hat{\mu}_{i'}|$, which provide a good indication of the observed range for discrepancies. We can now fix Δ^μ equal to some suitable quantile of the distribution of the $\hat{\Delta}_{ii'}$s. Typically the estimated clustering structure will be fairly sensitive to this choice. We recommend trying a few quantiles, e.g. between the 2nd and the 6th decile, to gauge sensitivity of the inferences. The numerical assessment of b_j^μ can now be obtained solving formula (8) of N&G (we set $p_0 = 0.95$). Specifically one has:

$$b_j^\mu = \frac{a_j^\mu}{2} (\Delta^\mu)^2 \left\{ F^{-1}_{2a_j^\mu} \left(\frac{1+p_0}{2} \right) \right\}^{-2},$$

where $F_{2a_j^\mu}$ is the distribution function of the Student distribution with $2a_j^\mu$ degrees of freedom. Since we chose to fix the a_j^μs all equal, it follows that the b_j^μs will be also equal, i.e. $b_j^\mu = b^\mu$.

There remains two more coefficients to fix, namely a^{τ^μ} and b^{τ^μ}, i.e. those governing the distribution of τ^μ. The former is set equal to 3 again, while the latter can be obtained from equation (9) of N&G. We disregard the trivial solution $b^{\tau^\mu} = b^\mu$, so that b^{τ^μ} will always be larger than b^μ: this guarantees that we are largely uninformative about the component locations θ_j^μ. This completes the procedure as far as the distributions of the quantities corresponding to the μ-parameters are concerned. The procedure is now repeated twice, with obvious modifications, in order to specify the distributions of the ψ- and ξ-parameters.

We close this section by remarking that choosing the coefficients d_j, m_j, a_j and b_j (we omit superscripts for simplicity) all equal among themselves, as we recommend, implies that the mixture labels are not uniquely identified. This label-switching problem has recently attracted the attention of many researchers (Celeux *et al.*, 2000, Stephens, 2000, and Frühwirth-Schnatter, 2001). In order to overcome this difficulty, some identifying constrains need to be imposed. This is done in a post-processing "after simulation" stage by ordering the labels according to the values taken by the parameters θ_j or σ_j or w_j. In our real-data setting, using the parameter θ_j, removed the label-switching problem in the sense that the posterior distribution of θ_j, σ_j and w_j did not exhibit multimodalities. This situation occurs for the μ, ψ and ξ parameters. We finally remark, however, that this problem does not affect the predictive distribution.

4 Application to insurance claim data

The methodology described in the previous section is now illustrated using the insurance claim data set described in Subsection 2.1. For comparison purposes we set the same threshold $u = 2.5$ for each claim-type, as in S&G, thus reducing the effective number of claim-types down to six as mentioned in the Introduction.

The MCMC-based results we present are based on $500,000$ sweeps, with a burn-in of $20,000$. Simulation time was close to 4 hours on a Pentium III 850 MHz. Values were recorded at the rate of 1 every 100 to dampen serial correlation. Convergence was assessed using diagnostics implemented in the package BOA, see Smith (2001). These diagnostics were performed on the parameters k^μ, k^ψ and k^ξ as well as for the parameters of selected models having high posterior probability. From these it appears that the number of iterations could have been safely halved without appreciably modifying the results and convergence issues.

We now describe in detail the specification of the coefficients in the prior distribution (4) following closely our discussion in Section 3.

Table 2 reports for three levels of the control parameter Δ, namely the three quartiles of the empirical distribution of the $\hat{\Delta}_{ii'}$ and the values of the coefficients b and b^τ (for simplicity, here and in the Tables, we omit the superscript indicating the parameter under consideration). While increasing Δ makes the values of b and b^τ increase too, we remark that the ratio of between-components variance $\sqrt{\frac{1}{\tau}}$ and within-components variance σ^2, as measured by the ratio of their expectations or the expectations of their ratios, is essentially unaffected, see Table 3.

Table 2 and Table 3 about here.

The value of m was derived in Section 3, obtaining: $m^\mu = 8.3112$, $m^\psi = 1.1864$ and $m^\xi = 0.7864$.

4.1 Clustering structures

From the MCMC output, effectively based on $5,000$ values, one can make inferences about the clustering structure, through the frequency distributions of the vector of partitions. For each of the main parameters of interest, μ, $\log\psi$ and ξ, Tables 4, 5 and 6 report the five most visited partitions for each of the three choices of the control parameter Δ indicated in Table 2. As anticipated in the previous section, the higher the value of Δ, the greater the posterior probability attached to partitions of smaller degree. For example, with reference to the location parameter μ, we notice in Table 4 that the exchangeability partition is virtually negligible for the two smaller values of Δ^μ, but receives three quarters of the probability under the highest value. The most probable partition appears to be that wherein claim-type 1 is isolated while the remaining five types are clustered together. The persistence of this partition is fairly robust to choices of Δ^μ, always scoring highest or second highest; its weight is however strongly downgraded for the largest value of Δ^μ. Further aspects of the posterior distribution are reported in Table 7. For the intermediate value of Δ^μ, only 56 partitions were visited; of these 9 were visited only once and 28 less than 10 times each. The first 10 most visited partitions accounted for almost 90% of the whole posterior probability. In summary one can conclude that, with reference to the parameter μ, there exists overwhelming evidence that claim-type 1 exhibits a quite distinctive behavior, with mild evidence that claim-type 5 may be also considered on its own. Of course this conclusion is broadly consistent with the graphical display of parameter estimates appearing in Figure 5 of S&G. The advantage of our approach is that we provide the posterior probability distributions on the space of partitions.

Table 4 about here

As for the parameter $\log \psi$, see Table 5, again it appears that claim-type 1 belongs to a cluster of its own, provided we are fairly restrictive on the maximal discrepancy allowed. The exchangeability partition however is surely a better candidate here, as it shows up among the five most frequent partitions for all choice of Δ^{ψ}, although it never attains a posterior probability greater than 85%.

Table 5 about here

As far as the last parameter ξ is concerned, see Table 6, it is evident that the exchangeability partition scores always very high with a posterior probability ranging from 76% to 89%. Only very mild evidence is given to the possibility of two clusters. Interestingly, however, it is *not* claim-type 1 that behaves in a distinctive way.

Table 6 about here

A summary of the posterior distribution of the partitions for each parameter of interest and selected values of Δ is reported in Table 7.

Table 7 about here

It appears that the algorithm visits a fair number of potential partitions, but quickly settles down on a very few ones, namely those having greatest support. Indeed it is often the case that the five most visited partitions account for a very large portion of the total probability mass.

Table 8 reports the posterior distribution of the number of mixture components k, for each parameter of interest under the usual three choices of Δ. For simplicity we comment only on the intermediate panel, corresponding to a value of Δ equal to the median of the empirical distribution of the $\hat{\Delta}_{ii'}$. The mode of the distribution of k^{μ} is 2 with appreciable probabilities also on the values 3 and 4. On the other hand both k^{ψ} and k^{ξ} have their mode on 1, although the former presents non trivial masses also on 2 and 3. This of course is consistent with our previous analysis of the distribution of the partitions. Notice however that the distribution of k includes cases with "empty components", namely sweeps wherein the actual number of distinct values taken on by the allocation variables z is smaller than the current value of k. This situation is quite typical in the Bayesian analysis of mixture models. Of course persistence of situations characterized by empty components should be a warning and must be carefully monitored. In our case this was not a problem: indeed the proportion of empty components was lower than that reported in Richardson and Green (1997). As to the proportion of split/combine moves that were accepted this is detailed in the last column of Table 8, and is well in agreement with similar studies using RJMCMC.

Table 8 about here

4.2 Parameter estimates

A simple representation of the posterior distribution of the main parameters is shown in Figures 2-4, for the usual three choices of the control parameter Δ and with stars marking MLEs. Posterior means, medians and 95% credible intervals along with standard deviations and interquartile ranges for the usual three choices of the control parameter Δ are shown in Table 9, 10 and 11. For simplicity we only comment results relative to intermediate value of Δ. It immediately appears from the boxplots that both parameters μ and $\log\psi$ of claim-type 1 are isolated with respect to the other five. In particular the credible interval of μ_1 does not overlap with those remaining, consistently with the findings of the clustering structure previously described. On the other hand the distribution of the parameter ξ appears essentially similar across claim-types.

Table 9, Table 10 and Table 11 about here

Figure 2, Figure 3 and Figure 4 about here

The control parameter Δ operates on the pattern of the shrinkage estimates both through the clustering structure describe in Subsection 4.1 as well as through the within- and between- variance. Two partly conflicting effects emerge as Δ increases: on the one hand the number of clusters tends to diminish, thus favoring aggregation of claim-types; on the other hand, because of the nature of our prior distributions, both variances tend to increase too, thus pushing parameter estimates towards MLEs. This phenomenon is best captured through the crossover diagrams depicted in Figure 5, 6 and 7 which show the sensitivity of the posterior means with respect to Δ. For example, looking at Figure 5, from bottom to top, one can see that estimates diverge as Δ approaches the 0.70 quantile of $\hat{\Delta}_{ii'}$; however as Δ increases further, the estimate for μ_1 is strongly shrunk towards the center since at the 0.75 quantile the exchangeability partition prevails. It is apparent that a "crossover" effect takes place for claim-type 2, 4 and 6, i.e. the lines connecting MLEs and Bayesian estimates intersect due to the shrinkage effect.

For the parameter $\log\psi$ the crossover diagram in Figure 6 is essentially interpretable along the lines described for μ. For the parameter ξ instead, posterior means diverge, from bottom to top, progressively since the exchangeability partition is always prevalent, see Table 6. The cross over effects still appear involving claim-type 4 and 6 and shrinkage is more pronounced when Δ corresponds to the third quartile.

Figure 5, Figure 6 and Figure 7 about here

For all parameters our results with Δ corresponding to the 0.75 quantile are broadly consistent with those obtained by S&G using a full exchangeability model, which in our setting can be reinterpreted as a one-component mixture prior distribution.

4.3 Prediction

Let Y_i^f denote a future exceedance over a threshold u relative to a year period for the ith claim-type, G the distribution function corresponding to the Poisson process of intensity (2) and $h\left(\mu, \log\psi, \xi \,|y\right)$ the posterior density of the parameters. The predictive distribution of Y^f conditional on y for a fixed claim-type (omitted from the notation) is given by

$$\begin{aligned} \Pr\left(Y^f < y^f \,|y\right) &= \int G\left(y^f \,|\mu, \log\psi, \xi\right) h\left(\mu, \log\psi, \xi \,|y\right) d\left(\mu, \log\psi, \xi\right) \\ &\approx \frac{1}{T}\sum_{t=1}^{T} G\left(y^f \left|(\mu, \log\psi, \xi)^{(t)}\right.\right), \end{aligned} \tag{5}$$

where $(\mu, \log\psi, \xi)^{(t)}$ represents the output at the tth iteration of the MCMC algorithm and T is the number of iterations. A plot of y_p^f against $\{-\log(1-p)\}$ on a logarithmic scale, where $\Pr\left(Y^f < y_p^f \,|y\right) = 1-p$, is called "predictive return level plot" and represents the equivalent of the conventional return level plot, see Coles (2001, Section 7.8). There are two ways to construct such a plot: one is to fix a grid of values y^fs and evaluate the G function at these points at each iteration t. Averaging over the MCMC output, this provides an estimate of G at each point of the grid and finally the return level plot is obtained by inversion. Clearly to achieve a reasonable quality of the plot, the grid must be sufficiently fine. An alternative approach is to assign at each iteration a probability $p^{(t)}$ (possibly drawn from a uniform distribution over the unit interval) and then to evaluate $y^f_{p^{(t)}} = G^{-1}\left(p^{(t)}\,|\cdots\right)$ whose empirical distribution function provides an estimate of G. Finally the predictive return level plot is obtained again by inversion. Figure 8 reports the predictive level plots based on MLE estimates, the full exchangeability model of S&G and the mixture models for the usual three choices of the control parameter Δ.

The major feature of the plots appears to be the striking difference between the MLE-based and the Bayesian predictive level plots. These differences are most notable for claim-type 1 and 2, where the MLE-based plots identify a much riskier situation. Conversely for claim-type 5 and 6, the Bayesian plots appear to indicate a considerable greater risk than the frequentist one. We finally stress that the mixture model exhibits some appreciable difference also with respect to the standard model used by S&G.

Figure 8 about here

5 Conclusions

Mixture models represent a flexible tool for Bayesian inference allowing for heterogeneity in the data while taking advantage of the borrowing strength effect. These aspects seem especially fruitful when examining extreme data relative to various categories of claims.

We have shown that current computational techniques allow to fit models and produce predictions with reasonable speed and accuracy. Our results indicate that Bayesian estimates and predictions can be appreciably different from those based on MLEs: as a consequence they can have significant impact also from a practical point of view. Within the Bayesian approach we have demonstrated that mixture models can be successfully used in place of the more traditional method based on full exchangeability.

Future work should extend the model presented here to a dynamic setting thus allowing for time dependent clustering effects.

Acknowledgments

We are indebted to Richard L. Smith for providing us with the data set on the insurance large claims, already analyzed in Smith and Goodman (2000). Leonardo Bottolo's research was supported by the University of Pavia and by a European Union TMR grant ERB-FMRX-CT96-0096 through the Athens University of Economics. Guido Consonni's and Antonio Lijoi's research was partially supported by the University of Pavia research project on the Statistical Analysis of Complex Systems.

Appendix: the Reversible Jump MCMC sampler

The Bayesian analysis described in Section 2 is implemented using the RJMCMC algorithm, introduced in Green (1995). We summarize below the computational scheme that we used. Let $M(A) := \#\{\text{claims whose size lies in } A\}$. For each A belonging to $\mathscr{B}(\mathbb{R})$, where $\mathscr{B}(\mathbb{R})$ is the σ-algebra of Borel, the random measure $M(\cdot)$ is assumed to be a Poisson process and, in particular, for any $y \geq u$, the expected value of $M((y, +\infty))$ coincides with Λ_y defined in (2), so that the likelihood function becomes

$$l\left(y \,|\mu, \psi, \xi\right) = e^{-N\sum_i \Lambda_{u_i}} \prod_{i\,l} \lambda_{y_{il}},$$

where $y = (y_{11}, ..., y_{1n_1}, y_{21}, ..., y_{In_I})$ are the observed exceedances over the claim-type specific threshold $u_i > 0$, $\lambda_x = -d\Lambda_x/dx$ and N is the number of years in which observations are collected, see, e.g., Basawa and Prakasa Rao (1993, Chapter 3). Hence

$$l\left(y \,|\mu, \psi, \xi\right) = e^{-N\sum_i \Lambda_{u_i}} \prod_{i=1}^{I} \frac{1}{\psi_i} \prod_{l=1}^{n_i} \left(1 + \xi_i \frac{y_{il} - \mu_i}{\psi_i}\right)^{-\frac{1}{\xi_i}-1} \mathbb{I}_{(0,+\infty)}(1 + \xi_i \psi_i^{-1}(y_{il} - \mu_i)),$$

where $\mathbb{I}_A$ stands for the indicator function of the set A. Henceforth we shall take for simplicity $u = u_i$. According to the distributional assumptions set forth in (4), one can easily determine the full conditional distributions for each of the vectors w^μ, z^μ, σ^μ, θ^μ, τ^μ given the observed data y and the remaining parameters. If $A_j^\mu := \{i : z_i^\mu = j\}$,

$n_j^\mu := \text{card}(A_j^\mu)$, $\bar{\mu}_j = \sum_{i \in A_j^\mu} \mu_i / n_j^\mu$ and ϕ denotes the density of a standard normal distribution, the full conditional distributions are described below

$$w^\mu \,|\cdots \sim Dir(d_1^\mu + n_1^\mu, ..., d_{k^\mu}^\mu + n_{k^\mu}^\mu) \tag{A.1}$$

$$\Pr(z_i^\mu = j \,|\cdots) = \frac{w_j^\mu \phi(\mu_i; \theta_j^\mu, \sigma_j^\mu)}{\sum_{l=1}^{k^\mu} w_l^\mu \phi(\mu_i; \theta_l^\mu, \sigma_l^\mu)} \tag{A.2}$$

$$(\sigma_j^\mu)^{-2} \,|\cdots \overset{ind}{\sim} Ga\left(a_j^\mu + \frac{n_j^\mu}{2}, b_j^\mu + \frac{1}{2} \sum_{i \in A_j^\mu} (\mu_i - \theta_j^\mu)^2 \right) \tag{A.3}$$

$$\theta_j^\mu \,|\cdots \overset{ind}{\sim} N\left(\frac{m_j^\mu \tau^\mu + n_j^\mu \bar{\mu}_j / (\sigma_j^\mu)^2}{\tau^\mu + n_j^\mu / (\sigma_j^\mu)^2}, \frac{(\sigma_j^\mu)^2}{n_j^\mu + \tau^\mu (\sigma_j^\mu)^2} \right) \tag{A.4}$$

$$\tau^\mu \,|\cdots \sim Ga\left(a^{\tau^\mu} + \frac{k^\mu}{2}, b^{\tau^\mu} + \frac{1}{2} \sum_{i=1}^{I} (\mu_i - \theta_i^\mu)^2 \right), \tag{A.5}$$

where $|\cdots$ stands for conditioning on y and all other random variables, save the one on the left hand side. The full conditional distributions referred to ψ and ξ are identical to the ones above, provided the superscript μ is replaced with ψ, respectively ξ.

The full conditional distribution for the parameters (μ, ψ, ξ) is analytically intractable and thus not available for sampling. This suggests to perform a Metropolis-Hastings step. Let $\eta_i = (\mu_i, \log \psi_i, \xi_i)$ and denote the proposal distribution with $q(\eta_i \,| e_i, v_i)$ which we take as a three-dimensional Normal law with mean e_i and variance-covariance matrix v_i. In order to increase the efficiency of the algorithm the matrix v_i is fixed as follows. Firstly, a maximum likelihood estimate $\hat{v}_i$ of v_i is obtained. Then a preliminary (random-walk) Metropolis-Hastings run is performed with the posterior distribution of η_i as the target distribution using, at this stage, $q(\eta_i \,\big|\, \eta_i^{(t)}, c_i \hat{v}_i)$ as the proposal distribution at the $(t+1)$th iteration, where c_i is a suitable tuning parameter. After running such a chain, one obtains a sample $\{\hat{\eta}_i^{(t)} : t \geq t_0\}$ from the posterior distribution of η_i and we set $v_i = c_i' \tilde{v}_i$, where $\tilde{v}_i$ denotes the corresponding sample variance-covariance matrix and c_i' another suitable tuning parameter.

Below we detail the steps of the sampling procedure. Superscripts are used to denote the iteration, while superscripts referring to the parameters μ, ψ, ξ are omitted for simplicity from the notation. Finally we set $A_j^{(t)} := \{i : z_i^{(t)} = j\}$.

1. `Fix initial values` $z^{(0)}, w^{(0)}, \theta^{(0)}, \sigma^{(0)}, \tau^{(0)}, \mu^{(0)}, \psi^{(0)}, \xi^{(0)}, k^{(0)}$.

2. Sample $z_i^{(t+1)}$ from

$$p(j) = \frac{w_j^{(t)} \phi(\mu_i^{(t)}; \theta_j^{(t)}, \sigma_j^{(t)})}{\sum_{l=1}^{k^{(t)}} w_l^{(t)} \phi(\mu_i^{(t)}; \theta_l^{(t)}, \sigma_l^{(t)})}, \qquad j = 1, ..., k^{(t)}, \quad i = 1, ..., I.$$

3. Sample $w^{(t+1)} = (w_1^{(t+1)}, ..., w_{k^{(t)}}^{(t+1)})$ from $Dir\left(d_1^{(t)} + n_1^{(t+1)}, ..., d_{k^{(t)}}^{(t+1)} + n_{k^{(t)}}^{(t+1)}\right)$, where $n_j^{(t+1)} := \text{card}\left(A_j^{(t+1)}\right)$.

4. Sample $\theta_j^{(t+1)}$ from

$$N\left(\frac{n_j^{(t+1)} \mu_j^{(t)} + \tau^{(t)} (\sigma_j^{(t)})^2 m_j}{n_j^{(t+1)} \tau^{(t)} (\sigma_j^{(t)})^2}, \frac{(\sigma_j^{(t)})^2}{n_j^{(t+1)} \tau^{(t)} (\sigma_j^{(t)})^2}\right).$$

5. Sample $(\sigma_j^{(t+1)})^{-2}$ from $Ga(a_j + n_j^{(t+1)}/2, b_j + s_j^{(t+1)})$, where

$$s_j^{(t+1)} = \frac{1}{2} \sum_{i \in A_j^{(t+1)}} (\mu_i^{(t)} - \theta_j^{(t+1)})^2.$$

6. Sample $\tau^{(t+1)}$ from $Ga(a^\tau + k^{(t)}/2,\, b^\tau + \sum_{i=1}^{I} (\mu_i^{(t)} - \theta_i^{(t+1)})^2/2)$.

7. Sample from the full conditional distribution of $\eta_i^{(t+1)}$ given the observations y and the (vector of) parameters

$$(z^{(t+1)}, w^{(t+1)}, \theta^{(t+1)}, \sigma^{(t+1)}, \tau^{(t+1)}, k^{(t)})$$

via a (random walk) Metropolis-Hastings step with proposal distribution

$$q(\eta_i^{(t+1)} \big| \eta_i^{(t)}, v_i).$$

8. Update k^μ, k^ψ, k^ξ via the RJMCMC.

As far as step 8 is concerned, updating is achieved by means of the RJMCMC algorithm employed in N&G, which is now described omitting as usual the superscript from the notation for simplicity. Suppose first that the parameters (d_j, a_j, b_j, m_j) differ across components, so that the corresponding labels are uniquely identified. The split/merge moves in the RJMCMC work as follows. Let k be the current number of components. A preliminary selection of a splitting move, with probability s_k, or of a merging move, with probability $c_k = 1 - s_k$ is made. Clearly, $s_1 = c_{k_{\max}} = 1$, whereas $s_k = c_k = 0.5$ for each $k = 2, ..., k_{\max} - 1$. If a split move is selected, then one randomly fixes a component, j^* say, to be split into two components, j_1 and j_2. Accordingly, the value of the parameters are set as follows

$$w_{j_1} = w_{j^*} u_1, \qquad w_{j_2} = w_j^*(1 - u_1)$$

$$\theta_{j_1} = \theta_{j^*} - u_2 \sigma_{j^*} \sqrt{w_{j_2}/w_{j_1}}, \qquad \theta_{j_2} = \theta_{j^*} + u_2 \sigma_{j^*} \sqrt{w_{j_1}/w_{j_2}}$$

$$\sigma^2_{j_1} = u_3(1 - u_2^2)\sigma^2_{j^*}(w_{j^*}/w_{j_1}), \qquad \sigma^2_{j_2} = (1 - u_3)(1 - u_2^2)\sigma^2_{j^*}(w_{j^*}/w_{j_2}),$$

where $u_1 \sim Beta(2, 2)$, $u_2 \sim 2Beta(2, 2) - 1$ and $u_3 \sim Beta(0, 1)$. Proceeding analogously as in N&G, we let j_1 coincide with j^* and take j_2 to be the $(k + 1)$th component in the mixture. The observations in j^* are reallocated to j_1 and to j_2 according to the following probabilities

$$\Pr(z_i = j_1) = \frac{p_{i1}}{p_{i1} + p_{i2}} = 1 - \Pr(z_i = j_2) \qquad \forall i \in A_{j^*},$$

and

$$p_{i1} = \frac{w_{j_1}}{\sigma_{j_1}} e^{-\frac{(\mu_i - \theta_{j_1})^2}{2\sigma^2_{j_1}}}, \qquad p_{i2} = \frac{w_{j_2}}{\sigma_{j_2}} e^{-\frac{(\mu_i - \theta_{j_2})^2}{2\sigma^2_{j_2}}} \qquad \forall i \in A_{j^*}. \tag{A.6}$$

If n_1^* and n_2^* denote the number of elements allocated in j_1 and in j_2, respectively, with $n_1^* + n_2^* = n_{j^*} = \text{card}\,(A_{j^*})$, then

$$P_{\text{alloc}} = \prod_{i \in A_{j_1}} p_{i1} \prod_{l \in A_{j_2}} p_{l2}$$

designates the probability of the specific reallocation obtained according to (A.6). The candidate state is accepted with probability $\min(1, R)$ where

$$R = \frac{\sigma_{j^*}^{n^*/2}}{\sigma_{j_1}^{n_{j_1}/2} \sigma_{j_2}^{n_{j_2}/2}} \exp\left\{ -\frac{1}{2} \left(\sum_{i \in A_{j_1}} \frac{(\mu_i - \theta_{j_1})^2}{\sigma_{j_1}} + \sum_{i \in A_{j_2}} \frac{(\mu_i - \theta_{j_2})^2}{\sigma_{j_2}} - \sum_{i \in A_{j_1} \cup A_{j_2}} \frac{(\mu_i - \theta_{j^*})^2}{\sigma_{j^*}} \right) \right\}$$

$$\times \frac{p(k+1)}{p(k)} \frac{\Gamma(\sum' d_j + d_{j_1} + d_{j_2})\Gamma(d_{j^*})}{\Gamma(\sum' d_j + d_{j^*})\Gamma(d_{j_1})\Gamma(d_{j_2})} \frac{w_{j_1}^{d_{j_1}+n_1^*-1} w_{j_2}^{d_{j_2}+n_2^*-1}}{w_{j^*}^{d_{j^*}+n_1^*+n_2^*-1}}$$

$$\times \sqrt{\frac{\tau_{j_1}\tau_{j_2}}{2\pi\tau_{j^*}}} \exp\left\{ -\frac{\tau_{j_1}(\theta_{j_1} - m_{j_1})^2 + \tau_{j_2}(\theta_{j_2} - m_{j_2})^2 - \tau_{j^*}(\theta_{j^*} - m_{j^*})^2}{2} \right\}$$

$$\times \frac{b_{j_1}^{a_{j_1}} b_{j_2}^{a_{j_2}}}{b_{j^*}^{a_{j^*}}} \frac{\Gamma(a_{j^*})}{\Gamma(a_{j_1})\Gamma(a_{j_2})} \frac{\sigma_{j^*}^{a_{j^*}+1}}{\sigma_{j_1}^{a_{j_1}+1} \sigma_{j_2}^{a_{j_2}+1}} \exp\left\{ -\frac{b_{j_1}}{\sigma_{j_1}} - \frac{b_{j_2}}{\sigma_{j_2}} + \frac{b_{j^*}}{\sigma_{j^*}} \right\}$$

$$\times \frac{c_{k+1}}{s_k P_{\text{alloc}}} \left(\frac{1}{2} g_{2,2}(u_1) g_{2,2}\left(\frac{u_2+1}{2}\right) g_{1,1}(u_3) \right)^{-1} w_{j^*} (1-u_2^2) \frac{\sigma_{j^*}^{3/2}}{u_1^{3/2}(1-u_1)^{3/2}}$$

where $\sum'$ extends over the indexes j referring to those components not affected by the split/merge move, whereas $g_{i,k}$ stands for the Beta density function with parameters i and k.

Conversely, suppose a merge move is selected. If we start from a situation in which $k+1$ components are present, two of them are selected as follows: j_1 is randomly chosen from the first k existing ones and j_2 is set equal to the $(k+1)$th. Hence, j_1 and j_2 merge to form a new single component, j^* say, with

$$w_{j^*} = w_{j_1} + w_{j_2}, \qquad \theta_{j^*} = \frac{w_{j_1}}{w_{j^*}} \theta_{j_1} + \frac{w_{j_2}}{w_{j^*}} \theta_{j_2}$$

and

$$\sigma_{j^*} = \frac{w_{j_1}}{w_{j^*}} (\theta_{j_1}^2 + \sigma_{j_1}) + \frac{w_{j_2}}{w_{j^*}} (\theta_{j_2}^2 + \sigma_{j_2}) - \theta_{j^*}^2.$$

The candidate state, j^*, is placed, within the list of components, in the place occupied by j_1. The move is accepted with probability equal to min(1, R^{-1}) where R is defined as above.

An implicit assumption underlying the algorithm we have just described is that the coefficients associated with any two distinct components j are different. If there are two distinct components with common coefficients, then the merge move differs allowing for a random choice of both merging components. Viceversa, the split move is not altered.

References

Basawa, I.V. and Prakasa Rao, L.S. (1980). *Statistical Inference for Stochastic Processes.* London: Academic Press.

Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory.* Chichester: Wiley.

Celeux, G., Hurn, M. and Robert, C.P. (2000). Computational and inferential difficulties with mixture posterior distributions. *J. Am. Statist. Assoc.*, **95**, 957-970.

Coles, S.G. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Berlin: Springer.

Coles, S.G. and Pericchi, L. (2001). Anticipating catastrophes through extreme value modelling. Manuscript, Dept. of Mathematics, Bristol University, U.K.

Coles, S.G. and Powell, E.A. (1996). Bayesian methods in extreme value modelling: a review and new developments. *Internat. Statist. Rev.*, **64**, 119-136.

Coles, S.G. and Tawn, J.A. (1996). A Bayesian analysis of extreme rainfall data. *Applied Statistics*, **45**, 463-478.

Consonni, G. and Veronese, P. (1995). A Bayesian method for combining results from several binomial experiments. *J. Am. Statist. Assoc.*, **90**, 935-944.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *J. Am. Statist. Assoc.*, **96**, 194-209.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711-732.

Green, P.J. and Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scand. J. Statist.*, **28**, 355-376.

Hartigan, J.A. (1990). Partition models. *Commun. Statist. Series A: Theory and Methods*, **19**, 2745-2756.

Malec, D. and Sedransk, J. (1992). Bayesian methodology for combining the results from different experiments when the specifications for pooling are uncertain. *Biometrika*, **79**, 593-601.

Nobile, A. and Green, P.J. (2000). Bayesian analysis of factorial experiments by mixture modelling. *Biometrika*, **87**, 15-35.

Richardson, S. and Green, P.J. (1997). On the Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. Roy. Statistic Soc. B*, **59**, 731-792.

Smith, B. (2001). Bayesian Output Analysis Program: version 1.0.0. Dept. of Biostatistics. University of Iowa, USA. `http://www.public-health.uiowa.edu/boa`

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

Smith, R.L. (1999). Bayesian and frequentist approaches to parametric predictive inference (with discussion). In *Bayesian Statistics 4*, J.M. Bernardo, J.O. Berger, A. P. Dawid and A.F.M. Smith (eds.). Oxford University Press, London, 589-612.

Smith, R.L. and Goodman, D.J. (2000). Bayesian risk analysis. In *Extremes and Integrated Risk Management*, P. Embrechts (ed.). Risk Books, London, 235-251.

Stephens, M. (2000). Dealing with label switching in mixture models. *J. Roy. Statist. Soc. B*, **62**, 795-809.

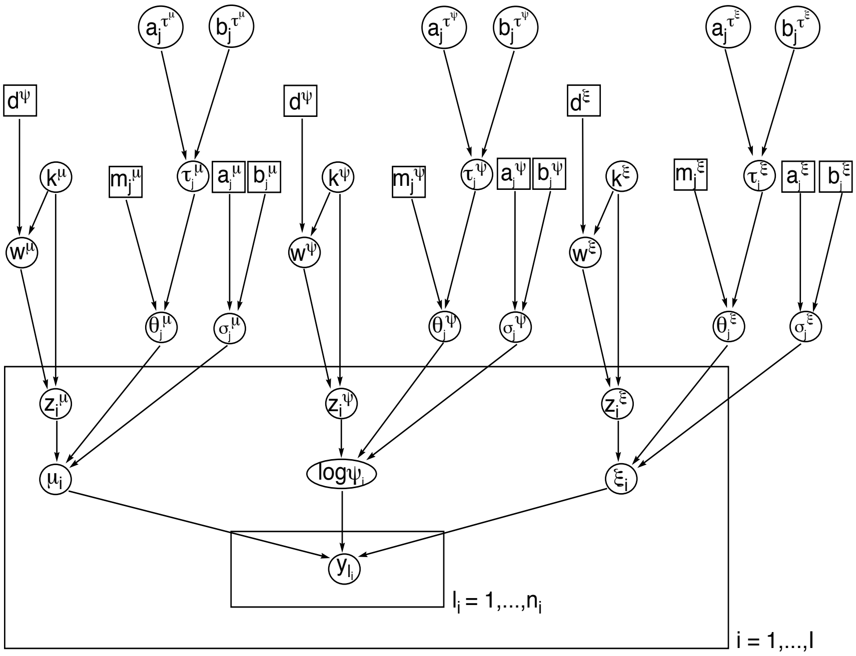

Figure 1: DAG corresponding to the joint distribution of all random quantities involved.

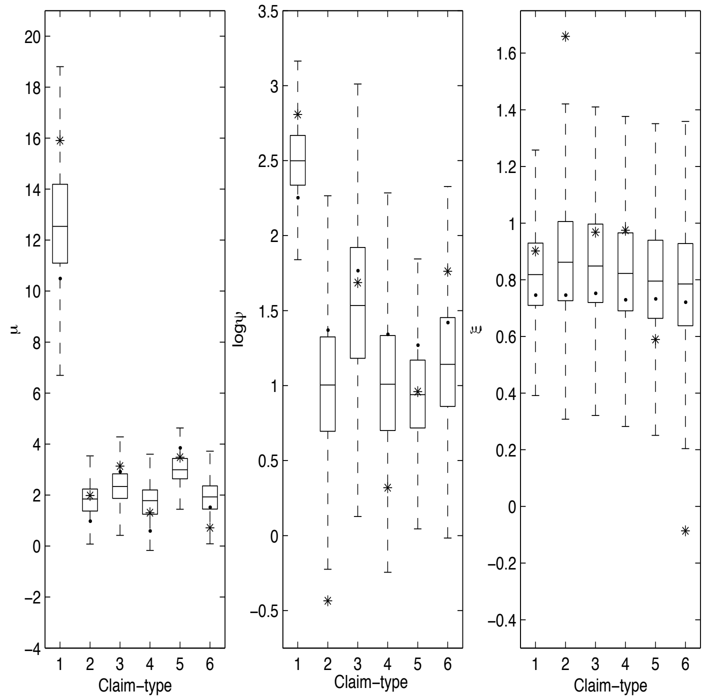

Figure 2: Boxplots of the posterior distribution for the parameters μ, $\log \psi$ and ξ. MLEs marked as stars and S&G's posterior medians marked as dots. Control parameter Δ corresponds to the 2nd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$.

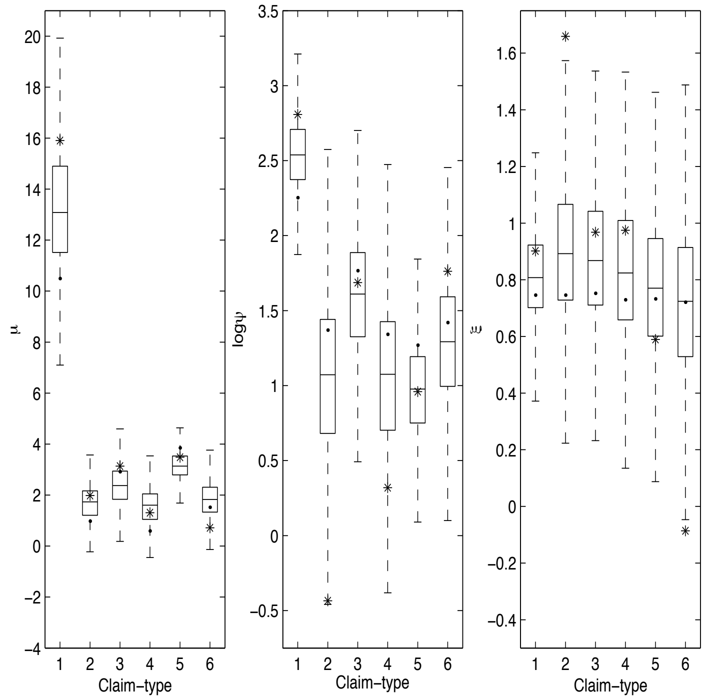

Figure 3: Boxplots of the posterior distribution for the parameters μ, $\log\psi$ and ξ. MLEs marked as stars and S&G's posterior medians marked as dots. Control parameter Δ corresponds to the 2nd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$.

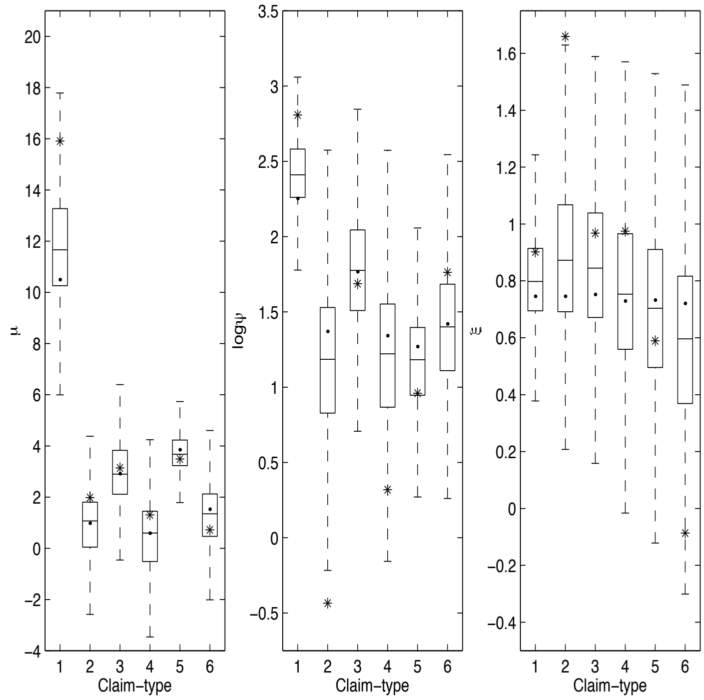

Figure 4: Boxplots of the posterior distribution for the parameters μ, $\log\psi$ and ξ. MLEs marked as stars and S&G's posterior medians marked as dots. Control parameter Δ corresponds to the 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$.

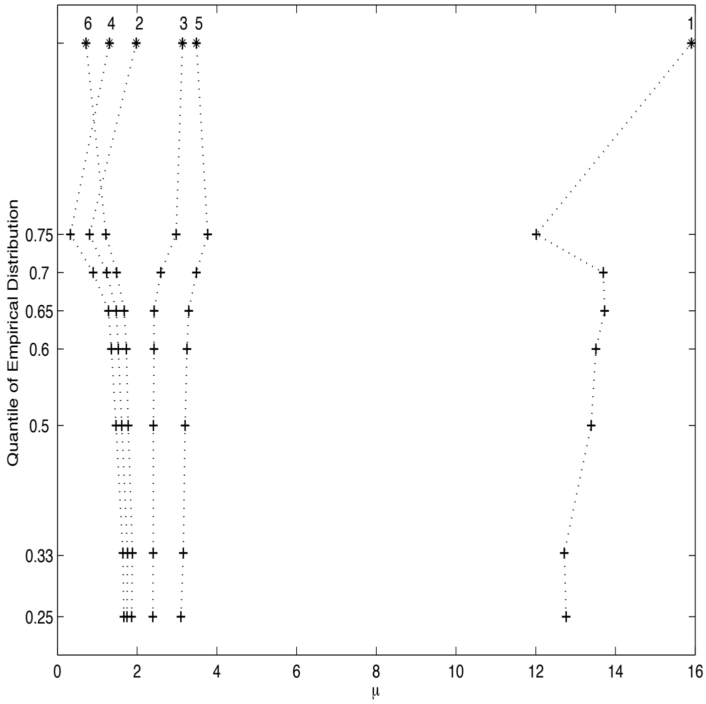

Figure 5: Crossover diagram for the parameter of location μ. MLEs marked as stars with claim-type numbers above. Posterior means marked as plus for various choices of the quantiles of empirical distribution of $\hat{\Delta}_{ii'}$.

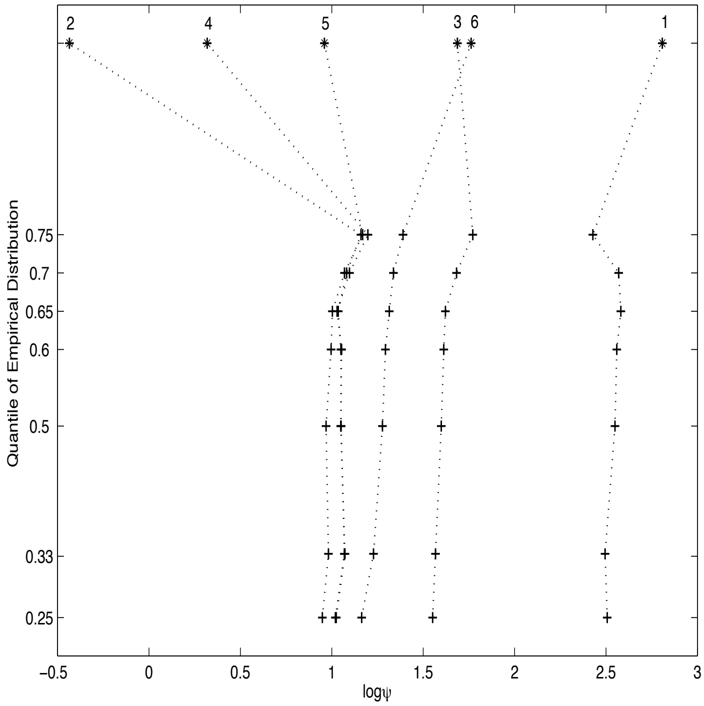

Figure 6: Crossover diagram for the parameter of scale $\log \psi$. MLEs marked as stars with claim-type numbers above. Posterior means marked as plus for various choices of the quantiles of empirical distribution of $\hat{\Delta}_{ii'}$.

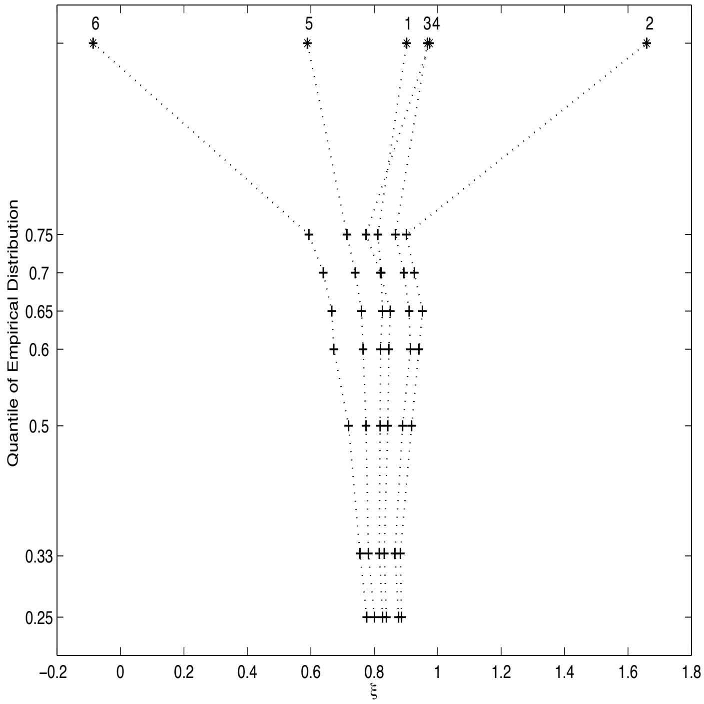

Figure 7: Crossover diagram for the parameter of shape ξ. MLEs marked as stars with claim-type numbers above. Posterior means marked as plus for various choices of the quantiles of empirical distribution of $\hat{\Delta}_{ii'}$.

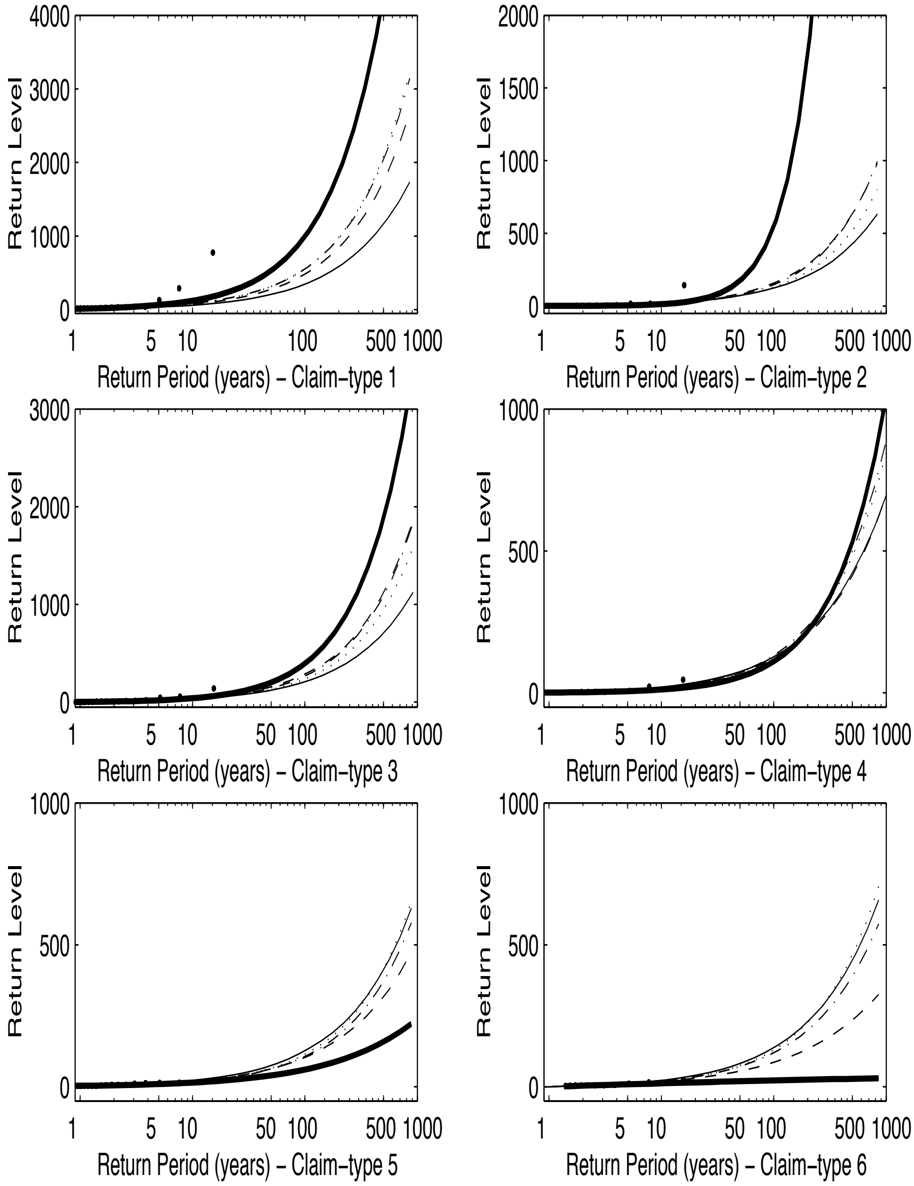

Figure 8: Predictive return level plots for the six claim-types. Solid bold line is MLE, solid line is S&G's full exchangeability model, dotted line is mixture model with Δ corresponding to the 1st quartile, dot-dashed line is mixture model with Δ corresponding to the 2nd quartile and dashed line is mixture model with Δ corresponding to the 3rd quartile. Dots correspond to empirical estimates based on the insurance claim data.

k	Cluster structure	Partitions	Prior prob.
1	(6)	111111	0.1034
2	$(5,1)$	111112, 111121, 111211, 112111, 121111, 211111	0.0223
	$(4,2)$	111122, 111212, 112112, 121112, 211112, 111221, 112121, 121121, 211121, 112211, 121211, 211211, 122111, 212111, 221111	0.0089
	$(3,3)$	111222, 112122, 121122, 122211, 112212, 121212, 122121, 122112, 121221, 112221	0.0067
3	$(4,1,1)$	111123, 111213, 112113, 121113, 211113, 111231, 112131, 121131, 211131, 112311, 121311, 211311, 123111, 213111, 231111	0.0058
	$(3,2,1)$	111223, 112123, 121123, 211123, 112213, 121213, 211213, 122113, 212113, 221113, 112231, 121231, 211231, 122131, 212131, 221131, 111232, 112132, 121132, 211132, 122311, 212311, 221311, 111322, 112312, 121312, 211312, 112321, 121321, 211321, 122311, 212311, 221311, 223111, 113122, 113212, 123112, 213112, 113221, 123121, 213121, 123211, 213211, 131122, 131212, 132112, 231112, 131221, 132121, 231121, 132211, 231211, 311122, 311212, 312112, 321112, 311221, 312121, 321121, 312211, 321211, 322111	0.0014
	$(2,2,2)$	112233, 121233, 122133, 122313, 112323, 122331	0.0012

Table 1: Prior probability induced by $w|k \sim Dir(1,\dots,1)$ on the space of partitions of $I = \{1,\dots,6\}$, classified by number of clusters ($k = 1,2,3$) and cluster structure.

Parameter	Δ	b	b^{τ}
μ	1.2361	0.3828	45.8944
$\log\psi$	0.7469	0.1398	16.7570
ξ	0.3624	0.0329	3.9445
μ	2.1794	1.1899	142.6683
$\log\psi$	1.3678	0.4687	56.2006
ξ	0.6911	0.1197	14.3469
μ	12.5064	39.1851	$4,698.2347$
$\log\psi$	1.9164	0.9200	110.3114
ξ	1.0048	0.2530	30.33293

Table 2: Values of b and b^{τ} for selected choices of Δ corresponding to the 1st, 2nd and 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$.

Parameter	$\mathrm{E}\left(\tau^{-1}\right)$	$\mathrm{E}\left(\sigma^{2}\right)$	$\mathrm{E}\left(\tau^{-1}\right)/\mathrm{E}\left(\sigma^{2}\right)$	$\mathrm{E}\left(\left(\tau\sigma^{2}\right)^{-1}\right)$
μ	22.9470	0.1914	119.89	179.84
$\log\psi$	8.3785	0.0699	119.86	179.80
ξ	1.9723	0.0164	120.26	179.84
μ	71.3340	0.5949	119.91	179.85
$\log\psi$	28.1000	0.2343	119.93	179.86
ξ	7.1735	0.0598	119.96	179.79
μ	$2,349.1000$	19.5935	119.89	179.85
$\log\psi$	55.1560	0.4600	119.90	179.86
ξ	15.1660	0.1265	119.89	179.84

Table 3: Between-component variance $\left(\tau^{-1}\right)$ and within-component variance $\left(\sigma^{2}\right)$, ratio of expected values and expected values of ratios for selected choices of Δ corresponding to the 1st, 2nd and 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$.

Δ^μ					
1.2361	211111	211131	312121	321211	211311
	[2,733]	[689]	[489]	[99]	[94]
	(0.5466)	(0.1378)	(0.0978)	(0.0198)	(0.0188)
2.1794	211111	211131	312121	211311	321211
	[3,258]	[365]	[261]	[129]	[99]
	(0.6516)	(0.0730)	(0.0522)	(0.0258)	(0.0198)
12.5064	1111111	211111	211121	212111	111211
	[3,765]	[631]	[51]	[38]	[33]
	(0.7530)	(0.1262)	(0.0102)	(0.0076)	(0.0066)

Table 4: The five most frequent partitions π^μ in the posterior sample ordered by number of visits for selected choices of Δ corresponding to the 1st, 2nd and 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$. Number of visits in squared brackets and posterior probability of visit in round brackets.

Δ^ψ					
0.7469	211111	111111	212111	213111	212112
	[1,517]	[789]	[619]	[161]	[111]
	(0.3034)	(0.1578)	(0.1238)	(0.0322)	(0.0222)
1.3678	111111	211111	212111	212112	211112
	[3,022]	[673]	[232]	[83]	[54]
	(0.6044)	(0.1346)	(0.0464)	(0.0166)	(0.0108)
1.9164	111111	211111	212111	212112	121111
	[4,272]	[132]	[72]	[36]	[33]
	(0.8544)	(0.0264)	(0.0144)	(0.0072)	(0.0066)

Table 5: The five most frequent partitions π^ψ in the posterior sample ordered by number of visits for selected choices of Δ corresponding to the 1st, 2nd and 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$. Number of visits in squared brackets and posterior probability of visits in round brackets.

Δ^{ξ}					
0.3624	111111	111112	121111	112111	111211
	[3,798]	[100]	[79]	[72]	[58]
	(0.7596)	(0.0200)	(0.0158)	(0.0144)	(0.0116)
0.6911	111111	111112	121111	111121	112111
	[4,209]	[74]	[72]	[52]	[51]
	(0.8418)	(0.0148)	(0.0144)	(0.0104)	(0.0102)
1.0048	111111	111112	111211	121111	111121
	[4,458]	[45]	[43]	[37]	[30]
	(0.8916)	(0.0090)	(0.0086)	(0.0074)	(0.0060)

Table 6: The five most frequent partitions π^{ξ} in the posterior sample ordered by number of visits for selected choices of Δ corresponding to the 1st, 2nd and 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$. Number of visits in squared brackets and posterior probability of visits in round brackets.

Parameter	# of visited partitions	# of visits and post. prob.			Post. prob. for the most visited partitions	
μ	50	1 visit	5	0.0010	5 most visited	0.8208
		<10 visits	23	0.0164	10 most visited	0.8908
		≥ 10 visits	27	0.9836		
$\log\psi$	155	1 visit	33	0.0066	5 most visited	0.6394
		<10 visits	102	0.0664	10 most visited	0.7060
		≥ 10 visits	53	0.9336		
ξ	137	1 visit	41	0.0082	5 most visited	0.8214
		<10 visits	106	0.0550	10 most visited	0.8672
		≥ 10 visits	31	0.9450		
μ	56	1 visit	9	0.0018	5 most visited	0.8224
		<10 visits	28	0.0164	10 most visited	0.8978
		≥ 10 visits	28	0.9836		
$\log\psi$	138	1 visit	53	0.0106	5 most visited	0.8128
		<10 visits	102	0.0494	10 most visited	0.8538
		≥ 10 visits	36	0.9506		
ξ	102	1 visit	32	0.0064	5 most visited	0.8916
		<10 visits	76	0.0348	10 most visited	0.9202
		≥ 10 visits	26	0.9652		
μ	81	1 visit	30	0.0060	5 most visited	0.9036
		<10 visits	56	0.0304	10 most visited	0.9282
		≥ 10 visits	25	0.9696		
$\log\psi$	89	1 visit	30	0.0060	5 most visited	0.9090
		<10 visits	66	0.0294	10 most visited	0.9350
		≥ 10 visits	23	0.9706		
ξ	82	1 visit	33	0.0066	5 most visited	0.9226
		<10 visits	63	0.0330	10 most visited	0.9434
		≥ 10 visits	19	0.9670		

Table 7: Features of the posterior distribution of partition for selected choices of Δ corresponding to the 1st, 2nd and 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$.

Parameter	$p(k\,\|y)$			Proportion of
	$k=1$	$k=2$	$k=3$	*split/combine*
	$k=4$	$k=5$	$k=6$	moves accepted
μ	0	0.3580	0.2738	0.0712
	0.1844	0.1134	0.0704	
$\log\psi$	0.1408	0.3780	0.2070	0.1434
	0.1252	0.0920	0.0570	
ξ	0.6810	0.1736	0.0746	0.1077
	0.0396	0.0184	0.0128	
μ	0.0004	0.4334	0.2528	0.1005
	0.1508	0.0956	0.0670	
$\log\psi$	0.5416	0.2390	0.1052	0.1573
	0.0562	0.0330	0.0250	
ξ	0.7586	0.1438	0.0524	0.1002
	0.0244	0.0134	0.0134	
μ	0.6740	0.1912	0.0730	0.1299
	0.0334	0.0176	0.0108	
$\log\psi$	0.7698	0.1404	0.0458	0.1143
	0.0236	0.0132	0.0072	
ξ	0.8032	0.1258	0.0398	0.0910
	0.0184	0.0096	0.0032	

Table 8: Posterior distribution of the number of clusters k for selected choices of Δ corresponding to the 1st, 2nd and 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$ and proportion of *split/combine* moves accepted.

1st quartile						
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Mean	12.7629	1.7470	2.3930	1.6680	3.0980	1.8614
	(2.3956)	(0.8271)	(0.8961)	(0.8363)	(0.7163)	(0.7882)
Median	12.5365	1.8457	2.3398	1.7791	2.9952	1.9278
	(2.2915)	(0.6415)	(0.7164)	(0.7034)	(0.5911)	(0.6741)
2.5 quantile	8.7595	-0.1793	0.7770	-0.3295	1.9742	0.0811
97.5 quantile	18.1972	3.0913	4.3144	3.0398	4.8654	3.2401
	$\log\psi_1$	$\log\psi_2$	$\log\psi_3$	$\log\psi_4$	$\log\psi_5$	$\log\psi_6$
Mean	2.5063	1.0196	1.5517	1.0240	0.9482	1.1634
	(0.2411)	(0.5077)	(0.4964)	(0.5022)	(0.3493)	(0.4533)
Median	2.4990	1.0044	1.5337	1.0091	0.9392	1.1415
	(0.2457)	(0.4669)	(0.5481)	(0.4698)	(0.3349)	(0.4392)
2.5 quantile	2.0469	0.0359	0.6514	0.0281	0.2679	0.3101
97.5 quantile	2.9969	2.0963	2.4821	2.0625	1.6608	2.0625
	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6
Mean	0.8261	0.8861	0.8766	0.8384	0.8005	0.7761
	(0.1633)	(0.2496)	(0.2494)	(0.2345)	(0.2254)	(0.2562)
Median	0.8182	0.8619	0.8486	0.8227	0.7951	0.7852
	(0.1632)	(0.2073)	(0.2052)	(0.2043)	(0.2045)	(0.2150)
2.5 quantile	0.5396	0.4989	0.4905	0.4298	0.3630	(0.2362
97.5 quantile	1.6608	2.0968	1.1816	1.4469	1.4509	1.2507
$\Pr(\xi > 1)$	0.1400	0.2566	0.2456	0.2014	0.1728	0.1600

Table 9: Posterior means, (standard deviations), medians, (interquartile ranges) and 95% credible intervals for the main parameters with control parameter Δ corresponding to the 1st quartile of the empirical distribution of $\hat{\Delta}_{ii'}$. For the parameter ξ the posterior probability $(\xi > 1)$ is also provided.

2nd quartile						
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Mean	13.3908	1.6157	2.4098	1.4721	3.2049	1.7769
	(2.6542)	(0.9151)	(0.9237)	(0.9307)	(0.6335)	(0.8612)
Median	13.0841	1.7346	2.3699	1.6025	3.1381	1.8261
	(2.5107)	(0.7090)	(0.8184)	(0.7409)	(0.5492)	(0.7256)
2.5 quantile	9.0633	-0.4719	0.7227	-0.7800	2.1618	-0.1195
97.5 quantile	19.5239	3.1270	4.3616	3.0209	4.6628	3.3240
	$\log\psi_1$	$\log\psi_2$	$\log\psi_3$	$\log\psi_4$	$\log\psi_5$	$\log\psi_6$
Mean	2.5488	1.0510	1.5984	1.0490	0.9686	1.2773
	(0.2502)	(0.5577)	(0.4191)	(0.5542)	(0.3384)	(0.4434)
Median	2.5387	1.0727	1.6101	1.0750	0.9769	1.2920
	(0.2483)	(0.5638)	(0.4158)	(0.5362)	(0.3275)	(0.4427)
2.5 quantile	2.0912	-0.0830	0.7420	-0.1098	0.2697	0.3566
97.5 quantile	3.0794	2.0622	2.3720	2.0578	1.6180	2.1051
	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6
Mean	0.8184	0.8979	0.8895	0.8428	0.7739	0.7190
	(0.1682)	(0.2857)	(0.2594)	(0.2778)	(0.2677)	(0.3034)
Median	0.8077	0.8923	0.8677	0.8242	0.7703	0.7240
	(0.1636)	(0.2506)	(0.2454)	(0.2599)	(0.2552)	(0.2855)
2.5 quantile	0.5218	0.4602	0.4527	0.3559	0.2686	0.0901
97.5 quantile	1.1737	1.1737	1.4537	1.4290	1.3157	1.2904
$\Pr(\xi > 1)$	0.1436	0.3434	0.3018	0.2602	0.1934	0.1664

Table 10: Posterior means, (standard deviations), medians, (interquartile ranges) and 95% credible intervals for the main parameters with control parameter Δ corresponding to the 2nd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$. For the parameter ξ the posterior probability $(\xi > 1)$ is also provided.

3rd quartile						
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Mean	12.0130	0.8082	2.9804	0.3262	3.7727	1.2162
	(2.5207)	(1.5168)	(1.4873)	(1.6122)	(0.7904)	(1.3855)
Median	11.6591	1.0653	2.8963	0.5957	3.6718	1.3477
	(2.2322)	(1.3013)	(1.2717)	(1.4573)	(0.7438)	(1.2272)
2.5 quantile	8.2074	-2.7998	0.0882	-3.6202	2.5047	-2.0260
97.5 quantile	18.0700	3.1544	6.1935	2.7172	5.5677	3.6067
	$\log\psi_1$	$\log\psi_2$	$\log\psi_3$	$\log\psi_4$	$\log\psi_5$	$\log\psi_6$
Mean	2.4277	1.1594	1.7710	1.1972	1.1685	1.3894
	(0.2441)	(0.5561)	(0.4042)	(0.5246)	(0.3395)	(0.4339)
Median	2.4098	1.1845	1.7758	1.2206	1.1821	1.4001
	(0.2383)	(0.5201)	(0.3966)	(0.5078)	(0.3339)	(0.4257)
2.5 quantile	1.9934	-0.0155	0.9483	0.0975	0.4788	0.5064
97.5 quantile	2.9561	2.1773	2.5495	2.1585	1.8134	2.1888
	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6
Mean	0.8114	0.9013	0.8665	0.7739	0.7139	0.5940
	(0.1651)	(0.3045)	(0.2795)	(0.3054)	(0.3079)	(0.3429)
Median	0.7981	0.8726	0.8450	0.7533	0.7039	0.5963
	(0.1627)	(0.2785)	(0.2725)	(0.3015)	(0.3075)	(0.3324)
2.5 quantile	0.5251	0.4088	0.3736	0.2455	0.1640	-0.0844
97.5 quantile	1.1736	1.6009	1.4748	1.4156	1.3366	1.2594
$\Pr(\xi > 1)$	0.1314	0.3260	0.2968	0.2184	0.1714	0.1146

Table 11: Posterior means, (standard deviations), medians, (interquartile ranges) and 95% credible intervals for the main parameters with control parameter Δ corresponding to the 3rd quartile of the empirical distribution of $\hat{\Delta}_{ii'}$. For the parameter ξ the posterior probability $(\xi > 1)$ is also provided.

Dipartimento di economia politica e metodi quantitativi
Università degli studi di Pavia

List of the lately published Technical Reports
(available at the web site: "http://economia.unipv.it/Eco-Pol/quaderni.htm").

Quaderni di Dipartimento

#	Date	Author(s)	Title
129	02-01	A.Lijoi	Approximating priors by finite mixtures of conjugate distributions for and exponential family
130	02-01	R.Lucchetti E.Rossi	Artificial Regression Testing in the GARCH-in-mean model
131	02-01	D.Sonedda	Employment Effects of Progressive Taxation in a Unionised Economy
132	04-01	P.L.Conti A.Lijoi F. Ruggeri	A Bayesian approach to the analysis of telecommunications systems performance
133	04-01	C. Bianchi	A Reappraisal of Verdoon's Law for the Italian Economy: 1951-1997
134	09-01	A.Roverato G.Consonni	Compatible Prior Distributions for DAG Models
135	11-01	D.Sonedda	On the dynamics of unemployment and labour tax progression
136	11-01	F.Chelli L.Rosti	Gender Discrimination, Entrepreneurial Talent and Self-Employment in Italy 'If you think you're so discriminated against, why don't you set up on your own?'
137	12-01	L. E. Nieto-Barajas I. Pruenster S.G. Walker	Normalized Random Measures driven by Increasing Additive Processes
138	02-02	A.Lijoi E.Regazzini	Means of a Dirichlet process and multiple hypergeometric functions
139	02-02	M.Maggi	Immunization in an affine term structure framework
140	02-02	F.Menoncin	Optimal Portfolio With Bechmark for Fund Managers
141	04-02	M.Balconi S.Breschi F.Lissoni	Il trasferimento di conoscenze tecnologiche dall'università all'industria in Italia: nuova evidenza sui brevetti di paternità dei docenti
142	04-02	M.Balconi S.Borghini A.Moisello	Ivory Tower vs Spanning University: il caso dell'Università di Pavia