Statistica Neerlandica (2004) Vol. 58, nr. 3, pp. 349–364

A bivariate Poisson count data model using conditional probabilities

Peter Berkhout[1]*

SEO Amsterdam Economics, University of Amsterdam, Roetersstraat 29, 1018 WB Amsterdam, The Netherlands

Erik Plug[2]

Department of Economics, University of Amsterdam, Scholar, Tinbergen Institute and IZA, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

The applied econometrics of bivariate count data predominantly focus on a bivariate Poisson density with a correlation structure that is very restrictive. The main limitation is that this bivariate distribution excludes zero and negative correlation. This paper introduces a new model which allows for a more flexible correlation structure. To this end the joint density is decomposed by means of the multiplication rule in marginal and conditional densities. Simulation experiments and an application of the model to recreational data are presented.

Key Words and Phrases: correlated count data, conditional modelling, bivariate Poisson distributions.

1 Introduction

For the estimation of discrete phenomena such as the number of vacations, career interruptions, scores of soccer games, number of children et cetera, the discrete Poisson distribution is commonly used. For single counts, application of this distribution is rather straightforward. For multiple counts, however, the application of the Poisson distribution is not that clear. A multivariate discrete distribution which allows for correlation is not readily available. The literature reports some studies (Gourieroux, Monfort and Trognon, 1984; Jung and Winkelmann, 1993; King 1989) using a bivariate Poisson distribution that builds on the model described by Johnson and Kotz (1969). The disadvantage of this particular

*We are grateful to Mars Cramer, Jacob Goerree, Hans van Ophem, Bernard van Praag, Piet Stam, Frank Windmeijer and two anonymous referees for their helpful comments on earlier versions of this paper, but we alone are responsible for the views expressed.

[1]peebee@seo.fee.uva.nl
[2]e.j.s.plug@uva.nl

distribution is that it does not allow for zero and negative correlation, and thus lacks generality.

In this paper we introduce a new econometric model based on conditioning Poisson distributions. The proposed model allows for positive as well as negative correlation. An application to tourist behavior is presented, where the counts represent the number of different recreational day-trips undertaken by an individual within a given time spell. In this particular application the sign of the correlation is of relevance. Let's consider day-trips as goods where the counts represent the number of times these goods are consumed. In economic terms these goods may either be complements or substitutes depending on whether the corresponding counts will show positive or negative correlation. Modelling counts simultaneously therefore means that a discrete joint density with flexible correlation is more appropriate.

There are a few other studies available that recognize this limitation and propose alternative bivariate count models that allow for more general correlation structures. As far as we know there are two alternative approaches. The first approach models dependence among counts through correlated random effects (CHIB and WINKELMANN, 2001; MUNKIN and TRIVEDI, 1999). The second approach models dependence through copula (VAN OPHEM, 1999; CAMERON, LI, TRIVEDI and ZIMMER, 2003). Another approach is introduced by CAMERON and JOHANSSON (1998). Although they model dependence through squared polynomial expansions, their paper differs because of its explicit focus on misdispersion.

The outline of the paper is as follows. In Section 2 we present the available Poisson models and discuss some of their merits and limitations. In Section 3 we introduce a new econometric count data model based on conditional and marginal Poisson distributions. In Section 4 we describe a simulation experiment and an application in recreational behavior using the bivariate Poisson models. Section 5 concludes.

2 Poisson models

Univariate Poisson model

Let Y be a random count variable, i.e. a random variable taking solely non-negative integers. Statistical theory provides several discrete distributions of which the Poisson distribution is frequently applied. Let Y follow a Poisson distribution

$$P(Y = y) = f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \quad \lambda > 0, \tag{1}$$

where y is the realized value of the random variable Y, taking the values 0, 1, 2, … and where λ is the parameter of the distribution. Both the mean and variance of the Poisson distribution are equal to λ. This parameter may in turn vary with regressor covariates x expressing observed heterogeneity among individuals. In common practice λ is defined as

$$\lambda = e^{x'\beta}, \tag{2}$$

where parameter β is the vector that measures the effect of x on λ. The exponent ensures that λ is positive. The loglikelihood of the sample reads as

$$\log L(y, x, \beta) = \sum_{i}^{n} y_i x_i' \beta - e^{x_i' \beta} - \ln y_i! \tag{3}$$

Subscript i indexes the individual. This model has its merits. For one, the Poisson maximum likelihood estimator is robust to distributional misspecification, and maintains certain efficiency properties even when the distribution is not Poisson (WOOLDRIDGE, 2002). WINKELMANN and ZIMMERMANN (1995) argue that in many ways its relevancy for count data can be compared to the relevance of ordinary least squares for continuous data.

There are also limitations to Poisson distributions. The biggest limitation is that it has only one parameter defining all moments. Therefore, the ratio between the variance and mean is equal to one. This is mostly referred to as equidispersion, a feature which is seldom found in count data under investigation. The available data may reveal underdispersion (the mean exceeds the variance) or overdispersion (the variance exceeds the mean). Examples of misdispersion can be found in MULLAHY (1986) or HALL, GRILICHES and HAUSMAN (1986).

In the case of Poisson distributions, usually two sources are held responsible for misdispersion. The first source is simply misspecification and relates to the underlying Poisson assumption that successive events are independent. If not, the count data reveal under- or overdispersion indicating that the Poisson distribution is not the statistical tool to describe the available count data properly (CAMERON and TRIVEDI, 1986). The second source is unobserved heterogeneity, in which case the variance exceeds the mean, showing overdispersion. This can be overcome by introducing an error term ϵ in the parameter

$$\lambda = e^{x'\beta + \epsilon} = e^{x'\beta} u \quad \text{for } i = 1, \ldots, n. \tag{4}$$

For a specific choice of a simultaneous distribution for λ and ϵ the model can be written in closed form, resulting in the well-known Negative Binomial model (CAMERON and TRIVEDI, 1986).

Bivariate Poisson model

Multivariate modelling seems more appropriate if different discrete events, which are naturally related, are observed simultaneously. The problem is that a bivariate Poisson distribution which allows for dependence is not readily available. Several bivariate Poisson distributions have been proposed in statistical theory, see KOCHERLAKOTA and KOCHERLAKOTA (1992), but applied statistics have mainly focused on the trivariate reduction method described by JOHNSON and KOTZ (1969).

They allow three random variables V_1, V_2 and U to follow three independent Poisson distributions with the positive parameters λ_1, λ_2 and α respectively. With these three random variables, they construct two new random variables Y_1 and Y_2

$$Y_j = V_j + U \quad \text{for } j = 1, 2, \tag{5}$$

where their bivariate Poisson distribution reads as

$$P(Y_1 = y_1, Y_2 = y_2) = \mathrm{e}^{-\lambda_1 - \lambda_2 - \alpha} \sum_{l=0}^{\min(y_1, y_2)} \frac{\alpha^l}{l!} \frac{\lambda_1^{y_1 - l}}{(y_1 - l)!} \frac{\lambda_2^{y_2 - l}}{(y_2 - l)!}. \tag{6}$$

The covariance between Y_1 and Y_2 becomes

$$\mathrm{cov}(Y_1, Y_2) = \mathrm{cov}(V_1 + U, V_2 + U) = \mathrm{var}(U) = \alpha, \tag{7}$$

and thus the correlation between Y_1 and Y_2 equals

$$\mathrm{corr}(Y_1, Y_2) = \frac{\alpha}{\sqrt{(\lambda_1 + \alpha)(\lambda_2 + \alpha)}}. \tag{8}$$

It is easy to see that the correlation exceeds zero as both α and the denominator exceed zero. If U follows a Poisson distribution, parameter α is defined to be strictly positive. In fact, a zero α would leave (6) undefined. Allowing for heterogeneity, the positive parameters α, λ_1 and λ_2 may be defined as

$$\alpha = \mathrm{e}^{\gamma}, \lambda_{ij} = \mathrm{e}^{x_i' \beta_j} \quad \text{for } j = 1, 2 \text{ and } i = 1, \ldots, n. \tag{9}$$

As can be seen from equation (8) this means that the correlation of Y_1 and Y_2 may vary between individuals. Here, the loglikelihood takes the form

$$\log L(y_1, y_2, x, \lambda_1, \lambda_2, \alpha) = n\alpha - \sum_i \sum_j \mathrm{e}^{x_i' \beta_j} + \sum_i \log B_i, \tag{10}$$

where B_i equals

$$B_i = \sum_{l=0}^{\min(y_{i1}, y_{i2})} \frac{\alpha^l}{l!} \frac{\lambda_{i1}^{y_{i1} - l}}{(y_{i1} - l)!} \frac{\lambda_{i2}^{y_{i2} - l}}{(y_{i2} - l)!}. \tag{11}$$

For more explicit details we refer to JUNG and WINKELMANN (1993) and KING (1989). Likelihood properties are derived in the Appendix.

Its merit is the possible correlation of the counts. The main drawback of this particular bivariate Poisson model is that the magnitude of the correlation is restricted. It does not only exclude zero and negative correlation, it also excludes values higher than

$$\frac{\alpha}{\alpha + \min(\lambda_{i1}, \lambda_{i2})}. \tag{12}$$

Both restrictions render the model less applicable. Note that dispersion problems apply for this bivariate Poisson model as well. It turns out that the proposed solution with respect to the unobserved heterogeneity leads to severe computational problems in the bivariate case (JUNG and WINKELMANN 1993).

3 A bivariate count data model using conditional probabilities

In this section we derive an alternative simultaneous Poisson count data model, using conditional probabilities. The model can be used to estimate two correlated count data processes, allowing for negative as well as positive correlation and will be referred to as CPM (conditional Poisson model). The proposed bivariate model can be regarded as a special case of a general framework describing more than two counts simultaneously.

Conditional count model

Let Y_1 and Y_2 be two dependent random count variables. According to conditional probability theory, their joint density can be written as a product of a marginal and a conditional distribution. Hence

$$P(Y_1 = y_1, Y_2 = y_2) = f(y_1, y_2) = g_2(y_2 \mid y_1)g_1(y_1), \tag{13}$$

or

$$P(Y_1 = y_1, Y_2 = y_2) = f(y_1, y_2) = g_1^*(y_1 \mid y_2)g_2^*(y_2). \tag{14}$$

The decomposition of the joint density in marginals and conditional densities in the bivariate case can take two forms: the number of permutations. With two random variables Y_1 and Y_2, there are only two permutations possible. In general every decomposition of $f(y_1, y_2)$ will lead to different marginal and conditional distributions. If π is defined as a permutation indicator, taking the values 0 and 1, z_1 and z_2 can be written as the realized permutation of y_1 and y_2

$$z_1 = (1 - \pi)y_1 + \pi y_2, z_2 = \pi y_1 + (1 - \pi)y_2 \quad \text{where } \pi = 0, 1, \tag{15}$$

then the joint density can be written as

$$f(y_1, y_2) = f^{\pi}(z_1, z_2) = g_2^{\pi}(z_2 \mid z_1)g_1^{\pi}(z_1) \text{ for all } \pi. \tag{16}$$

Traditionally, econometric theory focuses mainly on the left hand side of equation (16) and assumptions are made about the shape of the joint density $f^{\pi}(z_1, z_2)$, thereby defining the marginal and the conditional distribution. Contrary to this approach, we shall make assumptions regarding the marginal and conditional distributions, which define the joint density. Let $g_1^{\pi}(z_1)$ and $g_2^{\pi}(z_2 \mid z_1)$ follow Poisson distributions for any permutation π, yielding different joint densities $f^{\pi}(z_1, z_2)$ for all permutations.

The next step is to choose a permutation. In general, we let the available data determine which permutation yields the 'best fitting' joint density, i.e. which joint density is most likely. However, if we have a clear idea of the causality relation between both counts, we can let the choice of permutation depend on the expected underlying causality structure.

Returning to (16), we assume that $g_j^{\pi}(\cdot)$ is Poisson distributed for all j and for all observations i in the sample. The marginal distribution of z_1 reads as

$$g_{i1}^{\pi}(z_{i1}) = \frac{e^{-\lambda_{i1}}\lambda_{i1}^{z_{i1}}}{z_{i1}!}, \tag{17}$$

where the parameter is defined as in equation (2). The conditional distribution of z_2 given z_1 is also Poisson distributed

$$g_{i2}^{\pi}(z_{i2} \mid z_{i1}) = \frac{e^{-\lambda_{i2}}\lambda_{i2}^{z_{i2}}}{z_{i2}!} \quad \text{for } i = 1, \ldots, n. \tag{18}$$

To allow for heterogeneity and interdependence between the random variables, the parameter λ_2 does not only depend on the observed characteristics x but also on the realized value z_1 of the first count. The fact that parameter λ_2 depends on the count z_1 implies that the variable z_2 is conditionally distributed $g_2^{\pi}(z_2 \mid z_1)$. The parameters read as

$$\lambda_{i1} = e^{x_i'\beta_1} \text{ and } \lambda_{i2} = e^{x_i'\beta_2 + \alpha z_{i1}}. \tag{19}$$

Now the joint distribution for any permutation z^{π} can be written as

$$f^{\pi}(z_1, z_2) = \frac{e^{z_1(x'\beta_1) - \exp(x'\beta_1) + z_2(x'\beta_2 + \alpha z_1) - \exp(x'\beta_2 + \alpha z_1)}}{z_1! z_2!}. \tag{20}$$

For exposition purposes we leave out all the individual specific indices. By definition, the marginal distribution of z_2 may be derived from this joint distribution by summing over z_1. We have not been able to find a closed form expression for this distribution, yet we've been able to define this discrete distribution by its factorial moments. The (r, s)th factorial moment of the joint distribution is derived as follows:

$$\begin{aligned}
&E(z_1(z_1-1)\ldots(z_1-r+1)z_2(z_2-1)\ldots(z_2-s+1)) \\
&= \sum_{z_1=r}^{\infty}\sum_{z_2=s}^{\infty} \frac{e^{z_1(x'\beta_1) - \exp(x'\beta_1) + z_2(x'\beta_2 + \alpha z_1) - \exp(x'\beta_2 + \alpha z_1)}}{(z_1-r)!(z_2-s)!} \\
&= \sum_{z_1=r}^{\infty} \frac{e^{z_1(x'\beta_1) - \exp(x'\beta_1)}}{(z_1-r)!} e^{-\exp(x'\beta_2+\alpha z_1)} \sum_{z_2=0}^{\infty} \frac{e^{(x'\beta_2+\alpha z_1)z_2}}{z_2!} e^{s(x'\beta_2+\alpha z_1)} \\
&= \sum_{z_1=r}^{f} \frac{e^{z_1(x'\beta_1) - \exp(x'\beta_1)}}{(z_1-r)!} e^{-\exp(x'\beta_2+\alpha z_1)} e^{s(x'\beta_2+\alpha z_1)} e^{\exp(x'\beta_2+\alpha z_1)} \\
&= \sum_{z_1=r}^{f} \frac{e^{z_1(x'\beta_1) - \exp(x'\beta_1)}}{(z_1-r)!} e^{s(x'\beta_2+\alpha z_1)} \\
&= \sum_{z_1=0}^{f} \frac{e^{-\exp(x'\beta_1)} e^{z_1(x'\beta_1+s\alpha)}}{z_1!} e^{rx'\beta_1+sx'\beta_2+rs\alpha} \\
&= e^{-\exp(x'\beta_1)+\exp(x'\beta_1+s\alpha)+rx'\beta_1+sx'\beta_2+rs\alpha} \\
&= \lambda_1^r e^{\lambda_1(\exp(s\alpha)-1)+sx'\beta_2+rs\alpha}.
\end{aligned} \tag{21}$$

Using equation (21) we can now calculate all moments of the joint distribution by choosing r and s carefully. From equation (21) we derive the expectation and variance of z_2

$$\mathrm{E}z_2 = \mathrm{e}^{\lambda_1(\exp(\alpha)-1)}\mathrm{e}^{x'\beta_2}, \tag{22}$$

and

$$\mathrm{var}(z_2) = \mathrm{E}z_2 + (\mathrm{E}z_2)^2(\mathrm{e}^{\lambda_1(\exp(\alpha)-1)^2} - 1). \tag{23}$$

These equations show that for any α other than 0 the variance is always greater than the mean. The marginal distribution accounts for overdispersion. The covariance of z_1 and z_2 is equal to

$$\mathrm{cov}(z_1, z_2) = \mathrm{E}z_1z_2 - \mathrm{E}z_1\mathrm{E}z_2 = \lambda_1\mathrm{E}z_2(\exp(\alpha) - 1), \tag{24}$$

and the correlation now follows by dividing by the product of the standard deviations of z_1 and z_2

$$\mathrm{corr}(z_1, z_2) = \frac{\lambda_1\mathrm{E}z_2(\exp(\alpha) - 1)}{\sqrt{\lambda_1\mathrm{E}z_2(1 + \mathrm{E}z_2(\mathrm{e}^{\lambda_1(\exp(\alpha)-1)^2} - 1))}}. \tag{25}$$

Equations (24) and (25) makes transparent that the correlation is positive (negative) whenever α is positive (negative). Notice that when α equals zero the model reverts to the bivariate model of two independent Poisson processes. In this situation we find with the help of equations (22), (23) and (24) that the expectation $\mathrm{E}z_2 = \mathrm{var}(z_2) = \mathrm{e}^{x'}\beta_2$ and that the $\mathrm{cov}(z_1, z_2)$ equals zero.

Taking the logarithm of this bivariate joint density $f^{\pi}(z_1, z_2)$ and summing over n individuals and j alternatives, the loglikelihood function for the sample equals

$$\begin{aligned}\log L(y_1, y_2, x, \theta, \pi) &= \sum_i \sum_j \log g_j^{\pi}(\cdot) \\ &= \sum_i z_{i1}x_i'\beta_1 + z_{i2}x_i'\beta_2 - \lambda_{i1} - \lambda_{i2} - \ln z_{i1}! - \ln z_{i2}! + \alpha z_{i1}z_{i2},\end{aligned} \tag{26}$$

where θ is the parameter vector containing β_1, β_2 and α. Clearly, $\log L$ is a function of dependent variables y_1, y_2 and π, since (z_1, z_2) is a permutation of (y_1, y_2). The loglikelihood depends also on exogenous regressor variables x, and parameter vector θ. To obtain estimates of the unknown parameters θ, we maximize $\log L$ with respect to θ and π. This means that we optimize the log likelihood for each permutation π, and then choose the *maximum maximorum*: the parameter θ that belongs to the joint density with the highest likelihood.

Note that the model can be estimated by using the standard Poisson model of which a standard procedure is readily available in most statistical packages. The value of the loglikelihood in (26) of permutation π is obtained by summing the loglikelihoods of the marginal Poisson regression of z_1 on x and the conditional Poisson regression of z_2 on z_1 and x.

We already mentioned the existence of two alternative approaches that develop a bivariate count distribution with flexible dependence. We will go through these recent studies briefly and discuss our contribution relative to their work. The first approach (MUNKIN and TRIVEDI, 1999; CHIB and WINKELMANN, 2001) adds correlated random effects to the Poisson parameters

$$\lambda_1 = e^{x'\beta_1+\epsilon_1} = e^{x'\beta_1}u_1, \lambda_2 = e^{x'\beta_2+\epsilon_2} = e^{x'\beta_2}u_2. \tag{27}$$

Dependence among counts is obtained through correlated u_1 and u_2. The disadvantage is that this approach requires computationally intensive simulation based estimation techniques. Our CPM approach does not. The second approach (VAN OPHEM, 1999; CAMERON, LI, TRIVEDI and ZIMMER, 2003) introduces copulas to model dependence between two nonnegative counts. In the bivariate Poisson case, a copula representation implies that two marginal Poisson distributions are transformed into two ordered response models with continuous distributions and then converted to a bivariate distribution that allows for dependence through some copula function. In the bivariate Poisson case, copula estimation is not computationally demanding. The disadvantage, however, is that a copula approach departs from Poisson modelling – each random count variable now follows a continuous marginal distribution – and thus loses some of the nice features that are typical to Poisson quasi maximum likelihood estimators (WOOLDRIDGE, 2002).

This paper continues with an empirical application on day-trip behavior.

4 Application and experiments

In this section we model trip frequencies as Poisson variables. This is not an uncommon procedure (BARMBY and DOORNIK 1989). We present estimation results on recreational behavior using the bivariate Poisson count models, and compare their performance. We further test the reliability of our bivariate Poisson count model using simulated experiments.

Data

We start with a brief description of the data. For a more comprehensive and detailed data description we refer to the Dutch Central Bureau of Statistics (1991). Estimates are based on a survey held by the Dutch Central Bureau of Statistics (CBS) in the period between September 1990 and August 1991. The survey focuses on recreational day-tripping behavior of the Dutch population. The survey was set up as a semi-panel, dividing the year into 25 two-week periods. On average, in each two-week period one thousand respondents reported all their day-trip activities in the referred period.

Day-trips are observed as counts over a two-week period. We assume that successive day-trips of one type take place independently. We illustrate the bivariate models by two day-trip counts. The first is the attendance of cultural events, that is, the number of visits to theater, cinema and concerts. The second count is the number of visits to tourist attractions, e.g., visits to the zoo, an amusement park, Keukenhof's flower exposition, etc. The counts of these two types of day trips are denoted as y_1 and y_2 respectively.

Taking into account partial nonresponse, the number of observations in our dataset amounts to 21,034. Table 1 presents descriptive statistics of the counts and observed individual characteristics. Both counts are overdispersed – the variance exceeds the mean. Overdispersion is more pronounced for tourist attractions. The correlation between counts is negative but small: −0.02. The observed determinants we use to explain variation in recreational behavior are age, gender, income and labor market status and seasonal fluctuations.

Estimates

In this subsection we describe the estimation results of three different bivariate count data models. Table 2 presents results of our conditional model, the Johnson and Kotz model and an independent bivariate Poisson model. All models estimate the considered day-trip frequencies in relation to characteristics: gender, age, income, labor market status and seasonal fluctuations.

We begin by discussing the parameter estimates attached to these observed characteristics. It turns out that all the estimates are practically identical over all four alternative models. Overall, we find that men tend to attend cultural events less often than women do. With respect to tourist attractions, gender differences are not observed. We find opposite age effects. With age visits to the theater, cinema and concerts increase while visits to tourist attractions fall. Income effects are relatively

Table 1. Descriptive statistics.

	Mean	SE	Min	Max
Counts				
Cultural events	0.098	0.350	0.000	6.000
Tourist attractions	0.102	0.413	0.000	14.000
Correlation between counts[a]	−0.021			
Observed characteristics				
Gender	0.488	0.499	0.000	1.000
Age measured in logs	3.210	0.967	0.693	4.290
Income measured in logs	7.437	1.224	4.615	8.699
Part-time worker	0.066	0.248	0.000	1.000
Full-time worker	0.324	0.468	0.000	1.000
Child	0.233	0.422	0.000	1.000
Mother	0.156	0.362	0.000	1.000
Inactive	0.143	0.350	0.000	1.000
Student	0.076	0.265	0.000	1.000
Summer	0.213	0.409	0.000	1.000
Autumn	0.264	0.440	0.000	1.000
Winter	0.262	0.439	0.000	1.000
Spring	0.261	0.439	0.000	1.000

[a]Sample correlation between counts.

Table 2. Day-trip estimates of cultural events (y_1) and tourist attractions (y_2).

	Cultural events		Tourist attractions		Cultural events		Tourist attractions	
	Conditional probability model $f(y_1)$, $f(y_2 \mid y_1)$				Conditional probability model $f(y_2)$, $f(y_1 \mid y_2)$			
Intercept[a]	−3.861	*0.275*[b]	−1.466	*0.246*	−3.833	*0.275*	−1.465	*0.246*
Gender	−0.409	*0.049*	−0.048	*0.047*	−0.409	*0.049*	−0.046	*0.047*
Ln age	0.373	*0.057*	−0.196	*0.036*	0.369	*0.057*	−0.199	*0.036*
Ln income	−0.080	*0.019*	0.036	*0.020*	−0.080	*0.019*	0.037	*0.020*
Part-time	0.581	*0.109*	−0.081	*0.121*	0.580	*0.109*	−0.085	*0.121*
Full-time	1.027	*0.086*	−0.047	*0.086*	1.026	*0.086*	−0.056	*0.086*
Child	0.934	*0.142*	0.734	*0.111*	0.939	*0.142*	0.727	*0.111*
Inactive	0.330	*0.100*	−0.556	*0.116*	0.329	*0.100*	−0.558	*0.116*
Student	1.537	*0.115*	−0.003	*0.131*	1.535	*0.115*	−0.021	*0.131*
Autumn	0.239	*0.070*	−0.835	*0.056*	0.229	*0.070*	−0.836	*0.056*
Winter	0.386	*0.068*	−2.080	*0.090*	0.371	*0.068*	−2.083	*0.090*
Spring	0.410	*0.068*	−0.531	*0.051*	0.402	*0.068*	−0.534	*0.051*
α			−0.107	*0.071*	−0.103	*0.065*		
Log-likelihood	−13546.1				−13545.9			
	Johnson and Kotz model $f(y_1, y_2)$				Independent model $f(y_1)\,f(y_2)$			
Intercept	−3.862	*0.328*	−1.465	*0.319*	−3.861	*0.275*	−1.465	*0.246*
Gender	−0.409	*0.055*	−0.046	*0.061*	−0.409	*0.049*	−0.046	*0.047*
Ln age	0.373	*0.060*	−0.199	*0.047*	0.373	*0.057*	−0.199	*0.036*
Ln income	−0.080	*0.023*	0.037	*0.028*	−0.080	*0.019*	0.037	*0.020*
Part-time	0.582	*0.118*	−0.085	*0.128*	0.581	*0.109*	−0.085	*0.121*
Full-time	1.028	*0.096*	−0.055	*0.100*	1.027	*0.086*	−0.056	*0.086*
Child	0.935	*0.161*	0.727	*0.149*	0.934	*0.142*	0.727	*0.111*
Inactive	0.331	*0.114*	−0.558	*0.143*	0.330	*0.100*	−0.558	*0.116*
Student	1.538	*0.135*	−0.022	*0.164*	1.537	*0.115*	−0.021	*0.131*
Autumn	0.239	*0.079*	−0.837	*0.070*	0.239	*0.070*	−0.836	*0.056*
Winter	0.386	*0.077*	−2.083	*0.110*	0.386	*0.068*	−2.083	*0.090*
Spring	0.410	*0.078*	−0.534	*0.063*	0.410	*0.068*	−0.534	*0.051*
γ[c]	−12.165	*0.853*						
Log-likelihood	−13547.2				−13547.2			

[a]The reference individual is a non-working housewife in the summer.
[b]Parameter estimates with standard errors are in italics.
[c]The α coefficient in the Johnson and Kotz model is modelled as e^{γ}. A high negative value of γ indicates a correlation of almost zero.

small. Compared with non-working mothers, we further find that paid workers (both part- and full-time) visit cultural events more often. Also, the non-participants on the labor market such as children, students and unemployed or retired laborers are more frequent visitors of cultural events. In particular, students show a high attendance frequency to cinema, theater and concerts. The expected number of visits by a student is about 5 ($e^{1.5}$) times higher than the expected number of visits by a person working in the household. With respect to tourist attractions, most labor market variables show a negative, though not always, significant sign. The seasonal effects are as one should expect. Whereas tourist attractions are visited more often in the summer and hardly at all in winter, attendance to cultural events is highest in spring and winter.

But what about the dependence – our prime motivation in writing this paper – between counts? If we look at the conditional probability models we find that in both cases the correlation parameter α is negative but only in the margin. The estimated parameters are significantly different from zero at a confidence level of 90%. Using confidence levels of 95% or higher, however, this observation no longer holds. Also the available likelihood ratio tests reveal that the independent model is statistically identical to the bivariate models that allow for a more flexible correlation structure. In line with α's that are marginally negative, the correlation evaluated in the sample mean is very small but negative and equals −0.01. If we look at the Johnson and Kotz model, and impose positive correlation ($\alpha = e^{\gamma}$), we find that the parameter γ is highly negative and that the corresponding correlation is almost 0. These results are identical to the results obtained using two independent Poisson distributions.

Small α's also explain why the coefficients observed across various specifications are almost but not exactly identical. This can be seen when we compare expectations of z_2 (expressed in logs) obtained under different permutations. For π equal to 0 and 1, we find expectations

$$E(z_2 \mid \pi = 0) = \lambda_{20} = e^{\lambda_{10}(\exp(\alpha_0)-1)} e^{x'\beta_{20}},$$

and

$$E(z_2 \mid \pi = 1) = \lambda_{21} = e^{x'\beta_{21}}.$$

If we assume that for an average person the estimated λ's are about the same – all variation should then run through the marginal effects – differences between coefficients across the two permutations read as

$$x'(\beta_{21} - \beta_{20}) = \lambda_{10}(e^{\alpha_0} - 1).$$

For an average person with $\lambda_{10} = 0.1$ and $\alpha_0 = 0.1$ we find that $\lambda_{10}(e^{\alpha_0} - 1) = 0.01$ which is about the same, but small, factor by which some of the marginal effects are affected.

A small experiment

To see whether differences between coefficients across specifications are more pronounced in the case of more serious correlation, we run a simulation experiment. This experiment is organized as follows. We generate 10,000 observations. We draw an explanatory variable x from a uniform [0,1] distribution and two count variables y_1 and y_2 from two Poisson distributions with Poisson parameters

$$\lambda_1 = \exp(a_0 + a_1 x),$$

and

$$\lambda_2 = \exp(b_0 + b_1 x + \alpha y_1).$$

The true values for the parameters $a_0 = b_0 = -1$ and $a_1 = b_1 = 1$. These values remain the same for each experiment. The correlation parameter takes the values $\alpha = -0.5, -0.05, 0.05$ and 0.5. We estimate our conditional Poisson model for both

Table 3. Simulation results using conditional probability models.

	$f(y_1)\,f(y_2 \mid y_1)$		$f(y_2)\,f(y_1 \mid y_2)$		$f(y_1)\,f(y_2 \mid y_1)$		$f(y_2)\,f(y_1 \mid y_2)$	
	$\alpha = -0.5$				$\alpha = -0.05$			
a_0	−1.000	*0.033*[a]	−0.885	*0.032*	−0.998	*0.026*	−0.983	*0.027*
a_1	0.998	*0.048*	1.131	*0.049*	0.997	*0.043*	1.027	*0.044*
b_0	−0.996	*0.032*	−1.119	*0.033*	−1.002	*0.030*	−1.017	*0.028*
b_1	0.994	*0.057*	0.744	*0.053*	1.005	*0.042*	0.974	*0.044*
α	−0.497	*0.022*	−0.472	*0.021*	−0.050	*0.017*	−0.050	*0.005*
Log-likelihood	−19149	*81.32*	−19165	*81.45*	−20489	*97.14*	−20489	*97.14*
n^b	100				48			
	$\alpha = 0.5$				$\alpha = 0.05$			
a_0	−0.995	*0.027*	−1.041	*0.031*	−0.999	*0.028*	−1.014	*0.028*
a_1	0.995	*0.039*	0.532	*0.106*	0.999	*0.042*	0.964	*0.043*
b_0	−0.999	*0.022*	−0.800	*0.023*	−1.001	*0.027*	−0.985	*0.027*
b_1	0.996	*0.031*	1.417	*0.033*	1.001	*0.045*	1.034	*0.042*
α	0.499	*0.010*	0.227	*0.049*	0.049	*0.015*	0.049	*0.015*
Log-likelihood	−22411	*127.80*	−23018		−20822	*104.66*	20822	*104.66*
n	100				56			

[a]Average parameter estimates with standard deviations are in italics.
[b]n indicates how many times likelihood differences assign the correct permutation.

permutations 100 times, and present mean and standard deviation of parameter estimates and the mean of likelihoods. We further show how many times the likelihoods correctly assign the correct specification. Results are given in Table 3. The structure is such that the first column always represents the correct specification. As expected, we find that in the case of serious correlation the estimates are sensitive to the permutation we estimate. In these experiments, however, there is no uncertainty about which permutation we have to choose. In case of high α's the likelihoods always correctly assign the correct model. In case of low α's it happens that the likelihoods wrongfully credit the misspecified Poisson model. But because the estimates turn out to be practically identical across permutations and very close to the true values, it does not invalidate our approach.

Based on these results we conclude the following. In the case of the recreational data under investigation, there is no clear preference in the direction of one of the described models. This is probably due to the fact that the observed counts show only a slight negative correlation. Simulation results bear this out as well. This does not mean, however, that this particular empirical illustration is without value. We show that even a small negative correlation is identified by the conditional Poisson model (CPM).

5 Discussion and concluding remarks

The analysis of count data has in recent years become common practice in econometrics. Theory, however, focuses mainly on the estimation of single counts.

Modelling multiple count data processes simultaneously remains difficult, since a flexible multivariate discrete joint density is not readily available. Literature reports some studies where the bivariate Poisson density described by Johnson and Kotz is applied. The main limitation of this bivariate distribution is that the correlation is strictly positive and bounded from above.

It is only recently that econometricians began to recognize this limitation and started to look for alternative bivariate count models that allow for more general correlation structures.

In this paper we present a new approach to estimating two count data processes simultaneously. Instead of choosing a bivariate joint density, our point of departure is a decomposition of the 'unknown' bivariate density in a marginal distribution and a conditional distribution according to the multiplication rule. The next step is to make assumptions regarding these two distributions. In this paper we assume that both the marginal and conditional distribution follow a Poisson distribution and thus implicitly define the shape of the joint density. We show that the joint density allows for a flexible correlation structure.

A The first and second derivatives of the loglikelihoods

The Johnson and Kotz loglikelihood

Using[1] the Johnson and Kotz bivariate Poisson distribution

$$P(y_1, y_2; \alpha, \theta_1, \lambda_2) = \mathrm{e}^{-\alpha-\lambda_1-\lambda_2} \sum_{l=0}^{\min(y_1,y_2)} \frac{\alpha^l}{l!} \frac{\lambda_1^{y_1-l}}{(y_1-l)!} \frac{\lambda_2^{y_2-l}}{(y_2-l)!} \tag{A.1}$$

where λ_1 and λ_2 are defined as

$$\lambda_1 = \mathrm{e}^{x'\beta_1}, \lambda_2 = \mathrm{e}^{x'\beta_2} \text{ and } \alpha = \mathrm{e}^{\gamma}. \tag{A.2}$$

If we define

$$B(k) = \sum_{l=k}^{\min(y_1,y_2)} \frac{\alpha^l}{l!} \frac{\lambda_1^{y_1-l}}{(y_1-l)!} \frac{\lambda_2^{y_2-l}}{(y_2-l)!} \tag{A.3}$$

then the partial derivative of the loglikelihood with respect to α is

$$\frac{\partial \ln P}{\partial \alpha} = -\alpha + \frac{B(1)}{B(0)}. \tag{A.4}$$

The derivatives with respect to the parameters β_1 andβ_2 are defined

$$\frac{\partial \ln P}{\partial \beta_j} = \left[y_j - \lambda_j - \frac{B(1)}{B(0)} \right] \quad x' \text{ for } j = 1, 2. \tag{A.5}$$

[1]In this Appendix we suppress subscript *i*.

In order to evaluate the Hessian matrix the second partial derivatives read as

$$\frac{\partial \ln^2 P}{\partial\alpha\partial\alpha} = -\alpha + \frac{B(2)}{B(0)} - \left(\frac{B(1)}{B(0)}\right)^2 \tag{A.6}$$

and

$$\frac{\partial \ln^2 P}{\partial\alpha\partial\beta_j} = \left[\frac{B(2)}{B(0)} - \left(\frac{B(1)}{B(0)}\right)^2\right]x' \quad \text{for } j = 1, 2 \tag{A.7}$$

and

$$\frac{\partial \ln^2 P}{\partial\beta_j\partial\beta_k} = \left[\frac{B(2)}{B(0)} - \left(\frac{B(1)}{B(0)}\right)^2\right]x'x \quad \text{for } j, k = 1, 2, j \neq k \tag{A.8}$$

and

$$\frac{\partial \ln^2 P}{\partial\beta_j\partial\beta_j} = \left[-\lambda_j + \frac{B(2)}{B(0)} - \left(\frac{B(1)}{B(0)}\right)^2\right]x'x \quad \text{for } j = 1, 2. \tag{A.9}$$

The conditional loglikelihood

Using the bivariate Poisson distribution proposed in this paper,

$$P(z_1, z_2; \alpha, \beta_1, \beta_2) = f(z_1, z_2) = g_2^{\pi}(z_2 \mid z_1)g_1^{\pi}(z_1)$$
$$= \frac{e^{z_1(x'\beta_1) - \exp(x'\beta_1) + z_2(x'\beta_2 + \alpha z_1) - \exp(x'\beta_2 + \alpha z_1)}}{z_1! z_2!}. \tag{A.10}$$

Given the parameters

$$\lambda_1 = e^{x'\beta_1} \text{ and } \lambda_2 = e^{x'\beta_2 + \alpha z_1} \tag{A.11}$$

the first partial derivatives read as

$$\frac{\partial \ln P}{\partial\alpha} = (z_2 - \lambda_2)z_1. \tag{A.12}$$

The derivatives with respect to the parameters β_1 and β_2 are defined as

$$\frac{\partial \ln P}{\partial\beta_1} = (z_1 - \lambda_1)x' \tag{A.13}$$

and

$$\frac{\partial \ln P}{\partial\beta_2} = (z_2 - \lambda_2)x'. \tag{A.14}$$

To evaluate the Hessian matrix, the second partial derivatives are defined

$$\frac{\partial \ln^2 P}{\partial\alpha\partial\alpha} = -\lambda_2 z_1^2 \tag{A.15}$$

and

$$\frac{\partial \ln^2 P}{\partial\alpha\partial\beta_1'} = 0, \frac{\partial \ln^2 P}{\partial\alpha\partial\beta_2'} = -\lambda_2 z_1 x' \tag{A.16}$$

and

$$\frac{\partial \ln^2 P}{\partial\beta_1\partial\beta_1'} = -\lambda_1 x'x, \frac{\partial \ln^2 P}{\partial\beta_2\partial\beta_2'} = -\lambda_1 x'x. \tag{A.17}$$

Finally,

$$\frac{\partial \ln^2 P}{\partial\beta_1\partial\beta_2'} = 0 \tag{A.18}$$

References

BARMBY, T. and J. DOORNIK (1989), Modelling trip frequency as a Poisson variable, *Journal of Transport Economics and Policy* **23**, 309–315.

CAMERON, A. C. and P. K. TRIVEDI (1986), Econometric models based on count data: comparisons and applications of some estimators and tests, *Journal of Applied Econometrics* **1**, 29–53.

CAMERON, A. C., T. LI, P. K. TRIVEDI and D. M. ZIMMER (2003), *A copula-based approach for the distribution of differences in counted outcomes: two case studies of health care utilization*, Department of Economics Discussion Paper, University of California, Davis.

CAMERON, A. C. and P. JOHANSSON (1998), *Bivariate count data regression using series expansions: with applications*, Department of Economics Discussion Paper, University of California, Davis.

CHIB, S. and R. WINKELMANN (2001), Markov chain Monte-Carlo analysis of correlated count data, *Journal of Business and Economic Statistics* **19**, 428–435.

Dutch Central Bureau of Statistics (1991), *Dagrecreatie 1990–1991*, CBS: Den Haag.

GOURIEROUX, C., A. MONFORT and A. TROGNON (1984), Pseudo maximum likelihood methods: applications to Poisson models, *Econometrica* **52**, 701–720.

HALL, B. H., Z. GRILICHES and J.A. HAUSMAN (1986), Patents and R and D: is there a lag? *International Economic Review* **27**, 265–283.

JOHNSON, N. L. and S. KOTZ (1969), *Distributions in statistics: discrete distributions*, New York, John Wiley & Sons.

JUNG, C. J. and R. WINKELMANN (1993), Two aspects of labor mobility: a bivariate Poisson regression approach, *Empirical Economics* **18**, 543–556.

KING, G. (1989), A seemingly unrelated Poisson regression model, *Sociological Methods and Research* **17**, 235–255.

KOCHERLAKOTA, S. and K. KOCHERLAKOTA (1992), *Bivariate discrete distributions*, New York: Marcel Dekker.

MULLAHY, J. (1986), Specification and testing of some modified count data models, *Journal of Econometrics* **33**, 341–341.

MUNKIN, M. K. and P. K. TRIVEDI (1999), Simulated maximum likelihood estimation of multivariate mixed-Poisson models, with applications, *Econometric Journal* **2**, 29–48.

VAN OPHEM, H. (1999), A general method to estimate correlated discrete random variables, *Econometric Theory* **15**, 228–237.

WINKELMANN, R. and K. F. ZIMMERMANN (1995), Recent developments in count data modelling: theory and application, *Journal of Economic Surveys* **9**, 1–24.
WOOLDRIDGE, J. M. (2002), *Econometric analysis of cross section and panel data*, The MIT Press, Cambridge, Massachusetts.

Received: August 2002. Revised: March 2004.