Bivariate Count Data Regression using Series Expansions: with Applications*

A. Colin Cameron
Department of Economics
University of California
Davis, CA 95616 USA
accameron@ucdavis.edu

Per Johansson
Centre for Research on Transportation and Society (CTS)
Dalarna University
781 88 Borlänge, SWEDEN
Per.Johansson@cts-fs1.du.se

June 4, 1998

Abstract

Most research on count data regression models, i.e. models for where the dependent variable takes only non-negative integer values or count values, has focused on the univariate case. Very little attention has been given to joint modelling of two or more counts. We propose parametric regression models for bivariate counts based on squared polynomial expansions around a baseline density. The models are more flexible than the current leading bivariate count model, the bivariate Poisson. The models are applied to data on the use of prescribed and nonprescribed medications.

* Authors email: accameron@ucdavis.edu and Per.Johansson@Cts.du.se

1. Introduction

Most research on count data regression models, i.e. models for the special case where the dependent variable takes only non-negative integer values or counts, has focused on the univariate case. Very little attention has been given to joint modelling of two or more counts. A review of the literature on multivariate count models is given in Gurmu and Trivedi (1994) and Cameron and Trivedi (1998, chapter 8).

Here we focus on the bivariate case. The leading model for bivariate counts is the bivariate Poisson, proposed by Holgate (1964) and presented in Johnson and Kotz (1969, pp. 297-300) and Gourieroux, Monfort and Trognon (1984). This model is applied by King (1989) to annual number of presidential vetoes of social welfare bills and defense bills, by Jung and Winkelmann (1993) to number of voluntary and involuntary changes, and by Ozuna and Gomez (1994) to number of trips to different recreational sites.

The bivariate Poisson is restrictive in that it limits the marginals to being Poisson, so that, conditional on regressors, the variance equals the mean. This restriction is generally violated by count data, the well-known overdispersion problem. Multivariate negative binomial models are presented in Kocherlakota and Kocherlakota (1993), but these models restrict the forms of correlation and do not appear to have been applied to the regression case.

Another model for multivariate counts is the mixed multinomial-Poisson model, see Terza and Wilson (1990) and Hausman, Leonard and McFadden (1995). This model is for counts on several types of events that are mutually exclusive and collectively exhaustive, and is therefore not applicable to all settings in which multivariate counts arise.

We present parametric models for bivariate count data based on a squared polynomial expansion around a given joint count density. The approach is based on the univariate count data model of Cameron and Johansson (1997); see also Gurmu (1997), Gurmu, Rilstone and Stern (1998) and Gurmu and Elder (1998) for a related approach. These papers are in turn based on models for continuous data, using a squared Hermite series expansion from a baseline normal density, developed by Gallant and Nychka (1987) and applied to finance data in many applications, beginning with Gallant and Tauchen (1989). For continuous data using Hermite expansions the models have been extended to the multivariate case. Other multivariate orthogonal series expansions were considered by Cameron and Trivedi (1993), but they considered only hypothesis tests of departures from independence and did not square the series expansion term.

Specific models vary with choice of baseline density, which can be the product of two univariate marginal densities or a more general bivariate density. In turn these densities may, for example, be Poisson or the more general negative binomial. Further variation can be obtained by choice of functional form of the polynomial.

In section 2 we present the model and its properties, with additional details provided in an appendix. A detailed simulation analysis is presented in section 3. Application to data on the number of prescribed medicines and number of nonprescribed medicines used is presented in section 4. Concluding remarks are made in section 5.

2. Models based on squared polynomial expansions

2.1. General approach

We begin with a general presentation for any type of multivariate data and baseline density, before specializing to bivariate count data with the product of two univariate Poisson densities as the baseline density. Consider a vector random variable $\mathbf{y}$ with baseline density $f(\mathbf{y}|\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a parameter vector. Let $h(\mathbf{y}|\mathbf{a})$ denote a polynomial function of $\mathbf{y}$, where $\mathbf{a}$ denotes the coefficients of the polynomial. The density based on a squared polynomial series expansion is

$$g(\mathbf{y}|\boldsymbol{\lambda},\mathbf{a}) = f(\mathbf{y}|\boldsymbol{\lambda}) \times \frac{h^2(\mathbf{y}|\mathbf{a})}{\eta(\boldsymbol{\lambda},\mathbf{a})}, \tag{2.1}$$

where $\eta(\boldsymbol{\lambda},\mathbf{a})$ is a normalizing constant term that ensures that the density $g(y|\boldsymbol{\lambda},\mathbf{a})$ sums to unity, and squaring the polynomial ensures that the density is nonnegative. Note that for the new density $g(\mathbf{y}|\boldsymbol{\lambda},\mathbf{a})$, the r^{th} moment of $\mathbf{y}$ will generally differ from the r^{th} moment of the baseline density. In particular, this is the case for the mean.

The univariate case was studied by Cameron and Johansson (1997), who let $h(y|\mathbf{a}) = \sum_{k=0}^{p} a_k y^k$ where $a_0 = 1$. Then $\eta(\boldsymbol{\lambda},\mathbf{a}) = \sum_{k=0}^{p}\sum_{l=0}^{p} a_k a_l m_{k+l}$, where $m_r \equiv m_r(\boldsymbol{\lambda})$ denotes the r^{th} non-central moment of the baseline density $f(y|\boldsymbol{\lambda})$. The moments of the scalar random variable y with density $g(y|\boldsymbol{\lambda},\mathbf{a})$ are then readily obtained from those of the baseline density $f(y|\boldsymbol{\lambda})$ as $E[y^r] = \sum_{k=0}^{p}\sum_{l=0}^{p} a_k a_l m_{k+l+r}/\eta_p(\boldsymbol{\lambda},\mathbf{a})$.

Note that (2.1) can be generalized to $g(\mathbf{y}|\boldsymbol{\lambda},\mathbf{a}) = f(\mathbf{y}|\boldsymbol{\lambda})h^*(\mathbf{y}|\mathbf{a})/\eta(\boldsymbol{\lambda},\mathbf{a})$, where $h^*(\cdot) \geq 0$. The attraction of choosing $h^*(\mathbf{y}|\mathbf{a})$ as the square of a polynomial function is that the normalizing constant $\eta(\boldsymbol{\lambda},\mathbf{a})$ is then relatively easy to find.

2.2. A simple bivariate model

For the bivariate case, $\mathbf{y} = (y_1, y_2)$, it is more difficult than in the univariate to simply present general results. Instead we consider in detail a specific simple model. The polynomial is of the form

$$h(\mathbf{y}|\mathbf{a}) = 1 + a y_1 y_2, \tag{2.2}$$

and the baseline joint density is the product of two marginal densities

$$f(\mathbf{y}|\boldsymbol{\lambda}) = f_1(y_1|\boldsymbol{\lambda}_1) f_2(y_2|\boldsymbol{\lambda}_2). \tag{2.3}$$

Then

$$g(\mathbf{y}|\boldsymbol{\lambda},\mathbf{a}) = f_1(y_1|\boldsymbol{\lambda}_1)f_2(y_2|\boldsymbol{\lambda}_2)\cdot\frac{(1+ay_1y_2)^2}{\eta(\boldsymbol{\lambda},\mathbf{a})}. \tag{2.4}$$

In the appendix it is shown that the **normalizing constant** is

$$\eta(\boldsymbol{\lambda},\mathbf{a}) = (1+2am_{11}m_{21}+a^2m_{12}m_{22}), \tag{2.5}$$

where $m_{jr} = E_{f_j}[y_j^r]$ is the r^{th} moment of y_j with respect to the baseline density.

Interest lies in both the marginal and conditional densities. The **marginal density** of y_1 is

$$g_1(y_1|\boldsymbol{\lambda},a) = \frac{(1+2ay_1m_{21}+a^2y_1^2m_{22})}{\eta(\boldsymbol{\lambda},\mathbf{a})}f_1(y_1|\boldsymbol{\lambda}_1), \tag{2.6}$$

with **marginal mean**

$$E[y_1] = \frac{(m_{11}+2am_{12}m_{21}+a^2m_{13}m_{22})}{\eta(\boldsymbol{\lambda},\mathbf{a})}. \tag{2.7}$$

The marginal mean will generally differ from m_{11}, the mean of y_1 in the baseline density $f_1(y_1|\boldsymbol{\lambda}_1)$.

The **conditional density** of y_1 given y_2 is

$$g_{1|2}(y_1|y_2,\boldsymbol{\lambda},a) = \frac{(1+2ay_1y_2+ay_1^2y_2^2)}{(1+2ay_2m_{11}+a^2y_2^2m_{12})}f_1(y_1|\boldsymbol{\lambda}_1) \tag{2.8}$$

with **conditional mean**

$$E[y_1|y_2] = \frac{(m_{11}+2ay_2m_{12}+a^2y_1^2m_{13})}{(1+2ay_2m_{11}+a^2y_2^2m_{12})}. \tag{2.9}$$

Thus the conditional mean of y_1 given y_2 is a quite nonlinear function of y_2.

Note that if we extend to the case where the baseline density is instead given by $f_1(y_1|y_2,\boldsymbol{\lambda}_1)f_2(y_2|\boldsymbol{\lambda}_2)$ then all of the above extends with $m_{2r} = E_{f_2}[y_2^r]$ as above but now $m_{1r} = E_{f_1|f_2}[y_1^r|y_2]$ which involves conditional moments for the baseline density.

For univariate Poisson baseline densities, λ_j is a scalar and $f_j(y_j|\lambda_j)$ is given by

$$f_j(y_j|\lambda_j) = \frac{e^{-\lambda_j}\lambda^{y_j}}{y_j!}, \qquad j=1,2, \quad y_j = 0,1,2,\ldots. \tag{2.10}$$

The normalizing constant $\eta(\boldsymbol{\lambda},\mathbf{a})$ defined in (2.5) and the moments $E[y_1]$ and $E[y_1|y_2]$ defined in (2.7) and (2.9) are evaluated at the moments $m_r(\lambda)$ of the Poisson, which can be obtained from the moment generating function using $m_r(\lambda) = \partial^r \exp(-\lambda+\lambda e^t)/\partial t^r \mid_{t=0}$ whose first three moments are

$$\begin{aligned} m_1 &= \lambda \\ m_2 &= \lambda+\lambda^2 \\ m_3 &= \lambda+3\lambda^2+\lambda^3 \end{aligned}$$

2.3. More general bivariate models

More general polynomial functions than (2.2) may be considered. In particular in sections 3 and 4 we consider

$$h_1(\mathbf{y}|\mathbf{a}) = 1 + a_1 y_1 + a_2 y_2 \tag{2.11}$$

and

$$h_2(\mathbf{y}|\mathbf{a}) = 1 + a_1 y_1 + a_2 y_2 + a_{11} y_1^2 + a_{22} y_2^2 + a_{12} y_{12}. \tag{2.12}$$

In simulations and applications below we use these two polynomials, with baseline density the product of two independent Poissons. The models based on (2.11) and (2.12) are then called the Bi1 and Bi2 models.

More general baseline densities may also be considered. For overdispersed data it is natural to use univariate negative binomial baseline densities. Then $\boldsymbol{\lambda}_j = (\mu_j, \alpha)$ is a two parameter vector and $f_j(y_j|\boldsymbol{\lambda}_j)$ in (2.3) is given by

$$f(y|\mu,\alpha) = \frac{\Gamma(y+\alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y}, \quad \alpha \geq 0,\ y = 0,1,2,..., \tag{2.13}$$

where for notational simplicity the subscript j, $j = 1, 2$, is suppressed on y and μ. This form of the negative binomial is often referred to as Negbin2, as the variance is a quadratic function of the mean: $E[y] = \mu$ and $V[y] = \mu + \alpha\mu^2$. Another possibility is to use the bivariate Poisson as the baseline density.

2.4. Estimation

We consider estimation by maximum likelihood, based on a sample of independent observations $\{(\mathbf{y}_1, \mathbf{x}_1), \ \ldots, \ (\mathbf{y}_n, \mathbf{x}_n)\}$ of size n. The regressors $\mathbf{x}_i$ introduced by letting $\boldsymbol{\lambda}_i$ and/or $\mathbf{a}_i$ be a function of these regressors. In the case where only $\boldsymbol{\lambda}_i$ is parameterized, as $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}(\mathbf{x}_i, \boldsymbol{\beta})$, the log-likelihood function is

$$\ln \mathcal{L}(\boldsymbol{\beta}, \mathbf{a}) = \sum_{i=1}^{n} \{\ln f(\mathbf{y}_i|\boldsymbol{\lambda}(\mathbf{x}_i, \boldsymbol{\beta})) + 2 \ln h(\mathbf{y}_i|\mathbf{a}) - \ln \eta(\boldsymbol{\lambda}(\mathbf{x}_i, \boldsymbol{\beta}), \mathbf{a})\}, \tag{2.14}$$

which is minimized with respect to $\boldsymbol{\beta}$ and $\mathbf{a}$.

2.5. Lagrange multiplier tests

A useful way to interpret the series expansion is to consider the lagrange multiplier (LM) test of $\mathbf{a} = \mathbf{0}$. The moment condition for this test in some gives the direction

of departure from the baseline density. It is helpful to replace $h^2(\mathbf{y}|\mathbf{a})$ in (2.1) by the more general function $h^*(\mathbf{y}|\mathbf{a})$. Then the log-likelihood (2.14) becomes

$$\ln \mathcal{L}(\boldsymbol{\beta}, \mathbf{a}) = \sum_{i=1}^{n} \{\ln f(\mathbf{y}_i|\boldsymbol{\lambda}(\mathbf{x}_i, \boldsymbol{\beta})) + \ln h^*(\mathbf{y}_i|\mathbf{a}) - \ln \eta(\boldsymbol{\lambda}(\mathbf{x}_i, \boldsymbol{\beta}), \mathbf{a})\}. \tag{2.15}$$

Consider a function that is a transformation of a function linear in the weights $\mathbf{a}$, that is,

$$h^*(\mathbf{y}_i|\mathbf{a}) = h^*(1 + \mathbf{a}'\mathbf{p}(\mathbf{y}_i)),$$

where $\mathbf{p}(\mathbf{y}_i)$ is a column vector of polynomials in $\mathbf{y}_i$. In the appendix it is shown that

$$\left.\frac{\partial \ln \mathcal{L}(\boldsymbol{\beta}, \mathbf{a})}{\partial \mathbf{a}}\right|_{\mathbf{a}=\mathbf{0}} = \sum_{i=1}^{n} \mathbf{p}(\mathbf{y}_i) - E_{f_i}[\mathbf{p}(\mathbf{y}_i)],$$

aside from a scalar multiple.

Thus the LM test is a conditional moment test of whether

$$E\left[\mathbf{p}(\mathbf{y}_i) - E_{f_i}[\mathbf{p}(\mathbf{y}_i)]\right] = 0, \tag{2.16}$$

that is, whether the expected value of $\mathbf{p}(\mathbf{y}_i)$ equals that specified in the baseline density.

For example, in the simple bivariate model (2.2)-(2.3) where $\mathbf{a}'\mathbf{p}(\mathbf{y}_i) = ay_1y_2$ the LM test is a test of whether $E[y_1y_2] = E[y_1]E[y_2]$, i.e. a test of independence.

As a second example, if the baseline density is the product of Poissons, and $\mathbf{a}'\mathbf{p}(\mathbf{y}_i) = a\{(y_1 - \lambda_1)^2 - y_1\}\{(y_1 - \lambda_2)^2 - y_1\}$, we get the independence test based on the second-order orthogonal polynomial proposed by Cameron and Trivedi (1993).

3. Simulation

3.1. Definition of models

It is first helpful to define the following bivariate models.

- IP (independent Poisson): bivariate density is the product of two marginal Poisson densities defined in (2.10).
- INB (independent Poisson): bivariate density is the product of two marginal negative binomial densities, defined in (2.13), where overdispersion is parameterized as the Negbin2 form with $V[y|\mathbf{x}] = E[y|\mathbf{x}] + \alpha(E[y|\mathbf{x}])^2$.
- IPP1: bivariate density is the product of two marginal PP1 densities, each a first-order expansion with univariate baseline density a Poisson density. This is the obvious specialization of the PP3 given below.

- IPP2: bivariate density is the product of two marginal PP2 densities, each a second-order expansion with univariate baseline density a Poisson density. This is the obvious specialization of the PP3 given below.

- IPP3: bivariate density is the product of two marginal PP3 densities, each a second-order expansion with univariate baseline density a Poisson density, with

$$\begin{aligned} g(\mathbf{y}|\boldsymbol{\lambda},\mathbf{a}) &= f_1(y_1|\lambda_1)(1+a_{11}y_1+a_{12}y_1^2+a_{13}y_1^3)^2/\eta_1(\lambda_1,\mathbf{a}_1) \\ &\quad \times f_2(y_2|\lambda_2)(1+a_{21}y_2+a_{22}y_2^2+a_{23}y_2^3)^2/\eta_2(\lambda_2,\mathbf{a}_2), \end{aligned} \tag{3.1}$$

  where $\eta_1(\lambda_1,\mathbf{a}_1)$ and $\eta_2(\lambda_2,\mathbf{a}_2)$ are normalizing constants.

- BP: bivariate Poisson density as implemented by King (1986), who calls this the Supreme model, with

$$g(\mathbf{y}|\boldsymbol{\lambda},\mathbf{a}) = \exp\left[\lambda_1+\lambda_2+\lambda_3\right] \sum_{l=0}^{\min(y_1,y_2)} \frac{\lambda_1^{y_1-l}\lambda_2^{y_1-l}\lambda_3^l}{(y_1-l)!(\lambda_2-l)!l!}. \tag{3.2}$$

  This model arises if $y_1 = z_1 + z_3$, $y_2 = z_2 + z_3$ and $z_j \sim P(\lambda_j)$, $j = 1, 2, 3$.

- Bi1: first-order expansion bivariate density with baseline density a product of two marginal Poisson densities, with

$$g(\mathbf{y}|\boldsymbol{\lambda},\mathbf{a}) = f_1(y_1|\lambda_1)f_2(y_2|\lambda_2)(1+a_1y_1+a_2y_2)^2/\eta(\lambda,\mathbf{a}). \tag{3.3}$$

- Bi2: second-order expansion bivariate density with baseline density a product of two marginal Poisson densities, with

$$g(\mathbf{y}|\boldsymbol{\lambda},\mathbf{a}) = f_1(y_1|\lambda_1)f_2(y_2|\lambda_2)(1+a_1y_1+a_1y_2+a_{11}y_1^2+a_{11}y_1^2+a_{12}y_1y_2)^2/\eta(\lambda,\mathbf{a}). \tag{3.4}$$

In all these models regressors are introduced via

$$\lambda_j = \exp(\mathbf{x}'\boldsymbol{\beta}_j), \qquad j = 1, 2. \tag{3.5}$$

For the BP model, following King (1986) who called this the Supreme model, λ_3 is a scalar parameter that does not vary with $\mathbf{x}$.

The models Bi1-Bi2 and IPP1-PP3 can be grouped as polynomial models. For the polynomial models the baseline densities are univariate Poisson

$$f_j(y_j|\lambda_j) = e^{-\lambda_j}\lambda_j^{y_j}/y_j!, \qquad j = 1, 2. \tag{3.6}$$

The IP model is therefore a special case of all the other models considered.

3.2. Data generating processes

The simulations are based on d.g.p'.s with no regressor (intercept-only) and one regressor (plus intercept). The single regressor x is drawn from the uniform $[0, 1]$ distribution and held fixed in all replications. The following d.g.p.'s were considered, with sample sizes $n = 50$ and $n = 200$.

- Model 1: $y_1 \sim P(1 + x)$, $y_2 \sim P(1 + x)$. This is the product of conditionally independent Poissons.
- Model 1: $y_1 \sim P(1+1.5x)$, $y_2 \sim P(1+1.5x)$. This is the same as model 1 except different slope coefficient (see below).
- Model 3: $y_1 \sim P(1+x+\omega_1)$, $y_2 \sim P(1+x+\omega_2)$, where ω_1 and ω_2 are exponentially distributed with parameter $1/2$. Since different draws of ω are used for y_1 and y_2 there is still conditional independence between y_1 and y_2.
- Model 4: $(y_1, y_2) \sim BP(\lambda_1, \lambda_2, \lambda_3)$, where $\lambda_1 = 1 + x$, $\lambda_2 = 1 + x$, $\lambda_3 = 1$. In this and subsequent models y_1 and y_2 are conditionally dependent.
- Model 5: $(y_1, y_2) \sim BP(\lambda_1, \lambda_2, \lambda_3)$, where $\lambda_1 = 1 + x$, $\lambda_2 = 1 + x$, $\lambda_3 = 1.25$. This is the same as model 1 except different parameter value for λ_3.
- Model 6: $y_1 \sim P(1 + x + \omega)$, $y_2 \sim P(1 + x + \omega)$, where $\omega = \kappa(\eta^2 - 1)$ and $\eta \sim N(0, 1)$. For $n = 50$, $\kappa = 0.25$, and for $n = 200$, $\kappa = 0.15$.
- Model 7: $y_1 \sim P(1+x+\omega)$, $y_2 \sim P(1+x+\omega)$, where $\omega = \kappa\eta$ and $\eta \sim$Uniform$[0, 1]$. For $n = 50$, $\kappa = 1.0$, and for $n = 200$, $\kappa = 0.70$.
- Model 8: $y_1 \sim P(1+x+\omega_1)$, $y_2 \sim P(1+x+\omega_2)$, where $\omega_1 = \kappa(\eta_1 + \eta_2 - 1)$ and $\omega_2 = -\kappa(\eta_1 + \eta_2)$, with $\eta_1 \sim$Uniform$[0, 1]$ and $\eta_2 \sim$Uniform$[0, 1]$. For $n = 50$, $\kappa = 3.0$, and for $n = 200$, $\kappa = 1.75$.

These d.g.p.'s were used by Cameron and Trivedi (1993) in the context of establishing the size and power of tests of independence based on estimation of univariate Poisson models. Here we consider estimation of more general bivariate models.

In addition we use as d.g.p. the Bi2 model.

- Bi2: The Bi2 model, defined by (3.4)−(3.6), for different values of the means (E_1 and E_2) the variance-mean ratio (R_1 and R_2 taking values 0.7, 1.0 and 2.0) and the correlation parameter ($\rho = -0.15$, 0.0, 0.15).

The following summary illustrates the ability of the new distribution to explain a wide range of data behavior.

Moment	Values	Explanation
$\rho = corr(y_1, y_2)$	$-0.15, 0.0, .15$	Correlation
$E_1 = E[y_1]$	0.5	Marginal mean for y_1
$R_1 = V[y_1]/E[y_1]$	$0.7, 1.0, 2.0$	Marginal variance-mean ratio for y_1
$E_2 = E[y_2]$	1.0	Marginal mean for y_2
$R_2 = V[y_2]/E[y_2]$	$0.7, 1.0, 2.0$	Marginal variance-mean ratio for y_2

When there are no regressors, so $\lambda_j = \exp(\mathbf{x}'\boldsymbol{\beta}_j) = \exp(\beta_{j0})$, $j = 1, 2$, values of $\mathbf{a} = (a_1, a_2, a_{11}, a_{22}, a_{12})$ and $\boldsymbol{\beta}_0 = (\beta_{10}, \beta_{20})$ are chosen to correspond to the specified values of E_1, E_2, R_1, R_2, ρ. Two normalizations are needed, usually a_{11} and a_{22}. The 27 different possibilities considered are summarized in Table 1.

When a scalar regressor is introduced, $\exp(\mathbf{x}'\boldsymbol{\beta}_j) = \exp(\beta_{j0} + \beta_{j1}x)$, $j = 1, 2$. Then the same values of $\mathbf{a}$ and $\boldsymbol{\beta}_0$ are used, and $\boldsymbol{\beta}_0 = (\beta_{10}, \beta_{20}) = (0.5, 0.5)$. For each n we perform $54 = 3 \times 9 \times 2$ experiments: for each of the three ρ values, nine experiments are performed, and simulations are done with and without regressors.

Some more detailed analysis is performed of model 6, with different parameter values. In particular both y_1 and y_2 are drawn from $P(\beta_0 + \beta_1 x + \omega)$, where $\beta_0 = 1$ and $\beta_1 = 0.0$, 0.3, 0.6 or 1, and ω has the same distribution as in model 6. There are four different d.g.p.'s corresponding to the different values of β_1.

3.3. Results of simulation

The models are estimated using analytical derivatives and the BFGS iterative method. Summary of results to come.

4. Application

We consider count data from a sample of single adults of size 5190 from the Australian Health Survey 1977-78. Several measures of health service utilization such as doctor visits, days in hospital and number of medicines taken were modelled in Cameron, Trivedi, Milne and Piggott (1988). In their analysis these counts were separately modelled by univariate negative binomial regression.

Here we consider joint modelling of pairs of such measures. In section 4.1 we model the number of prescribed medicines (PRESC) and non-prescribed (NONPRESC) medications used in the past two days.

The dependent and regressor variables are defined, and summary statistics given, in Table 2. Regressors can be grouped into four categories: (1) socioeconomic: SEX, AGE, AGESQ, INCOME; (2) health insurance status indicators: LEVYPLUS, FREEPOOR, FREEOTHER with LEVY (government Medibank health insurance) the omitted category; (3) recent health status measures: ILLNESS, ACTDAYS; and (4) long-term health status measures: HSCORE, CHCOND1, CHCOND2.

4.1. Prescribed and nonprescribed medicines

In this section we model the number of prescribed medicines (PRESC) and nonprescribed (NONPRESC) medications used in the past two days. The sample means and variances, given in Table 2, reveal sample means less than 1, and modest amounts of overdispersion with variance-mean ratios of, respectively, 2.32 and 1.42. The actual counts are given in Table 3. The two series are negatively correlated, with sample correlation coefficient equal to -0.044.

All the regression models in this section include as regressors the twelve variables summarized in Table 2, plus an intercept.

The standard modelling approach for these data is to estimate separate negative binomial models (INB) for PRESC and NONPRESC. The INB results are given in Table 5. To interpret coefficients note that for this model $E[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta})$ so $\partial E[y|\mathbf{x}]/\partial \mathbf{x} = \boldsymbol{\beta} E[y|\mathbf{x}]$. The coefficient β_j therefore is a semi-elasticity, giving the percentage change in $E[y|\mathbf{x}]$ due to a one-unit change in x_j. The last column of Table reports $\frac{1}{n}\sum_{i=1}^{n} \widehat{\beta}_j \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})$, the sample average effect on the conditional mean of a one-unit change in x_j.

The coefficients in Table 5 indicate that medicine usage is higher for women, increases with age (at a rate that is decreasing with age), and varies little with income. Compared to the reference group of people who get the default government-provided health insurance by paying an income levy, people who receive government coverage through low income have essentially the same level of use of medicines, people who receive government coverage through old-age, disability or veteran status use more prescribed medicines and fewer nonprescribed medicines, and people with private health insurance cover use more prescribed medicines and a similar number of nonprescribed medicines. Recent illness increases substantially use of both prescribed and nonprescribed medicines. A higher value of HSCORE, indicating lower self-assessed health status, leads to a small increase in use of medicines. Days of reduced activity and chronic conditions have great impact on the use of prescribed medicines but little impact on the use of nonprescribed medicines.

The overdispersion parameter (α) is highly statistically significant in both regressions, indicating the need to control for overdispersion and rejection of the Poisson model. The model explains prescribed medicines much better than nonprescribed medicines, with deviance R-squared (based on Poisson regression) of 0.41 for PRESCR and 0.06 for NONPRESC. Finally, inclusion of regressors increases the magnitude of the negative correlation between PRESCR and NONPRESC from -0.043 to -0.093.

In Table 6 results are given when the Bi2 model is estimated. This is a bivariate model based on a series expansion around a baseline density that is the product of two Poissons. The statistical significance of the coefficients is quite similar to that for the INB model, with similar qualitative conclusions to those already given for INB.

The models IP, INB, IPP1, IPP2, IPP3, BP, Bi1 and Bi2, defined in section 3.1, were all estimated. The first five models separately model PRESCR and NONPRESC,

while the last three models are bivariate models. Various information criteria to discriminate between the models are given in Table 4.

The IP model is nested within all models. The IPP1 model is nested within IPP2 and IPP3 and the IPP2 is nested within the PP2. The Bi1 model is nested within the Bi2 model. Using likelihood ratio tests leads to rejection of the IP model, and favoring the IPP2 model within the IPP class and the Bi2 model within the Biclass.

To compare non-nested models we use the log-likelihood with a penalty for the number of parameters that need to be estimated. These measures are the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the consistent Akaike information criterion (CAIC). They are defined by

$$\begin{aligned} \text{AIC} &= -2\ln \mathsf{L} + k \\ \text{BIC} &= -2\ln \mathsf{L} + k\ln n \\ \text{CAIC} &= -2\ln \mathsf{L} + (1+\ln n)\,k, \end{aligned}$$

where k is the number of parameters in the model and n is the number of observations. Models with low values for the information criteria are preferred.

Among the first six models the INB model is preferred using AIC, while the IPP3 model is preferred using BIC and CAIC, which give a larger penalty for additional parameters. Among the last three models the BP is preferred to Bi1, while the Bi2 is strongly preferred to BP. Across all models the Bi2 model is preferred.

In addition the models are compared by their ability to predict the empirical probability distribution of the count dependent variables. Let $Pr[y = j|\mathbf{x}] = f_j(\mathbf{x}, \boldsymbol{\theta})$, the conditional probability that the count equals j. Then Table 7 reports $\frac{1}{n}\sum_{i=1}^{n} f_j(\mathbf{x}_i, \widehat{\boldsymbol{\theta}})$, the sample average predicted probability that $y = j$. It is clear that IP does a poor job, confirming the need to model overdispersion. Both INB and Bi2 do quite well, with Bi2 performing a little better especially for $\Pr[PRESCR = 1]$.

Turning to the joint distribution, let $Pr[y_1 = j, y_2 = k|\mathbf{x}] = f_{jk}(\mathbf{x}, \boldsymbol{\theta})$, the conditional probability that the two counts equal j and k. Then Table 8 reports $\frac{1}{n}\sum_{i=1}^{n} f_{jk}(\mathbf{x}_i, \widehat{\boldsymbol{\theta}})$, the sample average predicted probability that $y_1 = j$ and $y_2 = k$. Again it is clear that IP does a poor job, and Bi2 performs a little better than INB.

5. Concluding Remarks

New parametric bivariate models for counts have been proposed. The simulations and application indicate the feasibility of using these models. In the application considerable gain comes from appropriately modelling overdispersion by any method, such as the negative binomial. This is better than using the bivariate Poisson model. Better still is to use the new model proposed, which is capable of simultaneously modelling overdispersion and correlation between the two variables of interest.

References

Cameron, A. C. and P. Johansson (1997), "Count Data Regressions using Series Expansions: with Applications", *Journal of Applied Econometrics,* 12, 203-223.

Cameron, A. C. and P. K. Trivedi (1993), "Tests of Independence in Parametric Models: With Applications and Illustrations", *Journal of Business and Economic Statistics*, 11, 29-43.

Cameron, A.C. and P.K. Trivedi (1998), *Regression Analysis of Count Data*, Econometric Society Monograph No. 30, Cambridge University Press, due to be released Summer 1998.

Cameron, A.C., P.K. Trivedi, F. Milne and J. Piggott (1988), "A Microeconometric Model of the Demand for Health Insurance and Health Care in Australia", *Review of Economic Studies*, 55, 85-106.

Gallant, A.R. and D.W. Nychka (1987), "Seminonparametric Maximum Likelihood Estimation", *Econometrica*, 55, 363-390.

Gallant, A.R. and G. Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Apllications", *Econometrica*, 57, 1091-1120.

Goffe, W. L., Ferrier, G.D. and J. Rogers (1994), "Global Optimization of Statistical Functions with Simulated Annealing", *Journal of Econometrics*, 60, 65-99.

Gourieroux, C., A. Monfort and A. Trognon (1984), "Pseudo Maximum Likelihood Methods: Applications to Poisson Models", *Econometrica*, 52, 681-700.

Gurmu, S. (1997), "Semiparametric Estimation of Hurdle Regression Models with an Application to Medicaid Utilization", *Journal of Applied Econometrics*, 12, 225-242.

Gurmu, S. and J. Elder (1998), "Estimation of Multivariate Count Regression Models: Applications to Health Care Utilization", Department of Economics, University of Virginia.

Gurmu, S., P. Rilstone and S. Stern (1998), "Semiparametric Estimation of Count Regression Models", *Journal of Econometrics*, forthcoming.

Gurmu, S. and P.K. Trivedi (1994), "Recent Developments in Models of Event Counts: A Survey", Discussion Paper No.261, Thomas Jefferson Center, University of Virginia, Charlottesville.

Hausman, J.A., G. K. Leonard and D. McFadden (1995), "A utility-consistent, combined discrete choice and count data model: Assessing Recreational Use Losses due to Natural Resource Damage", *Journal of Public Economics*, 56, 1-30.

Holgate, P. (1964), "Estimation for the Bivariate Poisson Distribution", *Biometrika*,

51, 241-244.

Johnson, N.L. and S. Kotz (1969), *Discrete Distributions*, Boston, MA: Houghton Mifflin.

Jung, C. J., and R. Winkelmann (1993), "Two Apects 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,* 33, 17, 235-255.

Kocherlakota, S. and K. Kocherlakota (1993), *Bivariate Discrete Distributions,* New York: Marcel Dekker.

Ozuna, T. and I.A. Gomez (1994), "Estimating a System of Recreational demand Functions using a Seemingly Unrelated Poisson Regression Approach", *The Review of Economics and Statistics*, 76, 356-360.

Terza, J.V. and P.W. Wilson (1990), "Analyzing Frequencies of Several Types of Events: A Mixed Multinomial-Poisson Approach", *The Review of Economics and Statistics*, 72, 108-115.

Appendix A: Derivation of Results

Derivation of Section 2.2 Results

While proof is for the discrete case, the same result holds for the continuous case. For the **normalizing constant**

$$\begin{aligned}
\eta(\boldsymbol{\lambda}, \mathbf{a}) &= \textstyle\sum_{y_1} \sum_{y_2} (1 + a y_1 y_2)^2 f_1(y_1|\boldsymbol{\lambda}_1) f_2(y_2|\boldsymbol{\lambda}_2) \\
&= \textstyle\sum_{y_1} \left(\sum_{y_2} (1 + 2a y_1 y_2 + a_1^2 a_2^2) f_2(y_2|\boldsymbol{\lambda}_2) \right) f_1(y_1|\boldsymbol{\lambda}_1) \\
&= \textstyle\sum_{y_1} (1 + 2a y_1 m_{21} + a^2 y_1^2 m_{22}) f_1(y_1|\boldsymbol{\lambda}_1) \\
&= (1 + 2a m_{11} m_{21} + a^2 m_{11} m_{22}),
\end{aligned}$$

where $m_{jr} = E_{f_j}[y_j^r]$ is the r^{th} moment of y_j with respect to the baseline density. Note that if we extend to the case where the baseline density is instead given by $f_1(y_1|y_2, \boldsymbol{\lambda}_1) f_2(y_2|\boldsymbol{\lambda}_2)$ then the above extends with $m_{2r} = E_{f_2}[y_2^r]$ but now $m_{1r} = E_{f_1|f_2}[y_1^r|y_2]$ which involves conditional moments for the baseline density.

The **marginal density** of y_1 is

$$\begin{aligned}
g_1(y_1|\boldsymbol{\lambda}, a) &= \textstyle\sum_{y_2} g(\mathbf{y}|\boldsymbol{\lambda}, a) \\
&= \textstyle\sum_{y_2} (1 + 2a y_1 y_2 + a^2 y_1^2 y_2^2) f_1(y_1|\lambda_1) f_2(y_2|\lambda_2) / \eta(\boldsymbol{\lambda}, \mathbf{a}) \\
&= (1 + 2a y_1 m_{21} + a^2 y_1^2 m_{22}) f_1(y_1|\lambda_1) / \eta(\boldsymbol{\lambda}, \mathbf{a})
\end{aligned}$$

The r^{th} **moment** of y_1 is

$$\begin{aligned}
E[y_1^r] &= \textstyle\sum_{y_1} y_1^r g_1(y_1|\boldsymbol{\lambda}, a) \\
&= \textstyle\sum_{y_1} y_1^r (1 + 2a y_1 m_{21} + a^2 y_1^2 m_{22}) f_1(y_1|\lambda_1) / \eta(\boldsymbol{\lambda}, \mathbf{a}) \\
&= \textstyle\sum_{y_1} (y_1^r + 2a y_1^{r+1} m_{21} + a^2 y_1^{r+2} m_{22}) f_1(y_1|\lambda_1) / \eta(\boldsymbol{\lambda}, \mathbf{a}) \\
&= (m_{1r} + 2a m_{1,r+1} m_{21} + a^2 m_{1,r+2} m_{22}) / \eta(\boldsymbol{\lambda}, \mathbf{a}).
\end{aligned}$$

The **conditional density** of y_1 given y_2 is

$$\begin{aligned}
g_{1|2}(y_1|y_2, \boldsymbol{\lambda}, a) &= g(y_1, y_2|\boldsymbol{\lambda}, a) / g_2(y_2|\boldsymbol{\lambda}, a) \\
&= \frac{(1+2a y_1 y_2 + a y_1^2 y_2^2) f_1(y_1|\boldsymbol{\lambda}_1) f_2(y_2|\boldsymbol{\lambda}_2)) / \eta(\boldsymbol{\lambda}, \mathbf{a})}{(1+2a y_2 m_{11} + a^2 y_2^2 m_{12}) f(y_2|\lambda_2) / \eta(\boldsymbol{\lambda}, \mathbf{a})} \\
&= \frac{(1 + 2a y_1 y_2 + a_1^2 a_2^2)}{(1 + 2a y_2 m_{11} + a^2 y_2^2 m_{12})} f_1(y_1|\boldsymbol{\lambda}_1).
\end{aligned}$$

The r^{th} **conditional moment** of y_1 given y_2 is

$$\begin{aligned} E[y_1^r|y_2] &= \textstyle\sum_{y_1} y_1^r g_{1|2}(y_1|y_2,\boldsymbol{\lambda},a) \\ &= \textstyle\sum_{y_1} \dfrac{(y_1^r + 2ay_1^{r+1}y_2 + a^2y_1^{r+2}y_2)}{(1+2ay_2m_{11}+a^2y_2^2m_{12})} f_1(y_1|\lambda_1) \\ &= \dfrac{(m_{1r}+2ay_2m_{1,r+1}+a^2y_2^2m_{1,r+2}^2)}{(1+2ay_2m_{11}+a^2y_2^2m_{12})}. \end{aligned}$$

Derivation of Section 2.5 Results

Differentiating (2.15) with respect to $\mathbf{a}$ yields

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{\beta},\mathbf{a})}{\partial \mathbf{a}} = \sum_{i=1}^{n}\{\frac{\partial h^*(\mathbf{y}_i|\mathbf{a})/\partial \mathbf{a}}{h^*(\mathbf{y}_i|\mathbf{a})} - \frac{\partial \eta_i(\boldsymbol{\beta},\mathbf{a})/\partial \mathbf{a}}{\eta_i(\boldsymbol{\beta},\mathbf{a})}\}.$$

Note that by construction the normalizing constant

$$\eta_i(\boldsymbol{\beta},\mathbf{a}) = E_{f_i}[h^*(\mathbf{y}_i|\mathbf{a})].$$

When we consider a function that is a transformation of a function linear in the weights $\mathbf{a}$, that is

$$h^*(\mathbf{y}_i|\mathbf{a}) = h^*(1+\mathbf{a}'\mathbf{p}(\mathbf{y}_i)),$$

where $\mathbf{p}(\mathbf{y}_i)$ is a column vector of polynomials in $\mathbf{y}_i$, it follows that

$$\frac{\partial h^*(\mathbf{y}_i|\mathbf{a})}{\partial \mathbf{a}} = h^{*\prime}(1+\mathbf{a}'\mathbf{p}(\mathbf{y}_i))\mathbf{p}(\mathbf{y}_i),$$

where $h^{*\prime}(\cdot)$ denotes the first derivative, and

$$\frac{\partial \eta_i(\boldsymbol{\beta},\mathbf{a})}{\partial \mathbf{a}} = \frac{\partial E_{f_i}[h^*(\mathbf{y}_i|\mathbf{a})].}{\partial \mathbf{a}} = E_{f_i}\left[\frac{\partial h^*(\mathbf{y}_i|\mathbf{a})}{\partial \mathbf{a}}\right] = E_{f_i}[h^{*\prime}(1+\mathbf{a}'\mathbf{p}(\mathbf{y}_i))\mathbf{p}(\mathbf{y}_i)].$$

Therefore

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{\beta},\mathbf{a})}{\partial \mathbf{a}} = \sum_{i=1}^{n}\{\frac{h^{*\prime}(1+\mathbf{a}'\mathbf{p}(\mathbf{y}_i))\mathbf{p}(\mathbf{y}_i)}{h^*(1+\mathbf{a}'\mathbf{p}(\mathbf{y}_i))} - \frac{E_{f_i}[h^{*\prime}(1+\mathbf{a}'\mathbf{p}(\mathbf{y}_i))\mathbf{p}(\mathbf{y}_i)]}{E_{f_i}[h^*(1+\mathbf{a}'\mathbf{p}(\mathbf{y}_i)]}\}.$$

Setting $\mathbf{a}=\mathbf{0}$ and simplifying yields

$$\left.\frac{\partial \ln \mathcal{L}(\boldsymbol{\beta},\mathbf{a})}{\partial \mathbf{a}}\right|_{\mathbf{a}=\mathbf{0}} = \frac{h^{*\prime}(1)}{h^*(1)} \cdot \sum_{i=1}^{n}\{\mathbf{p}(\mathbf{y}_i) - E_{f_i}[\mathbf{p}(\mathbf{y}_i)]\}.$$

Table 1: Bi2 model: correlation (ρ), dispersion ratios (R) and means (E) for various parameter values

Negative correlation: $\rho = -0.15$									
	E_1, E_2	R_1	λ_1	λ_2	a_1	a_2	a_{11}	a_{22}	a_{12}
$R_2 = 2$	.5, 1	.7	.081896	.439508	1.96028	−.353101	.25	−.5	−2.72337
		1	.180432	.835089	.510999	−612837	.25	.25	−.183852
		2	.650376	.725245	−.636298	−.499630	.25	.25	−.153765
$R_2 = 1$	.5, 1	.7	.055615	.448635	2.70949	.293915	.50	.25	1.09250
		1	.152824	.336132	.550754	.225607	.50	.50	.454137
		2	.383783	.682762	−.161606	.251300	-.50	.00	.238102
$R_2 = 0.7$	.5, 1	.7	.063032	.277207	3.08855	1.52426	.25	.25	3.09575
		1	.206324	.228654	.656464	1.47452	.25	.50	.655819
		2	.600454	.198941	−.268590	1.37489	−.25	.50	−.125128

Positive correlation: $\rho = 0.15$									
	E_1, E_2	R_1	λ_1	λ_2	a_1	a_2	a_{11}	a_{22}	a_{12}
$R_2 = 2$	.5, 1	.7	.0569265	.873774	1.28563	−.814245	.25	.25	.583622
		1	.162931	.805859	.0778445	−.742451	.25	.25	.360745
		2	.503011	.704533	−.684899	−.628238	.25	.25	.146498
$R_2 = 1$	.5, 1	.7	.318065	.839209	.834059	.0177319	−.50	.00	.230282
		1	.484645	1.25856	.267565	−.151779	−.25	.00	.115651
		2	.269653	1.25613	−.904233	−.131972	-.25	.00	.0531537
$R_2 = 0.7$	.5, 1	.7	.0575247	.268306	1.25571	.736058	.25	.25	4.89590
		1	.351952	.652238	−.0785113	1.18105	.00	−.50	.447888
		2	.259259	.411630	−.978133	1.20811	−.25	−.25	−1.64842

No correlation: $\rho = 0$									
	E_1, E_2	R_1	λ_1	λ_2	a_1	a_2	a_{11}	a_{22}	a_{12}
$R_2 = 2$	.5, 1	.7	.080993	.979559	3.68552	.442750	−.50	−.25	−2.24328
		1	.50	.366611	.00	.606779	.00	−.915115	.00
		2	.256516	.852704	−.606442	.225172	−.50	−.25	.00
$R_2 = 1$	.5, 1	.7	.304786	1.00	1.12114	.00	−.521258	.00	.00
		1	.50	1.00	.00	.00	.00	.00	.00
		2	.26787	1.00	.808913	.00	−.954750	.00	.00
$R_2 = 0.7$	.5, 1	.7	.143960	.706698	1.96595	1.32805	−.50	−5	.473791
		1	.50	.727695	.00	1.18664	.00	−.433797	.00
		2	.483804	.534965	−.904283	1.05825	.50	.00	.103759

Table 2: Health data: summary statistics for dependent and regressor variables

Variable	Definition	Mean	St.Dev.
PRESC	# of prescribed medications used in past 2 days	.863	1.415
NONPRESC	# of nonprescribed medications used in past 2 days	.356	.712
DOCTORCO	# of consultations with a doctor or speacialist in past 2 weeks	.302	.798
NONDOCCO	# of consultations with non-doctor health professionals (chemist, optician, physiotherapist, social worker, chiropodist, district community nurse or chiropractor) in past 2 weeks	.215	.965
SEX	= 1 if female	.521	.500
AGE	age in years divided by 100	.406	.205
AGESQ	AGE squared	.207	.186
INCOME	annual income in hundreds of dollars	.583	.369
LEVYPLUS	= 1 if private health insurance	.443	.497
FREEPOOR	= 1 if free government health insurance due to low income	.043	.202
FREEREPA	= 1 if free government health insurance due to old age, disability or veteran status	.210	.408
ILLNESS	number of illnesses in past two weeks	1.432	1.384
ACTDAYS	number of days of reduced activity in past two weeks due to illness or injury	.862	2.888
HSCORE	general health questionnaire score using Goldberg's method	1.218	2.124
CHCOND1	= 1 if chronic condition not limiting activity	.403	.491
CHCOND2	= 1 if chronic condition limiting activity	.117	.321

Table 3: Health data: actual frequency distributions (n = 5190)

Count	0	1	2	3	4	5	6	7	8	9
PRESCR	3085	994	509	276	157	80	40	23	26	0
NONPRESC	3814	1055	228	64	15	7	2	4	1	0

Table 4: Prescribed and nonprescribed medicines: information criteria for IP, INB, IPP1, IPP2, IPP3, BP, Bi1 and Bi2 models

Criteria	IP	INB	IPP1	IPP2	IPP3	BP	Bi1	Bi2
-2 ln L	19083.7	18746.6	19083.7	18722.0	18722.0	19067.8	19083.7	18709.4
AIC	19109.7	18774.6	19111.7	18752.0	18754.0	19094.8	19111.7	18740.4
BIC	19306.2	18986.1	19323.3	18978.6	18995.7	19298.8	19323.3	18974.6
CAIC	19332.2	19014.1	19351.3	19008.6	19027.7	19325.8	19351.3	19005.6

Table 5: Prescribed and nonprescribed medicines: parameter estimates, standard errors, t-statistics, and sample average effect on conditional mean of a one unit change in regressor, for independent negative binomial (INB) model

(1) Prescribed medicines				
Variable	Coefficient	Standard error	t-statistic	dE[y|x]/dx
ONE	-2.7282	0.1477	-18.48	
SEX	0.5509	0.0425	12.97	0.482
AGE	2.1621	0.7317	2.95	1.890
AGESQ	-0.3518	0.7798	-0.45	-0.308
INCOME	0.0306	0.0649	0.47	0.027
LEVYPLUS	0.2666	0.0577	4.62	0.233
FREEPOOR	-0.0541	0.1357	-0.40	-0.047
FREEREPA	0.2916	0.0689	4.24	0.255
ILLNESS	0.2069	0.0132	15.65	0.181
ACTDAYS	0.0344	0.0050	6.89	0.030
HSCORE	0.0203	0.0081	2.71	0.019
CHCOND1	0.7725	0.0505	15.30	0.675
CHCOND2	1.0201	0.0620	16.46	0.892
α	0.2981	0.0307	9.70	

(2) Nonprescribed medicines				
Variable	Coefficient	Standard error	t-statistic	dE[y|x]/dx
ONE	-2.3234	0.1928	-12.05	
SEX	0.2494	0.0579	4.30	0.089
AGE	4.7731	1.0787	4.43	1.700
AGESQ	-6.0050	1.2157	-4.94	-2.139
INCOME	0.1086	0.0855	1.27	0.039
LEVYPLUS	-0.0424	0.0656	-0.65	-0.015
FREEPOOR	-0.0200	0.1403	-0.14	-0.007
FREEREPA	-0.2872	0.1044	-2.75	-0.102
ILLNESS	0.2071	0.0208	9.94	0.074
ACTDAYS	0.0038	0.0091	0.42	0.001
HSCORE	0.0293	0.0123	2.38	0.010
CHCOND1	0.1565	0.0631	2.48	0.056
CHCOND2	-0.0028	0.0941	-0.03	-0.001
α	0.7390	0.0827	8.93	

Table 6: Prescribed and nonprescribed medicines: parameter estimates, standard errors and t-statistics from Bi2 model

Variable	Coefficient	Standard error	t-statistic
ONE	-2.4532	0.1336	-18.36
SEX	0.3894	0.0332	11.72
AGE	2.1944	0.5476	4.01
AGESQ	-0.8933	0.5717	-1.56
INCOME	0.0102	0.0489	0.21
LEVYPLUS	0.2196	0.0466	4.71
FREEPOOR	-0.0363	0.1109	-0.33
FREEREPA	0.2144	0.0531	4.04
ILLNESS	0.1632	0.0102	15.94
ACTDAYS	0.0242	0.0034	7.04
HSCORE	0.0175	0.0057	3.06
CHCOND1	0.6146	0.0438	14.02
CHCOND2	0.7931	0.0508	15.60
ONE	-2.2932	0.1999	-11.47
SEX	0.2396	0.0465	5.16
AGE	3.8696	0.8521	4.54
AGESQ	-4.6244	0.9688	-4.77
INCOME	0.0912	0.0678	1.35
LEVYPLUS	-0.0018	0.0514	-0.03
FREEPOOR	-0.0220	0.1112	-0.19
FREEREPA	-0.1925	0.0835	-2.31
ILLNESS	0.1871	0.0170	11.01
ACTDAYS	0.0077	0.0070	1.09
HSCORE	0.0240	0.0093	2.59
CHCOND1	0.1856	0.0507	3.67
CHCOND2	0.1074	0.0747	1.44
a1	-0.1689	0.0265	-6.37
a11	0.1237	0.0172	7.18
a2	-0.1509	0.0492	-3.07
a22	0.1375	0.0224	6.14
a12	-0.1135	0.0185	-6.15

Table 7: Prescribed and nonprescribed medicines: actual probabilities and predicted probabilities from Independent Poisson, Independent negative binomial and Bi2 models

(1) Prescribed medicines				
Count	Actual	IP	INB	Bi2
0	0.5944	0.5491	0.5805	0.5901
1	0.1915	0.2435	0.2233	0.2040
2	0.0981	0.1057	0.0927	0.0892
3	0.0532	0.0507	0.0449	0.0511
4	0.0302	0.0253	0.0239	0.0310
5	0.0154	0.0127	0.0134	0.0177
6	0.0077	0.0064	0.0079	0.0092
7	0.0044	0.0032	0.0048	0.0044
8	0.0050	0.0016	0.0030	0.0019

(2) Nonprescribed medicines				
Count	Actual	IP	INB	Bi2
0	0.7349	0.7085	0.7377	0.7335
1	0.2033	0.2369	0.1948	0.1993
2	0.0439	0.0463	0.0492	0.0502
3	0.0123	0.0072	0.0130	0.0130
4	0.0029	0.0010	0.0037	0.0032
5	0.0013	0.0000	0.0011	0.0007
6	0.0004	0.0000	0.0004	0.0001
7	0.0008	0.0000	0.0001	0.0000
8	0.0002	0.0000	0.0000	0.0000

Table 8: Prescribed and nonprescribed medicines: Joint probabilites (1) actual; and (2) predicted from independent Poissons (IP model), (3) predicted from independent Poissons (INB model), (4) predicted from Bi2 model

(1) Actual joint probabilities								
Prescrib \ Nonprescrib	$y_2 = 0$	$y_2 = 1$	$y_2 = 2$	$y_2 = 3$	$y_2 = 4$	$y_2 = 5$	$y_2 = 6$	$y_2 = 7$
$y_1 = 0$	0.4295	0.1283	0.0266	0.0071	0.0013	0.0008	0.0004	0.0002
$y_1 = 1$	0.1393	0.0389	0.0094	0.0023	0.0006	0.0004	0.0000	0.0006
$y_1 = 2$	0.0738	0.0183	0.0033	0.0017	0.0008	0.0002	0.0000	0.0000
$y_1 = 3$	0.0408	0.0098	0.0021	0.0004	0.0000	0.0000	0.0000	0.0000
$y_1 = 4$	0.0252	0.0027	0.0017	0.0004	0.0002	0.0000	0.0000	0.0000
$y_1 = 5$	0.0119	0.0025	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000
$y_1 = 6$	0.0060	0.0015	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
$y_1 = 7$	0.0033	0.0012	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$y_1 = 8$	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

(2) Independent Poissons (IP) for joint distribution						
Prescrib \ Nonprescrib	$y_2 = 0$	$y_2 = 1$	$y_2 = 2$	$y_2 = 3$	$y_2 = 4$	$y_2 = 5$
$y_1 = 0$	0.3978	0.1255	0.0224	0.0031	0.0004	0.0000
$y_1 = 1$	0.1709	0.0586	0.0119	0.0019	0.0003	0.0000
$y_1 = 2$	0.0725	0.0262	0.0057	0.0010	0.0002	0.0000
$y_1 = 3$	0.0342	0.0129	0.0029	0.0005	0.0001	0.0000
$y_1 = 4$	0.0168	0.0066	0.0016	0.0003	0.0000	0.0000
$y_1 = 5$	0.0082	0.0034	0.0008	0.0002	0.0000	0.0000
$y_1 = 6$	0.0041	0.0018	0.0004	0.0001	0.0000	0.0000
$y_1 = 7$	0.0020	0.0009	0.0002	0.0000	0.0000	0.0000

(3) Independent negative binomials (INB) for joint distribution						
Prescrib \ Nonprescrib	$y_2 = 0$	$y_2 = 1$	$y_2 = 2$	$y_2 = 3$	$y_2 = 4$	$y_2 = 5$
$y_1 = 0$	0.4352	0.1102	0.0263	0.0065	0.0017	0.0005
$y_1 = 1$	0.1633	0.0441	0.0115	0.0031	0.0009	0.0003
$y_1 = 2$	0.0667	0.0187	0.0051	0.0015	0.0005	0.0002
$y_1 = 3$	0.0319	0.0092	0.0026	0.0008	0.0002	0.0001
$y_1 = 4$	0.0168	0.0050	0.0014	0.0004	0.0001	0.0000
$y_2 = 5$	0.0094	0.0029	0.0008	0.0003	0.0001	0.0000
$y_2 = 6$	0.0054	0.0017	0.0005	0.0002	0.0001	0.0000
$y_2 = 7$	0.0033	0.0010	0.0003	0.0001	0.0000	0.0000

(4) Bi2 model for joint distribution						
Prescrib \ Nonprescrib	$y_2 = 0$	$y_2 = 1$	$y_2 = 2$	$y_2 = 3$	$y_2 = 4$	$y_2 = 5$
$y_1 = 0$	0.4259	0.1219	0.0320	0.0081	0.0018	0.0004
$y_1 = 1$	0.1524	0.0378	0.0099	0.0029	0.0008	0.0002
$y_1 = 2$	0.0686	0.0157	0.0035	0.0010	0.0003	0.0001
$y_1 = 3$	0.0392	0.0094	0.0019	0.0004	0.0001	0.0000
$y_1 = 4$	0.0232	0.0063	0.0012	0.0002	0.0001	0.0000
$y_2 = 5$	0.0128	0.0039	0.0008	0.0001	0.0000	0.0000
$y_2 = 6$	0.0065	0.0022	0.0005	0.0001	0.0000	0.0000
$y_2 = 7$	0.0030	0.0011	0.0003	0.0000	0.0000	0.0000