OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 62, 4 (2000)
0305-9049

SEEMINGLY UNRELATED NEGATIVE BINOMIAL REGRESSION†

Rainer Winkelmann

I. INTRODUCTION

The seemingly unrelated Poisson (SUP) model (King (1989a), see also Aptech (1994)) is a model for bi- or multi-variate counted outcomes. Correlation between equations is introduced through a convolution structure with a common additive factor. Previous economic applications include the joint analysis of voluntary and involuntary labour mobility (Jung and Winkelmann, 1993) and the number of recreational trips to each of two sites (Ozuna and Gomez, 1994).

From a methodological point of view, these studies have been subject to a serious limitation of the seemingly unrelated Poisson model, namely its inability to account for over-dispersion, or extra-Poisson variation in the data. But this phenomenon is widespread in economic applications, as is well documented for univariate models (Winkelmann, 2000), and the usefulness of the SUP model is thus limited. Indeed, much of the recent literature on multivariate count data has turned away from a convolution structure and rather introduced correlation by mixing over a multiplicative error (see, for instance, Cincera, 1998, Gurmu and Elder, 1998, Chib and Winkelmann, 1999). Over-dispersion arises naturally in this context.

In this paper, I propose an alternative simple new model for seemingly unrelated counts with over-dispersion. The distinctive feature of this model is that it does not abandon the basic convolution structure of the seemingly unrelated Poisson model but rather generalizes some of its assumptions in order to allow for over-dispersion. The new model has negative binomial marginals and is referred to as a 'seemingly unrelated negative binomial' (SUNB) model. The SUNB generalizes the SUP model by introducing one additional dispersion parameter, the statistical significance of which can be tested by standard tests.

II. MODELING CORRELATED COUNT DATA

Since the seemingly unrelated negative binomial model (SUNB) is a generalization of the seemingly unrelated Poisson model (SUP), and in

†Valuable comments by an anonymous referee and the editor are gratefully acknowledged.

order to establish notation, it is useful to begin with a short review of the SUP model. The results are relatively standard and discussed, among others, in King (1989a), Jung and Winkelmann (1993), and Gurmu and Elder (1998).

Let $z_i = (z_{i1}, \ldots, z_{iJ})'$ denote a vector of J counts each element of which is independently Poisson distributed with parameter $\lambda_{ij} = \exp(x'_{ij}\beta_j)$. Furthermore, u_i, a scalar random variable, is independently Poisson distributed with mean γ. Now consider the vector $y_i = (y_{i1}, \ldots, y_{iJ})'$ with typical element $y_{ij} = z_{ij} + u_i$, $j \leqslant J$ (u_i does not depend on covariates although this could be easily generalized as, for instance, in King, 1989b).

In order to establish the distribution function of a sum of two non-negative random variables, the probability generating function $P(s) = \mathrm{E}(s^X)$ can be used (Feller, 1968). For the Poisson distribution $P(s) = \mathrm{e}^{\lambda(s-1)}$, and under independence of z and u,

$$P_{z+u}(s) = P_z(s)P_u(s) = \mathrm{e}^{(\lambda+\gamma)(s-1)}$$

Thus, the marginal distribution of y_{ij} is Poisson with parameter $\lambda_{ij} + \gamma$. Moreover, let $\Lambda_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{iJ})$ and $\mathbf{1}$ a $(J \times 1)$ vector of ones. The covariance matrix of y_i can be written as

$$\mathrm{Var}(y_i) = \Lambda_i + \gamma \mathbf{1}\mathbf{1}', \tag{1}$$

The resulting model for correlated counts is sometimes referred to as 'multivariate Poisson distribution', although other joint distributions with Poisson marginals can be constructed (Kocherlakota and Kocherlakota, 1992).

The joint distribution function for cluster i takes the form

$$f(y_{i1}, \ldots, y_{iJ}) = \sum_{k=0}^{s_i} f(k) \prod_{j=1}^{J} f(y_{ij} - k) \tag{2}$$

where f is the univariate Poisson probability function and $s_i = \min(y_{i1}, \ldots, y_{iJ})$. The intuition behind this joint probability function is as follows. First, u_i cannot exceed any of the observed counts $(y_{i1}, \ldots, y_{iJ})$ because each count is the sum of u_i and a non-negative count z_{ij}. Hence, its upper bound is s_i. Secondly, the joint probability of observing y_i is the sum over the joint probabilities $f(u_i, y_{i1} - u_i, \ldots, y_{iJ} - u_i)$ for $u_i = 0, \ldots, s_i$. Third, it follows from independence that the joint probability can be factored such that

$$\begin{aligned} f(u_i, y_{i1} - u_i, \ldots, y_{iJ} - u_i) &= f(u_i, z_{i1}, \ldots, z_{iJ}) \\ &= f(u_i)f(z_{i1}) \cdots f(z_{iJ}) \end{aligned}$$

The model is completed by making the standard assumption of random sampling of n individuals such that

$$f(y_1, \ldots, y_n|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(y_i|x_i) \tag{3}$$

In order to fully understand the advantages and limitations of this model the following points should be noted. First, the model was presented within the framework of a seemingly unrelated regression. The model is more general though, as it is equally suited for other types of correlated counts, as may arise in any multivariate context, including genuine panel data. Second, the model produces a simple test of independence, as the joint density factors into the product of Poisson marginals for $\gamma = 0$.

Third, the parameterization used here differs slightly from the version in King (1989a) and Jung and Winkelmann (1993). There, $z_{ij} \sim \text{Poisson}(\lambda_{ij} - \gamma)$ such that $y_{ij} \sim \text{Poisson}(\lambda_{ij})$. This specification does not guarantee that the parameter of the z_{ij}-distribution is positive, causing both conceptual and potential numerical problems. Although the two models differ not only in their constant but also in the underlying assumption for the skedastic (variance) function, the interpretation of the regression parameters is the same in both parameterizations, as in either case $\partial \text{E}(y_{ij}|x_{ij})/\partial x_{ij} = \exp(x_{ij}'\beta_j)\beta_j$.

Finally, the covariance matrix is quite similar to the covariance matrix of the normal linear error-components model. In particular, covariances are forced to be non-negative and identical within the cluster (and within-cluster covariances are the same across clusters). Whether these assumptions are justified depends on the specifics of the application.

There is, however, another feature of the covariance matrix that has caused dissatisfaction with the SUP model. By assumption $\text{E}(y_{ij}|x_{ij}) = \text{Var}(y_{ij}|x_{ij})$. This property, also referred to as 'equi-dispersion', restricts the conditional expectation to be equal to the conditional variance whereas in most reported uses of count data regression models, the conditional variance appears to exceed the conditional expectation, a situation referred to as 'over-dispersion'.

In the next section, a new model is derived that generalizes the SUP model in order to allow for such over-dispersion.

III. MULTIVARIATE NEGATIVE BINOMIAL MODEL

The standard approach to introduce the univariate negative binomial model for the analysis of count data is by way of a mixture distribution (Hausman, Hall and Griliches, 1984, Winkelmann, 2000). This approach has intuitive appeal as it can be seen to follow from the presence of unobserved heterogeneity in the model. Let $z_{ij}|v_{ij}$ be Poisson distributed with parameter $\lambda_{ij}v_{ij}$ where v_{ij} has a gamma distribution $\Gamma(\alpha, \alpha)$ with mean $\text{E}(v_{ij}) = 1$ and variance $\text{Var}(v_{ij}) = \alpha^{-1}$. The marginal distribution of z_{ij} after integration over v_{ij} can be shown to be negative binomial with mean $\text{E}(z_{ij}) = \lambda_{ij}$ and variance $\text{Var}(z_{ij}) = \lambda_{ij} + \alpha^{-1}\lambda_{ij}^2$.

For our purposes, it is useful to re-parameterize the model such that $\alpha = \lambda_{ij}/\sigma$. This change does not affect the mean but the variance $\mathrm{Var}(z_{ij}) = \lambda_{ij}(1+\sigma)$ is now a linear function of the mean. This parameterization of the negative binomial distribution is known as the 'Negbin I' model. Its probability generating function is given by

$$P_z(s) = [1+\sigma(1-s)]^{-\lambda_{ij}/\sigma}$$

With similar reasoning, assume that u_i has a Negbin I distribution with probability generating function

$$P_u(s) = [1+\sigma(1-s)]^{-\gamma/\sigma}$$

Note that the two distributions share a common dispersion-parameter σ. Under independence, we obtain for $y_{ij} = z_{ij} + u_i$ that

$$\begin{aligned} P_y(s) &= [1+\sigma(1-s)]^{-\lambda_{ij}/\sigma}[1+\sigma(1-s)]^{-\gamma/\sigma} \\ &= [1+\sigma(1-s)]^{-(\lambda_{ij}+\gamma)/\sigma} \end{aligned} \tag{4}$$

But (4) is the probability generating function of a negative binomial distribution with expectation $\mathrm{E}(y_{ij}) = \lambda_{ij} + \gamma$ and variance $\mathrm{Var}(y_{ij}) = (\lambda_{ij} + \gamma)(1+\sigma)$. The covariances between y_{ij} and y_{ik}, $j \neq k$, are given by

$$\begin{aligned} \mathrm{Cov}(y_{ij}, y_{js}) &= \mathrm{Var}(u_i) \\ &= \gamma(1+\sigma) \end{aligned}$$

and the covariance matrix can be written in compact form as

$$\mathrm{Var}(y_i) = [\Lambda_i + \gamma \mathbf{1}\mathbf{1}'](1+\sigma) \tag{5}$$

This matrix differs from the covariance matrix of the SUP model by a proportionality factor $(1+\sigma)$. Thus, the SUNB model allows for overdispersion as long as $\sigma > 0$, and the SUP restriction $\sigma = 0$ can be tested by a likelihood-ratio test. The asymptotic distributions of the likelihood ratio test statistic is not chi-squared since the parameter is bounded from below at zero. Following Self and Liang (1987), it can be established that the asymptotic distribution is a 50/50 mixture of a $\chi^2_{(1)}$ distribution and the constant value 0.

The joint probability function of the SUNB model for cluster i is obtained along the lines of (2) where f now stands for the univariate Negbin I probability function. For instance, for $z_{ij} = y_{ij} - u_i$

$$f_{NB}(z_{ij}) = \frac{\Gamma(\lambda_{ij}/\sigma + z_{ij})}{\Gamma(\lambda_{ij}/\sigma)\Gamma(z_{ij}+1)} \left(\frac{1}{1+\sigma}\right)^{\lambda_{ij}/\sigma} \left(\frac{\sigma}{1+\sigma}\right)^{z_{ij}}. \tag{6}$$

The parameters of the model can be estimated by maximizing the corresponding log-likelihood function. The estimator is asymptotically normal distributed with mean $\theta = (\beta_1, \beta_2, \gamma, \sigma)'$ and covariance matrix equal to the inverse of minus the expected Hessian matrix of the log-likelihood function.

The presence of the gamma function in (6) tends to make the numerical evaluation of the log-likelihood function somewhat imprecise, in particular for large values that occur for instance whenever the model approaches the SUP model for small values of σ. It is preferably, then, to exploit the recursive property of the Gamma function and replace the difference $\ln\Gamma(\lambda_{ij}/\sigma + z_{ij}) - \ln\Gamma(\lambda_{ij}/\sigma)$ by the sum

$$\sum_{s=1}^{z_{ij}} \ln(z_{ij} - s + \lambda_{ij}/\sigma)$$

(set equal to zero for $z_{ij} = 0$). Another aspect of numerical convenience is to include the logarithms of the non-negative parameters γ and σ rather than their levels, which can always be done given the invariance property of the maximum likelihood estimator.

Finally, it should be pointed out that the SUNB model differs from the multivariate negative binomial model in Hausman, Hall and Griliches (1984) that was based on a multiplicative error structure. One of the advantages of the SUNB model over the multiplicative model is a gain in flexibility since the SUNB model allows for separate estimation of correlation and dispersion whereas in the earlier model correlation and dispersion were determined by the same single parameter.

IV. DISCUSSION

The derivation of the model depends on two particular assumptions: the overdispersion parameter σ is the same for all dependent count variables, and the conditional variance is proportional to the conditional mean. Both assumptions can in principle be tested using marginal models. A simple test of the first hypothesis is to estimate the marginal models with and without imposing equality of the dispersion parameter, and then conduct a likelihood ratio test. The second hypothesis can be tested, for instance, based on the Negbin$_k$ hyper-model that nests the Negbin-I and Negbin-II models for particular value of the hyper-parameter k (see Winkelmann, 2000, for further details).

If the SUNB model is misspecified, the maximum likelihood estimator for β is not consistent, even if the conditional expectation function is correctly specified. Already the univariate Negbin-I model does not meet the necessary condition for consistent estimation of a pseudo maximum-likelihood estimator (See Gourieroux, Monfort and Trognon, 1984a,b). The convolution structure of the joint model is another factor why correct specification is required. But this problem is not restricted to the SUNB model. It also arises for the SUP model where pseudo-maximum likelihood results are not available, and consistent estimation by maximum likelihood requires the correct specification of the model, although pseudo-likelihood results are available in this case for the marginal models. The SUNB model offers an important advantage over the SUP model, since consistent estimation becomes possible under the plausible scenario of over-dispersion, where the SUP model is misspecified. Under the assumptions of the model, joint estimation is best asymptotically normal.

Of course, the aforementioned tests may reveal that the SUNB model should not be applied. In this case, one would wish to have available an estimator that remains consistent under weaker assumptions. A robust seemingly unrelated model with varying overdispersion can be obtained by specifying the first two moments of the joint distribution of the dependent variables only. Consider, for instance, the following non-linear regression model with linear variance function (for notational convenience the bivariate case is adopted)

$$\mathrm{E}\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \exp(x'_{i1}\beta_1) \\ \exp(x'_{i2}\beta_2) \end{pmatrix}$$

$$\mathrm{Var}\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \exp(x'_{i1}\beta_1)(1+\sigma_{11}) & \sigma_{12} \\ \sigma_{12} & \exp(x'_{i2}\beta_2)(1+\sigma_{22}) \end{pmatrix}$$

The model can be consistently estimated by generalized non-linear least squares (GNLS). First-step estimates are obtained by minimizing

$$\sum_{i=1}^{n}[(y_{i1} - \exp(x'_{i1}\beta_1))^2 + (y_{i2} - \exp(x'_{i2}\beta_2))^2] \tag{7}$$

If b_1 and b_2 denote the estimates, residuals can be obtained as

$$\hat{\omega}_{i1} = y_{i1} - \exp(x'_{i1}b_1), \qquad \hat{\omega}_{i2} = y_{i2} - \exp(x'_{i2}b_2)$$

For consistent estimators of σ_{ij}, i, $j = 1$, 2, apply ordinary least squares to

$$\hat{\omega}^2_{i1} - \exp(x'_{i1}b_1) = \sigma_{11}\exp(x'_{i1}b_1) + \text{error}$$

$$\hat{\omega}^2_{i2} - \exp(x'_{i2}b_2) = \sigma_{22}\exp(x'_{i2}b_2) + \text{error}$$

$$\hat{\omega}_{i1}\hat{\omega}_{i2} = \sigma_{12} + \text{error}$$

Finally, the (feasible) GNLS estimator is the solution to

$$\min\sum_{i=1}^{n}(y_{i1} - \exp(x'_{i1}b_1),\ y_{i2} - \exp(x'_{i2}b_2))\hat{V}_i^{-1}\begin{pmatrix} y_{i1} - \exp(x'_{i1}b_1) \\ y_{i2} - \exp(x'_{i2}b_2) \end{pmatrix} \tag{8}$$

where

$$\hat{V}_i = \begin{pmatrix} \exp(x_{i1}'b_1)(1+\hat{\sigma}_{11}) & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \exp(x_{i2}'b_2)(1+\hat{\sigma}_{22}) \end{pmatrix}$$

The asymptotic covariance of the GNLS estimator can be consistently estimated as (see, for instance, Davidson and MacKinnon, 1993)

$$\widehat{\text{Var}}(b) = \sum_{i=1}^{n}(H_i'\hat{V}_i^{-1}H_i)^{-1} \tag{9}$$

where

$$H_i = \begin{pmatrix} \exp(x_{i1}'b_1)x_{i1} & 0 \\ 0 & \exp(x_{i2}'b_2)x_{i2} \end{pmatrix}$$

V. CONCLUSION

This paper has introduced a new econometric model for correlated count data. The model has Negbin I-type marginal distributions. It generalizes the seemingly unrelated Poisson regression model as it allows for overdispersion. Under correct specification, the model parameters can be estimated by maximum likelihood and used, for instance, to predict joint or conditional probabilities.

One assumption of the model is that the dispersion parameter is the same for all counts. This assumption may be most reasonable in the context of panel data where it mimics the usual covariance structure of the linear random effects model. The assumption can be tested, and in case of rejection, the model should be estimated semi-parametrically by generalized non-linear least squares.

IZA Bonn and CEPR London

Date of Receipt of Final Manuscript: February 2000.

REFERENCES

Aptech (1994). *GAUSS Applications User Manual: Maximum Likelihood*, Aptech Systems: Maple Valley, WA.

Chib, S. and Winkelmann, R. (1999). 'Modeling and Simulation-Based Analysis of Correlated Count Data', mimeo, Washington University.

Cincera, M. (1998). 'Patents, R&D, and Technological Spillovers at the Firm Level: Some Evidence from Econometric Count Models for Panel Data', *Journal of Applied Econometrics*, Vol. 12, pp. 265–80.

Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, Wiley, New York.

Gourieroux, C., Monfort, A. and Trognon, A. (1984a). 'Pseudo Maximum Likelihood Methods: Theory', *Econometrica*, Vol. 52, pp. 681–700.

Gourieroux, C., Monfort, A. and Trognon, A. (1984b). 'Pseudo Maximum Likelihood Methods: Applications to Poisson Models', *Econometrica*, Vol. 52, pp. 701–21.

Gurmu, S. and Elder, J. (1998). 'Estimation of multivariate count regression models with applications to health care utilization', mimeo, University of Virginia.

Hausman, J., Hall, B. H. and Griliches, A. (1984). 'Econometric Models for Count Data with an Application to the Patents-R&D Relationship', *Econometrica*, Vol. 52, pp. 909–38.

Jung, R. C. and Winkelmann, R. (1993). 'Two aspects of labor mobility: a bivariate Poisson regression approach', *Empirical Economics*, Vol. 18, pp. 543–56.

King, G. (1989a). 'A seemingly unrelated Poisson regression model', *Sociological Methods and Research*, Vol. 17, pp. 235–55.

King, G. (1989b). 'Event Count Models for International Relations: Generalizations and Applications', *International Studies Quarterly* (June), Section 6.

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

Ozuna, T. and Gomez, I. A. (1994). 'Estimating a System of Recreation Demand Functions Using a Seemingly Unrelated Poisson Regression Approach', *Review of Economics and Statistics*, Vol. 76, pp. 356–60.

Self, S. G. and Liang, K. (1987). 'Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions', *Journal of the American Statistical Association*, Vol. 82, pp. 605–10.

Winkelmann, R. (2000). *Econometric Analysis of Count Data*, 3rd ed., Springer, Heidelberg.