Functional Form and Heterogeneity in Models for Count Data

William Greene*
Department of Economics, Stern School of Business,
New York University,
April, 2007

Abstract

This study presents several extensions of the most familiar models for count data, the Poisson and negative binomial models. We develop an encompassing model for two well known variants of the negative binomial model (the NB1 and NB2 forms). We then propose some alternative approaches to the standard log gamma model for introducing heterogeneity into the loglinear conditional means for these models. The lognormal model provides a versatile alternative specification that is more flexible (and more natural) than the log gamma form, and provides a platform for several "two part" extensions, including zero inflation, hurdle and sample selection models. We also resolve some features in Hausman, Hall and Griliches's (1984) widely used panel data treatments for the Poisson and negative binomial models that appear to conflict with more familiar models of fixed and random effects. Finally, we consider a bivariate Poisson model that is also based on the lognormal heterogeneity model. Two recent applications have used this model. We suggest that the correlation estimated in their model frameworks is an ambiguous measure of the correlation of the variables of interest, and may substantially overstate it. We conclude with a detailed application of the proposed methods using the data employed in one of the two aforementioned bivariate Poisson studies.

JEL classification: C14; C23; C25

Keywords: Poisson regression; Negative binomial; Panel data; Heterogeneity; Lognormal; Bivariate Poisson; Zero inflation; Two part model; Hurdle model

---

* 44 West 4th St., Rm. 7-78, New York, NY 10012, USA, Telephone: 001-212-998-0876; e-mail: wgreene@stern.nyu.edu, URL www.stern.nyu.edu/~wgreene.

1 Introduction

Models for count data have been prominent in many branches of the recent applied literature, for example, in health economics (e.g., in numbers of visits to health facilities[1]) management (e.g., numbers of patents[2]) and industrial organization (e.g., numbers of entrants to markets[3]). The foundational building block in this modeling framework is the Poisson regression model.[4] But, because of its implicit restriction on the distribution of observed counts – in the Poisson model, the variance of the random variable is constrained to equal the mean – researchers routinely employ more general specifications, usually the negative binomial (NB) model which is the standard choice for a basic count data model.[5] There are also many applications that extend the Poisson and NB models to accommodate special features of the data generating process, such as hurdle effects,[6] zero inflation[7] and sample selection.[8] The basic models for panel data, fixed and random effects, have also been extended to the Poisson and NB models for counts.[9] Finally, there have been several proposals for extending the Poisson model to bivariate and multivariate settings.[10] This list includes a substantial fraction of the received extensions of the basic Poisson and NB models. There have, however, been scores of further refinements and extensions that are documented in a huge literature and several book length treatments such as Cameron and Trivedi (CT) (1998), Winkelmann (2003) and Hilbe (2007).

This paper will survey some practical extensions of the Poisson and NB models that practitioners can employ to refine the specifications or broaden their reach into new situations. We will also resolve some apparent inconsistencies of the panel data models with other more familiar results for the linear regression model.

- There are two well known, nonnested forms of the negative binomial model, denoted NB1 and NB2 in the literature. [See CT (1986).] Researchers have typically chosen one form or the other (typically NB2), but not generally formed a preference for one or the other. We propose an encompassing model that nests both of them parametrically and allows a statistical test of the two functional forms against a more general alternative.
- The NB model arises as the result of the introduction of log gamma distributed unobserved heterogeneity into the loglinear Poisson mean. A lognormal model provides a suitable alternative specification that is more flexible than the log gamma form, and provides a platform for several useful extensions, including hurdle, zero inflation, and sample selection models. We will develop this alternative to the NB model, then show how it can be used to accommodate in a natural fashion, e.g., sample selection, hurdle effects, and a new model for zero inflation.
- The most familiar panel data treatments, fixed effects (FE) and random effects (RE), for count models were proposed by Hausman, Hall and Griliches (HHG) (1984). The Poisson FE model is particularly simple to analyze, and has long been recognized as one of a very few

[1] Contoyannis, Jones and Rice (2004), Munkin and Trivedi (1999), Riphahn, Wambach and Million (RWM) (2003). See, as well, Cameron and Trivedi (2005).

[2] Hausman, Hall and Griliches (1984) and Wang, Cockburn and Puterman (1998).

[3] Asplund and Sandin (1999).

[4] HHG (1984), Cameron and Trivedi (1986, 1998), and Winkelmann (2003).

[5] The NB model is by far the most common specification. See Hilbe (2007). The latent class (finite mixture) and random parameters forms have also been employed. See, e.g., Wang et al., op. cit.

[6] See, e.g., Mullahy (1986), Rose, Martin, Wannameuhler and Plikaytis (2006) and Yen and Adamowicz (1994) on separately modeling participation and usage.

[7] See, e.g., Heilbron (1994) and Lambert (1992) on industrial processes, Greene (1994) on credit defaults and Zorn (1998) on Supreme Court Decisions.

[8] See, e.g., Greene (1995) on derogatory credit reports and Terza (1998).

[9] See, again, HHG (1984) on the relationship between patents and research and development.

[10] See King (1989), Munkin and Trivedi (1999) and Riphahn, Wambach and Million (RWM) (2003).

known models in which the incidental parameters problem [see Neyman and Scott (1948) and Lancaster (2000)] is, in fact, not a problem. The same is not true of the NB model. Researchers are sometimes surprised to find that the HHG formulation of the FE NB model allows an overall constant – a quirk that has also been documented elsewhere [see Allison (2000) and Allison and Waterman (2002), for example]. We resolve the source of the ambiguity, and consider the difference between the HHG FE NB model and a 'true' FE NB model that appears in the familiar index function form. The true FE NB model has not been used by applied researchers, probably because of the absence of a computational method. We have developed a method of computing the true FE NB model that allows a comparison to the HHG formulation.

The familiar RE Poisson model using a log gamma heterogeneity term produces the NB model. We consider the lognormal model as an alternative, again, as a vehicle for more interesting specifications, and compare it to the HHG formulation. The HHG RE NB model is also unlike what might seem the natural application in which the heterogeneity term appears as an additive common effect in the conditional mean. Once again, this was a practical solution to the problem. The lognormal model provides a means of specifying the RE NB model in a natural index function form. We will develop this model, and, once again, compare it to the HHG formulation.

- Two recent applications, Munkin and Trivedi (1999) and RWM (2003), have used a form of the bivariate Poisson model in which the correlation is introduced through additive correlated variables in the conditional mean functions. Both of these studies have misinterpreted (and overstated) the correlation coefficient estimated in their model frameworks. What they have specified is correlation between the logs of the conditional mean functions. How this translates to correlation between the count variables themselves is quite unclear. We will examine this in detail.

The study is organized as follows: Section 2 will detail the basic modeling frameworks for count data, the Poisson and NB models and will propose models for observed and unobserved heterogeneity in count data. This section will suggest a parameterization of the of the NB model that introduces measured heterogeneity into the scaling parameter. We then develop the NBP model to encompass NB1 and NB2. Finally, we propose the lognormal model as an alternative to the log gamma model that produces the NB specification. Section 3 will extend the lognormal model to several two part models. Section 4 will examine the fixed and random effects models for panel data. Section 5 will consider applications of the Bivariate Poisson model. The various model extensions proposed are applied to the RWM panel data on health care utilization in Section 6. Some conclusions are drawn in Section 7.

2 Basic Functional Forms for Count Data Models

This section details the basic functional forms for count data models. The literature abounds with alternative models for counts – see, e.g., CT (1998) and Winkelmann (2003). However, the Poisson and a few forms of the negative binomial model overwhelmingly dominate the received applications. [See, as well, Hilbe (2007).] We will summarize the basic forms of the model and propose a few extensions that provide the departure point for more elaborate two part models in Part 3.

2.1 The Poisson Regression Model

The canonical regression specification for a variable Y that is a count of events is the Poisson regression, (2.1-1) $$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\Gamma(1+y_i)}, \lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta}), y_i = 0,1,\ldots, i = 1,\ldots,N\ ,$$

where $\mathbf{x}_i$ is a vector of covariates and, $i = 1,\ldots,N$, indexes the N observations in a random sample. For reasons that will emerge below, we explicitly assume that there is a constant term in the model. (The regression model is developed in detail in a vast number of standard references such as CT (1986, 1998, 2005), Winkelmann (2003) and Greene (2008), so we will refer the reader to one of these sources for background results.) The Poisson model has the convenient feature that

(2.1-2) $$E[y_i|\mathbf{x}_i] = \lambda_i.$$

It has the undesirable characteristic that

(2.1-3) $$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i.$$

This is the 'equidispersion' aspect of the model. Since observed data will almost always display pronounced *over*dispersion, analysts typically seek alternatives to the Poisson model, such as the negative binomial model described below.

Estimates of the parameters of the model using a sample of N observations on $(y_i,\mathbf{x}_i)$, $i = 1,\ldots,N$, are obtained by maximizing the log likelihood function,[11]

(2.1-4) $$\ln L = \sum_{i=1}^{N} \left[y_i(\alpha + \mathbf{x}_i'\boldsymbol{\beta}) - \lambda_i - \ln\Gamma(1+y_i)\right].$$

The likelihood equations take the characteristically simple form[12]

(2.1-5) $$\partial \ln L/\partial\begin{pmatrix}\alpha\\ \boldsymbol{\beta}\end{pmatrix} = \sum_{i=1}^{N}\begin{pmatrix}1\\ \mathbf{x}_i\end{pmatrix}(y_i - \lambda_i) = \sum_{i=1}^{N}\begin{pmatrix}1\\ \mathbf{x}_i\end{pmatrix}e_i = \mathbf{0}.$$

The partial effects in the Poisson model are

(2.1-6) $$\partial E[y_i|\mathbf{x}_i]/\partial\mathbf{x}_i = \lambda_i\boldsymbol{\beta} = \mathbf{g}_\mathbf{x}.$$

The delta method can be used for inference about the partial effects. The necessary Jacobian is

(2.1-7) $$\mathbf{J} = \frac{\partial\mathbf{g}_\mathbf{x}}{\partial(\alpha,\boldsymbol{\beta}')} = \mathbf{g}_\mathbf{x}(1,\mathbf{x}_i') + \lambda_i\left[\underline{\mathbf{0}}\ \mathbf{I}\right],$$

where $\underline{\mathbf{0}}$ indicates a conformable column vector of zeros. The estimator of the asymptotic covariance for $\mathbf{g}_\mathbf{x}$ evaluated at a particular $(1,\mathbf{x}_i)$ (or the sample mean, $(1,\overline{\mathbf{x}})$ would be

[11] The conditions on the data generating mechanism for $\mathbf{x}_i$ that are necessary for the MLE to be well behaved and to have the familiar properties of consistency, asymptotic normality, efficiency and invariance to one to one transformations are all assumed, and will not be treated separately. The assumptions are carried through to the other models discussed below. Aside from some complications arising from the need to approximate certain integrals by quadrature or simulation, the models examined here are all amenable to straightforward maximum likelihood estimation.

[12] Estimation and inference for the Poisson regression model are discussed in standard sources such as CT (1998) and Greene (2008).

(2.1-8) $$\text{Est.Asy.Var}\left[\hat{\mathbf{g}}_{\mathbf{x}}\right] = \hat{\mathbf{J}}\left(Est.Asy.Var\left[\hat{\alpha},\hat{\boldsymbol{\beta}}\right]\right)\hat{\mathbf{J}}',$$

where "^" indicates a matrix or vector evaluated at the maximum likelihood estimates. In the various developments below, we will present the only elements of the Jacobians, **J**, for each estimator of the partial effects. Computation of asymptotic covariance matrices follow along these lines in all cases.

2.2 The Negative Binomial and Poisson Lognormal Regression Models

As noted in (2.1-2) and (2.1-3), the Poisson model imposes the (usually) transparently restrictive assumption that the conditional variance equals the conditional mean. The typical alternative is the negative binomial (NB) model. The model can be motivated as an attractive functional form simply in its own right that allows overdispersion. However, it is useful for present purposes to obtain the specification through the introduction of unobserved heterogeneity in the Poisson regression model. We consider two possible cases, the conventional approach based on the log gamma distribution and, we will argue, a more flexible approach based on the lognormal distribution.

2.2.1 The Negative Binomial Model

To introduce latent heterogeneity into the count data model, we write

(2.2-1) $$E[y_i|\mathbf{x}_i,\varepsilon_i] = \exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i) = h_i\lambda_i,$$

where $h_i = \exp(\varepsilon_i)$ is assumed to have a one parameter gamma distribution, $G(\theta,\theta)$ with mean 1 and variance $1/\theta = \kappa$. That is

(2.2-2) $$f(h_i) = \frac{\theta^\theta \exp(-\theta h_i) h_i^{\theta-1}}{\Gamma(\theta)}, h_i \geq 0, \theta > 0.$$ [13]

The nonzero mean of ε_i will be absorbed in the constant term of the index function. Making the change of variable to $\varepsilon_i = \ln h_i$ produces the log gamma variate with density

(2.2-3) $$f(\varepsilon_i) = \frac{\theta^\theta \exp\left[-\theta\exp(\varepsilon_i)\right]\left[\exp(\varepsilon_i)\right]^\theta}{\Gamma(\theta)}, -\infty < \varepsilon_i < \infty, \theta > 0.$$

It will be useful for the empirical results below to obtain the mean and variance of the random variable ε_i. The end result is

$$E[\varepsilon_i] = \psi(\theta) - \ln\theta,$$

$$\text{Var}[\varepsilon_i] = \psi'(\theta),$$

[13] This general approach is discussed at length by Gourieroux, Monfort and Trognon (1984), CT (1986, 1997), Winkelmann (2003) and HHG (1984).

where $\psi(\theta)$ is the digamma function, $d\ln\Gamma(\theta)/d\theta$ and $\psi'(\theta)$ is the trigamma function, $d^2\ln\Gamma(\theta)/d\theta$. To prove this, we will, we will use an indirect method of derivation so as to employ some simple known results. For convenience, we drop the observation subscript. Taking logs in (2.2-2),

$$\ln f(h) = \theta\ln\theta - \ln\Gamma(\theta) - \theta h + (\theta - 1)\ln h.$$

The density in (2.2-2) is "regular" according by the Fischer criteria for the properties of maximum likelihood estimation. [See Greene (2008, Ch. 16).] Thus,

$$\begin{aligned} \mathrm{E}[\partial \ln f(h)/\partial\theta] &= 1 + \ln\theta - \psi(\theta) - E[h] + E[\ln h] \\ &= 0. \end{aligned}$$

We know that $E[h]$ equals 1 from earlier results for the gamma distribution, so the first part of the result for $E[\varepsilon_i]$ follows immediately, since $\ln h = \varepsilon$. For the second result, we know from the information matrix equality that

$$\begin{aligned} \mathrm{Var}[\partial \ln f(h)/\partial\theta] &= -E[\partial^2 \ln f(h)/\partial\theta^2] \\ &= \psi'(\theta) - 1/\theta. \end{aligned}$$

But, $\quad \mathrm{Var}[\partial \ln f(h)/\partial\theta] = \mathrm{Var}[h] + \mathrm{Var}[\ln h] - 2\mathrm{Cov}[h,\ln h]$

and $\quad \mathrm{Var}[h] = 1/\theta,$

so $\quad \psi'(\theta) - 1/\theta = \mathrm{Var}[\ln h] + 1/\theta - 2\mathrm{Cov}[h,\ln h].$

We need $\mathrm{Cov}[h,\ln h]$ to obtain $\mathrm{Var}[\ln h] = \mathrm{Var}[\varepsilon]$;

$$\begin{aligned} \mathrm{Cov}[h,\ln h] &= E[h \ln h] - E[h]E[\ln h] \\ &= E[h \ln h] - 1\times(\psi(\theta) - \ln\theta). \end{aligned}$$

Once again reverting to the gamma density for h,

$$\begin{aligned} E[h\ln h] &= \int_0^\infty h\ln h \frac{\theta^\theta \exp(-\theta h) h^{\theta-1}}{\Gamma(\theta)} \mathrm{d}h \\ &= \int_0^\infty \ln h \frac{\theta^\theta \exp(-\theta h) h^{\theta}}{\Gamma(\theta)} \mathrm{d}h \\ &= \int_0^\infty \ln h \frac{\theta^{\theta+1} \exp(-\theta h) h^{(\theta+1)-1}}{\Gamma(\theta+1)} \mathrm{d}h. \end{aligned}$$

We have used the recursion $\Gamma(\theta+1) = \theta\Gamma(\theta)$ in the third line,. The third line gives $E[\ln h]$ when h has a gamma($\theta,\theta+1$) density, so it follows from our earlier result that $E[h\ln h] = \psi(\theta+1) - \ln\theta$. Collecting terms,

$$\psi'(\theta) - 2/\theta = \mathrm{Var}[\ln h] - 2[(\psi(\theta+1) - \ln\theta) - (\psi(\theta) - \ln\theta)].$$

Finally, we use the recursion $\psi(\theta+1) = \psi(\theta) + 1/\theta$. [See Abramovitz and Stegun (1971).] Inserting this in the line above produces the final result for Var[ln$h$] = Var[ε] = $\psi'(\theta)$.

The conditional Poisson regression model is, therefore,

(2.2-4) $$\text{Prob}[Y = y_i \mid \mathbf{x}_i, \varepsilon_i] = \frac{\exp\left[-\exp(\varepsilon_i)\lambda_i\right][\exp(\varepsilon_i)\lambda_i]^{y_i}}{\Gamma(1+y_i)}, \lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta}), y_i = 0,1,...$$

The unconditional density, that is, conditioned only on $\mathbf{x}_i$, is obtained by integrating ε_i out of the joint density. That is,

(2.2-5) $$\begin{aligned}\text{Prob}[Y = y_i \mid \mathbf{x}_i] &= \int_{\varepsilon_i} \text{Prob}[Y = y_i \mid \mathbf{x}_i, \varepsilon_i]\ f(\varepsilon_i) d\varepsilon_i \\ &= \int_0^\infty \frac{\exp(-\lambda_i \exp(\varepsilon_i))(\lambda_i \exp(\varepsilon_i))^{y_i}}{\Gamma(1+y_i)} \frac{\theta^\theta \exp(-\theta \exp(\varepsilon_i))\left[\exp(\varepsilon_i)\right]^\theta}{\Gamma(\theta)} d\varepsilon_i.\end{aligned}$$

At this point, it is convenient to make the change of variable back to $h_i = \exp(\varepsilon_i)$. Then, the conditional density is

(2.2-6) $$\text{Prob}[Y = y_i \mid \mathbf{x}_i, h_i] = \frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(1+y_i)}, \lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta}), y_i = 0,1,...$$

and the unconditional density is

(2.2-7) $$\begin{aligned}\text{Prob}[Y = y_i \mid \mathbf{x}_i] &= \int_{h_i} \text{Prob}[Y = y_i \mid \mathbf{x}_i, h_i]\ f(h_i) dh_i \\ &= \int_0^\infty \frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(1+y_i)} \frac{\theta^\theta \exp(-\theta h_i) h_i^{\theta-1}}{\Gamma(\theta)} dh_i \\ &= \frac{\theta^\theta \lambda_i^{y_i}}{\Gamma(1+y_i)\Gamma(\theta)} \int_0^\infty \exp(-h_i(\lambda_i+\theta)) h_i^{\theta+y_i-1} dh_i \\ &= \frac{\theta^\theta \lambda_i^{y_i}}{\Gamma(1+y_i)\Gamma(\theta)} \frac{\Gamma(\theta+y_i)}{(\lambda_i+\theta)^{\theta+y_i}}.\end{aligned}$$

Defining $r_i = \theta/(\theta+\lambda_i)$ produces

(2.2-8) $$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\Gamma(\theta+y_i) r_i^\theta (1-r_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta)}, y_i = 0,1,...,\theta > 0,$$

which is the probability density function for the negative binomial distribution.

The conditional mean and variance of the NB random variable relate to the Poisson moments as follows:

(2.2-9) $E[y_i|\mathbf{x}_i] = \lambda_i$,

(2.2-10) $\partial E[y_i|\mathbf{x}_i]/\partial \mathbf{x}_i = \lambda_i\boldsymbol{\beta} = \mathbf{g}_\mathbf{x}$, (2.2-11) $$\mathbf{J} = \frac{\partial \mathbf{g_x}}{\partial(\alpha,\boldsymbol{\beta}',\theta)} = \mathbf{g_x}(1,\mathbf{x}_i',0) + \lambda_i\left[\underline{\mathbf{0}}\ \mathbf{I}\ \underline{\mathbf{0}}\right],$$

(the same as in the Poisson model) and

(2.2-12) $$\begin{aligned} \text{Var}[y_i|\mathbf{x}_i] &= \lambda_i\,[1 + (1/\theta)\lambda_i] \\ &= \lambda_i\,[1 + \kappa\,\lambda_i] \end{aligned}$$

where $\kappa = \text{Var}[h_i]$.

Maximum likelihood estimation of the parameters of the NB model ($\alpha,\boldsymbol{\beta},\theta$) is straightforward, as documented in, e.g., Greene (2007). Inference proceeds along familiar lines.[14] Inference about the specification, specifically the presence of overdispersion, is the subject of a lengthy literature, as documented, e.g., in CT (1990, 1998, 2005) and Hilbe (2007).

2.2.2 Poisson Lognormal Mixture Model

Consider, instead, introducing the heterogeneity in (2.2-1) as a normally distributed variable with mean zero and standard deviation σ, which we introduce into the model explicitly by standardizing ε_i. Then, the Poisson model is

(2.2-13) $$P(y_i|\mathbf{x}_i,\varepsilon_i) = \frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(1+y_i)},\ h_i\lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i),\ \varepsilon_i \sim \text{N}[0,1].$$

The unconditional density would be

(2.2-14) $$P(y_i|\mathbf{x}_i) = \int_{-\infty}^{\infty} \frac{\exp[-\exp(\sigma\varepsilon_i)\lambda_i][\exp(\sigma\varepsilon_i)\lambda_i]^{y_i}}{\Gamma(1+y_i)}\phi(\varepsilon_i)d\varepsilon_i,$$

where here and in what follows, $\phi(\varepsilon_i)$ denotes the standard normal density. The unconditional log likelihood function is

(2.2-15) $$\begin{aligned} \ln L &= \sum_{i=1}^{N} \ln P(y_i \mid \mathbf{x}_i) \\ &= \sum_{i=1}^{N} \ln\left\{\int_{-\infty}^{\infty} \frac{\exp[-\exp(\sigma\varepsilon_i)\lambda_i][\exp(\sigma\varepsilon_i)\lambda_i]^{y_i}}{\Gamma(1+y_i)}\phi(\varepsilon_i)d\varepsilon_i\right\}. \end{aligned}$$

Maximum likelihood estimates of the model parameters are obtained by maximizing the unconditional log likelihood function with respect to the model parameters (α, $\boldsymbol{\beta}$, σ).

The integrals in the log likelihood function do not exist in closed form. The quadrature based approach suggested by Butler and Moffitt(1982) is a convenient method of approximating them. Let

$$v_i = \varepsilon_i / \sqrt{2}$$

---

[14] It is common to base inference about the parameters on 'robust' covariance matrices (the familiar 'sandwich' estimator). See, e.g., Stata (2006). Since the model has been obtained through the introduction of latent heterogeneity, which is now explicitly accounted for; it is unclear what additional specification failure the MLE (or pseudo-MLE) would be robust to. See Freedman (2006).

and

$$\omega = \sigma\sqrt{2}\,.$$

After making the change of variable from ε_i to v_i and reparameterizing the probability, we obtain

$$\text{(2.2-16)} \qquad P(y_i|\mathbf{x}_i) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-v_i^2)\, P(y_i|\mathbf{x}_i,v_i)\mathrm{d}v_i$$

where the conditional mean is now $E[y_i|\mathbf{x}_i,v_i] = \exp(\alpha+\boldsymbol{\beta}'\mathbf{x}_i + \omega v_i)$. Maximum likelihood estimates of $(\alpha,\boldsymbol{\beta}, \omega)$ are obtained by maximizing the reparameterized log likelihood. In this form, $\ln L$ can be approximated by Gauss-Hermite quadrature. The approximation is

$$\text{(2.2-17)} \qquad \ln L_Q = \sum\nolimits_{i=1}^{N} \ln\left[\frac{1}{\sqrt{\pi}}\sum\nolimits_{h=1}^{H} W_h P(y_i \mid \mathbf{x}_i, V_h)\right],$$

where V_h and W_h are the nodes and weights for the quadrature. The BHHH estimator of the asymptotic covariance matrix for the parameter estimates is a natural choice given the complexity of the function. The first derivatives must be approximated as well. To save some notation, denote the individual terms summed in the log likelihood as $\ln L_{Q,i}$. We also use the result that $\partial P(.,.)/\partial z = P\times\partial \ln P(.,.)/\partial z$ for any argument z which appears in the function. Then,

$$\text{(2.2-18)} \qquad \partial \ln L_Q / \partial\begin{pmatrix}\alpha\\ \boldsymbol{\beta}\\ \omega\end{pmatrix} = \sum\nolimits_{i=1}^{N} \frac{1}{L_{Q,i}}\frac{1}{\sqrt{\pi}}\sum\nolimits_{h=1}^{H} W_h P(y_i \mid \mathbf{x}_i, V_h)\left\{y_i - \left[\exp(\omega V_h)\lambda_i\right]\right\}\begin{pmatrix}1\\ \mathbf{x}_i\\ V_h\end{pmatrix}.$$

The estimate of σ is recovered from the transformation $\sigma = \omega/\sqrt{2}$.

Simulation is another effective approach to maximizing the log likelihood function. [See Train (2003) and Greene (2008).] In the original parameterization in (2.2-13), the log likelihood function is

$$\text{(2.2-19)} \qquad \ln L = \sum\nolimits_{i=1}^{N} \ln \int_{-\infty}^{\infty} P(y_i|\mathbf{x}_i,\varepsilon_i)]\ \phi(\varepsilon_i)d\varepsilon_i\,.$$

The simulated log likelihood would be

$$\text{(2.2-20)} \qquad \ln L_S = \sum\nolimits_{i=1}^{N} \ln \frac{1}{M}\sum\nolimits_{m=1}^{M} P(y_i|\mathbf{x}_i,\sigma\varepsilon_{im})]$$

where ε_{im} is a set of M random draws from the standard normal population. [We would propose to improve this part of the estimation by using Halton sequences, instead. See Train (2003, pp. 224-238) and Greene (2008).] Extensive discussion of maximum simulated likelihood estimation appears in Gourieroux and Monfort (1996), Munkin and Trivedi (1999), Train (2003) and Greene (2008).][15] Derivatives of the simulated log likelihood for the ith observation are

[15] One could preserve the log gamma specification by drawing h_{im} from a gamma(1,1) population and using the logs in the simulation, rather than using draws from N[0,1] for w_{im}. This approach, which obviates deriving the unconditional distribution analytically, was used in Munkin and Trivedi (1999).

(2.2-21) $$\partial \ln L_{S,i} / \partial \begin{pmatrix} \alpha \\ \beta \\ \sigma \end{pmatrix} = \frac{1}{\ln L_{S,i}} \frac{1}{M} \sum_{m=1}^{M} P(y_i \mid \mathbf{x}_i, \varepsilon_{im})[y_i - \exp(\sigma\varepsilon_{im})\lambda_i] \begin{pmatrix} 1 \\ \mathbf{x}_i \\ \varepsilon_{im} \end{pmatrix}.$$

The mean and variance of the lognormal variable are

(2.2-22) $$E[\exp(\sigma\varepsilon_i)] = \exp(\sigma^2/2)$$

and

$$\begin{aligned} \text{Var}[\exp(\sigma\varepsilon_i)] &= \text{E}[\exp(\sigma\varepsilon_i)^2] - \{\text{E}[\exp(\sigma\varepsilon_i)]\}^2 \\ &= \exp(\sigma^2)[\exp(\sigma^2) - 1]. \end{aligned}$$

The conditional mean in the Poisson lognormal model is

(2.2-23) $$E[y_i \mid \mathbf{x}_i, \varepsilon_i] = \lambda_i \exp(\sigma\varepsilon_i).$$

It follows that

(2.2-24) $$\begin{aligned} E[y_i \mid \mathbf{x}_i] &= E_\varepsilon[E[y_i \mid \mathbf{x}_i, \varepsilon_i]] \\ &= \lambda_i \exp(\sigma^2/2) \\ &= \exp[(\alpha + \sigma^2/2) + \mathbf{x}_i'\boldsymbol{\beta}]. \end{aligned}$$

To obtain the unconditional variance, we use

(2.2-25) $$\text{Var}[y_i \mid \mathbf{x}_i] = \text{E}_{\varepsilon i}[\text{Var}[y_i \mid \mathbf{x}_i, \varepsilon_i] + \text{Var}_{\varepsilon i}[E[y_i \mid \mathbf{x}_i, \varepsilon_i]].$$

Combining the results above, we find

(2.2-26) $$\begin{aligned} \text{Var}[y_i \mid \mathbf{x}_i] &= \lambda_i \exp(\sigma^2/2)\{1 + \lambda_i \exp(\sigma^2/2)[\exp(\sigma^2)\text{-}1]\} \\ &= E[y_i \mid \mathbf{x}_i, \varepsilon_i]\{1 + \tau\, E[y_i \mid \mathbf{x}_i, \varepsilon_i]\},\ \tau = [\exp(\sigma^2)\text{-}1]. \end{aligned}$$

Thus, the variance in the lognormal model has the same quadratic form as that in the negative binomial model in (2.2-12).

For the log gamma model, the partial effects and Jacobian have the same form as in the Poisson model;

(2.2-27) $$\begin{aligned} &\mathbf{g}_\mathbf{x} = \exp(\sigma^2/2)\lambda_i \\ &\mathbf{J} = \frac{\partial \mathbf{g}_\mathbf{x}}{\partial(\alpha, \boldsymbol{\beta}', \sigma)} = \mathbf{g}_\mathbf{x}(1, \mathbf{x}_i', \sigma) + \exp(\sigma^2/2)\lambda_i \left[\underline{\mathbf{0}} \quad \mathbf{I} \quad \underline{\mathbf{0}}\right]. \end{aligned}$$

One could argue that the lognormal model is a more natural specification. If the heterogeneity captures the aggregate of individually small influences, then an appeal to the central limit theorem would motivate the normal distribution more than the log gamma. [See Winkelmann (2003).] The attraction in this development is the ease with which the normal mixture model can be extended and adapted to new models and formulations, such as the two part models below. The log gamma model that underlies the familiar negative binomial specification provides no means doing so. [See, as well, RWM (1003, p. 395) and Million (1998).]

2.3 Observed and Unobserved Heterogeneity in the NB Model – A Heterogeneous NB Model

The negative binomial random variable with density in (2.2-8) is heteroscedastic as can be seen in (2.2-12). However, the scedastic function is a simple function of the mean λ_i; $\text{Var}[y_i|\mathbf{x}_i] = \lambda_i[1+(1/\theta)\lambda_i]$. In this model, $1/\theta$ represents a scaling parameter. A logical extension of the model is to allow this parameter to be heterogeneous, in the form[16]

(2.3-1) $$\theta_i = \exp(\mathbf{z}_i'\gamma).$$

The conditional mean and partial effects in (2.2-9)-(2.2-11) are unchanged by this modification of the model. This is a straightforward extension of the NB model. Maximum likelihood estimation and inference are routine.

Assuming that the NB model is the functional form of choice, not through the introduction of heterogeneity, but as the base model in its own right, one might be tempted to (re)introduce latent heterogeneity in the conditional mean function as in (2.2-1). The full model would be

(2.3-2)
$$\text{Prob}[Y = y_i|\mathbf{x}_i,\varepsilon_i] = \frac{\Gamma(\theta_i + y_i)r_i^{\theta_i}(1-r_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta_i)}, y_i = 0,1,...,\theta_i > 0,$$

$$r_i = \theta_i/(\theta_i + h_i\lambda_i),$$

$$\theta_i = \exp(\mathbf{z}_i'\gamma),$$

$$h_i = \exp(\varepsilon_i).$$

The distribution of ε_i remains to be specified. Assuming, once again, a log gamma distribution, $G(\mu,\mu)$, does not produce the same benefit as before, since the functional form of the NB model is not conjugate with respect to a gamma(μ,μ) model for h_i. The normal distribution would provide a useful alternative, though the model would still have to be estimated by maximum simulated likelihood or by using quadrature to eliminate the open form integral. The simulated log likelihood for this extended NB model for a sample of n observations would be

[16] This specification has been labeled the 'generalized negative binomial model.' [See, e.g., Stata (2006) and Econometric Software (2007).] However, that term was applied to a much earlier (no longer current) model [see Amid (1978) and Jain and Consul (1971)]. What Stata calls the 'generalized NB model is more appropriately labeled the heterogeneous NB model, so we will use that label here. What might rightfully be called *the* generalized NB model would be the NB P model developed in the next section. However, the name NegBin P, or NB P, will turn out to be much more useful as a descriptor of the model. We note, finally, there is a well established 'generalized Poisson model,' $P(y_i|\mathbf{x}_i)=(\lambda_i/a_i)^{yi}\,[a_i/\Gamma(1+y_i)]\,\exp[-\lambda_i(1+\theta y_i)/a_i]$, $a_i = 1+\theta\lambda_i$, that reverts to the Poisson model if $\theta = 0$. [See Econometric Software (2007), Sec. 24.4.2, and Wong and Famoye (1997).] However, this generalization of the Poisson model has no obvious connection to the generalizations of the NB model we consider here.

$$\ln L_S(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma},\sigma) = \sum_{i=1}^{N} \ln \frac{1}{M}\sum_{m=1}^{M} \frac{\Gamma(\theta_i + y_i) r_{im}^{\theta_i}(1-r_{im})^{y_i}}{\Gamma(1+y_i)\Gamma(\theta_i)},$$

(2.3-3)

$$r_{im} = \frac{\exp(\mathbf{z}_i'\boldsymbol{\gamma})}{\exp(\mathbf{z}_i'\boldsymbol{\gamma}) + \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \sigma w_{im})} = \frac{\theta_i}{\theta_i + h_{im}\lambda_i},$$

where M is the number of draws for the simulation, $w_{im} \sim N[0,1]$, $m = 1,...,M$ and σ is the standard deviation of the latent variable, $\varepsilon_i = \sigma w_i$. [Further results on simulation appear below in Section 3.1.

From the general form of the model, we have

(2.3-4) $$E[y_i|\mathbf{x}_i,\varepsilon_i] = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i).$$

Since ε_i is unobserved, in order to obtain the conditional mean function, we seek

(2.3-5) $$E[y_i|\mathbf{x}_i] = E_\varepsilon\{E[y_i|\mathbf{x}_i,\varepsilon_i]\} = \lambda_i E_\varepsilon[\exp(\varepsilon_i)] = \lambda_i E[h_i].$$

The result is straightforward for the two cases we have considered,. For a Gamma(μ,μ) model, the same result as before is obtained, since $E[\exp(\varepsilon_i)] = E[h_i] = 1$. Thus, in this case, we still have $E[y_i|\mathbf{x}_i] = \lambda_i$, though the unconditional distribution is not the negative binomial. Likewise, if the heterogeneity is assumed to be normally distributed, then the conditional mean is, once again, $\exp(\sigma^2/2)\lambda_i$.

Since the NB model ultimately arises from the introduction of latent heterogeneity into the Poisson model, arguably, the NB model with latent heterogeneity is overspecified. There could be different explanations for a finding of a 'significant' estimate of σ (or, $1/\mu$). It could be explained in terms of functional form of the assumed distribution of h_i in the Poisson model, or misspecification of the Poisson or NB models, themselves.. If the assumptions of the Poisson model with log-gamma heterogeneity are all correct, then it would seem that σ should equal zero by construction.

2.4 The NEGBIN P Model

The negative binomial model in (2.2-8) was labeled the NEGBIN 2 (NB2) model by CT (1986), in reference to the appearance of the quadratic term for λ_i in the conditional variance function:

(2.4-1) $$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i + (1/\theta_i)\lambda_i^2 = \lambda_i[1 + (1/\theta_i)\lambda_i]$$
$$= \lambda_i + \kappa_i \lambda_i^2,\ \kappa_i = \exp(-\mathbf{z}_i'\boldsymbol{\gamma}).$$

[We have also incorporated (2.3-1).] CT (1986) suggested a reparameterization of the model,

(2.4-2) $$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i + \kappa_i\lambda_i^1 = \lambda_i[1 + \kappa_i],$$

and label the resulting specification NB1. The model is obtained by replacing θ_i with $\theta_i\lambda_i$ in (2.2-8). After simplification, we obtain the density for NB1,

$$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \frac{\Gamma(\theta_i\lambda_i + y_i) q_i^{\theta_i\lambda_i}(1-q_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta_i\lambda_i)},\ y_i = 0,1,...,$$

(2.4-3)

$$q_i = \frac{1}{1+\theta_i}.$$

The authors note in (1998) that other exponents would be possible. [See their p. 73 and (3.26).] By replacing θ_i with $\theta_i\lambda_i^{2-P}$, we obtain the NEGBIN P, or NBP model,

$$\text{(2.4-4)} \qquad \begin{aligned} &\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \frac{\Gamma(\theta_i\lambda_i^{2-P} + y_i)s_i^{\theta_i\lambda_i^{2-P}}(1-s_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta_i\lambda_i^{2-P})}, y_i = 0,1,..., \\ &s_i = \frac{\lambda_i}{\lambda_i + \theta_i\lambda_i^{2-P}}. \end{aligned}$$

(The log likelihood function and its derivatives are given in Appendix A.) The NB1 and NB2 models are the special cases of P =1 and P = 2. The conditional mean in this model is still λ_i, so the partial effects are still those in (2.2-9), while the conditional variance is

$$\text{(2.4-5)} \qquad \text{Var}[y_i|\mathbf{x}_i] = \lambda_i\,[1 + (1/\theta_i)\lambda_i^{P-1}].$$

CT (1998) focus on the P = 1 and P = 2 forms, but suggest that the "generalized event count model" (see their Section 4.4.1) does include the NEGBIN P as a special case. (CT (1986) also mentions the possibly of this extension of the model, but does not develop it at any length.) The GEC model [Winkelmann and Zimmermann (1991, 1995), King (1989)] which does include NEGBIN P is sufficiently cumbersome to have greatly restricted its general use. The NEGBIN P model achieves somewhat less of the generality of the GEC model, but is much simpler to implement.[17] An application appears below.[18]

Since the NB1 and NB2 models are not nested, there is no simple parametric test that one can employ to choose between them. (E.g., CT do not express a preference for either one or the other in (1986) or (1998); they merely note the difference. They do state "[T]he NB2 MLE [is] favored by econometricians and the NB1 GLM [generalized linear model] [is] used extensively by statisticians." This appeal to the estimation algorithm appears to be the closest to a preference for one or the other as appears in the recent literature. On the other hand, the various references to GEC and NBP models do suggest an attraction to a more general specification than NB1 or NB2.

For choosing statistically between NB1 and NB2, the models are nonnested and essentially equivalently parameterized, so a direct test is precluded. However, one possibility is the Vuong (1989) test based on

$$\text{(2.4-6)} \qquad \begin{aligned} &V = \frac{\sqrt{n}\;\bar{m}}{s_m}, \\ &m_i = \ln L_i(NB2) - \ln L_i(NB1). \end{aligned}$$

When the underlying conditions for its validity are met, the Vuong test statistic has a limiting standard normal distribution. Large positive values would favor NB2. We have found in applications that this statistic is rarely outside the inconclusive region (-1.96 to +1.96) for this model. It may be that NB1 and NB2 are not sufficiently different to enable a distinction on this basis. Since the NBP model does nest both of them, it provides a partial solution to the specification

[17] Winkelmann and Zimmermann (1995) develop a maximum likelihood estimator for the equivalent of the NEGBIN P model, but their formulation adds what appears to be a considerable yet unnecessary layer of difficulty to the derivation. In applications, the direct MLE based on (27) appears to be quite well behaved.

[18] The GEC model allows underdispersion as well as overdispersion and, as such, is more general than the NEGBIN P form. Overdispersion is the more common problem to be solved with an extended functional form.

problem. For example, in our application below, a simple likelihood ratio test rejects both the NB1 and NB2 null hypotheses.

3. Two Part Models

This section develops three "two part" extensions of the count data models, a model for sample selection, the zero inflated Poisson model (ZIP) (and the ZINB model), and a hurdle model. Each of these models consists of an equation for "participation" and a model for the event count that is conditioned on the outcome of the first decision. The third part of each specification is the observation mechanism that links the participation equation and the count outcome model. The sample selection model appears first in Greene (1995, 1997), Terza (1998) and Greene (2006), and is included here for completeness and to develop the platform for the other two. The ZIP and ZINB models are also established, e.g., by Heilbron (1992), Lambert (1994) and Greene (1994). The following presents an extension of this model to allow correlation between the regime and the count variable. The hurdle model [Mullahy (1986)] has been widely used, e.g., in health economics. The extensions of the ZIP/ZINB and hurdle models proposed here also allows correlation across the two equations.

3.1 Sample Selection

The generic sample selection model builds on (2.1-1) (Poisson) or (2.2-8) (NB) with latent heterogeneity in (2.3-1),

(3.1-1a)	$d_i^* = \mathbf{w}_i'\boldsymbol{\delta} + u_i \quad u_i \sim N[0,1]$,	
	$d_i = \mathbf{1}(d_i^* > 0)$,	
	$\text{Prob}(d_i = 0\|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\boldsymbol{\delta})$	(generic, binary choice, zero),
	$\text{Prob}(d_i = 1\|\mathbf{w}_i) = \Phi(\mathbf{w}_i'\boldsymbol{\delta})$	(probit selection equation),
(3.1-1b)	$y_i \| \mathbf{x}_i,\varepsilon_i \sim P(y_i \| \mathbf{x}_i,\varepsilon_i)$	(conditional on ε_i Poisson or NB model),
	$E[y_i\| \mathbf{x}_i,\varepsilon_i] = \exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i) = h_i\lambda_i$,	(conditional mean with heterogeneity),
	$\varepsilon_i \sim N[0,1]$	
(3.1-1c)	$[u_i,\varepsilon_i] \sim N_2[(0,1),(1,1),\rho)]$	(selection effect),
	$y_i,\mathbf{x}_i$ are observed only when $d_i = 1$.	(observation mechanism).

(We use the notation $N_2[(\mu_1,\mu_2),(\sigma_1^2,\sigma_2^2),\rho]$ to denote the bivariate normal distribution with correlation ρ.) "Selectivity" is transmitted through the correlation parameter ρ. Drawing on the results of Heckman (1979), it is tempting to estimate this model in the same fashion as in the linear case by (a) fitting the probit model by MLE and computing the inverse Mills ratio, $\hat{\psi}_i = \phi(\mathbf{w}_i'\hat{\boldsymbol{\delta}})/\Phi(\mathbf{w}_i'\hat{\boldsymbol{\delta}})$, for each observation in the selected subsample, then (b) adding $\hat{\psi}_i$ to the right hand side of the Poisson or NB model and fitting it by MLE, adding a Murphy and Topel (2002) correction to the estimated asymptotic covariance matrix. However, this would be inappropriate for this case (and other nonlinear models):

- The impact on the conditional mean in the Poisson model will not take the form of an inverse Mills ratio. That is specific to the linear model. (See Terza (1998) for a development in the context of the exponential regression. The result is given below.)

- The dependent variable, conditioned on the sample selection, is unlikely to have the Poisson or NB distribution in the presence of the selection. That would be needed to use this approach. Note that this even appears in the canonical linear case. The normally distributed disturbance in the absence of sample selection has a nonnormal distribution in the presence of selection. That is the salient feature of Heckman's development.

We develop, instead, a full information maximum likelihood estimator. [See Greene (1997).]

The log likelihood function for the full model is the joint density for the observed data. When d_i equals one, $(y_i,\mathbf{x}_i,d_i,\mathbf{w}_i)$ are all observed. To obtain the unconditional, joint discrete density, $P(y_i,d_i = 1|\mathbf{x}_i,\mathbf{w}_i)$ we proceed as follows:

(3.1-2) $$P(y_i, d_i = 1|\mathbf{x}_i,\mathbf{w}_i) = \int_{-\infty}^{\infty} P(y_i, d_i = 1 \mid \mathbf{x}_i, \mathbf{w}_i, \varepsilon_i)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i,$$

where $\phi(\varepsilon_i)$is the standard normal density. Conditioned on ε_i, d_i and y_i are independent, so,

(3.1-3) $$P(y_i,d_i = 1|\mathbf{x}_i,\mathbf{w}_i,\varepsilon_i) = P(y_i|\mathbf{x}_i,\varepsilon_i)\mathrm{Prob}(d_i = 1|\mathbf{w}_i,\varepsilon_i).$$

The first part, $P(y_i|\mathbf{x}_i,\varepsilon_i)$ is the conditional Poisson or NB density in (3.1-1). By joint normality,

(3.1-4) $$f(u_i|\varepsilon_i) = N[\rho\varepsilon_i, (1-\rho^2)],$$

so

(3.1-5) $$u_i = \rho\varepsilon_i + v_i\sqrt{1-\rho^2} \text{ where } v_i \sim N[0,1] \perp \varepsilon_i .$$

Therefore, using (3.1-1a),

(3.1-6) $$d_i^* = \mathbf{w}_i'\delta + \rho\varepsilon_i + v_i\sqrt{1-\rho^2}$$

so

(3.1-7) $$\mathrm{Prob}(d_i = 1|\mathbf{w}_i,\varepsilon_i) = \Phi\left([\mathbf{w}_i'\delta + \rho\varepsilon_i]/\sqrt{1-\rho^2}\right).$$

Combining terms, the unconditional joint density is obtained by integrating ε_i out of the conditional density. Thus,

(3.1-8) $$P(y_i,d_i = 1|\mathbf{x}_i,\mathbf{w}_i)=\int_{-\infty}^{\infty} P(y_i|\mathbf{x}_i,\varepsilon_i)\ \Phi\left([\mathbf{w}_i'\delta + \rho\varepsilon_i]/\sqrt{1-\rho^2}\right)\ \phi(\varepsilon_i)d\varepsilon_i .$$

By exploiting the symmetry of the normal cdf

(3.1-9) $$\mathrm{Prob}(d_i = 0|\mathbf{x}_i,\mathbf{w}_i,\varepsilon_i) = \Phi\left(-[\mathbf{w}_i'\delta + \rho\varepsilon_i]/\sqrt{1-\rho^2}\right)$$

and

(3.1-10) $$\mathrm{Prob}(d_i = 0|\ \mathbf{x}_i,\mathbf{w}_i) = \int_{-\infty}^{\infty} \Phi\left(-[\mathbf{w}_i'\delta + \rho\varepsilon_i]/\sqrt{1-\rho^2}\right)\ \phi(\varepsilon_i)d\varepsilon_i .$$

Expressions (3.1-8) and (3.1-10) can be combined by using the symmetry of the normal cdf,

(3.1-11) $$P(y_i,d_i|\mathbf{x}_i,\mathbf{w}_i) =\int_{-\infty}^{\infty} [(1- d_i) + d_i\,P(y_i|\mathbf{x}_i,\varepsilon_i)]\,\Phi\left((2d_i-1)[\mathbf{w}_i'\delta + \rho\varepsilon_i]/\sqrt{1-\rho^2}\right)\ \phi(\varepsilon_i)d\varepsilon_i ,$$

where for d_i equal to zero, $P(y_i,d_i \mid \mathbf{x}_i,\mathbf{w}_i)$ is just Prob($d_i = 0|\mathbf{w}_i$).

Maximum likelihood estimates of the model parameters are obtained by maximizing the unconditional log likelihood function,

$$\text{(3.1-12)} \qquad \ln L = \sum_{i=1}^{N} \ln P(y_i,d_i \mid \mathbf{x}_i,\mathbf{w}_i),$$

with respect to the model parameters (α, $\boldsymbol{\beta}$, σ, $\boldsymbol{\delta}$, ρ). We now consider how to maximize the log likelihood. Butler and Moffitt's (1982) quadrature based approach suggested in Section 2.2.2 is a convenient method. Let

$$\begin{aligned} v_i &= \varepsilon_i/\sqrt{2}\,, \\ \omega &= \sigma\sqrt{2}\,, \\ \text{(3.1-13)} \qquad \tau &= \sqrt{2}\left(\rho/\sqrt{1-\rho^2}\right), \\ \boldsymbol{\eta} &= [1/\sqrt{1-\rho^2}\,]\boldsymbol{\delta}. \end{aligned}$$

After making the change of variable from ε_i to v_i and reparameterizing the probability, we obtain

$$\text{(3.1-14)} \qquad P(y_i,d_i = 1|\mathbf{x}_i,\mathbf{w}_i) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-v_i^2)\, P(y_i|\mathbf{x}_i,v_i)\Phi(\mathbf{w}_i'\boldsymbol{\eta} + \tau v_i)\, dv_i,$$

where the conditional mean is now $E[y_i|\mathbf{x}_i,v_i] = \exp(\alpha+\boldsymbol{\beta}'\mathbf{x}_i + \omega v_i)$. Maximum likelihood estimates of (α,$\boldsymbol{\beta}$, ω, $\boldsymbol{\eta}$, τ) are obtained by maximizing the reparameterized log likelihood.[19] The Gauss-Hermite approximation is

$$\text{(3.1-15)} \qquad \ln L_Q = \sum_{i=1}^{N} \ln\left[\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H} W_h\left[(1-d_i)+d_i P(y_i \mid \mathbf{x}_i, V_h)\right]\Phi\left[(2d_i-1)\left(\mathbf{w}_i'\boldsymbol{\eta}+\tau V_h\right)\right]\right],$$

where V_h and W_h are the nodes and weights for the quadrature. The BHHH estimator of the asymptotic covariance matrix for the parameter estimates is a natural choice given the complexity of the function. The first derivatives must be approximated as well. For convenience, let

$$\begin{aligned} P_{ih} &= P(y_i| \mathbf{x}_i, V_h), \\ \text{(3.1-16)} \qquad \Phi_{ih} &= \Phi[(2d_i - 1)(\mathbf{w}_i'\boldsymbol{\eta} + \tau V_h)] \quad \text{(normal CDF)}, \\ \phi_{ih} &= \phi[(2d_i - 1)(\mathbf{w}_i'\boldsymbol{\eta} + \tau V_h)] \quad \text{(normal density)}, \end{aligned}$$

and to save some notation, denote the individual terms summed in the log likelihood as $\ln L_{Q,i}$. Then,

$$\partial \ln L_Q / \partial\begin{pmatrix}\alpha \\ \boldsymbol{\beta} \\ \omega\end{pmatrix} = \sum_{i=1}^{N} \frac{d_i}{L_{Q,i}}\frac{1}{\sqrt{\pi}}\sum_{h=1}^{H} W_h \Phi_{ih} P_{ih}\left[\frac{\partial \ln P(y_i \mid \mathbf{x}_i, V_h)}{\partial(h_h\lambda_i)}\right](h_h\lambda_i)\begin{pmatrix}1 \\ \mathbf{x}_i \\ V_h\end{pmatrix},$$

[19] The dispersion parameter, θ (or the heterogeneous version, θ_i) would appear in the parameter vector and in the derivatives in (3.1-17) and (3.1-19) if the (heterogeneous) NB model were used here instead of the Poisson.

(3.1-17) $h_h = \exp(\omega V_h)$,

$$\partial \ln L_Q / \partial \begin{pmatrix} \boldsymbol{\eta} \\ \tau \end{pmatrix} = \sum_{i=1}^{N} \frac{1}{L_{Q,i}} \frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} W_h \phi_{ih} [(1-d_i) + d_i P_{ih}] \begin{pmatrix} \mathbf{w}_i \\ V_h \end{pmatrix}.$$

Estimates of the structural parameters, ($\boldsymbol{\delta}$,ρ,σ) and their standard errors can be computed using the transformations shown above and the delta method or the method of Krinsky and Robb (1986).

Simulation can also be used to maximizing the log likelihood function. [See Train (2003) and Greene (2008).] Using the original parameterization of the conditional mean function. The simulated log likelihood based on (3.1-11) is

(3.1-18) $$\ln L_S = \sum_{i=1}^{N} \ln \frac{1}{M} \sum_{m=1}^{M} [(1-d_i) + d_i\, P(y_i|\mathbf{x}_i,\sigma\varepsilon_{im})]\ \Phi[(2d_i - 1)(\mathbf{w}_i'\boldsymbol{\eta} + \tau\varepsilon_{im})]$$

where ε_{im} is a set of M random draws from the standard normal population (or transformations of a Halton sequence). Derivatives of the simulated log likelihood for the ith observation are

$$\partial \ln L_{S,i} / \partial \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \\ \sigma \end{pmatrix} = \frac{d_i}{\ln L_{S,i}} \frac{1}{M} \sum_{m=1}^{M} \Phi_{im} P_{im} \left[\frac{\partial \ln P(y_i \mid \mathbf{x}_i, \sigma\varepsilon_{im})}{\partial (h_{im}\lambda_i)} \right] (h_{im}\lambda_i) \begin{pmatrix} 1 \\ \mathbf{x}_i \\ \varepsilon_{im} \end{pmatrix},$$

(3.1-19) $h_{im}=\exp(\sigma\varepsilon_{im})$,

$$\partial \ln L_{S,i} / \partial \begin{pmatrix} \boldsymbol{\eta} \\ \tau \end{pmatrix} = \frac{1}{\ln L_{S,i}} \frac{1}{M} \sum_{m=1}^{M} \phi_{im} [(1-d_i) + d_i P_{im}] \begin{pmatrix} \mathbf{w}_i \\ \varepsilon_{im} \end{pmatrix},$$

where Φ_{im}, ϕ_{im} and P_{im} are defined as in (3.1-16) using ε_{im} in place of V_h.

The sample selection alters the conditional mean as follows: [See Terza (1985).] From (3.1-1b), the overall mean is

(3.1-20) $$E[y_i|\mathbf{x}_i] = E_\varepsilon\{E[y_i|\mathbf{x}_i,\varepsilon_i]\} = \exp(\sigma^2/2)\lambda_i.$$

However,

(3.1-21)) $$\begin{aligned} E[y_i|\mathbf{x}_i,\mathbf{w}_i,d_i = 1] &= \lambda_i\, \mathrm{E}[\exp(\sigma\varepsilon_i)|\mathbf{w}_i,d_i = 1] \\ &= \lambda_i \frac{\exp((\rho\sigma)^2/2)\Phi(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})}{\Phi(\mathbf{w}_i'\boldsymbol{\delta})}. \end{aligned}$$

This greatly complicates the partial effects;

(3.1-22) $$\begin{aligned} \frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{w}_i, d_i = 1]}{\partial \mathbf{x}_i} &= \lambda_i \left[\frac{\exp((\rho\sigma)^2/2)\Phi(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})}{\Phi(\mathbf{w}_i'\boldsymbol{\delta})} \right] \boldsymbol{\beta} \\ \frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{w}_i, d_i = 1]}{\partial \mathbf{w}_i} &= \lambda_i \left(\frac{\exp((\rho\sigma)^2/2)}{\Phi(\mathbf{w}_i'\boldsymbol{\delta})} \right) \left[\phi(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta}) - \phi(\mathbf{w}_i'\boldsymbol{\delta}) \left(\frac{\Phi(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})}{\Phi(\mathbf{w}_i'\boldsymbol{\delta})} \right) \right] \boldsymbol{\delta}. \end{aligned}$$

The effects are added for variables that $\mathbf{x}_i$ and $\mathbf{w}_i$ variables in common. These might be logically labeled the direct and indirect effects, since the latter arise only due to the effect of the selection. Note that the large bracketed term in the indirect effect equals zero if ρ equals zero. Jacobians of the partial effects for use in obtaining standard errors are given in Appendix B.

3.2 Zero Inflation

The zero inflated Poisson and NB (ZIP and ZINB) models can be viewed as partial observation models or latent class models of a sort.[20] The familiar structure of the model is

(3.2-1a) $d_i^* = \mathbf{w}_i'\delta + u_i,$

$d_i = \mathbf{1}(d_i^* > 0),$

$\text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\delta)$ (regime selection equation),

$\text{Prob}(d_i = 1|\mathbf{w}_i) = 1 - \Pi_0(\mathbf{w}_i'\delta)$ (regime selection equation),

(3.2-1b) $y_i^* |\mathbf{x} \sim P(y_i^* \mid \mathbf{x}_i)$ (latent Poisson or NB model),

$E[y_i^*| \mathbf{x}_i] = \exp(\alpha+\mathbf{x}_i'\beta) = \lambda_i,$ (conditional mean),

(3.2-1c) $y_i = d_i\, y_i^*$ and $\mathbf{x}_i$ are observed (observation mechanism).

Thus, if d_i equals zero, then the observed y_i equals zero regardless of the latent value of y_i^*. If d_i equals one, the Poisson or NB variable (which might then still equal zero) is observed. The joint density for y_i and d_i is derived as follows;

(3.2-2)
$$\text{Prob}(y_i = 0|\mathbf{x}_i,\mathbf{w}_i,d_i = 0) = 1;\ \text{Prob}(d_i = 0|\mathbf{x}_i,\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\delta),$$
$$\text{Prob}(Y_i = y_i\, |\mathbf{x}_i,\mathbf{w}_i,d_i = 1) = P(y_i^*| \mathbf{x}_i);\ \text{Prob}(d_i = 1|\mathbf{x}_i,\mathbf{w}_i) = 1 - \Pi_0(\mathbf{w}_i'\delta).$$

Combining terms, the joint density is

(3.2-3)
$$P(y_i,d_i| \mathbf{x}_i,\mathbf{w}_i) = P(y_i| \mathbf{x}_i,\mathbf{w}_i,d_i)P(d_i|\mathbf{x}_i,\mathbf{w}_i)$$
$$= (1 - d_i)\, \Pi_0(\mathbf{w}_i'\delta) + d_i\, [1 - \Pi_0(\mathbf{w}_i'\delta)]P(y_i^*| \mathbf{x}_i).$$

The conditional mean function is

(3.2-4)
$$\sum_d \sum_y y_i\, P(y_i,d_i|\mathbf{x}_i,\mathbf{w}_i) = E[y_i|\mathbf{x}_i,\mathbf{w}_i] = [1 - \Pi_0(\mathbf{w}_i'\delta)]\lambda_i,$$

so the partial effects are

(3.2-5)
$$\partial E[y_i|\mathbf{x}_i,\mathbf{w}_i]/\partial \mathbf{x}_i = \lambda_i\, [1 - \Pi_0(\mathbf{w}_i'\delta)]\beta = \mathbf{g}_\mathbf{x}$$

and

$$\partial E[y_i|\mathbf{x}_i,\mathbf{w}_i]/\partial \mathbf{w}_i = -\lambda_i\, [d\Pi_0(\mathbf{w}_i'\delta)/d(\mathbf{w}_i'\delta)]\delta = \mathbf{g}_\mathbf{w}$$

with

[20] See Heilbron (1992), Lambert (1992), Greene (1994) and Zorn (1998).

$$
(3.2\text{-}6)\qquad
\begin{aligned}
\frac{\partial \mathbf{g}_{\mathbf{x}}}{\partial(\alpha,\boldsymbol{\beta}')} &= \mathbf{g}_{\mathbf{x}}(1,\mathbf{x}_i') + \lambda_i[1-\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})]\left[\underline{\mathbf{0}}\quad \mathbf{I}\right]\\
\frac{\partial \mathbf{g}_{\mathbf{x}}}{\partial\boldsymbol{\delta}'} &= -\lambda_i \frac{\mathrm{d}\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})}{\mathrm{d}\left(\mathbf{w}_i'\boldsymbol{\delta}\right)}\boldsymbol{\beta}\mathbf{w}_i'\\
\frac{\partial \mathbf{g}_{\mathbf{w}}}{\partial(\alpha,\boldsymbol{\beta}')} &= \mathbf{g}_{\mathbf{w}}(1,\mathbf{x}_i')\\
\frac{\partial \mathbf{g}_{\mathbf{w}}}{\partial\boldsymbol{\delta}'} &= -\lambda_i\left[\frac{\mathrm{d}^2\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})}{\mathrm{d}\left(\mathbf{w}_i'\boldsymbol{\delta}\right)^2}\boldsymbol{\delta}\mathbf{w}_i' + \frac{\mathrm{d}\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})}{\mathrm{d}\left(\mathbf{w}_i'\boldsymbol{\delta}\right)}\mathbf{I}\right]
\end{aligned}
$$

A variety of specifications have appeared in the literature. As noted, the event count model could be Poisson, NB, or something else, though these two are the only forms that have been used.[21] The form of $\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})$ is often based on the logistic probability model, though the probit model is equally common. Finally, a rarely used specification proposed in Lambert (1992) is the ZIP(τ) model in which $\Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) = 1-F[\tau(\alpha+\mathbf{x}_i'\boldsymbol{\beta})]$ for the same α and $\boldsymbol{\beta}$ that appear in the count model, and unrestricted scale parameter τ. This form is extremely restrictive and difficult to motivate.

The zero inflation model accommodates data such as the count of doctor visits that we will examine in the applications in Section 6. Figure 2 below gives a histogram for this variable.[22] The conspicuous spike at zero in this variable is decidedly nonPoisson.[23] The preponderance of zeros in these data might be motivated by the possibility that the population consists of "healthy" individuals who never need to visit the doctor (or refuse to do so), and "less healthy" individuals who may or may not visit the doctor, depending on circumstances.

The latent class interpretation of the model suggests a two level decision process, the regime and the event count. (The hurdle model of the next section might be a yet more natural candidate for this interpretation.) The ZIP and ZINB models have been widely used in a variety of applications. A common element throughout is the assumption that the latent effects in the regime equation and the count outcome are uncorrelated. The model developed in the preceding section can be adapted to allow this correlation to be unrestricted. The extended model would be

(3.2-7a) $d_i^* = \mathbf{w}_i'\boldsymbol{\delta} + u_i,\ u_i \sim N[0,1]$,

$d_i = \mathbf{1}(d_i^* > 0)$,

$\text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\boldsymbol{\delta})$

$\text{Prob}(d_i = 1|\mathbf{w}_i) = \Phi(\mathbf{w}_i'\boldsymbol{\delta})$ (probit regime selection equation),

(3.2-7b) $y_i^* |\mathbf{x}_i,\varepsilon_i \sim P(y_i^* |\mathbf{x}_i,\varepsilon_i)$ (Poisson or NB model with heterogeneity),

$E[y_i^*| \mathbf{x}_i,\varepsilon_i] = \exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta}+\sigma\varepsilon_i) = h_i\lambda_i$ (heterogeneous conditional mean),

$[u_i,\varepsilon_i] \sim N_2[(0,1),(1,1),\rho]$,

---

[21] Greene, Harris, Hollingsworth and Maitra (2007) have adapted the zero inflation model developed here to an ordered probit specification. Econometric Software, Inc. (2003) includes a zero inflated gamma model. [See Winkelmann (2003) for discussion of the gamma model for count data.]

[22] The sample size is 27,326. To help format the figure, we have dropped 196 observations (0.7% of the sample) for which Docvis is greater than 30.

[23] Greene (2008, Section 16.9.5.b) suggests a "geometric" count data model, $P(y_i|\mathbf{x}_i)=\theta_i(1-\theta_i)^{y_i}$, where $\theta_i=1/(1+\lambda_i)$ and $\lambda_i = \exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta})$ for these data. The fit of the geometric model to the zero heavy variable is dramatically better than that for the Poisson or NB models.

(3.2-7c) $y_i = d_i y_i^*$ and $(\mathbf{x}_i,\mathbf{w}_i)$ are observed (observation mechanism).

It is now straightforward to adapt the derivation of the preceding section to this model. The conditional (on ε_i) zero inflated Poisson probability joint density function for y_i and d_i would be

(3.2-8) $$P(y_i, d_i \mid \mathbf{x}_i,\mathbf{w}_i,\varepsilon_i) = (1-d_i)\text{Prob}(d_i = 0 \mid \mathbf{w}_i,\varepsilon_i) + [1 - \text{Prob}(d_i = 0 \mid \mathbf{w}_i,\varepsilon_i)]P(y_i|\mathbf{x}_i,\varepsilon_i)$$

$$= (1-d_i)\Phi\left[\frac{-(\mathbf{w}_i'\boldsymbol{\delta}+\rho\varepsilon_i)}{\sqrt{1-\rho^2}}\right] + \Phi\left[\frac{(\mathbf{w}_i'\boldsymbol{\delta}+\rho\varepsilon_i)}{\sqrt{1-\rho^2}}\right]\frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(y_i+1)}$$

where, once again, $h_i\lambda_i = \exp(\sigma\varepsilon_i)\exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta}) = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i)$. (The ZINB model is obtained by the corresponding replacement of $P(y_i|\mathbf{x}_i,\varepsilon_i)$ in (3.2-8). As before, maximum likelihood estimates of the parameters of the model are obtained by maximizing the unconditional log likelihood. It is convenient to reparameterize the model. Then,

(3.2-9) $$\ln L = \sum_{i=1}^{N} \ln\int_{-\infty}^{\infty} (1-d_i)\Phi\left[-\frac{(\mathbf{w}_i'\boldsymbol{\delta}+\rho\varepsilon_i)}{\sqrt{1-\rho^2}}\right] + \Phi\left[\frac{(\mathbf{w}_i'\boldsymbol{\delta}+\rho\varepsilon_i)}{\sqrt{1-\rho^2}}\right]\frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(y_i+1)}\,\phi(\varepsilon_i)d\varepsilon_i$$

$$= \sum_{i=1}^{N} \ln\int_{-\infty}^{\infty} \left\{\begin{array}{l}(1-d_i)\Phi\left[-\mathbf{w}_i'\boldsymbol{\eta}-\tau\varepsilon_i\right] \\ +\Phi\left[\mathbf{w}_i'\boldsymbol{\eta}+\tau\varepsilon_i\right]\dfrac{\exp[-\exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta}+\sigma\varepsilon_i)][\exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta}+\sigma\varepsilon_i)]^{y_i}}{\Gamma(y_i+1)}\end{array}\right\}\phi(\varepsilon_i)d\varepsilon_i.$$

At this point, the specification differs only slightly from the formulation in the preceding section, in (3.1-11). Either quadrature or simulation can be used to maximize the likelihood function, with the corresponding adaptation of either (3.1-15) for the quadrature approach or (3.1-18) for the simulation based estimator.

For convenience, let $A_i = \dfrac{(\mathbf{w}_i'\boldsymbol{\delta}+\rho\varepsilon_i)}{\sqrt{1-\rho^2}}$. The conditional mean function for the ZIP model with latent heterogeneity is

(3.2-10) $$E[y_i \mid \mathbf{x}_i,\mathbf{w}_i,,\varepsilon_i] = \Phi\left[A_i\right]\exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta}+\sigma\varepsilon_i)$$

The observable counterpart is

(3.2-11) $$E[y_i \mid \mathbf{x}_i,\mathbf{w}_i] = \lambda_i\int_{-\infty}^{\infty}\Phi\left[A_i\right]\exp(\sigma\varepsilon_i)\,\phi(\varepsilon_i)d\varepsilon_i$$

which must be computed either by simulation or quadrature. The partial effects are computed likewise. For the variables in the primary equation,

(3.2-12) $$\partial E[y_i \mid \mathbf{x}_i,\mathbf{w}_i]/\partial\mathbf{x}_i = \lambda_i\,\boldsymbol{\beta}\int_{-\infty}^{\infty}\Phi\left[A_i\right]\exp(\sigma\varepsilon_i)\,\phi(\varepsilon_i)d\varepsilon_i = \mathbf{g}_{\mathbf{x}}.$$

For the variables in the regime equation,

(3.2-13) $$\partial E[y_i \mid \mathbf{x}_i,\mathbf{w}_i]/\partial\mathbf{w}_i = \lambda_i\left(\frac{1}{\sqrt{1-\rho^2}}\right)\boldsymbol{\delta}\left\{\int_{-\infty}^{\infty}\phi\left[A_i\right]\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i\right\} = \mathbf{g}_{\mathbf{w}}.$$

Note that in the expression above, if the correlation, ρ, equals zero, then the conditional mean for the (only) heterogeneous ZIP model becomes

$$\text{(3.2-14)} \qquad E[y_i|\, \mathbf{x}_i,\mathbf{w}_i] = \lambda_i\ \Phi\left(\mathbf{w}_i'\delta\right)\int_{-\infty}^{\infty}\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i$$

$$= \lambda_i\ \Phi\left(\mathbf{w}_i'\delta\right)\exp(\sigma^2/2)$$

and the partial effects simplify considerably. Jacobians for these vectors of partial effects are given in Appendix C.

The original ZIP or ZINB model is restored if ρ equals zero *and* σ equals zero. A ZIP model with heterogeneity in $P(y_i|\mathbf{x}_i,\varepsilon_i)$ results if only ρ equals zero. (A nonzero ρ with a zero σ is internally inconsistent.) The ZIP model with heterogeneity has appeared elsewhere in the literature, in the form of random effects in a panel data application (see Hur (1998), Hall (2000) and Xie, Hur and McHugo (2006)]. This appears to be the first application that relaxes the restriction of zero correlation across the two equations.

3.3 Hurdle Models

The hurdle model is also a two part decision model. The first part is a participation equation and the second is an event count, conditioned on participation. Formally, the model can be constructed as follows:

(3.3-1a)
$$d_i^* = \mathbf{w}_i'\delta + u_i,$$
$$d_i = \mathbf{1}(d_i^* > 0),$$
$$\text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\delta) \quad \text{(hurdle equation)},$$
$$\text{Prob}(d_i = 1|\mathbf{w}_i) = \Phi(\mathbf{w}_i'\delta) \quad \text{(probit hurdle model)},$$

(3.3-1b)
$$y_i\, |\mathbf{x}_i,(d_i = 0) = \text{unobserved} \quad \text{(nonparticipation)},$$
$$y_i\, |\mathbf{x}_i,(d_i = 1) \sim P{+}(y_i \mid \mathbf{x}_i) \quad \text{(truncated Poisson or NB model given participation)}.$$

(Note, we distinguish between y_i = "unobserved" and y_i = 0 in the nonparticipation case.) The central feature of the model is the effect of the hurdle decision on the event count equation, which we denote $P{+}(y_i \mid \mathbf{x}_i)$. If $d_i = 1$, then by the construction, $y_i > 0$. Thus, the resulting count model has the truncated form. [See Terza (1985), Econometric Software (2007), Greene (2008).] The underlying motivation is similar to the latent class interpretation in the preceding section.

To obtain a likelihood for the hurdle model, we first obtain the joint density for y_i and d_i in this specification. Since nonzero values of y_i are only observed when $d_i = 1$, we can write

(3.3-2)
$$\text{Prob}(y_i \text{ is unobserved}|\, \mathbf{x}_i,d_i = 0) = 1,\ \text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\delta)$$

$$P(y_i|\mathbf{x}_i,d_i = 1) = P{+}(y_i \mid \mathbf{x}_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\left[1-\exp(-\lambda_i)\right]\Gamma(1+y_i)},\ y_i = 1,2,...,$$

$$\text{Prob}(d_i = 1|\mathbf{w}_i) = 1 - \Pi_0(\mathbf{w}_i'\delta)$$

Combining terms in the familiar fashion and once again, maintaining the Poisson model for convenience, we have

(3.3-3) $$P(y_i, d_i \mid \mathbf{x}_i, \mathbf{w}_i) = (1\text{-}d_i)\,\Pi_0(\mathbf{w}_i'\delta) + d_i \frac{[1-\Pi_0(\mathbf{w}_i'\delta)]}{[1-\exp(-\lambda_i)]}\frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\Gamma(1+y_i)}, y_i = 1,2,...$$

(Note, again, it is understood that in the $d_i = 0$ regime, y_i is unobserved; it is not assumed to be equal to zero.) The log likelihood takes a convenient form for this case. Taking the two parts separately, we find

(3.3-4)
$$\begin{aligned}\ln L &= \sum\nolimits_{d_i=0} \ln\Pi_0(\mathbf{w}_i'\delta) + \sum\nolimits_{d_i=1} \ln[1-\ln\Pi_0(\mathbf{w}_i'\delta)] - \ln[1-\exp(-\lambda_i)] + \ln P(y_i \mid \mathbf{x}_i)\\ &= \left\{\sum\nolimits_{d_i=0} \ln\Pi_0(\mathbf{w}_i'\delta) + \sum\nolimits_{d_i=1} \ln[1-\ln\Pi_0(\mathbf{w}_i'\delta)]\right\} + \\ &\quad \left\{\sum\nolimits_{d_i=1} \ln P(y_i \mid \mathbf{x}_i) - \ln[1-\exp(-\lambda_i)]\right\}\end{aligned}$$

The first term in braces is the log likelihood for the binary choice model (probit or logit) for d_i. The second term is the log likelihood for the truncated (at zero) Poisson (or NB) model. Thus, the hurdle model can be estimated in two independent parts. (This will not be true when we extend the model below.) The conditional mean function in the truncated Poisson model is

(3.3-5) $$E[y_i \mid \mathbf{x}_i, d_i = 1) = \frac{\exp(\alpha + \mathbf{x}_i'\beta)}{1-P(0 \mid \mathbf{x}_i)}.$$

Therefore,

(3.3-6) $$E[y_i \mid \mathbf{x}_i, \mathbf{w}_i] = \sum\nolimits_{d_i=0}^{1}\sum\nolimits_{y_i=1}^{\infty} y_i P(y_i, d_i \mid \mathbf{x}_i, \mathbf{w}_i) = \frac{[1-\Pi_0(\mathbf{w}_i'\delta)]}{[1-\exp(-\lambda_i)]}\lambda_i.$$

As usual, the alteration of the distribution carries through to the partial effects;

(3.3-7)
$$\partial E[y_i \mid \mathbf{x}_i, \mathbf{w}_i, d_i]/\partial \mathbf{x}_i = \frac{[1-\Pi_0(\mathbf{w}_i'\delta)]}{[1-\exp(-\lambda_i)]}\left(1-\frac{\lambda_i\exp(-\lambda_i)}{1-\exp(-\lambda_i)}\right)\lambda_i\boldsymbol{\beta} = \mathbf{g}_\mathbf{x},$$
$$\partial E[y_i \mid \mathbf{x}_i, \mathbf{w}_i, d_i]/\partial \mathbf{w}_i = \frac{-\mathrm{d}\Pi_0(\mathbf{w}_i'\delta)/\mathrm{d}(\mathbf{w}_i'\delta)}{[1-\exp(-\lambda_i)]}\lambda_i\boldsymbol{\delta} = \mathbf{g}_\mathbf{w}.$$

Derivatives of these partial effects for use in computing standard errors are given in Appendix D.

To relax the restriction that the two decisions are uncorrelated, we use the same device as before and now assume joint normality for the underlying heterogeneity. The extended model is

(3.3-8a)	$d_i^* = \mathbf{w}_i'\delta + u_i,\ u_i \sim N[0,1]$,		
	$d_i = \mathbf{1}(d_i^* > 0)$,		
	$\text{Prob}(d_i = 0 \mid \mathbf{w}_i) = 1 - \Phi(\mathbf{w}_i'\delta)$	(probit hurdle equation),	
(3.3-8b)	$y_i \mid \mathbf{x}_i, \varepsilon_i, (d_i = 0) =$ unobserved	(nonparticipation),	
	$y_i \mid \mathbf{x}_i, \varepsilon_i, (d_i = 1) \sim P+(y_i \mid \mathbf{x}_i, \varepsilon_i)$	(truncated Poisson or NB model),	
	$[u_i, \varepsilon_i] \sim N[(0,1),(1,1),\rho]$.		

Using the same devices as in the earlier derivations, we have, (3.3-9) $\quad P(y,d_i|\mathbf{x}_i,\mathbf{w}_i,\varepsilon_i) = (1\text{-}d_i)\ \Phi\left[-\mathbf{w}_i'\boldsymbol{\eta}-\tau\varepsilon_i\right] +$

$$d_i\frac{\Phi\left[\mathbf{w}_i'\boldsymbol{\eta}+\tau\varepsilon_i\right]}{\left[1-\exp(-\exp(\sigma\varepsilon_i)\lambda_i)\right]}\frac{\exp(-\exp(\sigma\varepsilon_i)\lambda_i)\left[\exp(\sigma\varepsilon_i)\lambda_i\right]^{y_i}}{\Gamma(1+y_i)}$$

and, finally,

(3.3-10)) $\quad \ln L = \sum_{i=1}^{N} \ln\int_{-\infty}^{\infty} P(y_i,d_i \mid \mathbf{x}_i,\mathbf{w}_i,\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i\,.$

For analysis of the partial effects, the conditional mean will be

(3.3-11) $$E[y_i|\mathbf{x}_i,\mathbf{w}_i] = \int_{-\infty}^{\infty}\Phi\left(\frac{\mathbf{w}_i'\boldsymbol{\delta}+\rho\varepsilon_i}{\sqrt{1-\rho^2}}\right)\frac{\left[\exp(\sigma\varepsilon_i)\lambda_i\right]}{\left[1-\exp(-\exp(\sigma\varepsilon_i)\lambda_i)\right]}\phi(\varepsilon_i)d\varepsilon_i\,.$$

From this point, estimation and analysis of the partial effects proceeds in the same fashion as in the preceding two sections. Note, in all three cases, the differences in the models consists of the density for y_i in the template function as in (3.1-15) or (3.1-18). The partial effects and derivatives for this model are extremely cumbersome. They are presented in Appendix E.

4. Models for Panel Data

We consider the most familiar treatments for panel data, the fixed and random effects models. For each of these, a separate set of results for Poisson and NB models have come into common use. These build on the familiar treatments in the linear model, but for the treatments in common use, only the Poisson FE model follows along the familiar lines.

4.1 Fixed Effects Poisson and NB Models

The now standard Poisson fixed effects model,

(4.1-1) $$P(y_{it}|\mathbf{x}_{it}) = \frac{\exp(-\lambda_{it})\lambda_{it}^{y_{it}}}{\Gamma(1+y_{it})},\lambda_{it}=\exp(\alpha_i+\mathbf{x}_{it}'\boldsymbol{\beta}),$$

is one of only a few known cases in which maximization of the full log likelihood with respect to $(\alpha_i, i=1,...,N, \boldsymbol{\beta})$ produces the numerically identical result for $\boldsymbol{\beta}$ as maximization of the conditional log likelihood based on

(4.1-2) $$P(y_{i1},...,y_{iT}|\Sigma_{t=1}^{T}y_{it},\mathbf{X}_i) = \frac{\Gamma\left(1+\Sigma_{t=1}^{T}y_{it}\right)}{\prod_{t=1}^{T}\Gamma(1+y_{it})}\prod_{t=1}^{T}A_{it}^{y_{it}}$$

where

(4.1-3) $$A_{it}=\frac{\lambda_{it}}{\Sigma_{t=1}^{T}\lambda_{it}}=\frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}{\Sigma_{t=1}^{T}\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}.$$

[See Lancaster (2000).] Note that the conditional log likelihood does not involve the constant terms, α_i. Nonetheless, the $\boldsymbol{\beta}$ that maximizes (with the solutions for α_i) the unconditional log likelihood, (4.1-4) $\ln L = \sum_{i=1}^{N}\sum_{t=1}^{T} \ln P(y_{it}|\mathbf{x}_{it})$

is numerically identical to the maximizer of the conditional log likelihood,

(4.1-5) $\ln L_C = \sum_{i=1}^{N} \ln P(y_{i1},...,y_{iT}|\Sigma_{t=1}^{T} y_{it},\mathbf{X}_i)$.

This is the commonly used form of the Poisson model that is built into widely used commercial software such as Stata, SAS and LIMDEP.

To set the stage for the development below, consider the implication of a time invariant variable in $\mathbf{x}_{it}$. If the conditional mean is written in the implied form

(4.1-6) $\lambda_{it} = \exp\left(\Sigma_{i=1}^{N}\alpha_i f_{it} + z_i\delta + \mathbf{x}_{it}'\boldsymbol{\beta}\right)$

where f_i are the individual specific dummy variables, then it becomes obvious on inspection that the model is fundamentally unidentified. For each i, the T observations, $\mathbf{z}_i$ is are a multiple of the T nonzero observations in variable $\mathbf{f}_i$. The NT observations in the column vector $\mathbf{z}$ can be replicated by a linear combination of the N dummy variables. The impact of this form of multicollinearity on a nonlinear model can be seen as follows: The log likelihood function for the fixed effects Poisson regression model is

(4.1-7) $\ln L = \sum_{i=1}^{N}\sum_{t=1}^{T}\ln P(y_{it} \mid \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})$

where the density appears in (1) The likelihood equation for a time invariant variable $x_{it,k} = z_i$ is

(4.1-8) $\partial\ln L/\partial\beta_k = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \lambda_{it})z_i$

and the likelihood equation for each of the fixed effects coefficients is

(4.1-9) $\partial\ln L/\partial\alpha_i = \sum_{t=1}^{T}(y_{it} - \lambda_{it})$

[see (2.1-5)]. Therefore, at any parameter vector, $\partial\ln L/\partial\beta_k = \Sigma_{i=1}^{N} z_i\partial\ln L/\partial\alpha_i$. The $N+K$ columns of the derivatives of the log likelihood function are linearly dependent; the model is not identified. For example, this precludes separate estimation of the fixed effects model as shown with an additional overall constant. (This merely reinforces the widely understood principle that fixed effects models cannot include time invariant independent variables.

Hausman, Hall and Griliches (1984) (HHG) report the following conditional density for the fixed effects negative binomial (FENB) model:

(4.1-10) $p\left(y_{i1}, y_{i2}, \dots, y_{iT_i} \mid \mathbf{X}_i, \Sigma_{t=1}^{T} y_{it}\right) = \dfrac{\Gamma(1+\Sigma_{t=1}^{T} y_{it})\Gamma(\Sigma_{t=1}^{T}\lambda_{it})}{\Gamma(\Sigma_{t=1}^{T} y_{it} + \Sigma_{t=1}^{T}\lambda_{it})}\prod_{t=1}^{T}\dfrac{\Gamma(y_{it}+\lambda_{it})}{\Gamma(1+y_{it})\Gamma(\lambda_{it})}$,

which is free of the fixed effects. This is the default FENB formulation used in popular software packages such as Stata, SAS and LIMDEP. Researchers accustomed to the earlier admonishment that fixed effects models cannot contain overall constants or time invariant covariates are sometimes surprised to find (perhaps accidentally) that this fixed effects model allows both. [This issue is explored at length in Allison (2000) and Allison and Waterman (2002).] The resolution of this apparent contradiction is that the HHG FENB model is not obtained by shifting the conditional mean function by the fixed effect, $\ln\lambda_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i$, as it is in the Poisson model and in other familiar models. Rather, the HHG model is obtained by building the fixed effect into the model as an individual specific θ_i in the Negbin 1 form in (2.4-3). In the two negative binomial models, the conditional mean functions are

$$\text{(4.1-11)} \quad \begin{aligned} &\text{NB1(HHG):}\ \ E[y_{it} \mid \mathbf{x}_{it}] = \phi_{it} = \exp(\alpha + \mathbf{x}_{it}'\boldsymbol{\beta}), \\ &\text{NB2(FE):}\ \ \ \ E[y_{it} \mid \mathbf{x}_{it}] = \exp(\alpha_i)\phi_{it} = \lambda_{it} = \exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}), \end{aligned}$$

Thus, the conditional mean function in the HHG model is homogeneous. The fixed effect in the model is introduced through the scaling parameter, θ_i, which enters the conditional variance of the random variable;

$$\text{(4.1-12)} \quad \begin{aligned} &\text{NB1(HHG):}\ \ \text{Var}[y_{it} \mid \mathbf{x}_{it}] = \phi_{it}[1 + (1/\theta_i)], \\ &\text{NB2(FE):}\ \ \ \ \text{Var}[y_{it} \mid \mathbf{x}_{it}] = \lambda_{it}[1 + (1/\theta)\lambda_{it}]. \end{aligned}$$

The relationship between the mean and the variance is quite different for the two models. For estimation purposes, one can explain the apparent contradiction noted earlier by observing that in the NB1 formulation, the individual effect is built into the scedastic (scaling) function, not the conditional mean. (In principle, given this finding, one could have a second set of fixed effects, in the mean of the HHG model.) Greene (2007a) analyzes the more familiar, FENB2 form with the same treatment of λ_{it}. Estimates for both models appear below.

Theory does not provide a reason to prefer the NB1 formulation over the more familiar NB2 model. The NB1 form does extend beyond the interpretation of the fixed effect as carrying only the sum of all the time invariant heterogeneity into the conditional mean function. The appearance of $\ln\theta_i$ in the conditional mean is an artifact of the exponential mean form; θ_i is a scaling parameter in this model. In its favor, the HHG model, being conditionally independent of the fixed effects, finesses the incidental parameters problem – the estimator of $\boldsymbol{\beta}$ in this model is consistent. This is not the case for the FENB2 form. But, it remains unclear what role the fixed effects play in this model, and how they relate to the fixed effects in other familiar treatments.

The conditional NB1 model obviates brute force maximization of the unconditional NB2 (or NB1) log likelihood function with respect to $\boldsymbol{\beta}$ and all N constants α_i, which is a significant practical advantage (notwithstanding the incidental parameter problem). However, Greene (2004) provides a solution to this problem that enables the computation even with large N. The estimates below are based on this method.

4.2 Random Effects Models

The random effects Poisson model can be formed by writing

$$\text{(4.2-1)} \qquad \lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i)$$

where u_i is independent of $\mathbf{x}_{it}$. Under the assumption that u_i has a log gamma density with $\exp(u_i) \sim G(\theta,\theta)$ as earlier in the cross section case, the unconditional joint density for individual i is

$$\text{(4.2-2)} \qquad P(y_{i1},y_{i2},\ldots,y_{iT}|\mathbf{X}_i) = \frac{\left[\Pi_{t=1}^{T}\lambda_{it}^{y_{it}}\right]\Gamma\left[\theta + \Sigma_{t=1}^{T}y_{it}\right]}{\Gamma(\theta)\left[\Pi_{t=1}^{T}\Gamma(1+y_{it})\right]\left[\left(\Sigma_{t=1}^{T}\lambda_{it}\right)^{\sum_{t=1}^{T}y_{it}}\right]} Q_i^{\theta}(1-Q_i)^{\sum_{t=1}^{T}y_{it}}$$

where $$Q_i = \frac{\theta}{\theta + \Sigma_{t=1}^{T}\lambda_{it}}.$$

This is a negative binomial distribution for $Y_i = \Sigma_{t=1}^{T} y_{it}$ with mean $\Lambda_i = \Sigma_{t=1}^{T} \lambda_{it}$. As noted earlier, the choice of the log gamma formulation is motivated by mathematical convenience, not by an appeal to an underlying model of heterogeneity. The Poisson RE model could also be specified with lognormal heterogeneity. Analysis would follow precisely along the lines of Section 2.3. The joint probability would be computed from

(4.2-3)
$$P(y_{i1},...,y_{iT} \mid \mathbf{X}_i) = \int_{u_i} \prod_{t=1}^{T} \frac{\exp(-\exp(u_i)\lambda_{it})(\exp(u_i)\lambda_{it})^{y_{it}}}{\Gamma(1+y_{it})} f(u_i)du_i$$
$$= \prod_{\mathrm{t}=1}^{\mathrm{T}} \left[\frac{\lambda_{it}^{y_{it}}}{\Gamma(1+y_{it})}\right] \int_{u_i} \exp\left[-\exp(u_i)\Sigma_{t=1}^{T}\lambda_{it}\right]\left[\exp(u_i)\right]^{\Sigma_{t=1}^{T} y_{it}} f(u_i)du_i.$$

This function and its derivatives can be approximated using either quadrature or simulation.

Like the fixed effects model, introducing random effects into the negative binomial model adds some additional complexity. Since the negative binomial model derives from the Poisson model by adding latent heterogeneity to the conditional mean, adding a random effect to the negative binomial model could be viewed as introducing the heterogeneity a second time. However, an approach that would preserve the form of the model would be to begin with a Poisson model and write

(4.2-4) $$\lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + u_i),$$

where both ε_{it} and u_i are log gamma distributed with parameters θ and μ, respectively. This would correspond to the mixed negative binomial model at the end of Section 2.3 [and the model used in RWM (2003) – see (5-12) below]. Departing from (4.2-4), if it is assumed that ε_{it} has the $G(\theta,\theta)$ distribution assumed in Section 2 and u_i has a normal distribution, then we obtain a "true" random effects model that parallels the fixed effects treatments developed earlier. The conditional negative binomial model will result from

(4.2-5) $$P(y_{it}|\mathbf{x}_{it},u_i) = \int_{\varepsilon_{it}} P(y_{it} \mid x_{it},\varepsilon_i,u_i) f(\varepsilon_{it}) d\varepsilon_{it}.$$

Changing the variable to $h_{it} = \exp(\varepsilon_{it})$ and integrating over h_{it} instead produces the negative binomial model with conditional mean $E[y_{it}|\mathbf{x}_{it},u_i] = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i)$ and dispersion parameter θ. The resulting conditional density is

(4.2-6)
$$P(y_{it}|\mathbf{x}_{it},u_i) = \frac{\Gamma(\theta+y_{it})}{\Gamma(1+y_{it})\Gamma(\theta)} r_{it}^{\theta}(1-r_{it})^{y_{it}},$$
$$\lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta}),$$
$$r_{it} = \theta/(\theta + \exp(u_i)\lambda_{it}).$$

We can then estimate the parameters by forming the conditional (on u_i) log likelihood and integrating u_i out either by quadrature or simulation.

Hausman et al.'s (1984) random effects negative binomial model is a hierarchical model that derives from a heterogeneous Poisson model. The mean in the Poisson model is $\exp(u_i)\lambda_{it}$

where $\exp(u_i)$ has $G(\theta,\theta)$ density. This produces the NB kernel. The unconditional distribution is obtained by treating $p_{it} = [\exp(u_i)\lambda_{it}]/[\Sigma_t \exp(u_i)\lambda_{it}]$ as a random vector with Dirichlet mixing distribution. Each pair of means, $\mu_{it} = \exp(u_i)\lambda_{it}$ $\mu_{is} = \exp(u_i)\lambda_{is}$ is such that $\mu_{it}/(\mu_{it}+\mu_{is})$ has a beta distribution with parameters a and b. The resulting unconditional density is

$$p(y_{i1}, y_{i2}, ..., y_{iT} \mid \mathbf{X}_i) = \frac{\Gamma(a+b)\Gamma\left(a + \Sigma_{t=1}^{T}\lambda_{it}\right)\Gamma\left(b + \Sigma_{t=1}^{T} y_{it}\right)}{\Gamma(a)\Gamma(b)\Gamma\left(a + \Sigma_{t=1}^{T}\lambda_{it} + b + \Sigma_{t=1}^{T} y_{it}\right)}. \tag{4.2-7}$$

This is the common form of the RENB model that is incorporated in several contemporary computer packages. As before, the relationship between the heterogeneity and the conditional mean function is unclear, and there is no obvious interpretation of the hyperparameters a and b. The parameters are simpler to interpret in the effects model in (4.2-4), where the estimated standard deviation of u_i can be directly interpreted against the other parameters in the model. Moreover, the HHG model does not admit of a ready test of the homogeneous model. It is unclear what the implication of $a = b = 0$ would be. Estimates of the two forms of the random effects model are presented in Section 6 for a comparison.

5. The Bivariate Poisson Model

There have been a variety of proposals for a bivariate (or multivariate) count data model. The earliest form is that of Kocherlakota and Kocherlakota (1992) which is based on the trivariate reduction method. Let z_1, z_2 and u denote three Poisson distributed random variables. Then, the observed random variables are

(5-1) and

$$y_1 = z_1 + u$$
$$y_2 = z_2 + u$$

have a bivariate Poisson distribution with covariance equal to $\mathrm{Var}[u]$. This model does produce a pair of correlated Poisson variables, however the correlation must be positive, which severely limits the generality of this specification. (For the outcomes examined in Section 6, doctor visits and hospital visits, a negative correlation would not be surprising.) [24] Munkin and Trivedi (1999) present a survey of other approaches.

Two recently developed approaches considered here [Munkin and Trivedi (1999) and RWM (2003)] build the bivariate model into latent heterogeneity structures, as employed in the various models proposed above. These allow the sign of the correlation to vary. However, they shift the impact of the bivariate distribution from the variables of interest, as in the trivariate model above, to the unobservables in the conditional mean function. The bivariate count outcomes model is still preserved. However, the estimated correlations in these models do not provide a clear picture of the implied correlations between the outcome variables that was the objective to begin with. As a general proposition, the correlation between the observed counts will be less, potentially far less, than the estimated correlation between the underlying unobserved heterogeneity.

The bivariate probit model specified in Munkin and Trivedi (1999) and Riphahn, Rambach and Million (2003) is

$$\exp(\mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \sigma_1\varepsilon_{1i}) = \lambda_{1i}\exp(\sigma_1\varepsilon_{1i})$$
$$\exp(\mathbf{x}_{2i}'\boldsymbol{\beta}_2 + \sigma_2\varepsilon_{2i}) = \lambda_{2i}\exp(\sigma_2\varepsilon_{2i})$$

[24] The trivariate reduction method was employed e.g., by Jung and Winkelmann (1993), Karlis and Ntzoufras (2003) and King (1989).

(5-2) $(\varepsilon_{1i},\varepsilon_{2i}) \sim N_2[(0,0),(1,1),\rho]$

$$P(y_{ji}|\mathbf{x}_{ji},\varepsilon_{ji}) = \frac{\exp(-\exp(\sigma\varepsilon_{ji})\lambda_{ji})(\exp(\sigma\varepsilon_{ji})\lambda_{ji})^{y_{ji}}}{\Gamma(1+y_{ji})}, j=1,2$$

Both studies build the empirical measurement of correlation of the two outcomes around the estimation of ρ. However, as we now demonstrate, the correlation coefficient, ρ, provides a misleading description of this correlation. Superficially, this is obvious from the construction. The coefficient ρ is not the correlation between y_{1i} and y_{2i}; it is the correlation between $\ln E[y_{1i}|\mathbf{x}_{1i},\varepsilon_{1i}]$ and $\ln E[y_{2i}|\mathbf{x}_{2i},\varepsilon_{2i}]$. How this relates to $\text{Corr}[y_{1i},y_{2i}|\mathbf{x}_{1i},\mathbf{x}_{2i}]$ is less than clear. To deduce this from the model specification, we proceed as follows:

$$(5\text{-}3)\quad \begin{aligned} Corr[y_{1i},y_{2i}\,|\,\mathbf{x}_{1i},\mathbf{x}_{2i}] &= \frac{Cov[y_{1i},y_{2i}\,|\,\mathbf{x}_{1i},\mathbf{x}_{2i}]}{\sqrt{Var[y_{1i}\,|\,\mathbf{x}_{1i}]}\sqrt{Var[y_{2i}\,|\,\mathbf{x}_{2i}]}} \\ &= \frac{E[Cov[y_{1i},y_{2i}\,|\,\mathbf{x}_{1i},\mathbf{x}_{2i},\varepsilon_{1i},\varepsilon_{2i}]] + Cov[E[y_{1i}\,|\,\mathbf{x}_{1i},\varepsilon_{1i}],E[y_{2i}\,|\,\mathbf{x}_{2i},\varepsilon_{2i}]]}{\sqrt{E[Var[y_{1i}\,|\,\mathbf{x}_{1i},\varepsilon_{1i}]]+Var[E[y_{1i}\,|\,\mathbf{x}_{1i},\varepsilon_{1i}]]}\sqrt{E[Var[y_{2i}\,|\,\mathbf{x}_{2i},\varepsilon_{2i}]]+Var[E[y_{2i}\,|\,\mathbf{x}_{2i},\varepsilon_{2i}]]}}. \end{aligned}$$

For convenience, let

(5-4) $\mu_{ji} = \lambda_{ji}\exp(\sigma_j^2/2)$.

The terms in the denominator were derived earlier. The unconditional variance is

$$(5\text{-}5)\quad \begin{aligned} \text{Var}[y_{ji}|\mathbf{x}_{ji}] &= \lambda_{ji}\exp(\sigma_j^2/2)[1 + \lambda_{ji}\exp(\sigma_j^2/2)[\exp(\sigma_j^2)-1]] \\ &= \mu_{ji}\,\{1 + \mu_{ji}[\exp(\sigma_j^2)-1]\}, j = 1,2. \end{aligned}$$

For the terms in the numerator, the first is zero, since conditioned on ε_{1i} and ε_{2i}, y_{1i} and y_{2i} (given $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$) are independent. Thus, what remains to derive is

$$(5\text{-}6)\quad \begin{aligned} Cov\{E[y_{1i}\,|\,\mathbf{x}_{1i},\varepsilon_{1i}],E[y_{2i}\,|\,\mathbf{x}_{2i},\varepsilon_{2i}]\} &= Cov\{[\lambda_{1i}\exp(\sigma_1\varepsilon_{1i})],[\lambda_{2i}\exp(\sigma_2\varepsilon_{2i})]\} \\ &= (\lambda_{1i}\lambda_{2i})Cov\{[\exp(\sigma_1\varepsilon_{1i})],[(\exp(\sigma_2\varepsilon_{2i})]\} \\ &= (\lambda_{1i}\lambda_{2i})\{E[\exp(\sigma_1\varepsilon_{1i})\exp(\sigma_2\varepsilon_{2i})] - E[\exp(\sigma_1\varepsilon_{1i})]E[\exp(\sigma_2\varepsilon_{2i})]\}. \end{aligned}$$

The two conditional means are the means for the univariate lognormals,

(5-7) $E[\exp(\sigma_j\varepsilon_{ji})] = \exp(\sigma_j^2/2)$.

The remaining term is straightforward;

$$(5\text{-}8)\quad \begin{aligned} E[\exp(\sigma_1\varepsilon_{1i})\exp(\sigma_2\varepsilon_{2i})] &= E[\exp(\sigma_1\varepsilon_{1i} + \sigma_2\varepsilon_{2i})] \\ &= \exp[(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)/2]. \end{aligned}$$

Combining terms and manipulating the expression produces

(5-9) $$Cov[y_{1i}, y_{2i} \mid x_{1i}, x_{2i}] = \left[\lambda_{1i} \exp(\sigma_1^2/2)\right]\left[\lambda_{2i} \exp(\sigma_2^2/2)\right]\left[\exp(\rho\sigma_1\sigma_2) - 1\right]$$
$$= \mu_{1i}\mu_{2i}\left[\exp(\rho\sigma_1\sigma_2) - 1\right].$$

Combining all terms and simplifying (slightly), we obtain the final result

(5-10) $$Corr(y_{1i}, y_{2i} \mid \mathbf{x}_{1i}, \mathbf{x}_{2i}) = \frac{\sqrt{\mu_{1i}\mu_{2i}}\left[\exp(\rho\sigma_1\sigma_2) - 1\right]}{\sqrt{1 + \mu_{1i}[\exp(\sigma_1^2) - 1]}\sqrt{1 + \mu_{2i}[\exp(\sigma_2^2) - 1]}}$$

How this relates to ρ is unclear. It has the same sign, but the magnitudes are likely to be essentially unrelated. We will examine it in the application below.

Munkin and Trivedi (1999) also develop a bivariate negative binomial model. The model is constructed as the joint density for two Poisson models with common heterogeneity, v_i. That is,

(5-11) $$y_{ji} \sim P(y_{ji} \mid \mathbf{x}_{ji}, v_i)$$
where
$$\lambda_{ji} = \exp(\alpha + \mathbf{x}_{ji}'\boldsymbol{\beta} + v_i),\ j = 1,2.$$

For $h_i = \exp(v_i)$ distributed with Gamma(θ,θ) distribution, this is precisely equivalent to the two period random effects Poisson model shown in Section 4.2. Although the functional form for the log likelihood function is known (see Section 4.2) (and it is given in the paper), the authors used simulation to estimate the parameters of the model. Since they assume the value of θ (one), the results cannot be compared to the Poisson – lognormal mixture model. One could, instead, form the bivariate NB model using correlated lognormal mixtures along the lines suggested at the end of Section 2.3. We leave this derivation for future research.

Finally, RWM (2003) extended this development to a panel data setting. Their random effects model is

$$\ln\lambda_{it,1} = \alpha_1 + \mathbf{x}_{it,1}'\boldsymbol{\beta}_1 + u_{i,1} + \varepsilon_{it,1} = \alpha_1 + \mathbf{x}_{\text{it},1}'\boldsymbol{\beta}_1 + v_{it,1},$$
$$\ln\lambda_{it,2} = \alpha_2 + \mathbf{x}_{it,2}'\boldsymbol{\beta}_2 + u_{i,2} + \varepsilon_{it,2} = \alpha_2 + \mathbf{x}_{it,2}'\boldsymbol{\beta}_2 + v_{it,2},$$
(5-12) $$(\varepsilon_{it,1},\varepsilon_{it,2}) \sim N_2[(0,0),(\sigma_1,\sigma_2),\rho],$$
$$(u_{i,1},u_{i,2}) \sim N_2[(0,0),(\omega_1,\omega_2),0].$$

The correlation between $\varepsilon_{it,1}$ and $\varepsilon_{it,2}$ creates the bivariate model. In the notation of our earlier formulation, the correlation of interest, between $v_{it,1}$ and $v_{it,2}$, is

(5-13) $$\rho_{12} = \frac{\rho\sigma_1\sigma_2}{\sqrt{\omega_1^2 + \sigma_1^2}\sqrt{\omega_2^2 + \sigma_2^2}}.$$

And the counterparts to σ_1 and σ_2 are the two terms in the denominator. The model also implies a T-variate Poisson-lognormal mixture model for each group for each of the two variables. The implied correlation is $\rho_{ts,j} = \omega_j^2/(\omega_j^2 + \sigma_j^2)$, $j = 1,2$. As they note and discuss, ρ is the correlation between the unique unobservable factors in the two equations. One could, however, misinterpret the magnitude of the value as representative of the correlation between the composed heterogeneity or, worse yet, between the outcome variables, themselves. For example, for their equation system applied to the males in their sample, they report $\rho = 0.599$, $\sigma_1 = 0.996$, $\sigma_2 = 1.244$, $\omega_1 = 0.795$ and $\omega_2 = 1.195$. The computation above produces $\rho_{12} = 0.276$. The calculation is relevant because the unobservable propensities are difficult to partition neatly into time varying and time invariant parts.

It is speculative to assume that ρ in isolation captures the full correlation of the unobservables apart from persistent, time invariant components (and leaves $u_{i,j}$ truly unexplained). We will revisit the computation of the implied correlation between the two outcomes below.

6. Applications

In "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation," Riphahn, Wambach and Million (2003) employed a part of the German Socioeconomic Panel (GSOEP) data set to analyze two count variables, *DocVis*, the number of doctor visits in the last three months and *HospVis*, the number of hospital visits in the last year. The authors employed a bivariate panel data (random effects) Poisson model to study these two outcome variables. A central focus of the investigation was the role of the choice of private health insurance in the intensity of use of the health care system, i.e., whether the data contain evidence of moral hazard. We will use these data to illustrate the model extensions described above.[25] The authors of this study presented estimates for the Poisson-lognormal model in Section 2.2.2 and the bivariate Poisson model in (5-12). We will analyze the single equation and two part models in some detail, but only analyze the correlation structure developed for the bivariate Poisson model in Section 5. (We have not proposed any extensions for this model; our analysis has only provided a more detailed interpretation of the existing model results.) In order to keep the amount of reported results to a manageable size, we will also restrict attention to *DocVis*, the count of doctor visits. Analysis of the count of hospital visits is left for further research.

6.1 The Data

The RWM data set is an unbalanced panel of 7,293 individual families observed from one to seven times. The number of observations varies from one to seven (1,525, 1,079, 825, 926, 1,051, 1000, 887) with a total number of observations of 27,326. The composition of the panel is shown in Figure 1.

The variables in the data file are listed in Table 1 with descriptive statistics for the full sample. They estimated separate equations for males and females and did not report any estimates based on the pooled data. Table 2 reports descriptive statistics for the two subsamples. The figures given all match those reported by RWM. (See their Table II, page 393.) The outcome variables of interest in the study were doctor visits in the last three months and number of hospital visits last year. Histograms for these variables for the full data set are shown in Figures 2 and 3. (Figure 1 was truncated at 20 visits. Figure 2 was truncated at 10. These remove about 200 observations from the sample used to form the figures.)

The base case count model used by the authors included the following variables in addition to the constant term:

$$\mathbf{x}_{it} = (Age, Agesq, HSat, Handdum, Handper, Married, Educ, Hhninc, Hhkids, Self, Civil, Bluec, Working, Public, AddOn)$$

and a set of year effects,

$$\mathbf{t} = (YEAR1985, YEAR1986, YEAR1987, YEAR1988, YEAR1991, YEAR1994).$$

[25] The raw data are published and available for download on the *Journal of Applied Econometrics* data archive website, The URL is given below Table 1.

The same specification was used for both *DocVis* and *HospVis*. We will use their specification in our count models. The estimated year effects are omitted from the reported results in the paper. The variables used in the participation equation in the two part models are discussed in Section 6.3.

6.2 Functional Forms and Heterogeneity

Table 3 presents estimates of the Poisson regression models for males and females. The pooled (across genders and across time) results appear in the first column. We tested for homogeneity of the coefficient vectors for males and females using a likelihood ratio test; the chi squared statistic is

$$\lambda_{LR} = 2[90097.4 - (42927.6+46275.1)] = 1789.4.$$

This is substantially larger than the critical chi squared with 16 degrees of freedom (26.30), so the hypothesis that the same model applies to males and females is rejected for the Poisson model. The Poisson specification is, itself, rejected in favor of a model with heterogeneity, so we repeated the homogeneity test with the log gamma (negative binomial) results. The log likelihood for the pooled data is -58082.0 – the pooled NB results are not shown – so the LR statistic for the NB model is 678.60, with 17 degrees of freedom. On this basis, we will not use the pooled data in any of the models estimated below. For brevity, we will present only the results for the males in the sample (n = 14,243). (Qualitative results for the two samples are the same. RWM do not pursue the differences in the results for males and females.)

The immediate impression is that the presence of public insurance and private addon insurance in the pooled model both have a significant influence on usage of physician visits. However, when the models are fit separately for males and femalse, the latter effect is dissipated. It appears that generally, the former effects disappears from the models that account for latent heterogeneity – of the four sets of results in Table 3, the effect of Addon remains significant only in the NB model for females.

As noted, the Poisson model is rejected based on the likelihood ratio test for either of the heterogeneity models (log gamma or lognormal) for both males and females. For the males, for example, for the negative binomial vs. the Poisson model, the chi squared is 2(42774.7 – 27480.4) = 30588.6, with one degree of freedom. Thus, the hypothesis is rejected. Similar results occur for the other three cases shown. The results are convincing that the Poisson model does not adequately account for the latent heterogeneity. The last four rows of Table 3 show the estimates of the parameters of the estimated distribution of latent heterogeneity. The estimated structural parameter is shown in boldface. The other values are derived as shown in the footnotes in the table. The two models produce similar results, however, the variance of the multiplicative heterogeneity (h_i) is substantially larger for the lognormal model. This is a reflection of the thick upper tail of the lognormal distribution. The overall impression of the distribution of ε_i might be a bit erroneous on this basis, as the mean of ε_i in the lognormal model is zero while the mean of ε_i in the log gamma model is $\psi(\theta) - \ln\theta = -1.09$ for the males. Thus, the range of variation of the centered variables in the two models is somewhat closer (though the lognormal still has the larger variance).

The third column of the two groups of estimates present the lognormal model as an alternative specification to the log gamma (negative binomial). These are the counterparts to RWM's results in their Table IV. Our estimates differ slightly; the difference appears small enough to be attributable to difference in the approximation methods. We used a 48 point Hermite approximation. RWM do not note what method they used for the heterogeneous Poisson model. They used a modification of the Hermite quadrature for the bivariate Poisson model. For example, for the log likelihood function, their reported value is -27411.4 vs. our -27408.6. The counterparts for females are -30213.4 for RWM and -30214.7 for ours. Based on the likelihoods, the lognormal model appears to be superior to the negative binomial model. Since the models are not nested, a direct test based on these values is inappropriate. The Vuong statistic suggested in (2.4-6) equals 2.329 in favor of the lognormal model.

Table 4 presents estimates of the parameters of the different specifications of the negative binomial model. The base case Poisson model corresponds to $P = 0$ in the encompassing NBP specification. Based on the likelihood ratio tests, any of the alternative specifications in the table, all of which nest the Poisson, will dominate it. As suggested earlier, NB1 and NB2 produce similar results, but nonetheless, are manifestly different specifications. The log likelihood for NB1 is significantly larger than that for NB2. However, as these two models are not nested, the LR test is inappropriate. Using the Vuong statistic in (2.2-6), we obtain a value of -1.63 in favor of NB1. In spite of the log likelihoods, this is in the inconclusive region. As expected, NBP produces a greater likelihood than either NB1 or NB2. Using a likelihood ratio statistic for testing against NB1, we obtain a chi squared of 207.8 with one degree of freedom. Thus, NB1 and, a fortiori, NB2 are rejected in favor of NBP for these data. The estimated standard error for the estimator of P for this model is 0.02293. The t (Wald) test against the null hypothesis that P equals 1.0 gives a statistic of 21.33, which, once again, would decisively reject the NB1 specification. The second rightmost column in Table 4 presents estimates of a heterogeneous model in which income and education influence the dispersion parameter in the NB2 model. The significantly negative coefficient on income indicates that increasing income increases the dispersion, since in this expanded model, $\kappa_i = (1/\theta)\exp(-\mathbf{z}_i'\boldsymbol{\gamma})$.

The last column in Table 4 presents of the heterogeneous NB2 model. This model specifies the NB2 functional form with, in addition,

$$\lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i)$$

where ε_i has a standard normal distribution. One way to view the model would be as a Poisson model with a compound disturbance in it,

$$\lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta} + \sigma_n\varepsilon_{ni} + \varepsilon_{gi})$$

where ε_{ni} is the standard normally distributed component and ε_{gi} is the log of h_i, which has the log gamma distribution that produces the NB model. If ε_{ni} and ε_{gi} are statistically independent, then the unconditional (on ε_{gi}) density will be the NB2 model, still with latent normally distributed heterogeneity. Though it might appear otherwise, there is no problem of identification in the model; the two variance components are identified through different features of the distribution; the variance of ε_{gi}, identified through θ, appears in the dispersion of y_i (and, indeed, in all higher moments). The estimation of this model is precisely analogous to estimation of the variance components in the stochastic frontier framework [See Aigner, Lovell and Schmidt (1977) and Greene (2006)], where the parameters are identifiable because of the different shapes of the distributions of the two random variables in the sum. Curiously, the estimate of the total variance of the heterogeneity in the compound model is smaller than that of the implied heterogeneity in NB2. Based on the two formulations above, we obtain, for the variance of $\ln h_i$ in NB2, $\psi'(0.5707) = 3.949$, and for the variance of $\sigma_n\varepsilon_n + \ln h_i$, $\psi'(0.9043) + 0.6961^2 = 2.393$. The respective standard deviations are 1.987 and 1.547

Based on the likelihood, the NB normal mixture model dominates NB2. The likelihood ratio statistic is 9.4, again with one degree of freedom. The mixture model does not dominate the NBP model. However, the models are not nested so the simple LR test is not usable.

RWM note based on comparing the Poisson-lognormal to the bivariate model that the significance and, in some cases, the signs of the coefficients change with the specification. We find generally, that this applies to the marginal variables, but that the pattern of significance of most of the variables in the equation is extremely stable. The very important exception is the variables that were the focus of the study, the insurance variables. What we find is that as the model is extended to account for latent heterogeneity, the importance of the private insurance variable diminishes consistently.

6.3 Two Part Models

6.3.1 Sample Selection

RWM used a type of selection model for *AddOn* (not the full information approach suggested here) to study the issue of adverse selection. They used a logit model for the choice of *AddOn*. We will use their specification for the participation equation, though we will be using a probit model throughout. The specification is

$$\mathbf{w}_{it} = (\textit{Constant, Handdum, Handper, Educ, Haupts, Reals, Abitur, Fachhs, Univ,}$$
$$\textit{Whitec, Married, Hhninc, Hhkids},$$
$$\mathbf{1}(30\leq Age\leq 34), \mathbf{1}(35\leq Age\leq 39), \mathbf{1}(40\leq Age\leq 44), \mathbf{1}(45\leq Age\leq 49), \mathbf{1}(50\leq Age\leq 54),$$
$$\mathbf{1}(55\leq Age\leq 59), \mathbf{1}(Age>59),$$
$$\textit{YEAR1985}, \textit{YEAR1986}, \textit{YEAR1987}, \textit{YEAR1988}, \textit{YEAR1991}, \textit{YEAR1994},$$

The authors' logit model also included a variable, *Number of health insurances* ("the number of private health insurance firms in an individual's state of residence"). This appears to be a variable that is not in the published data set. Moreover, it is unclear how the sample subsets for the decision variable were constructed; 9,274 of the 14,243 observations on men and 11,669 of the 13,083 women were used in this model.[26] We will use the variables listed in $\mathbf{w}_{it}$ above without a surrogate for the number of insurances for our sample selection approach and for our two part models. Since the issues of adverse selection and moral hazard are interesting ones in the study, we will take their approach in the sample selection model, but "select" on the *PUBLIC* variable for health insurance, purely for the sake of a numerical example. (Note that one must have the public insurance in order to obtain the addon insurance.)

The adverse selection issue turns on the endogeneity of the insurance coverage variable. As noted, the authors were interested in the marginal impact of the add-on insurance. (They found weak support for the adverse selection hypothesis.) To develop a numerical application, we have treated the entire insurance package, rather than just the add-on component. Thus, our "selection" model considers the possible endogeneity of PUBLIC. (One must purchase the public insurance to add the add-on.) Tables 5 and 6 present FIML estimates of the sample selection models for males and females. The hypothesis test turns on the estimated correlation, which is near zero and insignificant in the equation for males, but highly significant for females. The likelihood ratio test is carried out based on the likelihood function for the full model minus the sum of the two values for the equations with $\rho = 0$. The statistic is 10.6 for the females and only 0.16 for the males. The negative sign on the correlation indicates that the unobservable factors that increase the probability of purchasing the insurance are negatively correlated with the unobservable factors that increase demand on the health care system.

---

[26] RWM also report that the variable *Fachhs* is a perfect predictor of *Addon* in their restricted sample of males. We did not find this to be the case in the full data set, so we will not further restrict the specification.

6.3.2 Zero Inflation and Hurdle Models

Figure 2 is persuasive that the Poisson model probably does not assign sufficient mass to the zero outcome. The zero inflation model explicitly builds on the Poisson or NB model to shift the distribution toward the zero outcome. Tables 7 and 8 present four specifications of the ZIP model for the male subsample. The first model is the base case Poisson. The second is the conventional ZIP model proposed by Lambert (1992), Heilbron (1994) and Greene (1994). The Poisson model is not nested in the ZIP model; there is no parametric restriction on the ZIP model that produces the Poisson specification. Thus, an LR test is inappropriate. From Table 7, the difference of the two log likelihoods of roughly 700 is strongly suggestive. The Vuong statistic of 28.83 strongly favors the zero inflation model, as might be expected. Figure 4 compares the predictions from the ZIP model (the center bar in each cell) to the Poisson (the right bar) and the actual data. (Predictions for the two models are computed using the largest integer less than or equal to the predicted conditional mean.) For the large majority of the observations, that is, for the 0, 1, and 2 values, the ZIP model predicts substantially better than the Poisson model.

We note, in the first ZIP specification, in contrast to RWM's results, we find strong suggestion of moral hazard; that is, the coefficients on both PUBLIC and ADDON are strongly significant. Table 8 extends the model by adding unobserved heterogeneity to the Poisson part of the model. Endogeneity in this case would turn on the correlation between the latent heterogeneity in the regime equation (zero/not zero) and the count model. In the first set of results in Table 8, this correlation is assumed to be zero. In the second model, the correlation is unrestricted; the estimated value is 0.154. However, we do not find statistical evidence of endogeneity. The *t* statistic on the estimated correlation is only 0.14 and the LR statistic is only 1.6. A pattern that persists here as in the preceding specifications is that the statistical significance of the insurance indicators declines substantially when the model more explicitly accounts for latent heterogeneity. The persistent conclusion is that so far, the data do not contain evidence of moral hazard.

The hurdle model is related to the sample selection and zero inflation models. The applicable situation arises when one observes the data "on site." That is, by the nature of the observation mechanism, the count will be at least one. The model consists of a participation equation and the truncated (at zero) count model. Since that situation does not apply here (and as we have already used a large amount of space for this review), we will not pursue the hurdle model in this application.

6.4 The Bivariate Poisson Model

Result (5-10) provides the implied correlation between y_{i1} and y_{i2} in the bivariate Poisson model in which

$$\lambda_{i1} = \exp(\mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \sigma_1\varepsilon_{i1})$$

and

$$\lambda_{i1} = \exp(\mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \sigma_2\varepsilon_{i2})$$

where $(\varepsilon_{i1},\varepsilon_{i2}) \sim N_2[(0,1),(1,1),\rho]$.

Munkin and Trivedi (1999) used this specification in a model for the joint determination of the counts of emergency room visits and hospital visits for a sample of 4,406 elderly Americans drawn from the National Medical Expenditure Survey from 1987 and 1988. The authors report the estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and, in addition, $\hat{\sigma}_1 = 1.39$, $\hat{\sigma}_2 = 1.36$ and $\hat{\rho} = 0.92$. The last of these might lead one to suspect that emergency room visits and hospital visits were extremely highly correlated. However, as derived in Section 5, the 0.92 reflects only the correlation between the latent effects in the conditional means. In order to evaluate the correlation between the two outcomes, we propose to evaluate (5-10) at the sample observations, and then average the outcomes. However, without the Munkin and Trivedi in hand – we do have the RWM data, which we will examine below – we resort to an approximation. To a reasonable approximation, the sample average of λ_i evaluated at the individual data will equal the mean of the outcome variable. (The result is exact in the base case Poisson model – this is the likelihood equation for the constant term.) The authors report sample means of $\bar{y}_1 = 0.26$ and $\bar{y}_2 = 0.92$. Thus, in (5-10), we use

$$\mu_{i1} \approx \bar{y}_1 \exp(.5\times1.39^2) = 0.683162$$
$$\mu_{i2} \approx \bar{y}_2 \exp(.5\times1.36^2) = 0.756409,$$

and complete the computation with a hand calculator. The result is an estimated correlation of 0.668929, which is substantially less than 0.92. Munkin and Trivedi do not report the sample correlation of their two outcome variables, so we cannot measure the implied estimate against the sample statistic.

The RWM model is

(6-1)
$$\ln\lambda_{it,1} = \alpha_1 + \mathbf{x}_{it,1}'\boldsymbol{\beta}_1 + u_{i,1} + \varepsilon_{it,1} \;=\; \alpha_1 + \mathbf{x}_{it,1}'\boldsymbol{\beta}_1 + v_{it,1},$$
$$\ln\lambda_{it,2} = \alpha_2 + \mathbf{x}_{it,2}'\boldsymbol{\beta}_2 + u_{i,2} + \varepsilon_{it,2} \;=\; \alpha_2 + \mathbf{x}_{it,2}'\boldsymbol{\beta}_2 + v_{it,2},$$
$$(\varepsilon_{it,1},\varepsilon_{it,2}) \sim N_2[(0,0),(\sigma_1,\sigma_2),\rho],$$
$$(u_{i,1},u_{i,2}) \sim N_2[(0,0),(\omega_1,\omega_2),0].$$

for which we derived

(6-2)
$$\rho_{12} = \frac{\rho\sigma_1\sigma_2}{\sqrt{\omega_1^2+\sigma_1^2}\sqrt{\omega_2^2+\sigma_2^2}}.$$

The result in (5-10) can be used by using this expression for ρ and the two standard deviations, $\tau_1 = \sqrt{\omega_1^2+\sigma_1^2}$ and $\tau_2 = \sqrt{\omega_2^2+\sigma_2^2}$ for σ_1 and σ_2 in (5-10). The authors did not report the full set of estimated parameters (they omitted the coefficients on the year dummy variables). Rather than reestimate the full bivariate Poisson model, we proceeded as follows. Each of the equations in (6-1) can be consistently estimated in isolation. Moreover, we note that the marginal distribution of each of the observations, (i,t), in the sample, has a marginal Poisson distribution with normally distributed heterogeneity with mean zero and standard deviation τ_j. Thus, we estimated the four equations singly using the lognormal heterogeneity model discussed in Section 2.2.2. This provides consistent, albeit inefficient estimators of the parameters of the four equations. These are shown in Table 9. (RWM's counterparts are shown in Table 10 for comparison.). The estimated variance, σ^2 in each of these equations is an estimate of $\tau^2 = \sigma_\varepsilon^2 + \omega_u^2$ in the RWM model. This is also shown in Table 9. Only an estimate of ρ is needed to complete the calculations in (5-4), (5-10) and (5-13). We used the estimate of ρ reported in RWM for males and females, which appears in the last row of Table 9. For comparison purposes, we have decomposed the estimated variance from our estimates using the implied analysis of variance in RWM. The computations appear at the bottom of Table 9. The proportion denoted "p" in the table inferred from the RWM results is used to decompose the estimated variance from our model. With these statistics in hand, and with the estimated coevvicient vectors, we are able to compute the implied correlations for the two models (males and females). Using (5-13), we obtain individual specific estimates of the correlations of the outcome variables of 0.06938 for males and 0.05795 for females. These are an order of magnitude less than the estimate of ρ reported in the paper, and moreover, only about half of the actual correlation between the outcomes in the data.

6.5 Panel Data Models

Table 11 presents the estimated fixed and random effects Poisson models Based on the likelihood ratio test (which is valid in this case because the MLE is consistent), the “no effects” model is rejected convincingly. the chi squared statistic with (3,687-714) degrees of freedom is 41,156.36. The large degrees of freedom approximation in Greene (2008, result B-37) provides a standard normal test statistic of 209.79. (Note, there 3,687 individuals in the sample. However, 714 of them had zero visits in every period. These observations contribute a 1.00 to the likelihood function – Prob(y_{i1}=0,y_{i2}=0,...|$\Sigma_t\, y_{it}$=0) = 1, so constant terms cannot be estimated for them. The marked difference between the base case Poisson model (no effects) and the fixed effects estimates in the second column are to be expected. The random effects estimates in the third and fourth columns are quite similar. Two noticable differences are the coefficients on marital status and children in the household. Save for these, the Poisson random effects do not differ appreciably across the two platforms. The estimated variances of the heterogeneity are likewise quite similar. The similarities of the competing models does not carry over to the negative binomial specifications.

Estimates for the fixed and random effects negative binomial models appear in Table 12. The two sets of fixed effects estimates are quite different. The statistical significance and the signs of several of the coefficients change across the two specifications, including AGE, MARRIED, EDUC, CIVIL, and ADDON. The magnitude of several of the coefficients changes substantively, notably the coefficient on PUBLIC, which is five times larger in the “true” fixed effects estimates. The signs and statistical significance of the period effects reverse several times as well. The difference between the HHG and true FE models is that HHG builds the effects into the variance of the random variable, not the mean. Thus, we cannot conclude that the HHG estimator is a consistent estimator of a model that contains a heterogeneous mean. It is a consistent estimator in the context of a model with heterogeneous variance. We have convincing evidence from the Poisson model that there is substantial latent heterogeneity in the mean of the random variable. The log likelihood function for the “no effects” NB model falls to -27,480, which is thousands less than the log likelihood for either fixed effects specification. Thus, it is reasonable to conclude that the HHG estimator is at least potentially problematic. This finding does not weigh in favor of the true FE estimator, however. There is no minimally sufficient statistic for α_i in the NB2 model, so we are led to expect that the incidental parameters problem will surface in this setting. It remains to be investigated how substantial the biases (if there are any) will be, however. It seems unlikely that the simple proportional results widely known for the probit and logit models will carry over to this setting. The FE approach produces a bit of a Hobson’s choice. The HHG model does not actually build the heterogeneity into the mean of the random variable, so we might suspect that it suffers from an “omitted variable” problem. The true fixed effects estimates differ enough from the HHG estimates in this very large sample that one might suspect the appearance of the incidental parameters problem.

The random effects estimates for the NB models also differ substantially. In this case, however, there is no simple comparison one can draw. There are fewer sign changes, however, the magnitudes and statistical significance are surprisingly variable for a sample as large as this one. Once again, we suspect that the models differ in subtle, but significant structural ways. We have no way of interpreting the parameters of the beta distribution in the HHG model that implies a decomposition of the variance of the heterogeneity. For the lognormal model, we can decompose the variance as follows: The variance of the log gamma term is $\psi'(\theta) = \psi'(1/1.0192)$ = 1.681. The variance of the time invariant lognormal component is $.7979^2 = .637$. The total is thus 2.318. A counterpart that does not assume that the lognormal component is time invariant appears in the second to last column of Table 4. The same decomposition produces $\psi'(.9043)=1.909$ and $.6961^2 = .485$ for a nearly identical total of 2.394.

7. Conclusions

This study has proposed several extensions to some familiar models for count data including:

1. The NBP encompassing form for the negative binomial model;
2. The lognormal model as an alternative to the log gamma model for unobserved heterogeneity in count data models;
3. Extensions of some conventional two part models to allow for endogeneity of the participation decision in the first equation;
4. A detailed interpretation of some applications of the bivariate Poisson model;
5. Some alternative natural and convenient forms of the widely used forms of panel data models for count data.

We have also applied the techniques in an analysis of a large sample of German households.

The NBP variant of the negative binomial model is a convenient form that provides a means of formalizing the specification choice. Most received applications of the model have used the NB2 form. In a few other cases, such as HHG (1984), the NB1 model is used. In none of the cases, does the presentation provide a formal means of preferring one or the other. The NBP is an encompassing form that is simple to operationalize. In the application here (and in others we have considered), likelihood ratio tests suggest that the NBP form would be preferred to both NB1 and NB2.

The negative binomial has been used for a generation as the standard vehicle for introducing unobserved heterogeneity into loglinear count data models. The vast array of functional forms that appear in the literature, and the NB model itself, have largely been motivated by a desire to accommodate over or underdispersion. In fact, the Poisson form is probably unique in its restriction of the random variable to equidispersion. It is convenient, however, that the NB model also arises as a byproduct of the introduction of a particular form of latent heterogeneity – log gamma in distribution. A number of authors, e.g., Winkelmann (2003), Million (1998), RWM (2003), Greene (2008a), have suggested that the normal distribution would be a preferable platform on which to build the model. In addition to deriving from a natural assumption about the source of latent heterogeneity, a model based on the normal distribution provides a convenient setting in which to build useful extensions. We developed the methods for accommodating this form of heterogeneity in the count data model – this follows earlier applications such as Greene (1997), Munkin and Trivedi (1999) and RWM (2003). We then extended the lognormal model to several two part models, sample selection, zero inflation and hurdle models, to allow the participation to be endogenous. The development provides a unified framework that will accommodate other similar models with minimal change in the basic template.

One of the recent applications of the methods extended in this paper is in a type of bivariate count model. We found that in these models, the introduction of a "correlation coefficient" into the model within the conditional means provides only a partial indication of the degree of correlation between the outcome variables. We derived the relationship between the structural parameters and the reduced form correlation between the outcome variables in the bivariate Poisson model. In the application carried out in this paper, we find that the estimated correlation coefficient is far higher than the actual correlation of the variables in the model. Moreover, the implied correlation coefficient based on the model estimates, which is a function of the data and thus varies by observation, does a strikingly poor job of reproducing the actual, simple correlation of the outcome variables and, moreover, appears, on average in these data, to be a full order of magnitude less than the simple reported correlation coefficient. This calls into question the precise interpretation of this part of the model and whether this form of correlation is an effective approach to modeling the correlation across related count data outcome variables.

Finally, we examined some aspects of the most familiar forms of fixed and random effects models for count data. As in the earlier models, we find that the lognormal distribution provides a natural method of introducing time invariant heterogeneity into the equation. Likewise, we propose an alternative to the HHG fixed effects model. In this case, the results leave a choice to be made, and a point for further research. In the HHG fixed effects NB model, the fixed effects enter the model through the dispersion parameter rather than the conditional mean function. This has the implication that time invariant variables can coexist with the effects. This calls the interpretation of the heterogeneity in the model into question. We propose to apply the direct fixed effects approach suggested in Greene (2004) as an appropriate approach to introducing fixed effects into the NB model. While the proposed approach does parallel the treatment of fixed effects in other received models, like many of them, the specification may also suffer from the incidental parameters problem. In some specific cases, such as binary choice models, the MLE FE estimator has been found to exhibit a significant bias when T is small (as it is in our application). However, the negative binomial model remains to be examined. As shown in Greene (2004), not all estimators are biased away from zero, and some are (apparently) not biased at all. On the other hand, the HHG model provides a sufficient statistic for the fixed effects, so the estimator in their model would not exhibit an "incidental parameters problem." Because the conditional mean function in the HHG model remains homogeneous, however, one might expect a "left out variable" problem instead. We cannot characterize at this point which specification is likely to be more problematic in terms of the features of the population one is interested in studying. This remains an issue to be studied further.

Finally, the methods developed here were applied to the data set used in RWM(2003). Our results were largely similar to theirs. We do find that on the question of moral hazard – whether the presence of insurance appears positively to influence demand for health services – the apparent effect that shows up in the simple models (e.g., a pooled Poisson model) almost completely disappears when latent heterogeneity is formally introduced into the model.

APPENDIX A: Log Likelihood and Gradient for NBP model

The Negbin P model is obtained by replacing θ in NB2,

(A-1) $$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(1+y_i)} r_i^{\theta}(1-r_i)^{y_i}$$

where $r_i = \theta / (\theta + \lambda_i)$

with $\theta\lambda_i^{P\text{-}2.}$. For convenience, let $Q = P - 2$. Then, the density is

(A-2) $$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta\lambda_i^Q + y_i)}{\Gamma(\theta\lambda_i^Q)\Gamma(1+y_i)}\left(\frac{\theta\lambda_i^Q}{\theta\lambda_i^Q+\lambda_i}\right)^{\theta\lambda_i^Q}\left(\frac{\lambda}{\theta\lambda_i^Q+\lambda_i}\right)^{y_i}$$

Derivatives of $\ln L_i$ for the Negbin P model are straightforward, albeit tedious. We obtain them by writing the density as

(A-3) $$\ln L_i = \ln\Gamma(y_i + g_i) - \ln\Gamma(g_i) - \ln\Gamma(1+y_i) + g_i \ln r_i + y_i \ln(1-r_i)$$

where

$$g_i = \theta\,\lambda_i^Q \text{ and } w_i = g_i / (g_i + \lambda_i).$$

Then,

(A-4)
$$\partial\ln L_i / \partial\lambda_i = [\Psi(y_i + g_i) - \Psi(g_i) + \ln w_i]\,\partial g_i/\partial\lambda_i + [g_i/w_i - y_i/(1-w_i)]\,\partial w_i/\partial\lambda_i$$
$$\partial\ln L_i / \partial\theta = [\Psi(y_i + g_i) - \Psi(g_i) + \ln w_i]\,\partial g_i/\partial\theta + [g_i/w_i - y_i/(1-w_i)]\,\partial w_i/\partial\theta$$
$$\partial\ln L_i / \partial Q = [\Psi(y_i + g_i) - \Psi(g_i) + \ln w_i]\,\partial g_i/\partial Q + [g_i/w_i - y_i/(1-w_i)]\,\partial w_i/\partial Q.$$

The inner parts are:

(A-5)
$$\partial g_i / \partial\lambda_i = \theta Q\lambda_i^{Q-1} = (Q/\lambda_i)g_i$$
$$\partial g_i / \partial\theta = \lambda_i^Q = (1/\theta)g_i$$
$$\partial g_i / \partial Q = \theta\lambda_i^Q \log\lambda_i = \ln\lambda_i\, g_i$$
$$\partial w_i / \partial\lambda_i = [(Q-1)/\lambda_i]w_i(1-w_i)$$
$$\partial w_i / \partial\theta = (1/\theta)w_i(1-w_i)$$
$$\partial w_i / \partial Q = \log\lambda_i\, w_i(1-w_i)$$

Collecting terms, now, let

(A-6)
$$A_i = [\Psi(y_i + g_i) - \Psi(g_i) + \ln w_i]$$
$$B_i = [g_i(1 - w_i) - y_i w_i],$$

to obtain

(A-7) $$\partial \ln L_i / \partial\begin{pmatrix}\lambda_i\\ \theta\\ Q\end{pmatrix} = [A_i + B_i]\begin{pmatrix}Q/\lambda_i\\ 1/\theta\\ \log\lambda_i\end{pmatrix} - B_i\begin{pmatrix}1/\lambda_i\\ 0\\ 0\end{pmatrix}.$$

The final element needed is $\partial\ln L_i/\partial\boldsymbol{\beta} = (\partial\ln L_i/\partial\lambda_i)(\partial\lambda_i/\partial\boldsymbol{\beta})$ where $\partial\ln L_i/\partial\lambda_i$ appears above and $\partial\lambda_i/\partial\boldsymbol{\beta} = \lambda_i\mathbf{x}_i$. We use these and the BHHH estimator to compute the maximum likelihood estimates and their asymptotic standard errors for the NBP model. Good starting values for NBP iterative estimator are the NB2 estimates of β and θ with P=2 (Q = 0).

Appendix B Derivatives of Partial Effects in the Poisson Model with Sample Selection

The conditional mean function is

(B-1) $$E[y_i|\mathbf{x}_i,\mathbf{w}_i,d_i = 1] \;=\; \lambda_i \,\frac{\exp((\rho\sigma)^2/2)\Phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\mathbf{w}_i'\delta)}$$

The partial effects are

(B-2)
$$\frac{\partial E[y_i\,|\,\mathbf{x}_i,\mathbf{w}_i]}{\partial\mathbf{x}_i} = \lambda_i\left[\frac{\exp((\rho\sigma)^2/2)\Phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\mathbf{w}_i'\delta)}\right]\beta \;=\; \mathbf{g}_\mathbf{x}$$
$$\frac{\partial E[y_i\,|\,\mathbf{x}_i,\mathbf{w}_i]}{\partial\mathbf{w}_i} = \lambda_i\left(\frac{\exp((\rho\sigma)^2/2)}{\Phi(\mathbf{w}_i'\delta)}\right)\left[\phi(\rho\sigma+\mathbf{w}_i'\delta)-\phi(\mathbf{w}_i\delta)\left(\frac{\Phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\mathbf{w}_i'\delta)}\right)\right]\delta = \mathbf{g}_\mathbf{w}.$$

For the variables in the count model,

(B-3)
$$\frac{\partial\mathbf{g}_\mathbf{x}}{\partial(\alpha,\beta')} = \lambda_i\left[\frac{\exp((\rho\sigma)^2/2)\Phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\mathbf{w}_i'\delta)}\right]\left[\underline{\mathbf{0}}\quad\mathbf{I}\right]+\mathbf{g}_\mathbf{x}(1,\mathbf{x}_i')$$
$$\frac{\partial\mathbf{g}_\mathbf{x}}{\partial\rho} = \mathbf{g}_\mathbf{x}\left[\rho\sigma^2+\sigma\frac{\phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\rho\sigma+\mathbf{w}_i'\delta)}\right]$$
$$\frac{\partial\mathbf{g}_\mathbf{x}}{\partial\sigma} = \frac{\partial\mathbf{g}_\mathbf{x}}{\partial\rho}\frac{\rho}{\sigma}$$
$$\frac{\partial\mathbf{g}_\mathbf{x}}{\partial\delta'} = \lambda_i\exp(\rho\sigma)^2\left[\frac{\phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\mathbf{w}_i'\delta)}-\frac{\Phi(\rho\sigma+\mathbf{w}_i'\delta)\phi(\mathbf{w}_i'\delta)}{\left[\Phi(\mathbf{w}_i'\delta)\right]^2}\right]\beta\mathbf{w}'$$

For the variables in the selection equation,

(B-4)
$$\frac{\partial\mathbf{g}_\mathbf{w}}{\partial(\alpha,\beta')} = \mathbf{g}_\mathbf{w}(1,\mathbf{x}_i')$$
$$\frac{\partial\mathbf{g}_\mathbf{w}}{\partial\rho} = \mathbf{g}_\mathbf{w}\rho\sigma^2-\lambda_i\left(\frac{\exp((\rho\sigma)^2/2)}{\Phi(\mathbf{w}_i'\delta)}\right)\left[(\rho\sigma+\mathbf{w}_i'\delta)\phi(\rho\sigma+\mathbf{w}_i'\delta)+\phi(\mathbf{w}_i'\delta)\frac{\phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\mathbf{w}_i'\delta)}\right]\sigma\delta$$
$$\frac{\partial\mathbf{g}_\mathbf{w}}{\partial\sigma} = \frac{\partial\mathbf{g}_\mathbf{w}}{\partial\rho}\frac{\rho}{\sigma}$$
$$\frac{\partial\mathbf{g}_\mathbf{w}}{\partial\delta'} = -\frac{\phi(\mathbf{w}_i'\delta)}{\phi(\mathbf{w}_i'\delta)}\mathbf{g}_\mathbf{w}\mathbf{w}_i'+\lambda_i\exp(\rho\sigma)^2\left[\begin{array}{l}-(\rho\sigma+\mathbf{w}_i'\delta)\phi(\rho\sigma+\mathbf{w}_i'\delta)+\dfrac{\phi(\mathbf{w}_i'\delta)\phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\mathbf{w}_i'\delta)}+\\ \left(\mathbf{w}_i'\delta\right)\phi(\mathbf{w}_i'\delta)\dfrac{\Phi(\rho\sigma+\mathbf{w}_i'\delta)}{\Phi(\mathbf{w}_i'\delta)}-\\ \dfrac{\left[\phi(\mathbf{w}_i'\delta)\right]^2\Phi(\rho\sigma+\mathbf{w}_i'\delta)}{\left[\Phi(\mathbf{w}_i'\delta)\right]^2}\end{array}\right]\delta\mathbf{w}_i'$$

Appendix C Derivatives of Partial Effects of ZIP Model with Endogenous Zero Inflation.

Let $A(\varepsilon_i) = \dfrac{(\mathbf{w}_i'\boldsymbol{\delta} + \rho\varepsilon_i)}{\sqrt{1-\rho^2}}$ and $\Delta = \dfrac{1}{\sqrt{1-\rho^2}}$. The partial effects are

(C-1)
$$\partial E[y_i|\,\mathbf{x}_i,\mathbf{w}_i]/\partial\mathbf{x}_i = \lambda_i\,\boldsymbol{\beta}\int_{-\infty}^{\infty}\Phi\left[A(\varepsilon_i)\right]\ \exp(\sigma\varepsilon_i)\,\phi(\varepsilon_i)d\varepsilon_i = \mathbf{g}_{\mathbf{x}}.$$
$$\partial E[y_i|\,\mathbf{x}_i,\mathbf{w}_i]/\partial\mathbf{w}_i = \lambda_i\ \Delta\boldsymbol{\delta}\left\{\int_{-\infty}^{\infty}\phi\left[A(\varepsilon_i)\right]\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i\right\} = \mathbf{g}_{\mathbf{w}}.$$

The derivatives are

(C-2)
$$\frac{\partial\mathbf{g}_{\mathbf{x}}}{\partial(\alpha,\boldsymbol{\beta}')} = \mathbf{g}_{\mathbf{x}}(1,\mathbf{x}_i') + \lambda_i\left[\underline{\mathbf{0}}\quad\mathbf{I}\right]\int_{-\infty}^{\infty}\Phi\left[A(\varepsilon_i)\right]\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i$$
$$\frac{\partial\mathbf{g}_{\mathbf{x}}}{\partial\sigma} = \lambda_i\boldsymbol{\beta}\int_{-\infty}^{\infty}\Phi\left[A(\varepsilon_i)\right]\varepsilon_i\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i$$
$$\frac{\partial\mathbf{g}_{\mathbf{x}}}{\partial\boldsymbol{\delta}'} = \lambda_i\boldsymbol{\beta}\mathbf{w}_i'\Delta\int_{-\infty}^{\infty}\phi\left[A(\varepsilon_i)\right]\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i$$
$$\frac{\partial\mathbf{g}_{\mathbf{x}}}{\partial\rho} = \lambda_i\boldsymbol{\beta}\Delta\int_{-\infty}^{\infty}\phi\left[A(\varepsilon_i)\right]\exp(\sigma\varepsilon_i)\left[\varepsilon_i - \rho\Delta^2A(\varepsilon_i)\right]\phi(\varepsilon_i)d\varepsilon_i$$
$$\frac{\partial\mathbf{g}_{\mathbf{w}}}{\partial(\alpha,\boldsymbol{\beta}')} = \mathbf{g}_{\mathbf{w}}(1,\mathbf{x}_{\mathbf{i}}')$$
$$\frac{\partial\mathbf{g}_{\mathbf{w}}}{\partial\sigma} = \lambda_i\Delta\left\{\int_{-\infty}^{\infty}\phi\left[A(\varepsilon_i)\right]\varepsilon_i\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i\right\}$$
$$\frac{\partial\mathbf{g}_{\mathbf{w}}}{\partial\boldsymbol{\delta}'} = \lambda_i\boldsymbol{\delta}\mathbf{w}'\Delta^2{}_i\left\{\int_{-\infty}^{\infty} - A(\varepsilon_i)\phi\left[A(\varepsilon_i)\right]\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i\right\}$$
$$\frac{\partial\mathbf{g}_{\mathbf{w}}}{\partial\sigma} = \mathbf{g}_{\mathbf{w}}\rho\Delta^2 + \lambda_i\boldsymbol{\delta}\Delta^2\left\{\int_{-\infty}^{\infty} - A(\varepsilon_i)\phi\left[A(\varepsilon_i)\right]\left[\varepsilon_i - \rho\Delta^2A(\varepsilon_i)\right]\exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i\right\}$$

These must be approximated either by quadrature or by simulation. If ρ equals zero, then most of the preceding vanishes. The conditional mean is $= \lambda_i\Phi\left(\mathbf{w}_i'\boldsymbol{\delta}\right)\exp(\sigma^2/2)$ and the partial effects are

(C-3)
$$\mathbf{g}_{\mathbf{x}} = \lambda_i\boldsymbol{\beta}\,\Phi\left(\mathbf{w}_i'\boldsymbol{\delta}\right)\exp(\sigma^2/2),$$
$$\frac{\partial\mathbf{g}_{\mathbf{x}}}{\partial(\alpha,\boldsymbol{\beta}',\sigma)} = \mathbf{g}_{\mathbf{x}}(1,\mathbf{x}_i',\sigma),$$
$$\frac{\partial\mathbf{g}_{\mathbf{x}}}{\partial\boldsymbol{\delta}'} = \lambda_i\boldsymbol{\beta}\mathbf{w}_i'[\phi(\mathbf{w}_i'\boldsymbol{\delta})\exp(\sigma^2/2)]$$
$$\mathbf{g}_{\mathbf{w}} = \lambda_i\boldsymbol{\delta}\,\phi\left(\mathbf{w}_i'\boldsymbol{\delta}\right)\exp(\sigma^2/2),$$
$$\frac{\partial\mathbf{g}_{\mathbf{w}}}{\partial(\alpha,\boldsymbol{\beta}',\sigma)} = \mathbf{g}_{\mathbf{w}}(1,\mathbf{x}_i',\sigma),$$
$$\frac{\partial\mathbf{g}_{\mathbf{w}}}{\partial\boldsymbol{\delta}'} = -\mathbf{g}_{\mathbf{w}}\mathbf{w}_i'(\mathbf{w}_i'\boldsymbol{\delta}) + \lambda_i\phi(\mathbf{w}_i'\boldsymbol{\delta})\exp(\sigma^2/2)\mathbf{I}$$

Appendix D Derivatives of Partial Effects in Hurdle Models

We assume that the hurdle equation is a probit model. Adaptation to a logit hurdle equation requires substitution of $\Lambda(\mathbf{w}_i'\delta)$ for $\Phi(\mathbf{w}_i'\delta)$, $\{\Lambda(\mathbf{w}_i'\delta)[1-\Lambda(\mathbf{w}_i'\delta)]\}$ for $\phi(\mathbf{w}_i'\delta)$ and $[1-2\Lambda(\mathbf{w}_i'\delta)]\{\Lambda(\mathbf{w}_i'\delta)[1-\Lambda(\mathbf{w}_i'\delta)]\}$ for $-(\mathbf{w}_i'\delta)\phi(\mathbf{w}_i'\delta)]$ in what follows. The conditional mean is

(D-1) $$E[y_i|\mathbf{x}_i,\mathbf{w}_i] = \frac{\Phi(\mathbf{w}_i'\delta)\lambda_i}{[1-\exp(-\lambda_i)]}.$$

The partial effects are

$$\partial E[y_i|\mathbf{x}_i,\mathbf{w}_i,d_i]/\partial\mathbf{x}_i = \frac{\Phi(\mathbf{w}_i'\delta)}{[1-\exp(-\lambda_i)]}\left(1-\frac{\lambda_i\exp(-\lambda_i)}{1-\exp(-\lambda_i)}\right)\lambda_i\boldsymbol{\beta} = \mathbf{g}_\mathbf{x}$$

(D-2) $$\partial E[y_i|\mathbf{x}_i,\mathbf{w}_i,d_i]/\partial\mathbf{w}_i = \frac{\phi(\mathbf{w}_i'\delta)}{[1-\exp(-\lambda_i)]}\lambda_i\boldsymbol{\delta} = \mathbf{g}_\mathbf{w}$$

The derivatives are cumbersome. We proceed as follows: Write

$$c(\lambda_i) = \lambda_i / [1-\exp(-\lambda_i)].$$

Then

$$\begin{aligned}
dc(\lambda_i)/d\lambda_i &= c'(\lambda_i)\\
&= (1/\lambda_i)\{c(\lambda_i) - \exp(-\lambda_i)[c(\lambda_i)]^2\},\\
d^2c(\lambda_i)/d\lambda_i^2 &= c''(\lambda_i)\\
&= -(1/\lambda_i)\,c'(\lambda_i) + (1/\lambda_i)[\,c'(\lambda_i) - 2c(\lambda_i)\,c'(\lambda_i)\exp(-\lambda_i)]\\
&= -(2/\lambda_i)\,c(\lambda_i)\,c'(\lambda_i)\exp(-\lambda_i)\\
\partial c(\lambda_i)/\partial\mathbf{x}_i &= c'(\lambda_i)(\partial\lambda_i/\partial\mathbf{x}_i) = c'(\lambda_i)\,\lambda_i\,\boldsymbol{\beta}.
\end{aligned} \quad \text{(D-3)}$$

Thus,

$$\begin{aligned}
E[y_i|\mathbf{x}_i,\mathbf{w}_i] &= \Phi(\mathbf{w}_i'\delta)\,c(\lambda_i)\,\boldsymbol{\beta},\\
\mathbf{g}_\mathbf{x} &= \Phi(\mathbf{w}_i'\delta)\,c'(\lambda_i)\lambda_i\boldsymbol{\beta},\\
\mathbf{g}_\mathbf{w} &= \phi(\mathbf{w}_i'\delta)\,c(\lambda_i)\,\boldsymbol{\delta}.
\end{aligned} \quad \text{(D-4)}$$

Then,

$$\begin{aligned}
\frac{\partial\mathbf{g}_\mathbf{x}}{\partial(\alpha,\boldsymbol{\beta}')} &= \Phi(\mathbf{w}_i'\delta)\lambda_i\{[c'(\lambda_i)+\lambda_i c''(\lambda_i)]\boldsymbol{\beta}(1,\mathbf{x}_i') + c'(\lambda_i)[\underline{\mathbf{0}}\quad \mathbf{I}]\},\\
\frac{\partial\mathbf{g}_\mathbf{x}}{\partial\boldsymbol{\delta}'} &= \phi(\mathbf{w}_i'\delta)c'(\lambda_i)\lambda_i\boldsymbol{\beta}\mathbf{w}_i',\\
\frac{\partial\mathbf{g}_\mathbf{w}}{\partial(\alpha,\boldsymbol{\beta}')} &= \phi(\mathbf{w}_i'\delta)c'(\lambda_i)\lambda_i\boldsymbol{\delta}(1,\mathbf{x}_i'),\\
\frac{\partial\mathbf{g}_\mathbf{w}}{\partial\boldsymbol{\delta}'} &= -(\mathbf{w}_i'\delta)\phi(\mathbf{w}_i'\delta)c(\lambda_i)\lambda_i\boldsymbol{\delta}\mathbf{w}_i'.
\end{aligned} \quad \text{(D-5)}$$

Appendix E Partial Effects and Derivatives of Partial Effects in Hurdle Models with Endogenous Participation

The conditional mean is

(E-1) $$E[y_i|\mathbf{x}_i,\mathbf{w}_i] = \int_{-\infty}^{\infty} \Phi\left(\frac{\mathbf{w}_i'\delta + \rho\varepsilon_i}{\sqrt{1-\rho^2}}\right)\frac{\left[\exp(\sigma\varepsilon_i)\lambda_i\right]}{\left[1-\exp(-\exp(\sigma\varepsilon_i)\lambda_i)\right]}\phi(\varepsilon_i)\mathrm{d}\varepsilon_i\,.$$

For convenience, let

(E-2) $$\mathrm{A}(\varepsilon_i) = \frac{(\mathbf{w}_i'\delta + \rho\varepsilon_i)}{\sqrt{1-\rho^2}}$$

$$h_i = \exp(\sigma\varepsilon_i)$$

so that

$$\partial h_i/\partial\sigma = \varepsilon_i h_i.$$

(E-3) Let $u_i = h_i\lambda_i$, $\partial u_i/\partial\lambda_i = h_i = \exp(\sigma\varepsilon_i)$, $\partial u_i/\partial\sigma = \varepsilon_i u_i$,

and $$a_i = a(u_i) = \frac{h_i\lambda_i}{\left[1-\exp(-h_i\lambda_i)\right]} = \frac{u_i}{1-\exp(-u_i)}\,.$$

Then

$$a_i'(u_i) = \frac{\mathrm{d}a(u_i)}{\mathrm{d}u_i} = \frac{a_i}{u_i}\big(1 - a_i\ \exp(-u_i)\big),$$

$$a_i''(u_i) = \frac{\mathrm{d}^2a(u_i)}{\mathrm{d}u_i^2} = \frac{a_i\ \exp(-u_i)}{u_i}(a_i^2 - 2a_i'),$$

$$\frac{\partial a(u_i)}{\partial\sigma} = \varepsilon_i u_i a_i',$$

(E-4) $$\frac{\partial a(u_i)}{\partial\lambda_i} = h_i a_i',$$

$$\frac{\partial^2 a(u_i)}{\partial\sigma^2} = \varepsilon_i(\varepsilon_i u_i a_i' + \varepsilon_i u_i u_i a_i'') = \varepsilon_i^2 u_i(a_i' + u_i a_i''),$$

$$\frac{\partial^2 a(u_i)}{\partial\lambda_i^2} = h_i^2 a_i'',$$

$$\frac{\partial^2 a(u_i)}{\partial\lambda_i\partial\sigma} = \varepsilon_i h_i a_i' + h_i a_i''\varepsilon_i u_i = \varepsilon_i h_i(a_i' + u_i a_i'').$$

The conditional mean function is

(E-5) $$E[y_i|\mathbf{x}_i,\mathbf{w}_i] = \int_{-\infty}^{\infty} \Phi\left[A(\varepsilon_i)\right]a(u_i)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i$$

and the partial effects are

(E-6)

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{w}_i]}{\partial \mathbf{x}_i} = \left\{ \int_{-\infty}^{\infty} \Phi\left[A(\varepsilon_i)\right]\left(a'(u_i)h_i\right)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \lambda_i \boldsymbol{\beta} = \mathbf{g}_{\mathbf{x}},$$

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{w}_i]}{\partial \mathbf{w}_i} = \left\{ \int_{-\infty}^{\infty} \phi\left[A(\varepsilon_i)\right]a(u_i)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \left(\frac{1}{\sqrt{1-\rho^2}} \right) \boldsymbol{\delta} = \mathbf{g}_{\mathbf{w}}.$$

Let $\Delta = \quad 1/\sqrt{1-\rho^1}$ The derivatives are

(E-7)

$$\frac{\partial \mathbf{g}_{\mathbf{x}}}{\partial(\alpha, \boldsymbol{\beta}')} = \left\langle \left\{ \int_{-\infty}^{\infty} \Phi\left[A(\varepsilon_i)\right]\left(a''(u_i)h_i^2\lambda_i\right)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \lambda_i \boldsymbol{\beta} + \mathbf{g}_{\mathbf{x}} \right\rangle (1, \mathbf{x}_i')$$
$$+ \left\{ \int_{-\infty}^{\infty} \Phi\left[A(\varepsilon_i)\right]\left(a'(u_i)h_i\right)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \lambda_i \left[\underline{\mathbf{0}} \quad \mathbf{I}\right]$$

$$\frac{\partial \mathbf{g}_{\mathbf{x}}}{\partial \sigma} = \left\{ \int_{-\infty}^{\infty} \Phi\left[A(\varepsilon_i)\right]\varepsilon_i h_i \left(a'(u_i) + a_i''(u_i)u_i\right)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \lambda_i \boldsymbol{\beta}$$

$$\frac{\partial \mathbf{g}_{\mathbf{x}}}{\partial \boldsymbol{\delta}} = \left\{ \int_{-\infty}^{\infty} \phi\left[A(\varepsilon_i)\right]\left(a'(u_i)h_i\right)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \Delta\lambda_i \boldsymbol{\beta}\mathbf{w}_i'$$

$$\frac{\partial \mathbf{g}_{\mathbf{x}}}{\partial \rho} = \left\{ \int_{-\infty}^{\infty} \phi\left[A(\varepsilon_i)\right]\left(\varepsilon_i - \rho\Delta\left[A(\varepsilon_i)\right]\right)\left(a'(u_i)h_i\right)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \Delta\lambda_i \boldsymbol{\beta}$$

$$\frac{\partial \mathbf{g}_{\mathbf{w}}}{\partial(\alpha, \boldsymbol{\beta}')} = \left\{ \int_{-\infty}^{\infty} \phi\left[A(\varepsilon_i)\right]\left(a'(u_i)h_i\right)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \lambda_i \Delta \boldsymbol{\delta}(1, \mathbf{x}_i')$$

$$\frac{\partial \mathbf{g}_{\mathbf{w}}}{\partial \sigma} = \left\{ \int_{-\infty}^{\infty} \phi\left[A(\varepsilon_i)\right]\left(a'(u_i)\varepsilon_i u_i\right)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \Delta \boldsymbol{\delta}$$

$$\frac{\partial \mathbf{g}_{\mathbf{w}}}{\partial \boldsymbol{\delta}'} = \left\{ \int_{-\infty}^{\infty} -\left[A(\varepsilon_i)\right]\phi\left[A(\varepsilon_i)\right]a(u_i)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \Delta^2 \boldsymbol{\delta}\mathbf{w}_i'$$

$$\frac{\partial \mathbf{g}_{\mathbf{w}}}{\partial \rho} = \left\{ \int_{-\infty}^{\infty} -\left[A(\varepsilon_i)\right]\phi\left[A(\varepsilon_i)\right]\left(\varepsilon_i - \rho\Delta\left[A(\varepsilon_i)\right]\right)a(u_i)\phi(\varepsilon_i)\mathrm{d}\varepsilon_i \right\} \Delta^2 \boldsymbol{\delta}$$

9. References

Abramovitz, M., and I. Stegun. *Handbook of Mathematical Functions*. New York: Dover Press, 1971.

Aigner, D., K. Lovell, and P. Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, 6, pp. 21-37, 1977.

Allison, P., "Problems with Fixed-Effects Negative Binomial Models," Manuscript, Department of Sociology, University of Pennsylvania, 2000.

Allison, P. and R. Waterman, "Fixed-Effects Negative Binomial Regression Models," Manuscript, Department of Sociology, University of Pennsylvania, 2002.

Amid, A., "A Note on the Moments of the Generalized negative Binomial Distribution and on Certain Properties of the Distribution," *SIAM Journal of Applied Mathematics*, 34, 2, 1978, pp. 223-224.

Asplund, M. and R. Sandin, "The Number of Firms and Production Capacity in Relation to Market Size," *The Journal of Industrial Economics*, Vol. 47, No. 1 (Mar., 1999), pp. 69-85

Berndt, E., B. Hall, R. Hall, and J. Hausman. "Estimation and Inference in Nonlinear Structural Models." *Annals of Economic and Social Measurement,* 3/4, 1974, pp. 653–665.

Butler, J. and Moffitt, R., "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model," *Econometrica*, 50, 1982, pp. 761-764.

Econometric Software, Inc., *LIMDEP Version 8.0 and NLOGIT Version 3.0*, Plainview, New York, 2003.

Econometric Software, Inc., *LIMDEP Version 9.0 and NLOGIT Version 4.0*, Plainview, New York, 2007.

Cameron, A., and P. Trivedi. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics,* 1, 1986, pp. 29–54.

Cameron, A., and P. Trivedi. "Regression Based Tests for Overdispersion in the Poisson Model." *Journal of Econometrics,* 46, 1990, pp. 347–364.

Cameron, C., and P. Trivedi. *Regression Analysis of Count Data.* New York: Cambridge University Press, 1998.

Cameron, C. and P. Trivedi, *Microeconometrics: Methods and Applications*, Cambridge University Press, Cambridge, 2005.

Contoyannis, C., A. Jones and N. Rice, "The Dynamics of Health in the British Household Panel Survey," *Journal of Applied Econometrics*, 19, 4, 2004, pp. 473-503.

Dimitris, K. and I. Ntzoufras, "Analysis of Sports Data By Using Bivariate Poisson Models," *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52, 3, 2003, pp. 381–393.

Freedman, D., "On The So-Called "Huber Sandwich Estimator" and "Robust Standard Errors"," *The American Statistician*, 60, 4, 2006, pp. 299-302.

Gourieroux, C., and A. Monfort. *Simulation-Based Methods Econometric Methods*. Oxford: Oxford University Press, 1996.

Gourieroux, C., A. Monfort, and A. Trognon. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica,* 52, 1984, pp. 701–720.

Greene, W., "A Statistical Model for Credit Scoring," Working Paper EC-92-29, Department of Economics, Stern School of Business, New York University, 1992, forthcoming in *Credit Risk: Quantitative Methods and Analysis*, Hensher, D. and S. Jones, eds., Cambridge University Press, 2007.

Greene, W. "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models." Working Paper No. EC-94-10, Department of Economics, Stern School of Business, New York University, 1994.

Greene, W., "Sample Selection in the Poisson Regression Model" Working Paper EC-95-06, Department of Economics, Stern School of Business, New York University, 1995.

Greene, W., "FIML Estimation of Sample Selection Models for Count Data," Working Paper EC-97-02, Department of Economics, Stern School of Business, New York University, 1997.

Greene, W., "The Behavior of the Fixed Effects Estimator in Nonlinear Models," *The Econometrics Journal* , 7, 1, 2004, pp. 98-119.

Greene, W., "A General Approach to Incorporating 'Selectivity' in a Model," Working Paper EC-06-10, Department of Economics, Stern School of Business, New York University, 2006.

Greene, W., *LIMDEP 9.0 Reference Guide,* Econometric Software, Inc., Plainview, 2007

Greene, W., *Econometric Analysis, 6th ed.*, Prentice Hall, Englewood Cliffs, 2008a.

Greene, W., "The Econometric Approach to Efficiency Analysis," forthcoming in H. Fried, C.A.K. Lovell and S. Schmidt, eds., *The Measurement of Productive Efficiency, 2nd Edition*, edited by Oxford University Press, (forthcoming) 2008b.

Greene, W., Harris, M., Hollingsworth, B. and P. Maitra, "A Bivariate Latent Class Correlated Generalized Ordered Probit Model with an Application to Modeling Observed Obesity Levels," Department of Econometrics, Monash University, Melbourne, manuscript, 2007.

Hall, D.B. (2000). Zero Inflated Poisson and Binomial Regression with Random Effects: A Case Study, *Biometrics*, 56, 2000, pp. 1030-1039.

Hausman, J., B. Hall, and Z. Griliches. "Economic Models for Count Data with an Application to the Patents–R&D Relationship." *Econometrica,* 52, 1984, pp. 909–938.

Heckman, J., "Sample Selection Bias As a Specification Error," *Econometrica*, 47, 1979, pp. 153-161.

Heilbron, D., "Zero Altered and Other Regression Models for Count Data with Added Zeros, *Biometrical Journal* 36, 1994, pp. 531-547.

Hilbe, Joseph M. (2007). *Negative Binomial Regression*, Cambridge, UK: Cambridge University Press.

Hur, K., "A Random Effects Zero Inflated Poisson Regression Model for Clustered Extra - Zero Counts," Ph.D. Dissertation, Dartmouth Medical School, 1998.

Jain, C. and P. Consul, "A generalized Negative Binomial Distribution," *SIAM Journal of Applied Mathematics*," 21, 4, 1971, pp. 501-503.

Jung, R. and R. Winkelmann, "Two Aspects of Labor Mobility: A Bivariate Poisson Regression Approach," *Empirical Economics*, 18, 1993, pp. 543-556.

King, G., "A Seemingly Unrelated Poisson Regression Model," *Sociological Methods and Research*, 17, 3, 1989, pp. 235-255.

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

Krinsky, I. and L. Robb. "On Approximating the Statistical Properties of Elasticities." *Review of Economics and Statistics*, 68, 4, 1986, pp. 715-719.

Lambert, D. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." *Technometrics,* 34, 1, 1992, pp. 1–14.

Lancaster, T., "The Incidental Parameters Problem since 1948 ", *Journal of Econometrics*, 95, 2, 2000, pp. 391-414.

Million, A., "Models for Correlated Count Data," Unpublished dissertation, University of Munich, 1998.

Mullahy, J. "Specification and Testing of Some Modified Count Data Models." *Journal of Econometrics,* 33, 1986, pp. 341–365.

Munkin, M. and P. Trivedi, "Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models, with Application," *Econometrics Journal*, 2, 1999, pp. 29-49.

Murphy, K., and R. Topel. "Estimation and Inference in Two Step Econometric Models." *Journal of Business and Economic Statistics,* 3, 1985, pp. 370–379, Reprinted, 20, 2002, pp. 88-97

Neyman, J. and E. Scott, "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1948, pp. 1-32.

Riphahn, R., A. Wambach, and A. Million, A., "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation" *Journal of Applied Econometrics*, Vol. 18, No. 4, 2003, pp. 387-405

Rose, C., W. Martin, K. Wannemuehler and B. Plikaytis, "On the Use of Zero-Inflated and Hurdle Models for Modeling Vaccine Adverse Event Count Data," *Journal of Biopharmaceutical Statistics*, 16, 4, 2006, 463-481.

Stata. *Stata User's Guide, Version 9*. College Station, TX: Stata Press 2006.

Terza, J. "A Tobit Type Estimator for the Censored Poisson Regression Model." *Economics Letters,* 18, 1985b, pp. 361–365.

Terza, J. "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects." *Journal of Econometrics,* 84, 1, 1998, pp. 129–154.

Train, K., *Discrete Choice Models with Simulation*, Cambridge University Press, Cambridge, 2003.

Vuong, Q., "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, 57, 1989, pp. 307-344.

Wang, P., I. Cockburn, and L. Puterman, "Analysis of Patent Data – A Mixed-Poisson-Regression-Model Approach," *Journal of Business and Economic Statistics*, 16, 1, 1998, pp. 27-41

Winkelmann, R., *Econometric Analysis of Count Data*, Springer Verlag, Heidelberg, 4th ed. 2003.

Winkelmann, R., "Health Care Reform and the Number of Doctor Visits – An Econometric Analysis", *Journal of Applied Econometrics*, Vol. 19, No. 4, 2004, pp. 455-472.

Winkelmann, R. and K. Zimmermann, "A New Approach for Modeling Economic Count Data," *Economics Letters*, 37, 1991, pp. 139-143.

Winkelmann, R. and K. Zimmermann, "Recent Developments in Count Data Modelling: Theory and Application," *Journal of Economic Surveys*, 9, 1, 1995, pp. 1-36.

Wong, W. and Famoye, F. [1997] 'Modeling Household Fertility Decisions with Generalized Poisson Regression,' *Journal of Population Economics*, 10, pp. 273-283.

Xie, H., K. Hur and G. McHugo, "Using Random-Effects Zero-Inflated Poisson Model to Analyze Longitudinal Count Data with Extra Zeros," Manuscript, Dartmouth Medical School, http://isi.cbs.nl/iamamember/CD2/pdf/582.PDF, 2006.

Yen, S. and V. Adamowicz, "Participation, Trip Frequency and Site Choice: a Multinomial-Poisson Hurdle Model of Recreation Demand," *Canadian Journal of Agricultural Economics*, 42, 1, 1994, pp. 65-76.

Zorn, C., "An Analytic and Empirical Examination of Zero-Inflated and Hurdle Poisson Specifications, *Sociological Methods and Research*, 26, No. 3, 1998, pp. 368-400.

Table 1. Variables in German Health Care Data File

Variable	Measurement	Mean	Standard Deviation
ID	household identification, 1,...,7293		
YEAR	calendar year of the observation	1987.82	3.17087
YEAR1984	dummy variable for 1984 observation	.141770	.348820
YEAR1985	dummy variable for 1984 observation	.138842	.345788
YEAR1986	dummy variable for 1984 observation	.138769	.345712
YEAR1987	dummy variable for 1984 observation	.134158	.340828
YEAR1988	dummy variable for 1984 observation	.164056	.370333
YEAR1991	dummy variable for 1984 observation	.158823	.365518
YEAR1994	dummy variable for 1984 observation	.123582	.329110
AGE	age in years	43.5257	11.3302
AGESQ**	age saquared/1000	2.02286	1.00408
FEMALE	female = 1; male = 0	.478775	.499558
MARRIED	married = 1; else = 0	.758618	.427929
HHKIDS	children under age 16 in the household = 1; else = 0	.402730	.490456
*HHNINC****	household nominal monthly net income, German marks / 10000	.352084	.176908
EDUC	years of schooling	11.3206	2.32489
WORKING	employed = 1; else = 0	.677048	.467613
BLUEC	blue collar employee = 1; else = 0	.243761	.429358
WHITEC	white collar employee = 1; else = 0	.299605	.458093
SELF	self employed = 1; else = 0	.0621752	.241478
CIVIL	civil servant = 1; else = 0	.0746908	.262897
HAUPTS	highest schooling degree is Hauptschul = 1; else = 0	.624277	.484318
REALS	highest schooling degree is Realschul = 1; else = 0	.196809	.397594
FACHHS	highest schooling degree is Polytechnical= 1; else = 0	.0408402	.197924
ABITUR	highest schooling degree is Abitur = 1; else = 0	.117031	.321464
UNIV	highest schooling degree is university = 1; else = 0	.0719461	.258403
HSAT	health satisfaction, 0 - 10	6.78543	2.29372
*NEWHSAT**,***	health satisfaction, 0 - 10	6.78566	2.29373
HANDDUM	handicapped = 1; else = 0	.214015	.410028
HANDPER	degree of handicap in pct, 0 - 100	7.01229	19.2646
DOCVIS	number of doctor visits in last three months	3.18352	5.68969
*DOCTOR***	1 if DOCVIS > 0, 0 else	629108	.483052
HOSPVIS	number of hospital visits in last calendar year	.138257	.884339
*HOSPITAL***	1 of HOSPVIS > 0, 0 else	.0876455	.282784
PUBLIC	insured in public health insurance = 1; else = 0	.885713	.318165
ADDON	insured by add-on insurance = 1; else = 0	.0188099	.135856

Data source: http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/.

From Riphahn, R., A. Wambach and A. Million "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation," *Journal of Applied Econometrics*, 18, 4, 2003, pp. 387-405.

Notes: * *NEWHSAT* = *HSAT*; 40 observations on HSAT recorded between 6 and 7 were changed to 7.

** Transformed variable not in raw data file.

*** Divided by 1,000 rather than 10,000 by RWM. We used this scale to ease comparison of coefficients.

Table 2. Descriptive Statistics by Gender

Variable	Males		Females	
	Mean	Standard Dev.	Mean	Standard Dev.
YEAR	1987.84	3.19003	1987.80	3.14985
YEAR1984	.141613	.348665	.141940	.349002
YEAR1985	.138875	.345828	.138806	.345757
YEAR1986	.138173	.345094	.139418	.346395
YEAR1987	.134171	.340848	.134144	.340820
YEAR1988	.162396	.368826	.165864	.371973
YEAR1991	.157551	.364332	.160208	.366813
YEAR1994	.127220	.333231	.119621	.324530
AGE	42.6526	11.2704	44.4760	11.3192
AGESQ	1.94628	.987385	2.10623	1.01543
FEMALE	.000000	.000000	1.00000	.000000
MARRIED	.765148	.423921	.751510	.432154
HHKIDS	.412975	.492386	.391577	.488122
HHNINC	.359054	.173564	.344495	.180179
EDUC	11.7287	2.43649	10.8764	2.10911
WORKING	.850312	.356777	.488420	.499885
BLUEC	.340237	.473805	.138730	.345677
WHITEC	.299937	.458246	.299243	.457944
SELF	.0856561	.279865	.0366124	.187815
CIVIL	.117812	.322397	.0277459	.164250
HAUPTS	.601137	.489682	.649469	.477155
REALS	.176086	.380907	.219369	.413835
FACHHS	.0536404	.225315	.0269051	.161812
ABITUR	.146949	.354068	.0844608	.278088
UNIV	.0961876	.294859	.0455553	.208527
HSAT	6.92436	2.25148	6.63417	2.32951
NEWHSAT	6.92459	2.25148	6.63441	2.32953
HANDDUM	.227295	.419007	.199559	.399538
HANDPER	8.13371	20.3288	5.79143	17.9562
DOCVIS	2.62571	5.21121	3.79080	6.11113
DOCTOR	.559503	.496464	.704884	.456112
HOSPVIS	.127782	.930209	.149660	.831416
HOSPITAL	.0779330	.268076	.0982191	.297622
PUBLIC	.861055	.345902	.912558	.282492
ADDON	.9175525	.131323	.0201789	.140617
Sample Size	14,243		13,083	

Table 3. Poisson Models and Heterogeneity in Poisson (t ratios in parentheses)

Variable	Pooled Poisson	Males			Females		
		Poisson	Log gamma	Lognormal	Poisson	Log gamma	Lognormal
Constant	2.639 (39.46)	2.771 (28.85)	3.1488 (13.74)	2.8079 (11.26)	2.546 (28.54)	3.0245 (15.03)	2.7556 (12.47)
AGE	-0.00732 (-2.64)	-0.02387 (-5.44)	-0.03983 (-4.07)	-0.05858 (-5.51)	-0.01320 (-3.64)	-0.03119 (-3.78)	-0.04485 (-4.90)
AGESQ	0.1407 (4.54)	0.3693 (7.45)	0.5467 (4.77)	0.7853 (6.45)	0.1794 (4.46)	0.3727 (4.02)	0.5421 (5.27)
HSAT	-0.2149 (-151.9)	-0.2253 (-104.1)	-0.2392 (-42.44)	-0.2650 (-50.93)	-0.2034 (-108.3)	-0.2080 (-47.30)	-0.2225 (-46.54)
HANDDUM	0.1011 (8.71)	0.06899 (4.09)	-0.02090 (-0.46)	-0.01093 (-0.23)	0.1379 (8.55)	0.1133 (2.79)	0.1011 (2.48)
HANDPER	0.001992 (10.73)	0.002858 (10.04)	0.006614 (8.05)	0.007398 (9.08)	0.002414 (9.48)	0.004359 (5.92)	0.004432 (6.14)
MARRIED	0.02058 (2.32)	0.05831 (3.89)	0.06582 (2.18)	0.1276 (3.67)	0.02718 (2.39)	0.02816 (1.13)	0.04590 (1.63)
EDUC	-0.01483 (-7.96)	-0.02348 (-8.43)	-0.02623 (-4.59)	-0.02297 (-3.43)	0.01473 (5.65)	0.007725 (1.36)	0.01318 (2.09)
HHNINC	-0.1729 (-7.27)	-0.2220 (-5.93)	-0.1917 (-2.48)	-0.1257 (-1.44)	-0.2063 (-6.53)	-0.1624 (-2.57)	-0.1417 (-1.92)
HHKIDS	-0.1108 (-12.86)	-0.07598 (-5.75)	-0.08440 (-3.32)	-0.09013 (-2.94)	-0.1338 (-11.63)	-0.1243 (-4.91)	-0.1360 (-4.81)
SELF	-0.2914 (-16.18)	-0.2110 (-8.98)	-0.2179 (-5.02)	-0.3590 (-6.81)	-0.2175 (-7.47)	-0.2424 (-4.51)	-0.2885 (-4.55)
CIVIL	-0.05026 (-2.64)	0.09144 (3.78)	0.08411 (1.56)	0.01916 (0.32)	-0.07113 (-1.91)	-0.01982 (-0.34)	-0.03188 (-0.39)
BLUEC	-0.08920 (-9.01)	0.01779 (1.24)	.03706 (1.20)	-0.03137 (-.93)	-0.03543 (-2.38)	-0.04010 (-1.31)	-0.09991 (-2.81)
WORKING	-0.07478 (-7.62)	-0.05539 (-3.17)	-0.01545 (-0.38)	0.03119 (0.78)	0.01490 (1.29)	0.03046 (1.23)	0.03851 (1.38)
PUBLIC	0.1145 (7.32)	0.1001 (4.27)	.09340 (1.83)	0.05150 (0.91)	0.1312 (6.22)	0.09530 (2.44)	0.08076 (1.72)
ADDON	0.06084 (2.39)	0.06655 (1.63)	0.05506 (0.50)	0.1954 (1.81)	0.02071 (0.63)	0.03088 (0.32)	0.1175 (1.25)
θ			**0.5707 (59.96)**			**0.8289 (64.44)**	
κ			1.7522 (59.96)			1.2064 (64.44)	
σ(ε)			1.9874 (72.19)	**1.2520 (104.61)**		1.4757 (84.33)	**1.0608 (114.80)**
σ(h)			1.3237 (119.92)	4.2651 (29.49)		1.1.0984 (128.89)	2.5325 (41.13)
ln *L*	-89641.2	-42774.7	-27480.4	-27408.6	-45900.2	-30262.3	-30214.7
n	27326	14243			13083		

Notes: Estimated coefficients for year dummy variables, excluding year 1984, are not reported.

θ = the estimated parameter for the log gamma (NB) model

κ = 1/θ = Var[*h*] for log gamma model.

$\sigma(\varepsilon) = \sqrt{\psi'(\theta)}$ = Var(lnh_i) for the log gamma model. Estimated directly for the lognormal model.

$\sigma(h) = \sqrt{\kappa}$ for the log gamma model, $\sqrt{\exp(\sigma^2)[\exp(\sigma^2)-1]}$ for the lognormal model.

Table 4. Specifications for the Negative Binomial Model (t ratios in parentheses)

Variable	Poisson	NB 1	NB 2	NB P	NB lognormal	Heterogeneous NB
Constant	2.771 (28.85)	2.7760 (14.06)	3.1488 (13.74)	3.0500 (13.14)	3.1751 (12.0)	3.1446 (13.72)
AGE	-0.02387 (-5.44)	-.04768 (-5.61)	-0.03983 (-4.07)	-.04679 (-4.76)	-.05635 (-4.86)	-.04018 (-4.08)
AGESQ	0.3693 (7.45)	.6340 (6.58)	0.5467 (4.77)	0.6373 (5.66)	.7600 (5.62)	.5507 (4.79)
HSAT	-0.2253 (-104.1)	-.1886 (-44.58)	-0.2392 (-42.44)	-.2279 (-46.17)	-.2558 (-39.16)	-.2396 (-42.30)
HANDDUM	0.06899 (4.09)	.02292 (0.67)	-0.02090 (-0.46)	.01660 (0.41)	-.01128 (-0.21)	-.02042 (-0.44)
HANDPER	0.002858 (10.04)	.004141 (7.33)	0.006614 (8.05)	.005031 (7.30)	.006995 (7.22)	.006631 (7.93)
MARRIED	0.05831 (3.89)	.1299 (4.50)	0.06582 (2.18)	.1139 (3.55)	.1054 (2.83)	.06592 (2.17)
EDUC	-0.02348 (-8.43)	-.009550 (-1.80)	-0.02623 (-4.59)	-.01794 (-2.86)	-.02083 (-3.06)	-.02595 (-4.46)
HHNINC	-0.2220 (-5.93)	-.07878 (-1.13)	-0.1917 (-2.48)	-.1462 (-1.78)	-.1470 (-1.63)	-.1673 (-2.30)
HHKIDS	-0.07598 (-5.75)	-.07435 (-2.95)	-0.08440 (-3.32)	-.08672 (-3.10)	-.08395 (-2.58)	-.08219 (-3.23)
SELF	-0.2110 (-8.98)	-.2439 (-5.56)	-0.2179 (-5.02)	-.2628 (-5.42)	-.2887 (-5.38)	-.2130 (-4.92)
CIVIL	0.09144 (3.78)	.02782 (0.60)	0.08411 (1.56)	.05148 (0.93)	.04781 (0.79)	.08376 (1.58)
BLUEC	0.01779 (1.24)	-.009478 (-0.35)	.03706 (1.20)	.005597 (0.17)	.01142 (0.32)	.04093 (1.32)
WORKING	-0.05539 (-3.17)	.01258 (0.37)	-0.01545 (-0.38)	-.001046 (-0.20)	.02404 (.50)	-.01888 (-0.46)
PUBLIC	0.1001 (4.27)	.06067 (1.38)	.09340 (1.83)	.07823 (1.50)	.06632 (1.16)	.09601 (1.91)
ADDON	0.06655 (1.63)	.1393 (1.72)	0.05506 (0.50)	.1363 (1.34)	.1250 (1.07)	.05276 (0.48)
θ		0.2058 (62.08)	0.5707 (59.96)	0.3460 (36.47)	0.9043 (18.17)	0.5112 (11.61)
κ	0.0000 (fixed)	4.8598 (62.08)	1.7522 (59.96)	2.8905 (36.47)	1.1058 (18.17)	1.9562 (11.61)
P	0.0000 (fixed)	1.0000 (fixed)	2.0000 (fixed)	1.4897 (64.96)	2.0000 (fixed)	2.0000 (fixed)
σ(ε)	0.0000 (fixed)	0.0000 (fixed)	0.0000 (fixed)	0.0000 (fixed)	0.6961 (19.39)	0.0000 (fixed)
HHNINC	0.0000 (fixed)	0.0000 (fixed)	0.0000 (fixed)	0.0000 (fixed)	0.0000 (fixed)	-.3388 (-3.24)
EDUC	0.0000 (fixed)	0.0000 (fixed)	0.0000 (fixed)	0.0000 (fixed)	0.0000 (fixed)	.000540 (0.071)
ln *L*	-42774.7	-27410.0	-27480.4	-27306.1	-27357.8	-27475.7

Notes: Estimated coefficients for year dummy variables, excluding year 1984, are not reported.
θ = the estimated parameter for the log gamma (NB) model
κ = 1/θ = Var[*h*] for log gamma model.

Table 5 Estimated Sample Selection Model, Males

Variable	ρ=0 - No Selection				Sample Selection			
	Probit - PUBLIC		Poisson - DOCVIS		Probit - PUBLIC		Poisson - DOCVIS	
Constant	3.1416	(11.24)	2.7833	(10.77)	3.1465	(11.24)	2.8533	(10.01)
AGE			-.06454	(-5.67)			-.06691	(-5.73)
AGESQ			.8689	(6.62)			.8974	(6.65)
HSAT			-.2548	(-44.51)			-.2565	(-45.15)
HANDDUM	.07328	(1.06)	.006177	(.12)	.07362	(1.06)	.008152	(.16)
HANDPER	-.0000	(-.001)	.007824	(9.28)	.0000	(-.001)	.007792	(9.24)
MARRIED	-.01347	(-.31)	.06478	(1.80)	-.01353	(-.33)	.06610	(1.83)
EDUC	-.1808	(-7.39)	-.01501	(1.91)	-.1813	(-7.09)	-.01681	(-1.58)
HHNINC	-1.024	(-11.94)	-.08076	(-.85)	-1.025	(-14.41)	-.08917	(-.85)
HHKIDS	.06995	(1.85)	-.02862	(-.89)	.07007	(1.88)	-.03045	(-.95)
SELF			-.3116	(-5.54)			-.3264	(-5.11)
CIVIL			.02953	(.38)			.01633	(.19)
BLUEC			.006345	(.18)			-.005491	(-.13)
WORKING			.03976	(.90)			.05594	(1.15)
YEAR1985	.04480	(.75)	.07375	(1.53)	.04418	(.73)	.07153	(1.49)
YEAR1986	-.01027	(-.17)	.1704	(3.52)	-.01094	(-.18)	.1694	(3.51)
YEAR1987	-.08772	(-1.14)	.07857	(1.38)	-.08815	(-1.14)	.07343	(1.30)
YEAR1988	-.06434	(-1.15)	.09794	(2.02)	-.06476	(-1.15)	.09290	(1.93)
YEAR1991	-.03130	(-.55)	.1077	(2.06)	-.03161	(-.56)	.1015	(1.95)
YEAR1994	.05901	(.98)	.3428	(6.52)	.05850	(.96)	.3397	(6.48)
AGE3034	-.2211	(-4.06)			-.2208	(-4.00)		
AGE3539	-.3089	(-6.02)			-.3089	(-6.0)		
AGE4044	.5164	(9.96)			.5160	(10.08)		
AGE4549	-.1473	(-2.73)			-.1468	(-2.78)		
AGE5054	.2120	(3.55)			.2113	(3.43)		
AGE5559	.4462	(6.53)			.4463	(6.23)		
AGE60UP	.5531	(7.25)			.5532	(7.04)		
HAUPTS	.3631	(3.36)			.3635	(3.51)		
REALS	-.3657	(-3.21)			-.3646	(-3.24)		
FACHHS	.1456	(.95)			.1476	(.94)		
ABITUR	.06202	(.40)			.06490	(.40)		
UNIV	.03097	(.31)			.03274	(.31)		
WHITEC	1.1305	(27.83)			1.1307	(30.23)		
σ			1.2377	(99.92)			1.2394	(97.08)
ρ	0.0000	(fixed)			.02246	(.23)		
ln L	-4294.89		-24044.20		-28339.01			
n	14243		12264		14243		12264	

Table 6 Estimated Sample Selection Model, Females (t ratios in parentheses)

Variable	ρ=0 - No Selection				Sample Selection			
	Probit - PUBLIC		Poisson - DOCVIS		Probit - PUBLIC		Poisson - DOCVIS	
Constant	3.0642	(11.76)	2.5493	(11.52)	3.0694	(11.40)	2.1856	(9.51)
AGE			-.04117	(-4.37)			-.03866	(-4.07)
AGESQ			.5218	(4.93)			.4857	(4.56)
HSAT			-.2215	(-45.51)			-.2197	(-43.31)
HANDDUM	.1104	(1.19)	.1001	(2.48)	.07195	(.85)	.1205	(2.43)
HANDPER	.003088	(1.97)	.004725	(6.64)	.0032745	(2.32)	.004808	(5.93)
MARRIED	.003124	(.95)	.05571	(1.91)	-.0027795	(-.06)	.03836	(1.28)
EDUC	-.1678	(-7.38)	.02748	(3.91)	-.1593	(-7.02)	.06077	(7.47)
HHNINC	-1.183	(-12.9)	-.1607	(-2.08)	-1.1620	(-15.00)	.08070	(.97)
HHKIDS	.08852	(1.95)	-.1306	(-4.44)	.08804	(1.89)	-.1255	(-4.23)
SELF			-.3444	(-4.78)			-.2340	(-3.14)
CIVIL			.2667	(1.58)			.4428	(2.80)
BLUEC			-.09816	(-2.73)			.03352	(.90)
WORKING			.04561	(1.61)			-.05897	(-1.88)
YEAR1985	-.00536	(-.07)	-.0279	(2.19)	.007422	(.11)	-.03663	(-.86)
YEAR1986	-.00267	(-.04)	.09287	(.028)	-.005941	(-.09)	.1289	(3.08)
YEAR1987	-.1215	(-1.21)	-.03763	(-.79)	-.08333	(-.89)	-.08684	(-1.54)
YEAR1988	-.0909	(-1.42)	-.1467	(-3.46)	-.09197	(-1.43)	-.1391	(-3.18)
YEAR1991	.03407	(.51)	-.06931	(-1.54)	.01679	(.26)	-.06926	(-1.51)
YEAR1994	.1665	(2.28)	.2642	(5.86)	.1555	(2.16)	.2202	(4.76)
AGE3034	-.1565	(-2.27)			-.1604	(-2.35)		
AGE3539	-.1581	(-2.49)			-.1642	(-2.57)		
AGE4044	.1901	(3.00)			.1809	(2.81)		
AGE4549	-.1017	(-1.56)			-.08129	(-1.24)		
AGE5054	.06098	(.89)			.07611	(1.14)		
AGE5559	.1323	(1.83)			.1231	(1.72)		
AGE60UP	.1221	(1.68)			.1095	(1.47)		
HAUPTS	.4737	(3.89)			.3792	(3.21)		
REALS	.2016	(1.56)			.1203	(.98)		
FACHHS	.4686	(2.66)			.3346	(1.97)		
ABITUR	.4092	(2.38)			.2705	(1.64)		
UNIV	-.1886	(1.70)			-.2206	(2.00)		
WHITEC	.9482	(18.99)			.9087	(18.98)		
σ			1.0560	(109.96)			1.0993	(84.92)
ρ	0.0000 (fixed)				-.5339 (-8.87)			
ln L	-3099.5		-27833.5		-30927.7			
n	13083		11939		13083		11939	

Table 7 Estimated Zero Inflated Poisson Model, Males (t ratios in parentheses)

Variable	No Zero Inflation		Zero Inflated Poisson (No Heterogeneity)	
	Probit – Zero State	Poisson - DOCVIS	Probit – Zero State	Poisson - DOCVIS
Constant		2.771 (28.85)	-.01432 (-.07)	2.5722 (57.10)
AGE		-.02387 (-5.44)		-.005731 (-3.21)
AGESQ		.3693 (7.45)		.1285 (6.34)
HSAT		-.2253 (-104.1)		-.1564 (-186.94)
HANDDUM		.06899 (4.09)	.1274 (3.58)	.07641 (12.53)
HANDPER		.00286 (10.04)	-.01374 (-15.65)	.00081 (8.34)
MARRIED		.05831 (3.89)	-.1228 (-3.79)	-.01578 (-2.71)
EDUC		-.02348 (-8.43)	-.01112 (-.65)	-.01896 (-14.04)
HHNINC		-.2220 (-5.93)	-.06780 (-.92)	-.2240 (-14.07)
HHKIDS		-.07598 (-5.75)	.07732 (2.67)	-.03190 (-6.19)
SELF		-.2110 (-8.98)		-.1084 (-10.49)
CIVIL		.09144 (3.78)		.1111 (10.16)
BLUEC		.01779 (1.24)		.04336 (7.21)
WORKING		-.05539 (-3.17)		-.05992 (-9.02)
PUBLIC		.1001 (4.27)		.07612 (7.07)
ADDON		.06655 (1.63)		-.07040 (-3.67)
YEAR1985		2.7719 (28.85)		.09084 (11.58)
YEAR1986		-.02387 (-5.44)		.1840 (23.93)
YEAR1987		.3693 (7.45)		.1150 (14.64)
YEAR1988		-.2253 (-104.1)		.00065 (.08)
YEAR1991		.06899 (4.09)		-.1058 (-12.99)
YEAR1994		.00286 (10.04)		.1810 (22.07)
AGE3034			.02866 (.69)	
AGE3539			.06424 (1.58)	
AGE4044			-.1434 (-3.45)	
AGE4549			.1521 (3.47)	
AGE5054			-.1562 (-3.47)	
AGE5559			-.1864 (-3.93)	
AGE60UP			-.3261 (-5.89)	
HAUPTS			.06593 (.79)	
REALS			.07693 (.85)	
FACHHS			.1143 (.95)	
ABITUR			.2539 (2.10)	
UNIV			.01076 (.13)	
WHITEC			-.005596 (-.21)	
σ				0.0000 (fixed)
ρ			(0.0000) (fixed)	
ln *L*	-42774.7		-35757.0	
n	14243		14243	
Voung Stat.	0.00		28.83	

Table 8 Estimated Zero Inflated Poisson Models with Latent Heterogeneity, Males (t ratios in parentheses)

Variable	Exogenous Zero Inflation		Endogenous Zero Inflation	
	Probit – Zero State	Poisson - DOCVIS	Probit – Zero State	Poisson - DOCVIS
Constant	-.3218 (-.95)	2.4564 (8.83)	-.3015 (-.93)	2.5220 (8.74)
AGE		-.02405 (-2.02)		-.02436 (-1.96)
AGESQ		.3650 (2.67)		.36660 (2.56)
HSAT		-.2310 (-41.48)		-.2312 (-39.85)
HANDDUM	.3784 (5.02)	.03262 (.72)	.3933 (5.12)	.03832 (.82)
HANDPER	-.02667 (-6.36)	.002292 (2.98)	-.02830 (-6.36)	.001453 (1.43)
MARRIED	-.2088 (-3.29)	.005705 (.14)	-.1916 (-3.03)	.007373 (.17)
EDUC	-.02912 (-.96)	-.01201 (-1.52)	-.02994 (-1.03)	-.01053 (-1.26)
HHNINC	-.2411 (-1.51)	-.2310 (-2.39)	-.2410 (-1.53)	-.2498 (-2.46)
HHKIDS	.1193 (2.05)	-.02546 (-.72)	.1105 (1.92)	-.02517 (-.65)
SELF		-.2350 (-4.22)		-.2474 (-4.33)
CIVIL		.07712 (1.23)		.06009 (.94)
BLUEC		.04352 (1.13)		.04617 (1.15)
WORKING		-.05616 (-1.21)		-.04573 (-.94)
PUBLIC		.06828 (1.21)		.04932 (.88)
ADDON		.05530 (.52)		.05892 (.56)
YEAR1985		.1125 (2.41)		.1237 (2.67)
YEAR1986		.2171 (4.67)		.2198 (4.73)
YEAR1987		.2142 (4.06)		.2299 (4.26)
YEAR1988		.05805 (1.26)		.06311 (1.37)
YEAR1991		.02592 (.53)		.02908 (.59)
YEAR1994		.3226 (6.43)		.3189 (6.21)
AGE3034	.05143 (.71)		.05491 (.80)	
AGE3539	.1164 (1.70)		.1168 (1.76)	
AGE4044	-.2525 (-3.57)		-.2409 (-3.43)	
AGE4549	.1770 (2.47)		.1735 (2.51)	
AGE5054	-.2060 (-2.49)		-.1892 (-2.41)	
AGE5559	-.1555 (-1.79)		-.1464 (-1.75)	
AGE60UP	-.4165 (-3.29)		-.3874 (-3.19)	
HAUPTS	.1363 (.83)		.1045 (.68)	
REALS	.1287 (.73)		.1034 (.63)	
FACHHS	.2027 (.91)		.1796 (.86)	
ABITUR	.4061 (1.84)		.3680 (1.76)	
UNIV	.07217 (.51)		.09214 (.69)	
WHITEC	-.03377 (-.63)		-.02770 (-.52)	
σ		.9875 (70.08)		.9902 (66.31)
ρ	0.0000 (fixed)		.1540 (.14)	
ln *L*	-27183.9		-27183.1	
n	14243		14243	
Vuong Stat.	24.16		24.17	

Table 9. Single Equations Estimates of Bivariate Poisson Models

Variable	Males		Females	
	DocVis	HospVis	DocVis	HospVis
Constant	2.95829046	-.71769354	2.65970461	-.92395935
AGE	-.06039694	-.01495334	-.03467601	-.04182897
AGESQ	.81046943	.17526254	.40417087	.32800832
HSAT	-.26979714	-.28467610	-.22492822	-.21283211
HANDDUM	-.10360835	-.15938867	.14118341	.04898526
HANDPER	.00811821	.00687895	.00436759	.01053000
MARRIED	.06801734	-.11266046	.09394893	-.03039990
EDUC	-.02141476	-.05611756	.00669951	-.02245035
HHNINC	-.09317739	.28561716	-.10306991	.45290684
HHKIDS	-.05833037	.06532961	-.18274059	.02830548
SELF	-.27162867	-.07241379	-.27422854	-.12053254
CIVIL	.03028441	-.15292197	-.00966680	.16538415
BLUEC	-.05264588	.19299252	-.09417967	-.31982329
WORKING	.00747418	-.29732586	.01608792	.01324764
PUBLIC	.04537642	-.25102523	.08274879	.07493233
ADDON	.15634672	.61629605	.11353610	.28858560
YEAR1984	.00000000	.00000000	.00000000	.00000000
YEAR1985	.01555697	.38790097	-.02109554	.18039499
YEAR1986	.14371463	-.03957923	.11756445	.28653104
YEAR1987	.16606851	.06845882	-.10021643	.12299139
YEAR1988	.04018257	-.05038515	-.16888748	.43457292
YEAR1991	.02525603	-.06140680	-.06445527	.44076348
YEAR1994	.28195404	.07614490	.26471138	.13759322
σ	1.23777937	1.78190925	1.04809696	1.47269236
σ^2	1.53209777	3.17228488	1.09850724	2.16882279
σ_ε	.96737031	1.28460990	.79772504	1.00736894
ω_u	.77219975	1.23371888	.67981026	1.07425817
ρ_{DW}	0.276		0.201	
Average Correlation	0.06938		0.05795	
Sample Corr(Doc,Hosp)	0.1477		0.1255	
RWM Reported Results				
$\sigma(\varepsilon_{it})$	0.996	1.244	0.822	1.053
$\omega(u_i)$	0.795	1.195	0.701	1.123
$\sigma^2(\varepsilon_{it})$	0.992	1.548	0.676	1.109
$\omega^2(u_i)$	0.632	1.428	0.491	1.261
$p=\sigma^2(\varepsilon_{it})/[\sigma^2(\varepsilon_{it})+\omega^2(u_i)]$	0.6108	0.5202	0.5793	0.4679
$\tau=[\sigma^2(\varepsilon_{it})+\omega^2(u_i)]^{1/2}$	1.274	1.725	1.080	1.540
ρ_{DH}	0.276		0.201	
ρ	0.490		0.386	

Table 10. RWM Estimated Bivariate Poisson Models

Variable	**Males**		**Females**	
	DocVis	**HospVis**	**DocVis**	**HospVis**
Constant	2.563	-0.206	2.423	-1.567
AGE	-0.060	-0.077	-0.040	-0.032
AGESQ	0.823	0.942	0.499	0.234
HSAT	-0.237	-0.243	-0.191	-0.196
HANDDUM	-0.029	-0.086	0.063	0.039
HANDPER	0.007	0.008	0.004	0.010
MARRIED	0.085	-0.054	0.009	-0.044
EDUC	-0.022	-0.051	0.014	-0.015
HHNINC	-0.090	0.375	-0.107	0.407
HHKIDS	-0.059	0.103	-0.117	0.073
SELF	-0.356	-0.196	-0.256	-0.117
CIVIL	-0.011	-0.086	-0.069	0.281
BLUEC	-0.029	0.173	-0.034	-0.320
WORKING	0.041	-0.026	0.002	-0.014
PUBLIC	0.075	-0.136	0.058	0.246
ADDON	0.090	0.549	0.096	0.219
σ (ε_{it})	0.996	1.244	0.822	1.053
ω (u_i)	0.795	1.195	0.701	1.123
ρ	0.490		0.386	

Table 11 Estimated Panel Data Poisson Models, Males (t ratios in parentheses)

Variable	Fixed Effects		Random Effects	
	No Effects	Unconditional FE	log gamma (NB)	lognormal
Constant	2.639 (39.46)		2.6369 (24.56)	2.0775 (19.39)
AGE	-.00732 (-2.64)	.0008051 (.06)	-.02950 (-7.56)	-.02694 (-6.96)
AGESQ	.1407 (4.54)	.4797 (4.42)	.4883 (10.94)	.5003 (11.39)
HSAT	-.2149 (-51.9)	-.1682 (-50.59)	-.1808 (-160.17)	-.1828 (-161.27)
HANDDUM	.1011 (8.71)	.003135 (.17)	-.001932 (-.24)	.000159 (.02)
HANDPER	.001992(10.73)	.0000 (.01)	.001630 (7.68)	.001198 (5.81)
MARRIED	.02058 (2.32)	-.01136 (-.34)	-.01282 (-1.22)	.03822 (3.55)
EDUC	-.01483 (-7.96)	-.06482 (-3.02)	-.03379 (-5.85)	-.03474 (-5.95)
HHNINC	-.1729 (-7.27)	-.1786 (-2.72)	-.1759 (-6.16)	-.2058 (-7.04)
HHKIDS	-.1108 (-12.86)	.04577 (1.95)	.007354 (.86)	-.01688 (-1.87)
SELF	-.2914 (-16.18)	-.03933 (-.71)	-.1372 (-7.39)	-.1517 (-8.38)
CIVIL	-.05026 (-2.64)	-.1375 (-2.01)	-.01156 (-.45)	-.01119 (-.43)
BLUEC	-.08920 (-9.01)	-.06725 (-2.18)	-.03458 (-2.63)	-.04332 (-3.34)
WORKING	-.07478 (-7.62)	.03806 (1.23)	.004875 (.37)	-.001994 (-.16)
PUBLIC	.1145 (7.32)	.1044 (2.30)	.1057 (5.53)	.1109 (5.80)
ADDON	.06084 (2.39)	-.04068 (-.73)	-.03437 (-1.19)	-.0343 (-1.21)
YEAR1985	2.639 (39.46)	.05690 (2.37)	.08268 (8.87)	.08383 (8.95)
YEAR1986	-.00732 (-2.64)	.1063 (3.53)	.1622 (18.82)	.1618 (18.86)
YEAR1987	.1407 (4.54)	.04392 (1.11)	.1145 (11.32)	.1109 (10.64)
YEAR1988	-.2149 (-151.9)	-.09314 (-1.94)	.01033 (1.00)	.002153 (.20)
YEAR1991	.1011 (8.71)	-.2429 (-3.10)	-.05520 (-4.22)	-.07157 (-5.64)
YEAR1994	.001992(10.73)	-.06790 (-.62)	.1985 (12.53)	.1713 (11.17)
κ			.9879 (38.57)	
σ				1.0051 (91.11)
$\ln L$	-42774.74	-21696.56	-32850.59	-32897.37
N	3687 (714 unusable in FE)		3687	
$\Sigma_i T_i$	14243		14243	

Table 12 Estimated Panel Data Negative Binomial Models, Males (t ratios in parentheses)

Variable	Fixed Effects		Random Effects	
	HHG	Unconditional FE	HHG	lognormal
Constant	1.2571 (4.04)		1.8500 (8.79)	2.8711 (9.85)
AGE	-.06890 (-5.23)	-.01465 (-.55)	-.06123 (-6.54)	-.04729 (-3.64)
AGESQ	.9328 (6.23)	.6122 (2.95)	.8085 (7.51)	.6677 (4.41)
HSAT	-.1461 (-26.53)	-.1858 (-27.74)	-.1839 (-42.07)	-.2287 (-37.96)
HANDDUM	-.02760 (-.74)	-.02142 (-.54)	-.01461 (-.43)	-.02789 (-.59)
HANDPER	.003961 (4.74)	.002916 (2.40)	.004813 (7.52)	.006229 (5.92)
MARRIED	.04188 (.97)	-.01870 (-.30)	.1158 (3.84)	.07753 (1.92)
EDUC	.04176 (4.09)	-.07045 (-2.02)	-.004814 (-.85)	-.02949 (-3.45)
HHNINC	-.006220 (-.07)	-.08619 (-.75)	-.04278 (-.59)	-.1071 (-1.15)
HHKIDS	.02149 (.63)	.03225 (.74)	-.05129 (-1.98)	-.05727 (-1.65)
SELF	-.2327 (-3.66)	-.3279 (-3.25)	-.2792 (-6.31)	-.3388 (-5.40)
CIVIL	-.09470 (-1.33)	-.3001 (-2.46)	.002865 (.06)	-.007380 (-.11)
BLUEC	-.1222 (-3.12)	-.1035 (-1.76)	-.05024 (-1.76)	-.02313 (-.55)
WORKING	.1358 (2.91)	.1051 (1.74)	.05998 (1.64)	.02431 (.48)
PUBLIC	.01414 (.22)	.07094 (.91)	.06681 (1.46)	.06861 (1.10)
ADDON	.1136 (1.06)	-.005359 (-.05)	.1273 (1.45)	.03729 (.32)
YEAR1985	.06908 (1.61)	.09386 (2.12)	.06592 (1.64)	.1147 (2.62)
YEAR1986	.1312 (3.15)	.1551 (2.84)	.1379 (3.57)	.2103 (4.87)
YEAR1987	.1025 (2.24)	.07871 (1.10)	.09462 (2.29)	.1335 (2.52)
YEAR1988	.06409 (1.55)	-.001798 (-.02)	.07583 (2.02)	.09372 (2.22)
YEAR1991	.06162 (1.41)	-.1119 (-.83)	.09586 (2.47)	.05652 (1.23)
YEAR1994	.2230 (4.83)	.07991 (.43)	.2544 (6.54)	.3137 (6.47)
κ		1.8131 (41.31)		1.0192 (50.76)
σ				.7979 (34.31)
a			3.1782 (21.53)	
b			6.2577 (17.94)	
ln L	-15690.87	-23000.24	-26824.63	-26881.20
N	3687 (714 have $\Sigma_t y_{it}$ = 0)		3687	
$\Sigma_i T_i$	14243		14243	

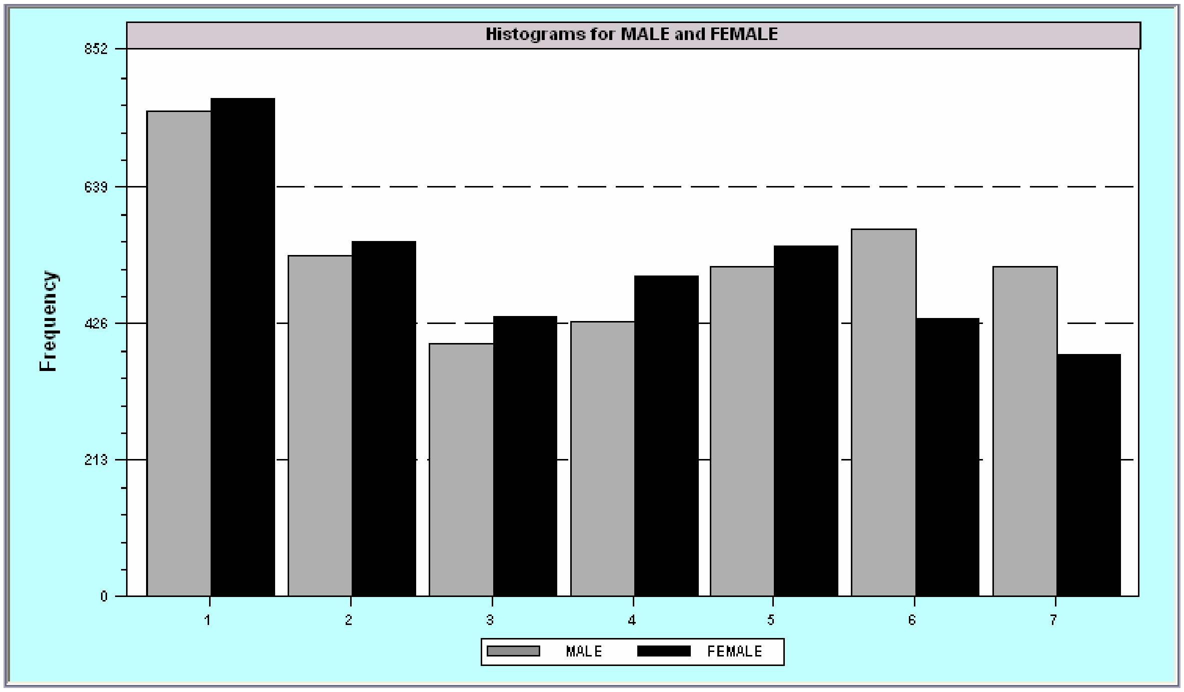

Figure 1 Group Sizes in RWM Panel Data

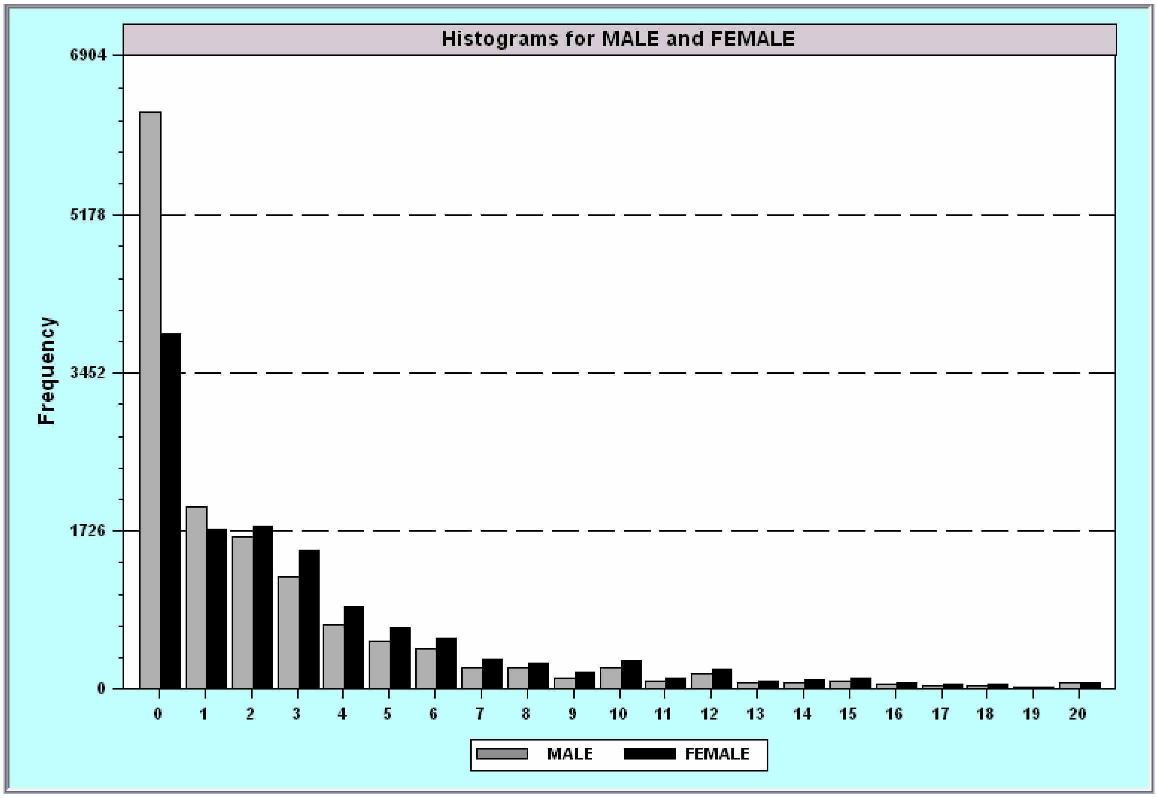

Figure 2   Histograms for DocVis

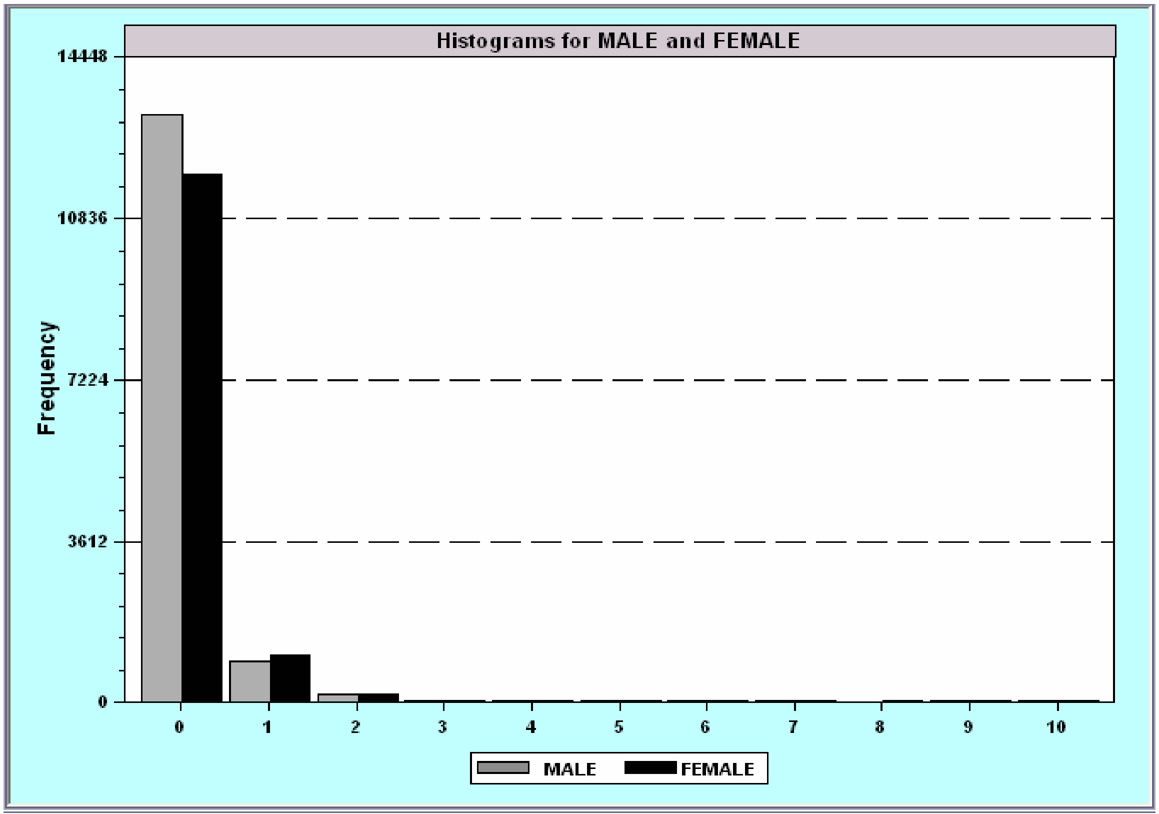

Figure 3 Histograms for HospVis

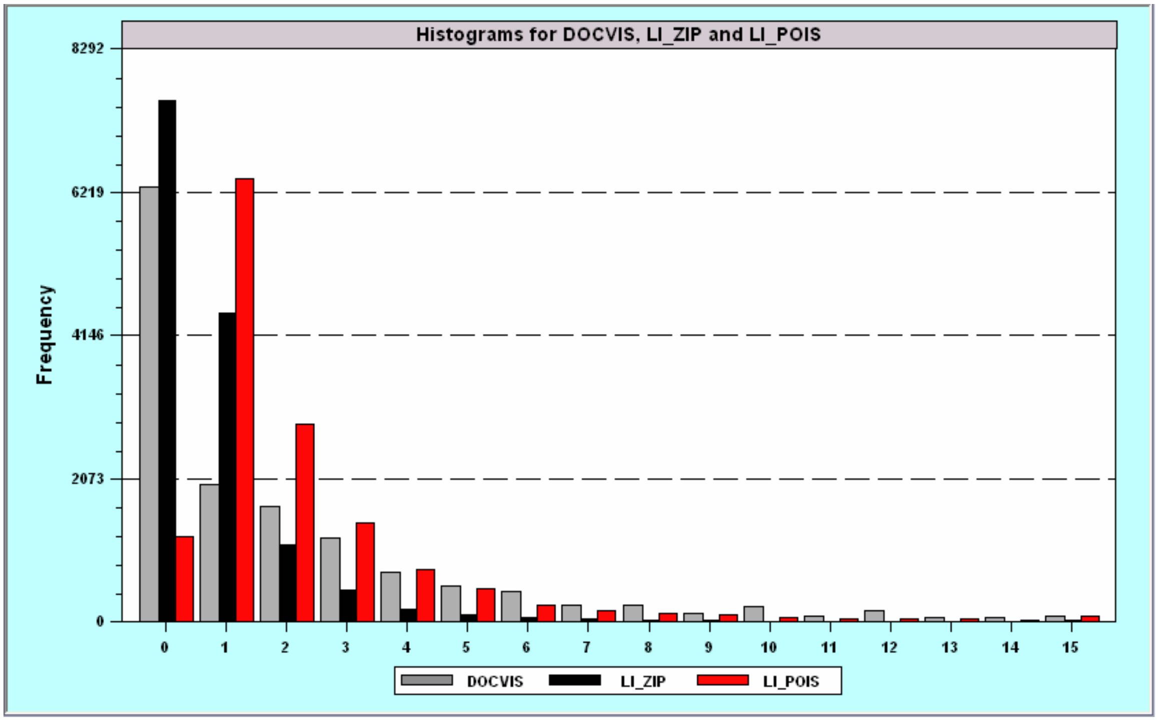

Figure 4 Predictions from ZIP and Poisson Models and Actual DocVis