Statistics and Computing **15**: 255–265, 2005

Multivariate Poisson regression with covariance structure

DIMITRIS KARLIS* and LOUKIA MELIGKOTSIDOU†

**Department of Statistics, Athens University of Economics and Business, 76, Patission Str., 10434, Athens, Greece*
karlis@aueb.gr
†*Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, United Kingdom*
l.meligotsidou@lancaster.ac.uk

Received June 2003 and accepted March 2005

In recent years the applications of multivariate Poisson models have increased, mainly because of the gradual increase in computer performance. The multivariate Poisson model used in practice is based on a common covariance term for all the pairs of variables. This is rather restrictive and does not allow for modelling the covariance structure of the data in a flexible way. In this paper we propose inference for a multivariate Poisson model with larger structure, i.e. different covariance for each pair of variables. Maximum likelihood estimation, as well as Bayesian estimation methods are proposed. Both are based on a data augmentation scheme that reflects the multivariate reduction derivation of the joint probability function. In order to enlarge the applicability of the model we allow for covariates in the specification of both the mean and the covariance parameters. Extension to models with complete structure with many multi-way covariance terms is discussed. The method is demonstrated by analyzing a real life data set.

Keywords: data augmentation, EM algorithm, Markov chain Monte Carlo, multivariate reduction, crime data

1. Introduction

Multivariate data analysis has a long history for continuous data based on the multivariate normal and related distributions. However, this is not true for discrete data. When treating multivariate count data, approximations by continuous multivariate normal models can be used, but they can be misleading especially when the observed means are not large and there are several zero counts. The multivariate Poisson distribution, while the most important among discrete multivariate distributions (see, e.g., Johnson, Kotz and Balakrishnan 1997), has several shortcomings for its application. The main drawback of the application of the multivariate Poisson distribution is the complicated form of the joint probability function.

Inferential procedures for a special case of the multivariate Poisson model, with a single common covariance term, are described in the recent papers of Tsionas (1999, 2001) and Karlis (2003). This model is rather restrictive for real applications since it assumes that all the pairs of variables have the same covariance.

In this paper we extend the above model by allowing for larger covariance structure between the variables. Namely we construct a model that allows for a different covariance for each pair of variables. The model is initially presented in its general form, with full (but perhaps unnecessarily large) structure. Then, we focus on a useful reduced model where only two-way covariance terms are used. We also include covariates in the model. Inference from the classical point of view through maximum likelihood (ML) estimation is proposed via an EM algorithm, while Bayesian inference is proposed via an MCMC algorithm. Both approaches make use of a data augmentation scheme based on multivariate reduction.

Potential application of the model can be made in a variety of disciplines where count data occur quite often, like epidemiology (e.g., incidences of different types of illness), marketing (purchases of different products), industrial control (different types of faults) etc. In all these circumstances traditional analysis using a multivariate normal approximation can be misleading due to the nature of the data (small marginal means with a lot of zero counts).

The advantages of our model are the following. Firstly, it generalizes the univariate Poisson model and therefore it is a standard reference model for multivariate count data. Secondly, it allows for realistic covariance structure among the variables. Furthermore, one can introduce covariates in the covariance terms in order to model the covariances in a flexible way (this is not true for other competing models). Finally, the proposed model is less complicated, and therefore less computationally demanding, than several competing models. Limitations of the model are the facts that it does not allow for overdispersion and negative correlation as, for example, the model of Chib and Winkelmann (2001). There are also some other models that allow for negative correlation (see van Ophem 1999, Berkhout and Plug 2004), which, however, are much more complicated and require special efforts for parameter estimation.

The remainder of the paper proceeds as follows. The multivariate Poisson distribution in its general form is described in Section 2. In Section 3 we describe our proposed models. ML estimation for these models is considered in Section 4, while Section 5 describes the Bayesian approach. Section 6 presents an application to real data, while Section 7 contains some concluding remarks.

2. The multivariate Poisson distribution

2.1. *General definition*

The derivation of the multivariate Poisson distribution is based on a general multivariate reduction scheme. Assuming Y_r, $r = 1, \ldots, k$, are independent univariate Poisson random variables, i.e. $Y_r \sim Po(\theta_r)$, $r = 1, \ldots, k$, then the general definition of the multivariate Poisson distribution is made through the vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_k)'$ and an $m \times k$ matrix $\boldsymbol{A}$, $m \leq k$, with 0 and 1 elements. Specifically, the vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)'$ defined as $\boldsymbol{X} = \boldsymbol{AY}$ follows a multivariate Poisson distribution.

An alternative expression for the multivariate Poisson random vector $\boldsymbol{X}$ arises if we consider each column of $\boldsymbol{A}$ as a vector ϕ_r, $r = 1, \ldots, k$. Then, $\boldsymbol{A} = [\phi_1 \ \phi_2 \ldots \phi_k]$ and hence $\boldsymbol{X} = \boldsymbol{AY} = \sum_{r=1}^{k} \phi_r Y_r$. In this framework, the variability of the random vector $\boldsymbol{X}$, which has the m-variate Poisson distribution, is explained through the variability of k independent univariate Poisson random variables. Note that the elements of $\boldsymbol{X}$ are dependent as indicated by the structure of the matrix $\boldsymbol{A}$.

The most general form of the multivariate Poisson distribution arises if the matrix $\boldsymbol{A}$ has the form $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots \boldsymbol{A}_m]$, where $\boldsymbol{A}_j$, $j = 1, \ldots, m$ is a sub-matrix of dimensions $m \times C_j^m$, where C_j^m is the number of combinations of picking j from m numbers, each column of $\boldsymbol{A}_j$ has exactly j ones and $(m - j)$ zeroes and no duplicate columns exist. Thus, $\boldsymbol{A}_m$ is the column vector of 1s, while $\boldsymbol{A}_1$ is the identity matrix of size $m \times m$. Then, the vector $\boldsymbol{Y}$ can also be written in the form $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2', \ldots, \boldsymbol{Y}_m')'$, where $\boldsymbol{Y}_j$ is a sub-vector of dimension C_j^m, $j = 1, \ldots, m$. Hence, the definition of the vector $\boldsymbol{X}$ becomes $\boldsymbol{X} = \sum_{r=1}^{k} \phi_r Y_r = \sum_{j=1}^{m} \boldsymbol{A}_j \boldsymbol{Y}_j$, which means that $\boldsymbol{X}$ is expressed as a sum of different vector-terms, which explain the variability of its dependent elements. These terms can be interpreted as main effects and two-way up to m-way covariance effects in an ANOVA like fashion.

Under the multivariate Poisson model the mean and the variance-covariance matrix of $\boldsymbol{X}$ are given by

$$E(\boldsymbol{X}) = \boldsymbol{A}\theta$$

and

$$\text{Var}\,(\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T$$

where $\boldsymbol{\Sigma} = \text{diag}\,(\theta_1, \theta_2, \ldots, \theta_k)$ is the variance-covariance matrix of $\boldsymbol{Y}$ (Σ is diagonal because of the independence of Y_i's). Each element of $\boldsymbol{X}$ marginally follows a univariate Poisson distribution. This general model has been theoretically described in Mahamunulu (1967) and Johnson, Kotz and Balakrishnan (1997) among others.

2.2. *The probability distribution*

A main problem, which limits the usage of multivariate distributions in general, is the complexity of calculating the probability distribution function. In the case of the multivariate Poisson distribution, the calculation of the probability mass function can be of great difficulty, as it often demands summations over high-dimensional spaces. If we consider an m-variate Poisson model, the calculation of the probability mass function requires $2^m - m - 1$ summations. For example, the probability function of the 3-variate Poisson distribution can be written with 4 nested sums (see also Mahamunulu 1967 for a symbolic presentation of the problem).

Formally, the definition of the multivariate Poisson distribution was made through a mapping $g : N^k \to N^m$, $k \geq m$, such that $\boldsymbol{X} = g(\boldsymbol{Y}) = \boldsymbol{AY}$. Hence, the joint probability of the m-vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)'$ is given by the sum of the joint probabilities of all k-vectors $\boldsymbol{y} = (y_1, y_2, \ldots, y_k)'$ such that $g(\boldsymbol{y}) = \boldsymbol{x}$. If $\boldsymbol{x} \in N^m$, let the set $g^{-1}(\boldsymbol{x}) \subset N^k$ denote the inverse image of $\boldsymbol{x}$ under g. The probability mass function of $\boldsymbol{X}$ is then defined as

$$P(\boldsymbol{X} = \boldsymbol{x}) = \sum_{\boldsymbol{y} \in g^{-1}(\boldsymbol{x})} P(\boldsymbol{Y} = \boldsymbol{y})$$

Since the elements of $\boldsymbol{Y}$ follow independently univariate Poisson distributions, we obtain that

$$P(\boldsymbol{X} = \boldsymbol{x}) = \sum_{\boldsymbol{y} \in g^{-1}(\boldsymbol{x})} \prod_{i=1}^{k} P(y_i; \theta_i). \tag{1}$$

It is clear that the calculation of these probabilities can be computationally expensive, since the summations needed might be exhausting in some cases, especially when the number of dimensions is large. However, computation of the probabilities can be accomplished via recursive schemes. Kano and Kawamura (1991) provided a general scheme for constructing recurrence relations for multivariate Poisson distributions. For computational details on the construction of a recursive scheme for the

calculation of the multivariate Poisson mass function see the Appendix.

3. Proposed models

3.1. *The multivariate Poisson model with two-way covariance structure*

The m-variate Poisson models derived by setting $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_m]$ have previously been used in the literature and the resulting distributions have been referred to as the multivariate Poisson distributions (see, e.g. Tsionas 2001, Karlis 2003). The limitation of these reduced models is that all the pairs of variables are forced to have the same covariance. To our knowledge, this class of models is the only one used in practice. Here we propose a more realistic class of multivariate Poisson models.

We consider the distributions of $\boldsymbol{X}$-vectors arising from a larger structure of the matrix $\boldsymbol{A}$ than that of the above reduced models, but still less complicated than the complete structure of Section 2.1. In our setting, only the main effects and the two-way covariance effects are considered, i.e. $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2]$. This formulation of the multivariate Poisson model provides a useful generalization of the univariate case. Not only does each element of $\boldsymbol{X}$ marginally follow a univariate Poisson distribution, but also the parameters of the joint distribution of $X_1, X_2, \ldots, X_m$ have an obvious interpretation, naturally extended from the univariate case. The model allows for a different covariance term for each pair of variables and thus it can be considered as a discrete counterpart of the multivariate normal distribution, suitable for multivariate count data.

For example, consider the case of the trivariate Poisson model with two-way covariance structure. This takes the form

$$\begin{aligned} X_1 &= Y_1 + Y_{12} + Y_{13} \\ X_2 &= Y_2 + Y_{12} + Y_{23} \\ X_3 &= Y_3 + Y_{13} + Y_{23} \end{aligned} \tag{2}$$

where $Y_i \sim Po(\theta_i)$, $i \in \{1, 2, 3\}$ and $Y_{ij} \sim Po(\theta_{ij})$, $i, j \in \{1, 2, 3\}$, $i < j$. Then, each X_i, $i \in \{1, 2, 3\}$ follows marginally a Poisson distribution with parameter $\theta_i + \theta_{ij} + \theta_{ik}$, $i, j, k \in \{1, 2, 3\}$, $i \neq j \neq k$. Recall that the parameter of the univariate Poisson distribution is its mean and its variance at the same time. Now, the random variables X_1, X_2, X_3 follow jointly a trivariate Poisson distribution with parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_{12,}\theta_{13}, \theta_{23})'$. The mean vector of this distribution is $\boldsymbol{A}\theta = (\theta_1 + \theta_{12} + \theta_{13}, \theta_2 + \theta_{12} + \theta_{23}, \theta_3 + \theta_{13} + \theta_{23})'$ and its variance-covariance matrix is given by

$$\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}' = \begin{bmatrix} \theta_1 + \theta_{12} + \theta_{13} & \theta_{12} & \theta_{13} \\ \theta_{12} & \theta_2 + \theta_{12} + \theta_{23} & \theta_{23} \\ \theta_{13} & \theta_{23} & \theta_3 + \theta_{13} + \theta_{23} \end{bmatrix}.$$

The parameters θ_{ij}, $i, j = 1, 2, 3$, $i \neq j$, have the straightforward interpretation of being the covariances between the variables X_i and X_j and, thus, we refer to them as the covariance parameters. The parameters θ_i, $i = 1, 2, 3$, appear only at the marginal means and variances and we refer to them as the mean parameters. The mean and the variance-covariance matrix for the m-variate Poisson distribution ($m > 3$) are defined in an analogous manner. For notational convenience we will refer to this model as $m - MP(\boldsymbol{\theta})$. It is clear that this model is more flexible and reasonable for real applications than the one with common covariance. For example, if the data refer to the number of purchases of different products, it is natural to assume that each pair of different products has a different covariance due to the intrinsic nature of these products instead of assuming that all the pairs have the same covariance.

3.2. *Multivariate Poisson regression*

In order to extend the applicability of the model we assume that the parameters θ_i (including both the mean and the covariance parameters) are functions of explanatory variables. Using the standard approach for univariate Poisson regression models we assume log-link functions between the explanatory variables and the parameters of interest.

We introduce some notation. Let the vector $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})'$, $i = 1, \ldots, n$, denote the i-th available m-variate observation. (Throughout the first subscript will denote the observation while the second one the variable.) For notational convenience define the set $\mathcal{S} = \mathcal{R}_1 \cup \mathcal{R}_2$, where $\mathcal{R}_1 = \{1, 2, \ldots, m\}$ and $\mathcal{R}_2 = \{ij,\ i, j = 1, \ldots, m, i < j\}$. The sets $\mathcal{R}_1$ and $\mathcal{R}_2$ contain the subscripts needed for the definition of the unobserved variables Y_j and the corresponding parameters θ_j, $j \in \mathcal{S}$. Note that the parameters θ_j, $j \in \mathcal{R}_1$ are the mean parameters while the parameters θ_j, $j \in \mathcal{R}_2$ are the covariance parameters as defined above. We will indicate all the parameters θ_j with only one subscript for notational convenience, but in order to make clear whether we refer to the mean or the covariance parameters we will refer to the sets $\mathcal{R}_1$ and $\mathcal{R}_2$ respectively. Otherwise we imply the entire vector of parameters. Note also that one may eliminate a particular covariance by removing the appropriate element of $\mathcal{S}$. This is equivalent to setting this parameter equal to 0.

Assume that $\boldsymbol{X}_i \sim m - MP(\boldsymbol{\theta}_i t_i)$, $i = 1, \ldots, n$, where t_i is an offset, such as a population or area, and $\boldsymbol{\theta}_i$ is the vector of all the parameters for the i-th observation, i.e. $\boldsymbol{\theta}_i = \{\theta_{ij};\ j \in \mathcal{S}\}$. This implies that the latent variables Y_{ij}, $j \in \mathcal{S}$, follow independently $Po(\theta_{ij} t_i)$ distributions, $i = 1, \ldots, n$. Assume further that the parameter θ_{ij} of the vector $\boldsymbol{\theta}_i$ depends on a vector of covariates $\boldsymbol{z}_{ij}$ of dimension p_j through a univariate Poisson regression structure $\log \theta_{ij} = \boldsymbol{z}_{ij}' \boldsymbol{\beta}_j$, $i = 1, \ldots, n$, $j \in \mathcal{S}$, where $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp_j})$ is a p_j-vector of regression coefficients associated with the j-th parameter. Note that the covariates associated with each parameter may be different.

Remark 1. The above general definition of the multivariate Poisson regression model allows using covariates associated with both the mean parameters $\theta_j,\ j \in \mathcal{R}_1$ and the covariance parameters $\theta_j,\ j \in \mathcal{R}_2$. Since the covariates for each parameter may be different, the model in fact treats means, variances and covariances separately. The ability to model the covariances as well as the means is an advantage of the proposed model as it allows for great modelling flexibility.

Remark 2. One may consider a reduced model with covariates associated only with the mean parameters. This model seems plausible when there is no need to consider an extremely large structure. Since the mean parameters also appear in the variances of the variables, the variances (but not the covariances) are also affected and thus the model is heteroscedastic (assuming different variance for each observation).

4. Maximum likelihood estimation via the EM algorithm

Bivariate Poisson regression models have been used in the past usually via numerical methods for maximizing the likelihood (see, e.g. Kocherlakota and Kocherlakota 2001) or a generalized least squares approach (see Ho and Singer 2001). Multivariate Poisson regressions are sparse (see, e.g., Li *et al.* 1999)

In the present paper we will describe ML estimation via an EM algorithm (for more details on the EM algorithm see McLachlan and Krishnan 1997). The multivariate reduction derivation of Section 2 allows for an easy missing data representation of the problem. The observed data are the vectors $\boldsymbol{x}_i$, while the missing (unobserved) data are the vectors $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ik})'$, $k = \text{card}(\mathcal{S})$. Note that for the full two-way covariance model $k = m(m-1)/2$, m being the dimension of the data. Our model has the form

$$\begin{aligned} \boldsymbol{X}_i &\sim m - MP(\boldsymbol{\theta}_i t_i), \quad i = 1, \ldots, n \text{ and} \\ \log(\theta_{ij}) &= \boldsymbol{z}_{ij}' \boldsymbol{\beta}_j \end{aligned} \tag{3}$$

with $j \in S$.

In fact, at the E-step one needs to calculate the conditional expectations of the latent variables $\boldsymbol{Y}_i$, $i = 1, \ldots, n$, given the observed data and the current values of the estimates. At the M-step one must maximize the complete data likelihood. Since the elements of $\boldsymbol{Y}_i$ follow independent Poisson distributions the M-step is reduced to fitting simple Poisson GLM models, using the conditional expectations obtained at the E-step as dependent variables and, for each parameter, the specified design matrix. Thus, the M-step is relatively easy and it can be accomplished via standard statistical packages offering GLM procedures. The remaining complication is the E-step where the calculation of the conditional expectations is computationally demanding.

A full description of the EM algorithm is the following:

Denote as $\boldsymbol{\Theta}^{(r)} = (\boldsymbol{\beta}_1{}^{(r)}, \boldsymbol{\beta}_2{}^{(r)}, \ldots, \boldsymbol{\beta}_k{}^{(r)})$, the vector of all the parameters after the r-th iteration.

- *E-Step:* Using the observed data and the current estimates $\boldsymbol{\Theta}^{(r-1)}$, calculate the pseudo values

$$\begin{aligned} s_{ij} &= E\left(Y_{ij} \mid \boldsymbol{X}_i, t_i, \boldsymbol{\Theta}^{(r-1)}\right) \\ &= \frac{\sum_{\boldsymbol{y}_i \in g^{-1}(\boldsymbol{x}_i)} y_{ij} \prod_{j=1}^{k} Po\left(y_{ij} \mid \theta_{ij}^{(r-1)} t_i\right)}{P\left(\boldsymbol{x}_i \mid \theta_{ij}^{(r-1)} t_i\right)}, \end{aligned}$$

  $i = 1, \ldots, n,\ \ j \in \mathcal{S}$, where $\theta_{ij}^{(r-1)} = \exp(\boldsymbol{z}_{ij}' \boldsymbol{\beta}_j{}^{(r-1)})$, $i = 1, \ldots, n,\ j \in \mathcal{S}$, $Po(\cdot \mid \theta)$ is the probability function of a Poisson distribution with parameter θ and $P(\cdot \mid \theta)$ is the joint probability function of an $m - MP(\boldsymbol{\theta})$ distribution. Details for the computation of the required probabilities via recursive relationships can be found in the Appendix.
- *M-Step:* Update the vector $\boldsymbol{\beta}_j$ by fitting a Poisson regression using the s_{ij}'s as dependent variables and the explanatory variables $\boldsymbol{z}_{ij}$, $i = 1, \ldots, n,\ j \in \mathcal{S}$.
- If some convergence criterion is satisfied stop iterating, else go back to the E-step for one more iteration.

Initial values can be obtained by fitting separate Poisson regressions assuming an independence model. This easily gives initial values for the regression coefficients of the mean parameters. Initial values for the regression coefficients of the covariance parameters can be chosen so as to reflect the observed covariances among data.

It is interesting to note that the E-step can be simplified using the recursive relationships used for the calculation of the probability mass function. In fact, the pseudo values computed in the E-step can be calculated within the recursion via the formula

$$s_{ij} = \frac{\theta_{ij} t_i P(\boldsymbol{X} = \boldsymbol{x}_i - \phi_r)}{P(\boldsymbol{X} = \boldsymbol{x}_i)}, \quad i = 1, \ldots, n, r = 1, \ldots, k.$$

The denominator is the probability of the observation given the current values for the parameters. Thus, in order to evaluate the conditional expectations we only need to calculate two probabilities. In addition, if one calculates the probability in the denominator, the probability of the numerator has already been calculated (as it is a byproduct of the recursion towards the probability of the denominator). This considerably reduces the calculations required for the computation of the conditional expectations.

If we partition the pseudo vector $\boldsymbol{s}_i$ according to the partition of $\boldsymbol{Y}_i$, i.e. $\boldsymbol{s}_i = (\boldsymbol{s}_{i1}', \boldsymbol{s}_{i2}')'$, where $\boldsymbol{s}_{i1}$ contains the pseudo values indicated by elements of $\mathcal{R}_1$ and $\boldsymbol{s}_{i2}$ contains the pseudo values indicated by elements of $\mathcal{R}_2$, then the calculation of $\boldsymbol{s}_{i2}$ suffices since $\boldsymbol{s}_{i1} = \boldsymbol{x}_i - \boldsymbol{A}_2 \boldsymbol{s}_{i2}$, $i = 1, \ldots, n$.

One may replace the closed form expectation in the E-step by simulation from the posterior distribution, following a Monte Carlo EM algorithm approach. This can reduce the effort to obtain the conditional expectations but, in practice, simulation from the conditional distribution can be time consuming.

5. Bayesian estimation

In this section we will describe Bayesian estimation via MCMC methods (Tierney 1994, Chib and Greenberg 1995, Gilks, Richardson and Spiegelhalter 1996) for the model with covariates and two-way covariance structure. Our approach is analogous to the one used for creating the EM algorithm in the previous section, based on a standard data augmentation scheme (Tanner and Wong 1987).

Using the notation of Section 2.1, we consider models defined by $\boldsymbol{X} = g(\boldsymbol{Y}) = \boldsymbol{AY}$, where $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2]$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')'$. For the i-th observation $\boldsymbol{x}_i$ the sub-vector $\boldsymbol{y}_{i2} = (y_{i,m+1}, \ldots, y_{ik})$, contains the missing data and, hence, $\boldsymbol{y}_{i1} = \boldsymbol{x}_i - \boldsymbol{A}_2\boldsymbol{y}_{i2}$, such that the vector $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}', \boldsymbol{y}_{i2}')'$ contains the complete data. The likelihood of the observed data $L(\boldsymbol{X} \mid \boldsymbol{\Theta})$, where $\boldsymbol{\Theta}$ denotes the vector of the unknown parameters of the model, is not tractable while the likelihood of the unobserved (complete) data $L(\boldsymbol{Y} \mid \boldsymbol{\Theta}) = L(\boldsymbol{X}, \boldsymbol{Y}_2 \mid \boldsymbol{\Theta})$ becomes easy to handle, as it involves Poisson likelihoods. Indeed,

$$L(\boldsymbol{Y} \mid \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{k} Po(y_{ij}; t_i \exp(\boldsymbol{z}_{ij}'\boldsymbol{\beta}_{ij})) I\{\boldsymbol{y}_i \in g^{-1}(\boldsymbol{x}_i)\} \right\},$$

where $I\{A\}$ is the indicator function of the event A.

If $\pi(\boldsymbol{\Theta})$ is a prior distribution for $\boldsymbol{\Theta}$, then the posterior of the whole set of parameters is given by $p(\boldsymbol{\Theta}, \boldsymbol{Y}_2 \mid \boldsymbol{X}) \propto L(\boldsymbol{X}, \boldsymbol{Y}_2 \mid \boldsymbol{\Theta})\pi(\boldsymbol{\Theta})$. This posterior can be explored via an MCMC scheme by successively sampling from the full conditional distributions of the unknown parameters including the missing data.

The choice of prior distributions for the regression parameters $\boldsymbol{\beta}_i$, $i = 1, \ldots, k$, can be discussed in the general framework of prior specification for GLM. A proper and possibly informative conjugate prior, most commonly the multivariate Normal distribution, can be used. In situations where strong prior information does not exist, the mean vector of the Normal distribution can be chosen to have all its elements set to zero, except for the intercept term (see Ntzoufras, Dellaportas and Forster 2003). On the other hand, there are two possible choices of an improper and non-informative prior distribution, namely the uniform prior, i.e. $\pi(\boldsymbol{\beta}) \propto 1$, and Jeffreys' invariant prior (for a thorough discussion of the use of Jeffreys' priors in Bayesian analysis of GLM see Ibrahim and Laud 1991).

Here we adopt the uniform prior, which seems to have some nice properties in our case. Firstly, it is a non-informative reference prior, particularly suitable for the regression coefficients for which prior information is not available. Secondly, as the corresponding posterior distribution is proportional to the likelihood, the posterior mode of $\boldsymbol{\beta}$ equals its MLE. Furthermore, the uniform prior doesn't depend on $\boldsymbol{\beta}$ as, in fact, is the case for Jeffreys' prior.

Under the assumption that the observed data $\boldsymbol{x}_i$, $i = 1, \ldots, n$ come from independent $m - MP(t_i \exp(\boldsymbol{z}_{i1}'\boldsymbol{\beta}_1), \ldots, t_i \exp(\boldsymbol{z}_{ik}'\boldsymbol{\beta}_k))$ distributions, the full conditional distribution of $\boldsymbol{Y}_{i2}$, the data augmentation parameters for $i = 1, \ldots, n$, has mass function

$$p(\boldsymbol{Y}_{i2} \mid \boldsymbol{X}_i, t_i, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) = \frac{\prod_{j=1}^{k} Po(y_{ij}; \theta_{ij} t_i)}{\sum_{\boldsymbol{y}_i \in g^{-1}(\boldsymbol{x}_i)} \prod_{j=1}^{k} Po(y_{ij}; \theta_{ij} t_i)},$$

for $\boldsymbol{y}_i \in g^{-1}(\boldsymbol{x}_i) \subset N^k$, where $\theta_{ij} = \exp(\boldsymbol{z}_{ij}'\boldsymbol{\beta}_j)$, for $i = 1, \ldots, n$, $j = 1, \ldots, k$, $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}', \boldsymbol{y}_{i2}')'$ and $\boldsymbol{y}_{i1} = \boldsymbol{x}_i - \boldsymbol{A}_2\boldsymbol{y}_{i2}$. For the posterior distribution of each $\boldsymbol{\beta}_j$, $j = 1, \ldots, m$ we have

$$p(\boldsymbol{\beta}_j \mid \boldsymbol{X}, \boldsymbol{Y}_2, t_i, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) \propto \exp\left\{ \sum_{i=1}^{n} \boldsymbol{y}_{ij}\boldsymbol{z}_{ij}'\boldsymbol{\beta}_j - t_i \exp(\boldsymbol{z}_{ij}'\boldsymbol{\beta}_j) \right\}.$$

The posterior distributions of the regression coefficients $\boldsymbol{\beta}_j$, $j = 1, \ldots, k$ have non-standard functional form and, therefore, these parameters can not be updated directly. Several simulation methods have been proposed to obtain samples from these densities, including adaptive rejection sampling (Dellaportas and Smith 1993, Gilks, Best and Tan 1995) and various Metropolis-Hastings algorithms (see Bennett, Racine-Poon and Wakefield 1996, for a comparative study of the methods). The Metropolis-Hastings methods are the most easily implemented and computationally efficient, especially if the asymptotic variance-covariance matrix evaluated at the MLEs is used in the proposal distribution.

We choose to use the random walk Metropolis-Hastings algorithm (RWM), in order to move from the draw $\boldsymbol{\beta}_j^{(r-1)}$ at the $(r-1)$-th iteration to $\boldsymbol{\beta}_j^{*(r)}$ at the r-th iteration, with transition density $q(\boldsymbol{\beta}_j, \boldsymbol{\beta}_j^*) \equiv N(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)$. The acceptance probability of the algorithm is given by

$$\alpha(\boldsymbol{\beta}_j, \boldsymbol{\beta}_j^*) = \min\left(1, \exp\left\{ \sum_{i=1}^{n} \boldsymbol{y}_{ij}[\boldsymbol{z}_{ij}'(\boldsymbol{\beta}_j^* - \boldsymbol{\beta}_j)] - t_i[\exp(\boldsymbol{z}_{ij}'\boldsymbol{\beta}_j^*) - \exp(\boldsymbol{z}_{ij}'\boldsymbol{\beta}_j)] \right\}\right).$$

In other words, at the r-th iteration $\boldsymbol{\beta}_j^{(r)}$ is set to be $\boldsymbol{\beta}_j^{*(r)}$ with probability $\alpha(\boldsymbol{\beta}_j, \boldsymbol{\beta}_j^*)$, otherwise the chain repeats the previous draw, i.e. $\boldsymbol{\beta}_j^{(r)} = \boldsymbol{\beta}_j^{(r-1)}$. This MCMC scheme is similar to the one used for the regrssion coefficients in the model of Chib and Winkelmann (2001).

The variance-covariance matrix of the proposal distribution is chosen to be $\boldsymbol{\Sigma}_j = c_j\boldsymbol{V}_j$, where $\boldsymbol{V}_j$ is the asymptotic variance-covariance matrix of the ML estimate of $\boldsymbol{\beta}_j$ and c_j is a scalar that determines the size of Metropolis-Hastings moves, for $j = 1, \ldots, k$. The values of c_j, $j = 1, \ldots, k$, should be adjusted after the burn-in period of the run, so that the acceptance rates of the Metropolis steps fall in the range (0.15,0.5) (see, e.g., Roberts, Gelman and Gilks 1997, Roberts and Rosenthal 2001). The ML estimates of the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$ can be used as starting values of the Gibbs sampler in order to achieve quick convergence. For a general discussion on inference about Poisson regression in the context of Bayesian GLM see Dey, Ghosh and Mallick (2000).

An interesting case arises if for some of the parameters we consider only constant covariates. For example, as mentioned earlier, one may use explanatory variables only for the mean parameters and not for the covariance parameters. If we denote by $\boldsymbol{\theta}_2$ the vector of parameters without covariates, then the entire vector of parameters is $\boldsymbol{\Theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_m, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_2 = (\theta_{m+1}, \ldots, \theta_k)$. For these parameters we may assume conjugate Gamma priors. Specifically, we assume that the prior for $\boldsymbol{\theta}_2$ is given by a product of $k - m$ independent gamma distributions, that is $\theta_i \sim Ga(a_i, b_i)$, $i = m+1, \ldots, k$, where $Ga(a, b)$ denotes the density of a gamma variate with mean a/b and variance a/b^2. The posterior distribution of $\boldsymbol{\theta}_2$ is then

$$p(\boldsymbol{\theta}_2 \mid \boldsymbol{X}, \boldsymbol{Y}_2, t) \equiv \prod_{j=m+1}^{k} Ga\left(a_j + \sum_{i=1}^{n} y_{ij}, b_j + \sum_{i=1}^{n} t_i\right).$$

This simplifies the estimation task considerably.

6. Application

The data concerns the counts of four different types of crime in a series of 49 Greek prefectures in 1997. The different crimes considered were rapes (X_1), arsons (X_2), manslaughter (X_3), smuggling of antiquities (X_4). The data were collected from the relevant publications of the National Statistical Service of Greece together with some characteristics of the prefectures reflecting the socio-economic situation of them. These were the Gross Domestic Product per capita in Euros for each prefecture (GDP), the unemployment rate of the prefecture (unem) and two dummy variables that show whether the prefecture is at the borders of the country (bord) (to account for the increased number of economic refugees crossing the borders) and whether the prefecture has at least one city with population larger than 150 thousands habitants (urb) to reflect the level of urbanization of the area.

Initially, we fitted two different models. The simple model assumes that the vector of observations (X_1, X_2, X_3, X_4) follows a 4-variate Poisson distribution with up to 2-way covariance structure without any covariates. The population of the area is used as an offset. This model has 10 parameters θ_i, $i \in \{1, 2, 3, 4, 12, 13, 14, 23, 24, 34\}$, where the parameters θ_{ij}, $i = 1, 2, 3,\ j = i+1, \ldots, 4$ reflect covariances.

The second model assumes that the parameters θ_i, $i = 1, 2, 3, 4$ are associated through a log-link function with a set of covariates (the variables described above). Note that we associate covariates only with the mean parameters (which in fact are also variance parameters for the Poisson case) in order not to impose too much structure. Models with covariates related to covariance parameters are discussed later. Both models were fitted using both the EM algorithm and the Gibbs sampler described in the relevant sections.

6.1. *ML estimation*

In order to fit the model via the ML method, the EM algorithm of Section 4 was used. Due to the large number of parameters, and in order to avoid problems with multiple maxima, we run univariate Poisson regressions and then used these estimates as initial values for the parameters. In order to start from several different points, we generated initial values from normal distributions located at the ML estimates of the simple univariate Poisson regressions and variances equal to the variances of the estimates from the univariate models. For the rest of the parameters (the covariance parameters) we used random initial values over the interval (0,5). One hundred sets of initial values were used with a simple convergence criterion, namely we stopped iterating when the relative change in the likelihood was smaller than 10^{-6}. Then from the solution with highest likelihood we kept iterating until the relative change was smaller than 10^{-12}. It is important to note that all runs coincided to the same solution (with small numerical perturbations). The same is true for a large number of simulated data sets used to validate the algorithm.

The results for the simple model and the one with covariates are given in Table 1. The maximum log-likelihood achieved for the simple model is -419.0839 while for the more complicated model is -401.7182 . A standard likelihood ratio test has a value of 34.74 for 16 degrees of freedom (p-value $= 0.004$) and thus the addition of covariates improved the model significantly.

In order to test separately for every covariate one may use standard Wald tests. The standard errors of the MLEs were calculated using the Jackknife (Efron and Tibshirani 1993). Alternatively one may use the standard approach via the second derivatives but in general this is quite cumbersome, while jackknife is relatively cheap because good initial values are available and the algorithm converges quite quickly. We kept all the covariates in the model for illustration purposes.

Examining the results, one can see that the urbanization variable has a negative sign for both rapes and arsons and a positive for manslaughters and smuggling. The covariate borders is significant for all the variables and it has a negative effect in manslaughters. Borders are also highly significant for smuggling, too, which has a clear explanation since this crime is related to border areas in general. For arsons the unemployment and the GDP variates are not significant. Note that usually arsons appear during summer and they are helped by the dry climate or farmer activities.

One may also observe that the MLEs for the last three covariance parameters for both models are zero, implying that the corresponding variables are uncorrelated. These parameters can be excluded from both models. The largest covariance was between rapes and manslaughters.

A model with a common covariance term led to a zero common covariance. This implies a model of independent Poisson regressions. Using such a model, any existing covariance between pairs of variables is entirely lost.

Table 1. *ML estimates derived via the EM algorithm for the parameters of the two models*

	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}
			Simple model without covariates			
MLE	0.9501	4.9387	1.6927	0.0000	0.0000	0.0000
SE	0.1691	0.3278	0.2921	0	0	0

	θ_1	θ_2	θ_3	θ_4
MLE	11.4741	11.5288	25.9213	11.1235
SE	0.3756	0.4595	0.5354	0.3665

	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}
			Model with covariates			
MLE	0.7813	4.0857	0.6911	0.0000	0.0000	0.0000
SE	0.1906	1.4119	0.1670	0.1906	0.1676	0.2497

	rapes		arsons		manslaughter		smuggling	
	β_1	S.E	β_2	SE	β_3	SE	β_4	SE
				Regression parameters				
constant	0.7485	0.0848	2.2908	0.0016	3.1422	0.0032	1.7952	0.0052
unemp	0.0629	0.0015	0.0089	0.0054	0.0143	0.0015	0.0123	0.0038
borders	0.0251	0.0036	0.0783	0.0051	−0.9831	0.0053	0.6628	0.0018
GDP	0.1117	0.0022	0.0046	0.0020	0.0099	0.0029	0.0335	0.0014
urb	−0.5308	0.0039	−0.1056	0.0022	0.0859	0.0010	0.1921	0.0022

Table 2. *Posterior summaries for the parameters of the two models*

	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}
			Simple model without covariates			
mean	0.9050	4.8513	1.5163	0.0693	0.0584	0.1189
variance	0.6132	3.9115	1.3926	0.0402	0.0284	0.1079

	θ_1	θ_2	θ_3	θ_4
mean	11.7581	11.3995	25.4810	11.1099
variance	6.8548	2.3742	7.4760	3.0791

	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}
			Model with covariates			
mean	0.9999	4.5859	0.6742	0.0369	0.0344	0.0697
Variance	0.6717	3.9759	1.0067	0.0217	0.0191	0.0683

	rapes		arsons		manslaughter		smuggling	
	β_1	var	β_2	var	β_3	var	β_4	var
				Regression parameters				
constant	1.1340	0.6277	2.3514	0.5530	3.3447	0.2700	1.8485	0.5373
unemp	0.0496	0.0004	0.0066	0.0005	0.0061	0.0003	0.0107	0.0004
borders	0.0815	0.1003	0.0020	0.0896	−0.9239	0.0799	0.6275	0.0710
GDP	0.0796	0.0054	−0.0033	0.0048	−0.0053	0.0024	0.0270	0.0047
urb	−0.5489	0.1791	−0.1240	0.1438	0.1251	0.0574	0.1419	0.1102

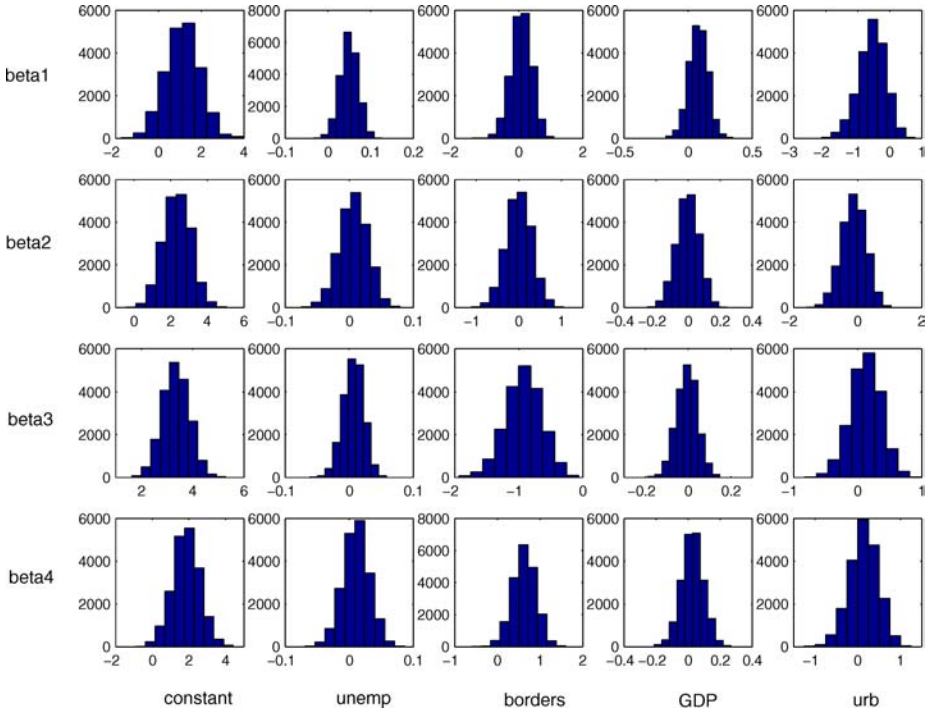

Fig. 1. *Histograms of the posteriors of the regression parameters*

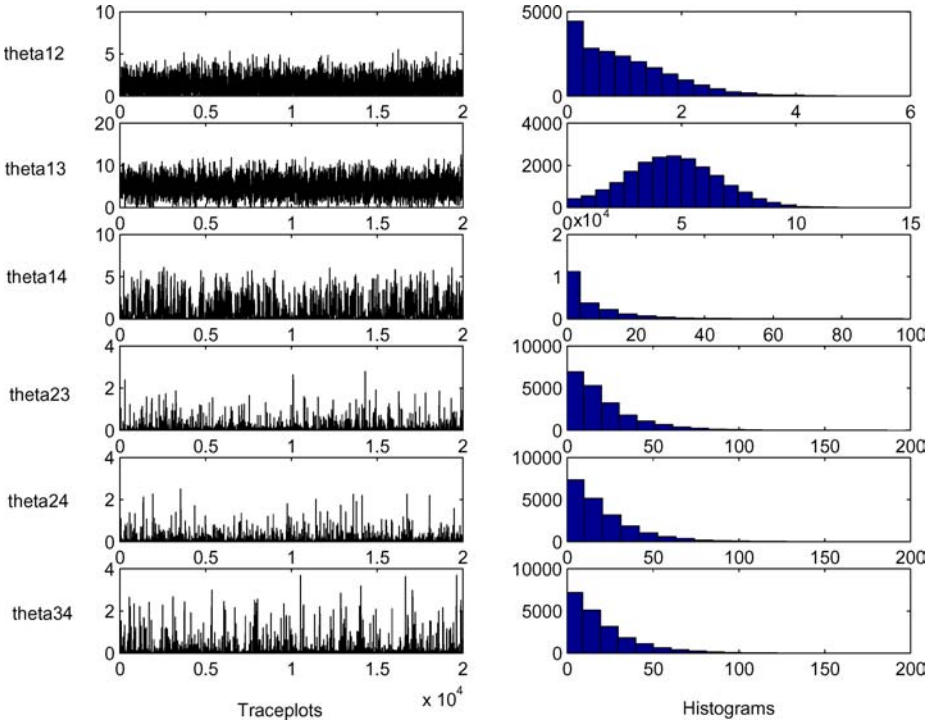

Fig. 2. *Trace plots and histograms of the posteriors of the covariance parameters*

6.2. *Bayesian estimation*

In the Bayesian framework, uninformative priors were used. For the first model without covariates independent conjugate gamma priors Ga(a, b) were used for each parameter θ_i, $i \in \{1, 2, 3, 4, 12, 13, 14, 23, 24, 34\}$.

For the second model the parameter vector is $(\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_4', \theta_{12}, \ldots, \theta_{34})$, where $\boldsymbol{\beta}_j$, $j = 1, \ldots, 4$ is the vector of regression coefficients for the GLM of the j-th variable and the θ_{ij}'s, $i = 1, 2, 3$, $j = i + 1, \ldots, 4$, are the covariances between the variables.

Table 3. *Log-likelihoods obtained by fitting a series of different models with different complexity*

Model	Log-likelihood	Parameters
MVP with 2-way structure without covariates	−419.0839	10
MVP with 2-way structure with covariates	−401.7182	26
Covariates in the covariance terms		
θ_{13} (all covariates)	−398.6605	30
θ_{13} (only unem)	−399.1394	27
θ_{12} (all covariates)	−400.7453	30
θ_{12} (only borders)	−401.2309	27

For the covariances, the priors were again independent conjugate gamma priors Ga(a, b). The values of the hyperparameters were chosen to be small, suggesting rather flat priors. Initially, all priors for the Poisson parameters were chosen to be Ga(0.1, 0.1). However, this resulted in poor mixing of the Markov chain as, for the sampled chains from the distributions of certain (covariance) parameters, long excursions in the neighborhood of zero were observed. This problem was solved taking slightly into account our ML estimates of the model parameters. We considered a little more informative priors, choosing $a = 0.5$ for those parameters for which MLEs gave strong evidence of being non-zero. The same policy was followed for the prior specification of the covariance parameters for the more complicated model with covariates. Trace plots for the covariance parameters can be seen in Fig. 2.

For the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_4$, non-informative uniform priors were used. The random walk Metropolis steps needed for these parameters were performed as described in Section 4, with the scalars that determine the size of the steps chosen after the burn-in to be, $c1 = c2 = c3 = c4 = 0.8$. The observed acceptance rates for the presented results were 0.4460, 0.4290, 0.4277 and 0.4247 respectively.

Several runs of the algorithm, from different starting values but with the same prior specification, led to identical posterior distributions of the parameters. The runs starting from the MLEs of the parameters for both models achieved extremely quick convergence (after 1000 iterations). Convergence was checked using different diagnostics, such as Raftery and Lewis, Geweke, Gelman and Rubin diagnostics, implemented by Matlab software, monitoring ergodic means and plotting the output of the runs.

A sample of size 20000 is used to summarize the posterior distribution of the parameters. The means and the variances of the chains are reported in Table 2 while histograms of the posterior distributions can be shown in Figs. 1 and 2. In Fig. 2, the last four histograms are in log-scale. The sampled chains for all the parameters had small autocorrelations and hence thinning of the chains was not needed.

For both models the three last covariance parameters have distributions with quite small means. These covariance parameters correspond to the pairs where the ML estimates were found to be 0, and indicate no significance covariance. Recall the use of informative priors for the Bayesian model in order to avoid trapping the chain to zero values. The results in general do not differ very much from those found via the ML approach. Bayesian model choice could be done via a standard approach like, e.g., Bayes factors but this is beyond the scope of the present paper.

6.3. *Models with covariates in covariance parameters*

We also run some models that allow for covariates in the covariance parameters. From the analysis in the previous section

Table 4. *Comparison of the fitted models. Model (1) is the MVP with two way covariance structure without covariates in the covariance parameters, and model (2) corresponds to the model where unemployment is used as a covariate in* θ_{13}

Model	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}
(1)	0.7813	4.0857	0.6911	0	0	0
(2)	0.7691	–	1.2402	0	0	0

	rapes β_1		arsons β_2		manslaughter β_3		smuggling β_4		θ_{13} β_{13}
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(2)
					Regression parameters				
constant	0.7485	1.5987	2.2908	2.3011	3.1422	3.6966	1.7952	1.7929	−0.9101
unemp	0.0629	0.0280	0.0089	0.0090	0.0143	−0.0100	0.0123	0.0115	0.1399
borders	0.0251	−0.1501	0.0783	0.0703	−0.9831	−1.1837	0.6628	0.6596	–
GDP	0.1117	0.0765	0.0046	0.0035	0.0099	−0.0141	0.0335	0.0299	–
urb	−0.5308	−0.6975	−0.1056	−0.1108	0.0859	0.1272	0.1921	0.1957	–

we found that some of the covariances were zero, so we used covariates only for the covariances with large values. One interesting thing is that since the marginal means are functions of both the mean parameters and the covariance parameters the use of the same covariates in both terms will cause problems in the interpretability of the effect of these covariates as they enter additively the mean in the log scale of each of the terms.

We fitted several models with covariates in the two larger covariances θ_{12} and θ_{13}; some of them are presented in Table 3. Judging by the LRT one can see that the model with only the unemployment as covariate in the covariance θ_{13} improves significantly the model. This is the best model found. In Table 4 one can see the estimated regression coefficients for this model in comparison with the coefficients derived when no covariates in θ_{13} had been used. As expected the coefficients for unemployment have changed (see, Table 5). Parameters related to other variables remain almost unchanged, however there is a change in the parameters related to variables X_1 (rapes) and X_3 (manslaughters) for which parameter θ_{13} reflects their covariance.

What the model really tells is that the covariance between rapes and manslaughters depends on the unemployment. When the unemployment increases the correlation between those two variable increases, too. Note also that the regression coefficients related to unemployment decreased implying that a large part of the effect in the mean is due to the covariance between the variables. This can be very helpful in predictions. Similar results were obtained when the Bayesian approach was employed.

7. Concluding remarks

In this paper an extended multivariate Poisson model was treated. This model allows for different covariances between the variables and thus it is more realistic than the model with only a common covariance term that has been used in the past. This model tries to mimic the multivariate normal specification that has tremendously influenced multivariate statistical analysis. To this direction our model has potential applications in many distinct disciplines where multivariate count data appears.

However, one must keep in mind that the covariance in the multivariate Poisson setting is always positive and hence the model cannot treat negative association between the variables. Moreover, the multivariate Poisson model treated in the present paper has Poisson marginal distributions and thus it cannot model overdispersion. The general approach and the data augmentation based on the multivariate reduction technique applies to more complicated structures with latent variables from overdispersed discrete distributions and, thus, our approach can be extended to cover overdispersion as well. Small amount of overdispersion can be introduced by considering inflated versions of multivariate Poisson regression models, like the models described in Karlis and Ntzoufras (2003) for the bivariate case.

In this paper we described numerical issues for carrying out inference for the proposed model via both classical ML method and the Bayesian approach. Within both methods the same data augmentation was used. An advantage of our model is that we can use covariates at the covariances of the model.

Putting aside philosophical issues for and/or against the Bayesian versus the classical approach, the two approaches share some common elements and each one has its pros and cons. For example, the Bayesian approach does not allow for exactly zero covariances that indicate the independence between two variables. On the other hand several other functions of the parameters are readily available from the MCMC output and can be used for inference. Computationally both approaches are similar. Usually the EM algorithm needs less iterations to the final solution at the cost of starting from different initial values in order to ensure that the maximum has been found.

Appendix: Calculation of the mass function via recursive relationships

First consider the case of the trivariate Poisson model with two-way covariance structure as given in (2). The probability mass function defined in (1) for the resulting trivariate Poisson distribution is analytically expressed as

$$P((\boldsymbol{X}=x)) = P(X_1=x_1, X_2=x_2, X_3=x_3)$$
$$= \exp(-(\theta_1+\theta_2+\theta_3+\theta_{12}+\theta_{13}+\theta_{23})) \times \sum_{(y_{12},y_{13},y_{23})\in C} \frac{\theta_1^{x_1-y_{12}-y_{13}}\theta_2^{x_2-y_{12}-y_{23}}\theta_3^{x_3-y_{13}-y_{23}}\theta_{12}^{y_{12}}\theta_{13}^{y_{13}}\theta_{23}^{y_{23}}}{(x_1-y_{12}-y_{13})!(x_2-y_{12}-y_{23})!(x_3-y_{13}-y_{23})!y_{12}!y_{13}!y_{23}!},$$

where the summation is over the set $C \subset N^3$ defined by

$$C = [(y_{12}, y_{13}, y_{23}) \in N^3 : \{y_{12}+y_{13} \le x_1\} \cap \{y_{12}+y_{23} \le x_2\} \cap \{y_{13}+y_{23} \le x_3\} \neq \emptyset].$$

A recursive scheme for the calculation of this mass function is the following:

- If at most one element of $\boldsymbol{x} = (x_1, x_2, x_3)'$ is non-zero, then

$$P(\boldsymbol{X}=\boldsymbol{x}) = \exp\left\{-\sum_{i<j}\theta_{ij}\right\}\prod_{i=1}^{3} Po(x_i;\theta_i),$$
$$i, j \in \{1,2,3\}.$$

- If only one of the x_i's is zero, i.e. $x_k = 0$, $k \in \{1,2,3\}$ and $x_i, x_j \neq 0$, $i, j \in \{1,2,3\}/\{k\}$, then

$$P(\boldsymbol{X}=\boldsymbol{x}) = \exp\{-\theta_{ik}-\theta_{jk}\}Po(x_k;\theta_k)BP(x_i, x_j;\theta_i, \theta_j, \theta_{ij}),$$

where BP denotes the bivariate Poisson distribution defined analogously (for more details, see Kocherlakota and Kocherlakota 1992).

- If all elements of $\boldsymbol{x} = (x_1, x_2, x_3)$ are non-zero, the recurrence relation is

$$\begin{aligned} x_1 P(\boldsymbol{X} = \boldsymbol{x}) &= \theta_1 P(X_1 = x_1 - 1, X_2 = x_2, X_3 = x_3) \\ &+ \theta_{12} P(X_1 = x_1 - 1, X_2 = x_2 - 1, X_3 = x_3) \\ &+ \theta_{13} P(X_1 = x_1 - 1, X_2 = x_2, X_3 = x_3 - 1), \end{aligned}$$

  Similar expressions can be derived by interchanging (x_1, x_2, x_3), resulting in three recursions.

The above recursive scheme can be generalized to more than three dimensions. Recall the alternative definition of the multivariate Poisson random vector $\boldsymbol{X}$ through the vectors ϕ_r, $r = 1, \ldots, k$, i.e. $\boldsymbol{X} = \sum_{r=1}^{k} \phi_r Y_r$ (Section 2.1.1). Let us define the vector $\boldsymbol{\theta}^* = (\theta_1 P(\boldsymbol{X} - \phi_1), \ldots, \theta_k P(\boldsymbol{X} - \phi_k))'$. Then, the m recurrence relations for the calculation of the probabilities of the m-variate Poisson distribution are given by the rows of the matrix equation

$$\boldsymbol{X} P(\boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{A}\boldsymbol{\theta}^*.$$

Note that, in the case of $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2]$, each row of $\boldsymbol{A}$ contains exactly m ones, so each recurrence relationship for the calculation of $P(\boldsymbol{X} = \boldsymbol{x})$ requires the computation of m previous probabilities. Obviously, as m increases the complexity of the model, and hence the computational effort, increases too. Note also that errors due to recursion are accumulated and, thus, for large m the scheme can be unstable. The solution is either to compute the probabilities analytically via (1) or to use some clever algorithm that reduces the number of recursions.

Acknowledgment

The authors thank Paul Fearnhead, Carmen Fernández and Ole Christensen for helpful comments on an earlier version of the present paper. They also thank two anonymous referees for constructive comments that improved the paper.

References

Bennett J.E., Racine-Poon A., and Wakefield J.C. 1996. MCMC for nonlinear hierarchical models. In: Gilks W.R., Richardson S., and Spiegelhalter D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman and Hall, New York, pp. 339–357.

Berkhout P. and Plug E. 2004. A bivariate Poisson count data model using conditional probabilities. Statistica Neerlandica 58: 349–364.

Chib S. and Greenberg E. 1995. Understanding the Metropolis-Hastings algorithm. The American Statistician 49: 327–335.

Chib S. and Winkelmann R. 2001. Markov chain Monte Carlo analysis of correlated count data. Journal of Business and Economic Statistics 19: 428–435.

Dellaportas P. and Smith A.F.M. 1993. Bayesian inference for generalized linear and proportional hazard models via Gibbs sampling. Applied Statistics 41: 443–459.

Dey D.K., Ghosh S.K., and Mallick B.K. 2000. Generalized Linear Models: A Bayesian Perspective. Marcel-Dekker, Inc., New York.

Efron B. and Tibshirani R.J. 1993. An Introduction to the Bootstrap, Chapman and Hall, New York.

Gilks W.R., Best N.G., and Tan K.K.C. 1995. Adaptive rejection metropolis sampling within gibbs sampling. Applied Statistics 44: 155–73.

Gilks W.R., Richardson S., and Spiegelhalter D.J. 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, New York.

Ho L.L. and Singer J.M. 2001. Generalized least squares methods for bivariate Poisson regression. Communications in Statistics, Theory and Methods 30: 263–277.

Ibrahim J.G. and Laud P.W. 1991. On Bayesian analysis of generalized linear models using Jeffreys's prior. Journal of the American Statistical Association 86: 981–986.

Johnson N., Kotz S., and Balakrishnan N. 1997. Discrete Multivariate Distributions, New York, Wiley.

Kano K. and Kawamura K. 1991. On recurrence relations for the probability function of multivariate generalized Poisson distribution. Communications in Statistics-Theory and Methods 20: 165–178.

Kocherlakota, S. and Kocherlakota, K. 1992. Bivariate Discrete Distributions, Marcel Dekker, New York.

Kocherlakota S. and Kocherlakota K. 2001. Regression in the bivariate Poisson distribution. Communications in Statistics-Theory and Methods 30: 815–827.

Karlis D. 2003. An EM algorithm for multivariate Poisson distribution and related models. Journal of Applied Statistics 30: 63–77.

Karlis D. and Ntzoufras J. 2003. Analysis of sports data using bivariate Poisson models. The Statistician 52: 381–393 .

Lee L.F. 2001. On the range of correlation coefficients of bivariate ordered discrete random variables. Econometrics Journal 17: 247–256

Li C.S, Lu J.C., Park J., Kim K., and Peterson J. 1999. Multivariate zero-inflated Poisson models and their applications. Technometrics 41: 29–38.

Mahamunulu D.M. 1967. A note on regression in the multivariate Poisson distribution. Journal of the American Statistical Association 62: 251–258

McLachlan G. and Krishnan T. 1997. The EM Algorithm and Extensions, Wiley, New York.

Ntzoufras I., Dellaportas P., and Forster J.J. 2003. Bayesian variable and link determination for generalised linear models. Journal of Statistical Planning and Inference 111: 165–180.

Roberts G.O., Gelman A., and Gilks W. 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. Annals of Applied Probability 7: 110-120.

Roberts G.O. and Rosenthal S. 2001. Optimal scaling for various Metropolis-Hastings algorithms. Statistical Science 16: 351–367.

Tanner M.A. and Wong W.H. 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82: 528–540.

Tierney L. 1994. Markov chains for exploring posterior distributions (with discussion). The Annals of Statistics 22: 1701–1762.

Tsionas E.G. 1999. Bayesian analysis of the multivariate Poisson distribution. Communications in Statistics-Theory and Methods 28: 431–451.

Tsionas E.G. 2001. Bayesian multivariate Poisson regression. Communications in Statistics—Theory and Methods 30: 243–255.

Van Ophem H. 1999. A general method to estimate correlated discrete random variables. Econometric Theory 15: 228–237