Econometric Theory, **15**, 1999, 228–237. Printed in the United States of America.

A GENERAL METHOD TO ESTIMATE CORRELATED DISCRETE RANDOM VARIABLES

HANS VAN OPHEM
University of Amsterdam
and
Tinbergen Institute

In this paper a method is presented to estimate correlated discrete random variables with known univariate distribution functions up to some parameters. We also present an empirical illustration on Dutch recreational data.

1. INTRODUCTION

In this paper, we consider multivariate count data models. We introduce a method to relate any two discrete random processes without having to assume zero correlation.

In a recent survey of count data, Winkelmann and Zimmermann (1995, p. 6) only marginally discuss multivariate count data models. They only mention the bivariate Poisson regression model already described in Johnson and Kotz (1969, pp. 297–300). This model is based on the trivariate reduction method: a bivariate distribution of stochastic variables Y_1 and Y_2 with Poisson marginals is generated by setting $Y_1 = V_1 + V_3$ and $Y_2 = V_2 + V_3$, where V_1, V_2, and V_3 are independently Poisson distributed. Applications of this model can be found in Gourieroux, Monfort, and Trognon (1984), King (1989), and Jung and Winkelmann (1993). An obvious disadvantage of this method is that the correlation is restricted to be strictly positive. Assuming that the parameters of the Poisson distribution of V_1, V_2, and V_3 are γ_1, γ_2, and γ_3, the correlation between Y_1 and Y_2 equals $\gamma_3/((\gamma_1 + \gamma_3)(\gamma_2 + \gamma_3))^{0.5}$, which cannot exceed $\gamma_3/(\gamma_3 + \min(\gamma_1, \gamma_2))$ (Johnson and Kotz, 1969, p. 298).

Berkhout and Plug (1996) overcome this problem by introducing a (conditional) bivariate Poisson model that is based on conditioning on one (say Y_1) of the two Poisson distributed random variables. The outcome of this random variable is then used as a factor influencing the expectation of the second Poisson

I am indebted to Peter Berkhout, Erik Plug, and Hessel Oosterbeek for discussions and support. The comments of a referee and the Co-Editor are gratefully acknowledged. The Foundation for Economic Research of the University of Amsterdam generously allowed me to make use of the dataset utilized in the empirical part of this paper. The data and computer programs used in this paper are available on request. Address correspondence to: Hans van Ophem, Institute of Actuarial Science and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.

 © 1999 Cambridge University Press 0266-4666/99 $9.50

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

distributed variable: $Y_2 \sim \text{Poisson}(\mu_2)$ with $\mu_2 = \exp(x'\beta_2 + \alpha Y_1)$, where x is a vector of explanatory variables and β and α are parameters to be estimated. They prove that if Y_1 and Y_2 conditional on Y_1 are Poisson distributed with expectations $\mu_1 = \exp(x'\beta_1)$ and $\mu_2 = \exp(x'\beta_2 + \alpha Y_1)$, $\text{cov}(Y_1, Y_2) = (\exp(\alpha) - 1)\exp((\exp(\alpha) - 1)\exp(x'\beta_1) + x'(\beta_1 + \beta_2))$. The sign of the correlation between Y_1 and Y_2 depends on the sign of α. An obvious disadvantage of this method is that two different estimates of the parameters in the model can be generated by conditioning on either one of the random variables. In particular two different estimates of the correlation between Y_1 and Y_2 can be obtained.

For two continuous random variables with known marginal distributions, the method proposed in Lee (1983) can be applied to generate a bivariate distribution in which the random variables are allowed to correlate. Assume that the continuous stochastic variables Y_1 and Y_2 have marginal cumulative distributions $F_1(\cdot)$ and $F_2(\cdot)$. Then the transformed variables

$$y_j = \Phi^{-1}(F_j(Y_j)) \qquad j = 1,2,$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard cumulative normal distribution function, are standard normal random variables. Following Lee (1983), the probability that Y_1 and Y_2 do not exceed ε_1 and ε_2, respectively, equals

$$\Pr(Y_1 \leq \varepsilon_1, Y_2 \leq \varepsilon_2) = H(\varepsilon_1, \varepsilon_2; \rho) = B(\Phi^{-1}(F_1(\varepsilon_1)), \Phi^{-1}(F_2(\varepsilon_2)); \rho),$$

where $B(\cdot, \cdot; \rho)$ is the bivariate cumulative normal distribution function with means 0, variances 1, and correlation ρ. The marginal distributions of this cumulative distribution function are $F_1(\cdot)$ and $F_2(\cdot)$.

In this paper we will generalize Lee's method to include broad classes of discrete random variables. In particular, the method can be applied to ordered discrete and count data.

Section 2 discusses the generalization of Lee's method to discrete variables. In Section 3 this method is applied to model bivariate count data. The empirical illustration deals with recreational activities. Section 4 concludes.

2. MODELING CORRELATED DISCRETE OBSERVATIONS

Consider two possibly correlated discrete stochastic variables Y_1 and Y_2 and denote the correlation by ρ_{12}. The variable Y_1 can take on values $1,2,\ldots,K$ and Y_2 $1,2,\ldots,M$, where K and M are finite or countable.[1] We assume that the possible outcomes are ranked in a logical manner. Furthermore, we assume that the marginal distributions of both Y_1 and Y_2 are known up to vectors of parameters θ_1 and θ_2. Denote the marginal distributions of Y_1 and Y_2 by

$$\Pr(Y_1 = k) = p_1(k; \theta_1) \qquad k = 1,\ldots,K,$$
$$\Pr(Y_2 = m) = p_2(m; \theta_2) \qquad m = 1,\ldots,M, \qquad \textbf{(1)}$$

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

To have a regular distribution function we require

$$\sum_{k=1}^{K} p_1(k;\theta_1) = 1 \quad \text{and} \quad p_1(k;\theta_1) \geq 0 \qquad k = 1,\ldots,K,$$

$$\sum_{m=1}^{M} p_2(m;\theta_2) = 1 \quad \text{and} \quad p_2(m;\theta_2) \geq 0 \qquad m = 1,\ldots,M. \tag{2}$$

Given the finite number of outcomes and under the assumption that θ_1 takes on a particular value, we can always determine numbers $\eta_1, \eta_2, \ldots, \eta_{K-1}$ such that

$$p_1(k;\theta_1) = \int_{\eta_{k-1}}^{\eta_k} \phi(u_1)\, du_1 \qquad k = 2,\ldots,K-1, \tag{3a}$$

where $\phi(\cdot)$ is the density of the standard normal distribution function. For observations $k = 1$ and $k = K$ we specify

$$p_1(1;\theta_1) = \int_{-\infty}^{\eta_1} \phi(u_1)\, du_1,$$

$$p_1(K;\theta_1) = \int_{\eta_{K-1}}^{+\infty} \phi(u_1)\, du_1. \tag{3b}$$

Equations (3a) and (3b), given some θ_1, define $\eta_1, \eta_2, \ldots, \eta_{K-1}$ uniquely, and η_k can be seen as a function of θ_1. The relation between the probabilities $p_1(k;\theta_1)$ and the η_k's can also be written as

$$p_1(k;\theta_1) = \Phi(\eta_k) - \Phi(\eta_{k-1}), \tag{4}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Note that

$$\Pr(Y_1 \leq \kappa) = \sum_{k=1}^{\kappa} p_1(k;\theta_1) = \Phi(\eta_\kappa), \tag{5}$$

where we define $\eta_K = +\infty$. Consequently

$$\eta_\kappa = \Phi^{-1}\left(\sum_{k=1}^{\kappa} p_1(k;\theta_1)\right). \tag{6a}$$

This relation defines η_k $(k = 1,\ldots,K-1)$ uniquely for any value of θ_1, so clearly η_k is a function of θ_1: $\eta_k(\theta_1)$.

For the distribution of Y_2 and given some value of θ_2 we define λ_μ in the same manner:

$$\lambda_\mu(\theta_2) = \Phi^{-1}\left(\sum_{m=1}^{\mu} p_2(m;\theta_2)\right) \qquad \mu = 1,\ldots,M-1, \tag{6b}$$

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

and $\lambda_M = +\infty$. Suppose that we have N observations to estimate the parameter vectors θ_1 and θ_2 with the method of maximum likelihood. For observation i we observe $Y_{1i} = k$ and $Y_{2i} = m (i = 1, \ldots, N, k = 1, \ldots, K, m = 1, \ldots, M)$. The likelihood contribution of this observation is

$$\ell_i = \Pr(Y_{1i} = k, Y_{2i} = m). \tag{7}$$

The problem now is that we do not want to assume that these stochastic variables are uncorrelated. The solution is to apply the method described in Lee (1983) to the present discrete case. Given (6a) and (6b) we can write

$$\Pr(Y_{1i} \leq k, Y_{2i} \leq m) = B(\eta_k, \lambda_m; \rho), \tag{8}$$

where $B(\cdot, \cdot; \rho)$ is the bivariate normal cumulative distribution function with expectations equal to 0, variances equal to 1, and correlation ρ. The marginal distributions of Y_{1i} and Y_{2i} remain the original distributions $p_1(k; \theta_1)$ and $p_2(m; \theta_2)$. The likelihood contribution of observation i equals

$$\begin{aligned}
\ell_i &= \Pr(Y_{1i} = k, Y_{2i} = m) \\
&= \Pr(Y_{1i} \leq k, Y_{2i} \leq m) - \Pr(Y_{1i} \leq k - 1, Y_{2i} \leq m) \\
&\quad - \Pr(Y_{1i} \leq k, Y_{2i} \leq m - 1) + \Pr(Y_{1i} \leq k - 1, Y_{2i} \leq m - 1) \\
&= B(\eta_k, \lambda_m; \rho) - B(\eta_{k-1}, \lambda_m; \rho) - B(\eta_k, \lambda_{m-1}; \rho) + B(\eta_{k-1}, \lambda_{m-1}; \rho)
\end{aligned} \tag{9}$$

To estimate θ_1, θ_2, and ρ the log-likelihood function needs to be maximized. Define a dummy variable $z_i(k, m)$ equal to 1 if we observe for observation i: $Y_{1i} = k$ and $Y_{2i} = m$. Then we maximize

$$\begin{aligned}
\log L(\theta_1, \theta_2, \rho) &= \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{m=1}^{M} z_i(k, m) \log(\Pr(Y_{1i} = k, Y_{2i} = m)) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{m=1}^{M} z_i(k, m) \log(B(\eta_k(\theta_1), \lambda_m(\theta_2); \rho) \\
&\quad - B(\eta_{k-1}(\theta_1), \lambda_m(\theta_2); \rho) - B(\eta_k(\theta_1), \lambda_{m-1}(\theta_2); \rho) \\
&\quad + B(\eta_{k-1}(\theta_1), \lambda_{m-1}(\theta_2); \rho))
\end{aligned} \tag{10}$$

To determine the relation between ρ and ρ_{12} consider (8):

$$\Pr(Y_{1i} \leq k, Y_{2i} \leq m) = B(\eta_k, \lambda_m; \rho) = \int_{-\infty}^{\eta_k} \int_{-\infty}^{\lambda_m} b(u_1, u_2; \rho)\, du_1\, du_2, \tag{11}$$

where $b(\cdot, \cdot; \rho)$ is the bivariate normal density with means 0, variances 1, and correlation ρ. Clearly ρ is the correlation between the underlying normally distributed variables. The term ρ_{12} is the correlation between Y_1 and Y_2, and it is not

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

equal to ρ in general. However, it is related. Consider the covariance between Y_1 and Y_2:

$$\operatorname{cov}(Y_1,Y_2) = \sum_{k=1}^{K}\sum_{m=1}^{M} km \Pr(Y_1 = k, y_2 = m) - \left(\sum_{k=1}^{K} k \Pr(Y_1 = k)\right)\left(\sum_{m=1}^{M} m \Pr(Y_2 = m)\right), \quad \textbf{(12)}$$

where only the first probability is a function of ρ. From (9) it can be simply deduced that if $\rho = 0$, $\operatorname{cov}(Y_1,Y_2) = 0$, and consequently $\rho_{12} = 0$. The expression $\rho = \pm 1$, indicating an exact linear relation between the underlying normally distributed variables, does not translate into ρ_{12} being equal to ± 1, as is obvious in the case that $K \neq M$. Consequently, in that case there exists some lower (>-1) and upper bound ($<+1$) of ρ_{12}. If $K = M$ and $Y_1 = Y_2$, clearly $\rho_{12} = \rho = 1$.

Although (12) gives the relation between $\operatorname{cov}(Y_1,Y_2)$ and ρ, it is very hard to find the properties of this relation. In particular, we would like to establish the result that ρ_{12} is an increasing function of ρ. The difficulty in proving this lies in the first probability on the right-hand side of (12). It is the only element depending on ρ, but as is clear from (9) it actually consists of four bivariate normal probabilities all depending on ρ. Given the analytical result that if $\rho_{12} = 0$, $\rho = 0$, we only need to show that EY_1Y_2 is an increasing function of ρ to establish the required result in one direction. This result was not obtained analytically, however.[2] To get some better insight, a numerical analysis was carried out. Table 1 contains the results. One thousand replications were performed for each entry (differing in ρ). Each replication represents two correlated draws of 10,000 observations from the Poisson distribution with expectation 1. For every observation and for each replication, two numbers were drawn from the normal distribution with correlation ρ. By means of the probabilities following from the Poisson distribution, these numbers were converted to correlated Poisson distributed stochastic variables. The mean sample correlations ($\hat{\rho}_{12}$) of the 1,000 replications of the 10,000 Poisson distributed pairs of observations created in this manner are presented in Table 1.

The results in Table 1 indicate that $|\rho_{12}|$ is always smaller than $|\rho|$ and that if ρ is positive (negative) the mean ρ_{12} is also positive (negative). Furthermore there is a clear correspondence between the magnitudes of ρ and ρ_{12}. Note that the estimated sample correlations do not appear to be symmetric around 0 as the input correlations are by definition. Another striking result is that the standard errors increase the closer one gets to 0. All in all, we can conclude that there appears to exist a one to one relation between ρ and ρ_{12}.

In the case that individuals are observed to make choices Y_{1i} and Y_{2i} it is customary to relate individual characteristics to the observed outcomes. The way to proceed in that case is to make the characteristics of the distribution function dependent on these individual characteristics: $\theta_{1i} = \theta_1(w_{1i}$ and $\theta_{2i} = \theta_2(w_{2i})$,

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

TABLE 1. Simulation results for the Poisson distribution with $\lambda = 1$

ρ	Mean $\hat{\rho}_{12}$	Standard error
−0.9	−0.502	(0.005)
−0.8	−0.474	(0.006)
−0.7	−0.441	(0.006)
−0.6	−0.400	(0.007)
−0.5	−0.351	(0.008)
−0.4	−0.296	(0.008)
−0.3	−0.232	(0.008)
−0.2	−0.160	(0.009)
−0.1	−0.081	(0.010)
0	0.000	(0.010)
0.1	0.083	(0.010)
0.2	0.167	(0.010)
0.3	0.247	(0.010)
0.4	0.322	(0.010)
0.5	0.392	(0.009)
0.6	0.454	(0.009)
0.7	0.510	(0.008)
0.8	0.559	(0.008)
0.9	0.602	(0.007)

where w_{1i} and w_{2i} are vectors of the individual characteristics, not necessarily equal. As a result η_k, λ_m, and ρ_{12} (cf. (6a), (6b), and (12)) differ across observations. Note that this conclusion holds even though ρ is constant across individuals. This is another way to show that $\rho \neq \rho_{12}$.

Another interesting feature of the method to link correlated stochastic variables discussed previously is that different distributions (discrete or continuous) can be combined. This is in contrast with the usual approaches (cf. Kocherlakota and Kocherlakota, 1992; Johnson and Kotz, 1969).

The original approach of Lee (1983) differs from the present method. Lee discussed how a polychotomous choice selectivity model can be estimated by combining a multinomial logit model and a linear equation. He made use of the well-known relation between utility maximization and the multinomial logit model if utility contains a stochastic component that has a type I extreme-value distribution. Lee was then in a position to again use his method for combining two continuous stochastic variables. In the present case things are quite similar except that an intermediate step is needed as a result of the absence of an underlying continuous process. In particular, in our case we do not need to make use of the random components of utility being type I extreme-value distributed, but we can start from the logit probabilities directly. These probabilities are related to the

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

normal distribution as is shown in (6). An ordered probitlike approach is used to relate a discrete process to the underlying continuous process.

The discussed approach can in theory be extended to combining more than two stochastic variables. The resulting problems will be of a computational nature: the bivariate distribution in (10) will become a multivariate normal distribution. For higher dimensions, this probability needs to be evaluated by simulation estimation techniques (see, e.g., Hajivassiliou and Ruud, 1994). Note that the inverse of the normal cumulative distribution function, needed to calculate η_k's and the λ_m's, remains one dimensional.

In principle it is possible to utilize another continuous density function in transforming the discrete probabilities to probabilities of a continuous stochastic variable (cf. (3)). However, the normal distribution has the advantage that it is easily generalized to combinations of correlated normally distributed random variables. For other distributions such a generalization is often troublesome (cf. Johnson and Kotz, 1972).

In the next section an empirical illustration on the basis of count data will be presented.

3. EMPIRICAL ILLUSTRATION

In this section we make use of the data of Berkhout and Plug (1996). Their empirical analysis deals with recreational or tourist behavior, in particular the recreational day-tripping behavior of the Dutch population. The data are from a survey held by the Dutch Central Bureau of Statistics (CBS) in the period September 1990–August 1991. This year was divided into 25 two-week periods. On average 1,000 respondents reported in each two-week period on all their day-trip activities in the preceding two weeks. A large number of different types of day trips are distinguished, but we will concentrate on sightseeing (the number of times the individual claimed to be engaged in sightseeing) and visits to tourist attractions (zoos, amusement parks, expositions). The total number of observations is 21,034. The estimates presented in this paper are based on a 10% random sample (2,103 observations). The counts range from 0 up to 4 (sightseeing, mean: 0.217, variance: 0.266) and 0 up to 5 (tourist attractions, mean: 0.088, variance: 0.128).

The distributions of Y_1 (sightseeing) and Y_2 (tourist attractions) will be assumed to be Poisson distributions. The corresponding probabilities for each possible outcome are

$$\text{Poisson distribution: } \Pr(Y_j = k) = \frac{\exp(-\lambda_j)\lambda_j^k}{k!} \qquad k = 0,1,2,\ldots, \tag{13}$$

where the λ_j's are positive numbers. The expectation and variance of Y_j are both equal to λ_j. To allow individual specific characteristics (represented by the vector X_i) to influence the probabilities we will specify $\lambda_j = \exp(X_i\beta_j)$.

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

The following twelve variables are used as explanatory variables: gender (reference females), log age, log income, five dummies representing labor market status (part-time worker, full-time worker, full-time mother, inactive, student (reference unemployed)), a dummy variable representing whether the respondent has children, and three dummies for seasonal effects (autumn, winter, spring (reference summer)).

Table 2 contains the estimation results based on maximizing the log-likelihood function (10) under the assumption that the counts are Poisson distributed with and without the restriction $\rho = 0$.[3]

The number of times an individual went sightseeing and the number of tourist attractions he or she visited during the two-week period before the moment the interview took place are strongly positively correlated. Apparently, sightseeing is often combined with visiting an attraction, or they are more or less considered to be the same kind of recreational activities. As was already noted in Section 2, the estimated correlation is not an estimate of ρ_{12}. To obtain an estimate of this correlation we estimated the covariance between Y_1 and Y_2 by calculating the sample mean of the observation specific covariance in (12). Dividing this number by the standard errors of Y_1 and Y_2 yields an estimate of ρ_{12} for every observation. The mean of these estimates is 0.394. The minimum and maximum values of ρ_{12} are 0.176 and 0.554. So, the actual correlation is smaller than the correlation between the underlying variables u_1 and u_2. The assumption of uncorrelated counts does not change the magnitude of the estimated coefficients very much, but the standard errors are considerably larger. At a 10% level, 17 out of 25 estimates are significantly different from zero in the case with correlation, whereas this number reduces to 11 out of 24 estimates in the uncorrelated case. The estimation results indicate that older people prefer to go sightseeing and the younger people prefer visiting tourist attractions. Sightseeing is more popular for individuals with higher incomes, whereas tourist attractions are more liked by individuals with lower incomes. Full-time workers spend less time on both recreational activities. A somewhat surprising result is that individuals having children do not significantly visit tourist attractions more frequently than individuals without children, but they do go sightseeing more often. The effect of the seasons is as we would expect. Summer is the reference category, and then both activities are more popular. Sightseeing or visiting tourist attractions in winter is much less liked.

4. CONCLUSION

In this paper a simultaneous model was presented that allows the estimation of the parameters of correlated discrete random variables with known univariate distributions. The univariate distribution functions are the marginals of the employed simultaneous distribution function. The method is a direct generalization of the one discussed in Lee (1983) for relating continuous variables. Combining Lee's approach and the method discussed here yields the possibility to combine

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press

TABLE 2. Estimation results on day trips (Poisson distributions)

	With correlation				Without correlation			
	Sightseeing		Tourist attractions		Sightseeing		Tourist attractions	
Intercept	−3.186	(0.540)	0.264	(0.705)	−3.264	(0.609)	−0.329	(0.797)
Gender	0.148	(0.104)	−0.188	(0.171)	0.099	(0.124)	−0.283	(0.187)
Log age	0.150	(0.087)	−0.262	(0.107)	0.128	(0.108)	−0.233	(0.130)
Log income	0.172	(0.049)	−0.150	(0.060)	0.193	(0.055)	−0.080	(0.062)
Part-time	0.116	(0.220)	−1.372	(0.677)	0.104	(0.236)	−1.433	(0.766)
Full-time	−0.375	(0.168)	−0.462	(0.270)	−0.344	(0.196)	−0.362	(0.310)
Inactive	−0.097	(0.171)	−0.626	(0.334)	−0.096	(0.196)	−0.597	(0.380)
Student	−0.147	(0.295)	−0.622	(0.378)	−0.213	(0.328)	−0.853	(0.449)
Children	0.616	(0.251)	0.413	(0.301)	0.655	(0.299)	0.561	(0.370)
Autumn	−0.207	(0.121)	−0.567	(0.170)	−0.220	(0.146)	−0.605	(0.194)
Winter	−0.902	(0.153)	−2.207	(0.358)	−0.896	(0.173)	−2.154	(0.362)
Spring	−0.058	(0.117)	−0.694	(0.193)	−0.036	(0.144)	−0.651	(0.219)
ρ	0.699 (0.028)							

Note: $\log L(\rho \neq 0) = -1{,}584.001$; $\log L(\rho = 0) = -1{,}744.992$; $N = 2{,}103$. Asymptotic standard errors are in parentheses.

any discrete and continuous random variables. In an empirical illustration the proposed method was shown to be feasible.

NOTES

1. For some discrete distributions, e.g., the Poisson distribution, the stochastic variables Y_{1i} and Y_{2i} are unbounded ($K, M = +\infty$). This does not cause any problems in the estimation of the model. The log-likelihood function only contains the probabilities of actual observations, and these are always bounded.

2. Lancaster and Hamdan (1964) (also see Kendall and Stuart, 1967, p. 561) discussed a related problem. They introduced the polychoric method for estimation of the correlation coefficient in an $r \times c$ contingency table. Under the assumption of a bivariate normal distribution, the estimate of the correlation is a solution of a nonlinear equation that can only be calculated with numerical means.

3. All estimations are carried out with GAUSS. The numerical approximation of the inverse of the cumulative normal distribution function is based on the algorithm presented in Kennedy and Gentle (1980, p. 95). It is rated to be accurate to 1×10^{-7} but seems to be even more accurate than that. For the case that $\rho = 0$, the results of the estimation with the use of the method discussed in this paper are compared to those of a direct estimation of the Poisson distribution (GAUSS-module COUNT). The results do not differ up to the sixth digit of each estimate.

REFERENCES

Berkhout, P.H.G. & E.J.S. Plug (1996) A Bivariate Poisson Count Data Model Using Conditional Probabilities. Tinbergen Institute Discussion paper TI 96-88/7, Amsterdam, Tinbergen Institute.

Gourieroux, C., A. Monfort, & A. Trognon (1984) Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica* 52, 701–720.

Hajivassiliou, V.A. & P.A. Ruud (1994) Classical estimation methods for LDV models using simulation. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2383–2441. Amsterdam: Elsevier Science B.V.

Johnson, L.J. & S. Kotz (1969) *Distributions in Statistics: Discrete Distributions*, vol. 1. New York: Wiley.

Johnson, L.J. & S. Kotz (1972) *Distributions in Statistics: Continuous Multivariate Distributions*, vol. 4. New York: Wiley.

Jung, C.J. & R. Winkelmann (1993) Two aspects of labor mobility: A bivariate Poisson regression approach. *Empirical Economics* 18, 543–556.

Kendall, M.G. & A. Stuart (1967) *The Advanced Theory of Statistics*, vol. 2, 2nd ed. London: Charles Griffin & Company.

Kennedy, W.J. & J.E. Gentle (1980) *Statistical Computing.* New York: Marcel Dekker.

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

Kocherlakota, S. & K. Kocherlakota (1992) *Bivariate Discrete Distributions.* New York: Marcel Dekker.

Lancaster, H.O. & M.A. Hamdan (1964) Estimation of the correlation coefficient in contingency tables with possibly nonmetrical characters. *Psychometrika* 29, 383–391.

Lee, L.F. (1983) Generalized econometric models with selectivity. *Econometrica* 51, 507–512.

Winkelmann, R. & K.F. Zimmermann (1995) Recent developments in count data modelling: Theory and application. *Journal of Economic Surveys* 9, 1–24.

https://doi.org/10.1017/S0266466699152058 Published online by Cambridge University Press