Empir Econ (2009) 36:133–173
DOI 10.1007/s00181-008-0190-5

ORIGINAL PAPER

Models for count data with endogenous participation

William Greene

Received: 15 May 2007 / Accepted: 15 October 2007 / Published online: 28 March 2008

Abstract We present several extensions of the Poisson and negative binomial models for count data based on the lognormal model for latent heterogeneity in the conditional means. The lognormal model provides a versatile specification that is more flexible than the familiar log gamma form, and provides a platform for several "two part" extensions that have appeared in the literature, including zero inflation, hurdle and sample selection models. We then extend these received two- part models by allowing for endogeneity of the participation equation. We conclude with a detailed application using the data employed in a recent study of the German health care system.

Keywords Poisson regression · Count data · Heterogeneity · Lognormal · Two part model

JEL Classification C14 · C23 · C25

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]

[1] Contoyannis (2004), Munkin and Trivedi (1999), Riphahn et al. (2003). See, as well, Cameron and Trivedi (2005).

W. Greene (✉)
Department of Economics, Stern School of Business, New York University,
44 West 4th St., Rm. 7–78, New York, NY 10012, USA
e-mail: wgreene@stern.nyu.edu
URL: http://www.stern.nyu.edu/∼wgreene

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 generally 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] Several of these have employed elements of the models analyzed here, such as Winkelmann (2004), who proposes the log-normal hurdle model.

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 analyze 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 study is organized as follows: Sect. 2 will detail the basic modeling frameworks for count data, the Poisson and NB models and will develop the lognormal models for unobserved heterogeneity in count data. Section 3 will extend the lognormal model to several two part models. The two part models are applied to the Riphahn et al. (2003) (RWM) panel data on health care utilization in Sect. 4. Some conclusions are drawn in Sect. 5.

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., Cameron and Trivedi (1998) (CT) 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.)

[2] Hausman et al. (1984) (HHG) and Wang et al. (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. (1998).

[6] See, e.g., Mullahy (1986), Rose et al. (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 (1997) on derogatory credit reports and Terza (1995, 1998) on sample selection in the Poisson model.

2.1 The poisson regression model

The canonical regression specification for a variable Y that is a count of events is the Poisson regression,

$$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\Gamma(1+y_i)}, \quad \lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta}),$$
$$y_i = 0, 1, \ldots, \quad i = 1, \ldots, N, \tag{1}$$

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).] The Poisson model has the convenient feature that

$$E[y_i|\mathbf{x}_i] = \lambda_i. \tag{2}$$

It has the undesirable characteristic that

$$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i. \tag{3}$$

This is the 'equidispersion' aspect of the model. Since observed data will almost always display pronounced *over* dispersion, analysis 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,[9]

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

The likelihood equations take the characteristically simple form[10]

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

[9] 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.

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

The partial effects in the Poisson model are

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

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

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

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, $(1, \bar{\mathbf{x}})$ would be

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

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) and (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

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

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

$$f(h_i) = \frac{\theta^\theta \exp(-\theta h_i) h_i^{\theta-1}}{\Gamma(\theta)}, \quad h_i \geq 0, \ \ \theta > 0.$$ [11]

[11] This general approach is discussed at length by Gourieroux et al. (1984), CT 1986, 1998, Winkelmann (2003) and HHG 1984.

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

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

The conditional Poisson regression model is

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

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

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

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

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

and the unconditional density is

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

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

$$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\Gamma(\theta + y_i)r_i^{\theta}(1 - r_i)^{y_i}}{\Gamma(1 + y_i)\Gamma(\theta)}, \quad y_i = 0, 1, \ldots, \quad \theta > 0, \tag{5}$$

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:

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

(the same as in the Poisson model) and

$$\begin{aligned}\text{Var}[y_i|\mathbf{x}_i] &= \lambda_i[1 + (1/\theta)\lambda_i] \\ &= \lambda_i[1 + \kappa\lambda_i]\end{aligned} \tag{6}$$

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 (2007c). Inference proceeds along familiar lines.[12] Inference about the specification, specifically the presence of overdispersion, is the subject of a lengthy literature, as documented, e.g., in Cameron and Trivedi (1990, 1998, 2005) and Hilbe (2007).

2.2.2 *Poisson lognormal mixture model*

Consider, instead, introducing the heterogeneity in (4) 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

$$P(y_i|\mathbf{x}_i, \varepsilon_i) = \frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(1 + y_i)}, \quad h_i\lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i), \quad \varepsilon_i \sim \text{N}[0, 1]. \tag{7}$$

The unconditional density would be

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

[12] 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).

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

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

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

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 $v_i = \varepsilon_i/\sqrt{2}$ and $\omega = \sigma\sqrt{2}$. After making the change of variable from ε_i to v_i and reparameterizing the probability, we obtain

$$P(y_i|\mathbf{x}_i) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v_i^2) P(y_i|\mathbf{x}_i, v_i) dv_i$$

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 $(\alpha, \boldsymbol{\beta}, \omega)$ are obtained by maximizing the reparameterized log likelihood. In this form, $\ln L$ can be approximated by Gauss-Hermite quadrature. (See Abramovitz and Stegun 1971.) The approximation is

$$\ln L_Q = \sum_{i=1}^{N} \ln \left[\frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} W_h P(y_i|\mathbf{x}_i, V_h) \right],$$

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,

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

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 (7), the log likelihood function is

$$\ln L = \sum_{i=1}^{N} \ln \int_{-\infty}^{\infty} P(y_i|\mathbf{x}_i, \varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i.$$

The simulated log likelihood would be

$$\ln L_S = \sum_{i=1}^{N} \ln \frac{1}{M} \sum_{m=1}^{M} P(y_i|\mathbf{x}_i, \sigma\varepsilon_{im})$$

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).[13] Derivatives of the simulated log likelihood for the ith observation are

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

The mean and variance of the lognormal variable are

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

and

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

The conditional mean in the Poisson lognormal model is

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

It follows that

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

[13] 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).

To obtain the unconditional variance, we use

$$\mathrm{Var}[y_i|\mathbf{x}_i] = E_{\varepsilon i}[\mathrm{Var}[y_i|\mathbf{x}_i, \varepsilon_i] + \mathrm{Var}_{\varepsilon i}[E[y_i|\mathbf{x}_i, \varepsilon_i]]. \tag{8}$$

Combining the results above, we find

$$\begin{aligned}\mathrm{Var}[y_i|\mathbf{x}_i] &= \lambda_i \exp(\sigma^2/2)\{1 + \lambda_i \exp(\sigma^2/2)[\exp(\sigma^2) - 1]\} \\ &= E[y_i|\mathbf{x}_i, \varepsilon_i]\{1 + \tau E[y_i|\mathbf{x}_i, \varepsilon_i]\}, \quad \tau = [\exp(\sigma^2) - 1].\end{aligned} \tag{9}$$

Thus, the variance in the lognormal model has the same quadratic form as that in the negative binomial model in (9). [As (8) suggests, since the conditional (on ε_i) mean equals the conditional variance, any multiplicative mixture with a constant variance will produce a counterpart to (9).]

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

Poisson model;

$$\begin{aligned}\mathbf{g_x} &= \exp(\sigma^2/2)\lambda_i \\ \mathbf{J} &= \frac{\partial \mathbf{g_x}}{\partial(\alpha, \boldsymbol{\beta}', \sigma)} = \mathbf{g_x}(1, \mathbf{x}_i', \sigma) + \exp(\sigma^2/2)\lambda_i \begin{bmatrix} \underline{\mathbf{0}} & \mathbf{I} & \underline{\mathbf{0}} \end{bmatrix}.\end{aligned}$$

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, Riphahn et al. (2003, p. 395), and Million (1998).]

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 in Terza (1995, 1998) and in Greene (1997, 2006, 2007a), 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 (1) (Poisson) or (5) (NB) with latent heterogeneity as in (4) and (7),

$$
\begin{aligned}
&d_i^* = \mathbf{w}_i'\boldsymbol{\delta} + u_i, u_i \quad \sim N[0, 1], && \text{(10a)}\\
&d_i = \mathbf{1}(d_i^* > 0),\\
&\text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) \quad \text{(generic, binary choice, zero)},\\
&\text{Prob}(d_i = 1|\mathbf{w}_i) = \boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta}) \quad \text{(probit selection equation)},\\
&y_i|\mathbf{x}_i, \varepsilon_i \sim P(y_i|\mathbf{x}_i, \varepsilon_i) \quad \text{(conditional on } \varepsilon_i \text{ Poisson or NB model)}, && \text{(10b)}\\
&E[y_i|\mathbf{x}_i, \varepsilon_i] = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i) = h_i\lambda_i, \quad \text{(conditional mean with heterogeneity)},\\
&\varepsilon_i \sim N[0, 1]\\
&[u_i, \varepsilon_i] \sim N_2[(0, 1), (1, 1), \rho] \quad \text{(selection effect)}, && \text{(10c)}\\
&y_i, \mathbf{x}_i \text{ are observed only when } d_i = 1. \quad \text{(observation mechanism)}.
\end{aligned}
$$

(We use the notation $\text{N}_2[(\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2), \rho]$ 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, $\hat{\psi}_i = \phi(\mathbf{w}_i'\hat{\boldsymbol{\delta}})/\boldsymbol{\Phi}(\mathbf{w}_i'\hat{\boldsymbol{\delta}})$, 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 (1995, 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.

The log likelihood function for the full model (see Terza (1995, 1998) and Greene (1997)) 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:

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

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

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

The first part, $P(y_i|\mathbf{x}_i, \varepsilon_i)$ is the conditional Poisson or NB density in (10). By joint normality,

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

so

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

Therefore, using (10a),

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

so

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

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

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

By exploiting the symmetry of the normal cdf

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

and

$$\text{Prob}(d_i = 0|\mathbf{x}_i, \mathbf{w}_i) = \int_{-\infty}^{\infty} \boldsymbol{\Phi}\left(-[\mathbf{w}_i'\boldsymbol{\delta} + \rho\varepsilon_i]/\sqrt{1-\rho^2}\right) \phi(\varepsilon_i)d\varepsilon_i. \quad (12)$$

Expressions (11) and (12) can be combined by using the symmetry of the normal cdf,

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

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

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

$$\ln L = \sum_{i=1}^{N} \ln \; P(y_i, d_i | \mathbf{x}_i, \mathbf{w}_i),$$

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

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

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

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

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 $(\alpha, \boldsymbol{\beta}, \omega, \boldsymbol{\eta}, \tau)$ are obtained by maximizing the reparameterized log likelihood.[14] The Gauss–Hermite approximation is

$$\ln \; L_Q = \sum_{i=1}^{N} \ln \left[\frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} W_h \left[(1 - d_i) + d_i P(y_i | \mathbf{x}_i, V_h) \right] \boldsymbol{\Phi} \left[(2d_i - 1) \left(\mathbf{w}_i' \boldsymbol{\eta} + \tau V_h \right) \right] \right], \tag{13}$$

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

$$\begin{aligned} P_{ih} &= P(y_i | \mathbf{x}_i, V_h), \\ \boldsymbol{\Phi}_{ih} &= \boldsymbol{\Phi}[(2d_i - 1)(\mathbf{w}_i' \boldsymbol{\eta} + \tau V_h)] \quad \text{(normal CDF)}, \\ \phi_{ih} &= \phi[(2d_i - 1)(\mathbf{w}_i' \boldsymbol{\eta} + \tau V_h)] \quad \text{(normal density)}, \end{aligned} \tag{14}$$

[14] The dispersion parameter, θ (or the heterogeneous version, θ_i) would appear in the parameter vector and in the derivatives in (16) and (18) if the (heterogeneous) NB model were used here instead of the Poisson.

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

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

$$h_h = \exp(\omega V_h), \tag{15}$$

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

Estimates of the structural parameters, $(\boldsymbol{\delta}, \rho, \sigma)$ 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 (13) is

$$\ln L_S = \sum_{i=1}^{N} \ln \frac{1}{M} \sum_{m=1}^{M} [(1 - d_i) + d_i P(y_i | \mathbf{x}_i, \sigma \varepsilon_{im})] \, \boldsymbol{\Phi} \left[(2d_i - 1) \left(\mathbf{w}_i' \boldsymbol{\eta} + \tau \varepsilon_{im} \right) \right] \tag{16}$$

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

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

$$h_{im} = \exp(\sigma \varepsilon_{im}), \tag{17}$$

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

where $\boldsymbol{\Phi}_{im}$, ϕ_{im} and P_{im} are defined as in (14) using ε_{im} in place of V_h.

The sample selection alters the conditional mean as follows: (See Terza 1995.) From (10b), the overall mean is

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

However,

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

This greatly complicates the partial effects;

$$\frac{\partial E[y_i|\mathbf{x}_i, \mathbf{w}_i, d_i = 1]}{\partial \mathbf{x}_i} = \lambda_i \exp\left[(\rho\sigma)^2/2\right] A_i \boldsymbol{\beta}$$
$$\frac{\partial E[y_i|\mathbf{x}_i, \mathbf{w}_i, d_i = 1]}{\partial \mathbf{w}_i} = \lambda_i \left(\frac{\exp((\rho\sigma)^2/2)}{\boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta})}\right) \left[\phi(\rho\sigma + w_i'\delta) - \phi(\mathbf{w}_i'\boldsymbol{\delta})A_i\right] \boldsymbol{\delta}.$$

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 A.

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.[15] The familiar structure of the model is

$$\begin{aligned}
&d_i^* = \mathbf{w}_i'\boldsymbol{\delta} + u_i, \\
&d_i = \mathbf{1}(d_i^* > 0), \\
&\text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) && \text{(regime selection equation)}, \\
&\text{Prob}(d_i = 1|\mathbf{w}_i) = 1 - \Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) && \text{(regime selection equation)}, \\
&y_i^*|\mathbf{x} \sim P(y_i^*|\mathbf{x}_i) && \text{(latent Poisson or NB model)}, \\
&E[y_i^*|\mathbf{x}_i] = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta}) = \lambda_i, && \text{(conditional mean)}, \\
&y_i = d_i y_i^* \text{ and } \mathbf{x}_i \text{ are observed} && \text{(observation mechanism)}.
\end{aligned}$$

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;

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

Combining terms, the joint density is

$$\begin{aligned}
P(y_i, d_i|\mathbf{x}_i, \mathbf{w}_i) &= P(y_i|\mathbf{x}_i, \mathbf{w}_i, d_i)P(d_i|\mathbf{x}_i, \mathbf{w}_i) \\
&= (1 - d_i)\Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) + d_i[1 - \Pi_0(\mathbf{w}_i'\boldsymbol{\delta})]P(y_i^*|\mathbf{x}_i).
\end{aligned}$$

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

The conditional mean function is

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

so the partial effects are

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

and

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

with

$$\begin{aligned}
\frac{\partial \mathbf{g_x}}{\partial(\alpha, \boldsymbol{\beta}')} &= \mathbf{g_x}(1, \mathbf{x}_i') + \lambda_i [1 - \Pi_0(\mathbf{w}_i' \boldsymbol{\delta})] \begin{bmatrix} \underline{\mathbf{0}} & \mathbf{I} \end{bmatrix} \\
\frac{\partial \mathbf{g_x}}{\partial \boldsymbol{\delta}'} &= -\lambda_i \frac{d\Pi_0(\mathbf{w}_i' \boldsymbol{\delta})}{d\left(\mathbf{w}_i' \boldsymbol{\delta}\right)} \boldsymbol{\beta} \mathbf{w}_i' \\
\frac{\partial \mathbf{g_w}}{\partial(\alpha, \boldsymbol{\beta}')} &= \mathbf{g_w}(1, \mathbf{x}_i') \\
\frac{\partial \mathbf{g_w}}{\partial \boldsymbol{\delta}'} &= -\lambda_i \left[\frac{d^2\Pi_0(\mathbf{w}_i' \boldsymbol{\delta})}{d\left(\mathbf{w}_i' \boldsymbol{\delta}\right)^2} \boldsymbol{\delta} \mathbf{w}_i' + \frac{d\Pi_0(\mathbf{w}_i' \boldsymbol{\delta})}{d\left(\mathbf{w}_i' \boldsymbol{\delta}\right)} \mathbf{I} \right]
\end{aligned}$$

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.[16] 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 Sect. 4. Figure 1 below gives a histogram for this variable.[17] The conspicuous spike at zero in this variable is decidedly nonPoisson.[18] 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

[16] Harris and Zhao (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.

[17] 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.

[18] Greene (2008, Sect. 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.

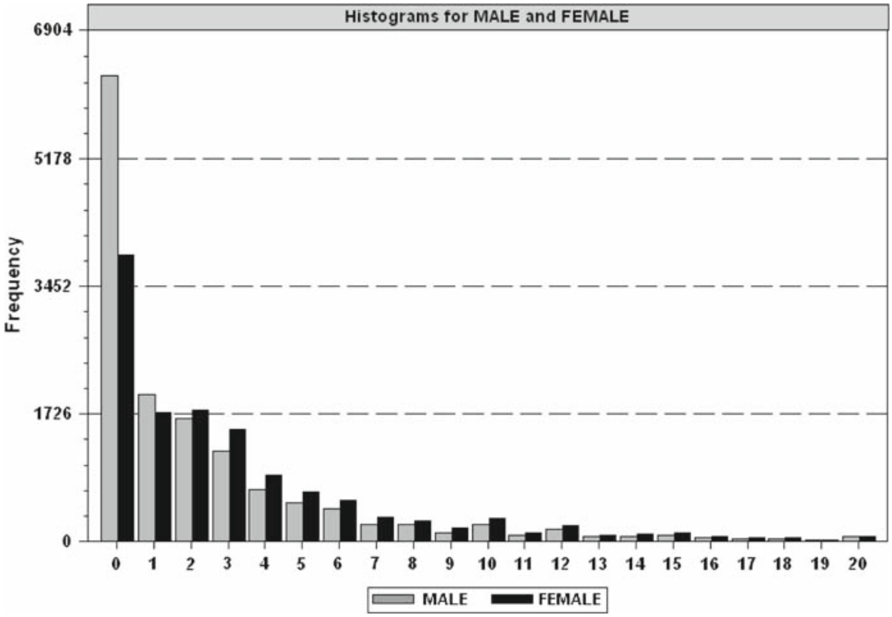

Fig. 1 Histograms for DocVis

(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

$$
\begin{aligned}
&d_i^* = \mathbf{w}_i'\boldsymbol{\delta} + u_i,\, u_i \sim N[0, 1], \\
&d_i = \mathbf{1}(d_i^* > 0), \\
&\text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) \\
&\text{Prob}(d_i = 1|\mathbf{w}_i) = \boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta}) && \text{(probit regime selection equation)}, \\
&y_i^*|\mathbf{x}_i, \varepsilon_i \sim P(y_i^*|\mathbf{x}_i, \varepsilon_i) && \text{(Poisson or NB model with heterogeneity)}, \\
&E[y_i^*|\mathbf{x}_i, \varepsilon_i] = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i) = h_i\lambda_i && \text{(heterogeneous conditional mean)}, \\
&[u_i, \varepsilon_i] \sim N_2[(0, 1), (1, 1), \rho], \\
&y_i = d_i y_i^* \text{ and } (\mathbf{x}_i, \mathbf{w}_i) \text{ are observed} && \text{(observation mechanism)}.
\end{aligned}
$$

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

$$
\begin{aligned}
P(y_i, d_i|\mathbf{x}_i, \mathbf{w}_i, \varepsilon_i) &= (1-d_i)\text{Prob}(d_i = 0|\mathbf{w}_i, \varepsilon_i) \\
&\quad +[1-\text{Prob}(d_i = 0|\mathbf{w}_i, \varepsilon_i)]P(y_i|\mathbf{x}_i, \varepsilon_i) \\
&= (1-d_i)\mathbf{\Phi}\left[\frac{-(\mathbf{w}_i'\boldsymbol{\delta} + \rho\varepsilon_i)}{\sqrt{1-\rho^2}}\right] \\
&\quad +\mathbf{\Phi}\left[\frac{(\mathbf{w}_i'\boldsymbol{\delta} + \rho\varepsilon_i)}{\sqrt{1-\rho^2}}\right]\frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(y_i+1)} \qquad (18)
\end{aligned}
$$

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 (18). 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,

$$
\begin{aligned}
\ln L &= \sum_{i=1}^{N} \ln \int_{-\infty}^{\infty} (1-d_i)\mathbf{\Phi}\left[-\frac{(\mathbf{w}_i'\boldsymbol{\delta} + \rho\varepsilon_i)}{\sqrt{1-\rho^2}}\right] + \mathbf{\Phi}\left[\frac{(\mathbf{w}_i'\boldsymbol{\delta} + \rho\varepsilon_i)}{\sqrt{1-\rho^2}}\right] \\
&\quad \times \frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(y_i+1)}\phi(\varepsilon_i)d\varepsilon_i \\
&= \sum_{i=1}^{N} \ln \int_{-\infty}^{\infty} \left\{ \begin{matrix} (1-d_i)\mathbf{\Phi}\left[-\mathbf{w}_i'\boldsymbol{\eta} - \tau\varepsilon_i\right] \\ +\mathbf{\Phi}\left[\mathbf{w}_i'\boldsymbol{\eta} + \tau\varepsilon_i\right] \frac{\exp[-\exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta}+\sigma\varepsilon_i)][\exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta}+\sigma\varepsilon_i)]^{y_i}}{\Gamma(y_i+1)} \end{matrix} \right\} \phi(\varepsilon_i)d\varepsilon_i.
\end{aligned}
$$

At this point, the specification differs only slightly from the formulation in the preceding section, in (13). Either quadrature or simulation can be used to maximize the likelihood function, with the corresponding adaptation of either (13) for the quadrature approach or (16) for the simulation based estimator.

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

$$
E[y_i|\mathbf{x}_i, \mathbf{w}_i, , \varepsilon_i] = \mathbf{\Phi}[A_i]\exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i)
$$

The observable counterpart is

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

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

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

For the variables in the regime equation,

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

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

$$\begin{aligned} E[y_i|\mathbf{x}_i, \mathbf{w}_i] &= \lambda_i \boldsymbol{\Phi}\left(\mathbf{w}_i'\boldsymbol{\delta}\right) \int_{-\infty}^{\infty} \exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i \\ &= \lambda_i \boldsymbol{\Phi}\left(\mathbf{w}_i'\boldsymbol{\delta}\right)\exp(\sigma^2/2) \end{aligned}$$

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

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; Xie et al. 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:

$$\begin{aligned} &d_i^* = \mathbf{w}_i'\boldsymbol{\delta} + u_i, \\ &d_i = \mathbf{1}(d_i^* > 0), \\ &\text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) \quad &&\text{(hurdle equation)}, \\ &\text{Prob}(d_i = 1|\mathbf{w}_i) = \boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta}) \quad &&\text{(probit hurdle model)}, \\ &y_i|\mathbf{x}_i, (d_i = 0) = \text{unobserved} \quad &&\text{(nonparticipation)}, \\ &y_i|\mathbf{x}_i, (d_i = 1) \sim P + (y_i \quad |\mathbf{x}_i) \quad &&\text{(truncated Poisson or NB model given participation)}. \end{aligned}$$

(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 $P+(y_i|\mathbf{x}_i)$. 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

$$\begin{aligned}
&\text{Prob}(y_i \text{ is unobserved } |\mathbf{x}_i, d_i = 0) = 1, \text{Prob}(d_i = 0|\mathbf{w}_i) = \Pi_0(\mathbf{w}_i'\boldsymbol{\delta})\\
&P(y_i|\mathbf{x}_i, d_i = 1) = P+(y_i|\mathbf{x}_i)\\
&\qquad\qquad = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\left[1-\exp(-\lambda_i)\right]\Gamma(1+y_i)}, \quad y_i = 1, 2, \ldots,\\
&\text{Prob}(d_i = 1|\mathbf{w}_i) = 1 - \Pi_0(\mathbf{w}_i'\boldsymbol{\delta}).
\end{aligned}$$

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

$$\mathrm{P}(y_i, d_i|\mathbf{x}_i, \mathbf{w}_i) = (1-d_i)\Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) + d_i \frac{\left[1-\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})\right]}{\left[1-\exp(-\lambda_i)\right]}\frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\Gamma(1+y_i)}, \quad y_i = 1, 2, \ldots$$

(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

$$\begin{aligned}
\ln L &= \sum_{d_i=0} \ln \Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) + \sum_{d_i=1} \ln[1 - \ln \Pi_0(\mathbf{w}_i'\boldsymbol{\delta})] - \ln[1-\exp(-\lambda_i)] + \ln P(y_i|\mathbf{x}_i)\\
&= \left\{\sum_{d_i=0} \ln \Pi_0(\mathbf{w}_i'\boldsymbol{\delta}) + \sum_{d_i=1} \ln[1 - \ln \Pi_0(\mathbf{w}_i'\boldsymbol{\delta})]\right\}\\
&\quad + \left\{\sum_{d_i=1} \ln P(y_i|\mathbf{x}_i) - \ln[1-\exp(-\lambda_i)]\right\}.
\end{aligned}$$

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

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

Therefore,

$$E[y_i|\mathbf{x}_i,\mathbf{w}_i] = \sum_{d_i=0}^{1}\sum_{y_i=1}^{\infty} y_i P(y_i,d_i|\mathbf{x}_i,\mathbf{w}_i) = \frac{\left[1-\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})\right]}{\left[1-\exp(-\lambda_i)\right]}\lambda_i.$$

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

$$\begin{aligned}
\partial E[y_i|\mathbf{x}_i,\mathbf{w}_i,d_i]/\partial\mathbf{x}_i &= \frac{\left[1-\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})\right]}{\left[1-\exp(-\lambda_i)\right]}\left(1-\frac{\lambda_i\exp(-\lambda_i)}{1-\exp(-\lambda_i)}\right)\lambda_i\boldsymbol{\beta} = \mathbf{g}_\mathbf{x},\\
\partial E[y_i|\mathbf{x}_i,\mathbf{w}_i,d_i]/\partial\mathbf{w}_i &= \frac{-d\Pi_0(\mathbf{w}_i'\boldsymbol{\delta})/d(\mathbf{w}_i'\boldsymbol{\delta})}{\left[1-\exp(-\lambda_i)\right]}\lambda_i\boldsymbol{\delta} = \mathbf{g}_\mathbf{w}.
\end{aligned}$$

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.[19] The extended model is

$$\begin{aligned}
&d_i^* = \mathbf{w}_i'\boldsymbol{\delta} + u_i,\ u_i \sim \mathrm{N}[0,1],\\
&d_i = \mathbf{1}(d_i^* > 0),\\
&\mathrm{Prob}(d_i = 0|\mathbf{w}_i) = 1-\boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta}) \qquad \text{(probit hurdle equation)},\\
&y_i|\mathbf{x}_i,\varepsilon_i,(d_i=0) = \text{unobserved} \qquad \text{(nonparticipation)},\\
&y_i|\mathbf{x}_i,\varepsilon_i,(d_i=1) \sim P+(y_i|\mathbf{x}_i,\varepsilon_i) \ \text{(truncated Poisson or NB model)},\\
&[u_i,\varepsilon_i] \sim N[(0,1),(1,1),\rho].
\end{aligned}$$

Using the same devices as in the earlier derivations, we have,

$$\begin{aligned}
\mathrm{P}(y,d_{ii}|\mathbf{x}_i,\mathbf{w}_i,\varepsilon_i) &= (1-d_i)\boldsymbol{\Phi}\left[-\mathbf{w}_i'\boldsymbol{\eta}-\tau\varepsilon_i\right]\\
&\quad + d_i\frac{\boldsymbol{\Phi}\left[\mathbf{w}_i'\boldsymbol{\eta}+\tau\varepsilon_i\right]}{\left[1-\exp(-\exp(\sigma\varepsilon_i)\lambda_i)\right]}\frac{\exp(-\exp(\sigma\varepsilon_i)\lambda_i)\left[\exp(\sigma\varepsilon_i)\lambda_i\right]^{y_i}}{\Gamma(1+y_i)}
\end{aligned}$$

and, finally,

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

[19] This model appears in Winkelmann (2004, pp. 462–463). Based on Smith and Moffatt's (1999) results, he suggests some skepticism about identification of ρ, though, as he notes, formally, the parameter is identified. We found no serious difficulties estimating the parameter in the application below, though the estimate was statistically insignificant.

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

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

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 (14) or (17). The partial effects and derivatives for this model are extremely cumbersome. They are presented in Appendix E.

4 Applications

In "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation", Riphahn et al. (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.[20] The authors of this study presented estimates for the Poisson-lognormal model in Sect. 2.2.2 and a bivariate Poisson model. We will analyze the single equation and two part models. (See Greene 2007b for results on the bivariate Poisson model.) 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.

4.1 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, 1,000, 887) with a total number of observations of 27,326. In order to illustrate the two part models, we have pooled the observations. RWM fit separate models for males and females. We have done likewise. 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, p. 393.) The outcome variables of interest in the study were doctor visits in the last three months and number of hospital

[20] 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.

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	3.1709
YEAR1984[a]	Dummy variable for 1984 observation	0.1418	0.3488
YEAR1985[a]	Dummy variable for 1985 observation	0.1388	0.3458
YEAR1986[a]	Dummy variable for 1986 observation	0.1388	0.3457
YEAR1987[a]	Dummy variable for 1987 observation	0.1342	0.3408
YEAR1988[a]	Dummy variable for 1988 observation	0.1641	0.3703
YEAR1991[a]	Dummy variable for 1991 observation	0.1588	0.3655
YEAR1994[a]	Dummy variable for 1994 observation	0.1236	0.3291
AGE	Age in years	43.5257	11.3302
AGESQ[a]	Age squared/1000	2.0229	1.0041
FEMALE	Female = 1; male = 0	0.4788	0.4996
MARRIED	Married = 1; else = 0	0.7586	0.4279
HHKIDS	Children under age 16 in the household = 1; else = 0	0.4027	0.4905
HHNINC[b]	Household nominal monthly net income, German marks/10000	0.3521	0.1769
EDUC	Years of schooling	11.3206	2.3249
WORKING	Employed = 1; else = 0	0.6770	0.4676
BLUEC	Blue collar employee = 1; else = 0	0.2438	0.4294
WHITEC	White collar employee = 1; else = 0	0.2996	0.4581
SELF	Self employed = 1; else = 0	0.0622	0.2415
CIVIL	Civil servant = 1; else = 0	0.0747	0.2629
HAUPTS	Highest schooling degree is Hauptschul = 1; else = 0	0.6243	0.4843
REALS	Highest schooling degree is Realschul = 1; else = 0	0.1968	0.3976
FACHHS	Highest schooling degree is Polytechnical= 1; else = 0	0.0408	0.1979
ABITUR	Highest schooling degree is Abitur = 1; else = 0	0.1170	0.3215
UNIV	Highest schooling degree is university = 1; else = 0	0.0719	0.2584
HSAT	Health satisfaction, 0–10	6.7854	2.2937
NEWHSAT[a,c]	Health satisfaction, 0–10	6.7857	2.2937
HANDDUM[d]	Handicapped = 1; else = 0	0.2140	0.4100
HANDPER	Degree of handicap in pct, 0–100	7.0123	19.2646
DOCVIS	Number of doctor visits in last three months	3.1835	5.6897
DOCTOR[a]	1 if DOCVIS > 0, 0 else	0.6291	0.4831
HOSPVIS	Number of hospital visits in last calendar year	0.1383	0.8843
HOSPITAL[a]	1 of HOSPVIS > 0, 0 else	0.0876	0.2828
PUBLIC	Iinsured in public health insurance = 1; else = 0	0.8857	0.3182
ADDON	Iinsured by add-on insurance = 1; else = 0	0.01881	0.1359

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

[a] Transformed variable not in raw data file

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

[c] *NEWHSAT* = *HSAT*; 40 observations on HSAT recorded between 6 and 7 were changed to 7

[d] 18 observations on dummy variable HANDDUM recorded between 0.15 and 0.17 were changed to 1.0

Table 2 Descriptive statistics by gender

Variable	Males		Females	
	Mean	SD	Mean	SD
YEAR	1987.84	3.1900	1987.80	3.14985
YEAR1984	0.1416	0.3487	0.1419	0.3490
YEAR1985	0.1389	0.3458	0.1388	0.3458
YEAR1986	0.1382	0.3451	0.1394	0.3464
YEAR1987	0.1342	0.3408	0.1341	0.3408
YEAR1988	0.1624	0.3688	0.1659	0.3720
YEAR1991	0.1576	0.3643	0.1602	0.3668
YEAR1994	0.1272	0.3332	0.1196	0.3245
AGE	42.6526	11.2704	44.4760	11.3192
AGESQ	1.9463	0.9874	2.1062	1.0154
FEMALE	0.0000	0.0000	1.0000	0.0000
MARRIED	0.7651	0.4239	0.7515	0.4322
HHKIDS	0.4130	0.4924	0.3916	0.4881
HHNINC	0.3591	0.1736	0.3445	0.1802
EDUC	11.7287	2.4365	10.8764	2.1091
WORKING	0.8503	0.3568	0.4884	0.4999
BLUEC	0.3402	0.4738	0.1387	0.3457
WHITEC	0.2999	0.4582	0.2992	0.4579
SELF	0.0857	0.2799	0.0366	0.1878
CIVIL	0.1178	0.3224	0.0278	0.1642
HAUPTS	0.6011	0.4897	0.6495	0.4772
REALS	0.1761	0.3809	0.2194	0.4138
FACHHS	0.0536	0.2253	0.0269	0.1618
ABITUR	0.1470	0.3541	0.0845	0.2781
UNIV	0.0962	0.2949	0.0456	0.2085
HSAT	6.9244	2.2515	6.6342	2.3295
NEWHSAT	6.9246	2.2515	6.6344	2.3295
HANDDUM	0.2273	0.4190	0.1996	0.3995
HANDPER	8.1337	20.3288	5.7914	17.9562
DOCVIS	2.6257	5.2112	3.7908	6.1111
DOCTOR	0.5595	0.4965	0.7049	0.4561
HOSPVIS	0.1278	0.9302	0.1497	0.8314
HOSPITAL	0.0779	0.2681	0.0982	0.2976
PUBLIC	0.8611	0.3459	0.9126	0.2825
ADDON	0.9176	0.1313	0.0202	0.1406
Sample size	14,243		13,083	

visits last year. As noted, we have analyzed only the first of these. A histograms for DocVis for the full data set is shown in Fig. 1. (The figure was truncated at 20 visits. This removes 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,

$$\mathbf{t} = (\textit{YEAR}\ 1985, \textit{YEAR}\ 1986, \textit{YEAR}\ 1987, \textit{YEAR}\ 1988, \textit{YEAR}\ 1991, \textit{YEAR}\ 1994).$$

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 Sect. 4.3.

4.2 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

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

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 females, 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 versus 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

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)	0.03706(1.20)	−0.03137(−0.93)	−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)	0.09340(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)

Table 3 continued

Variable	Pooled Poisson	Males			Females		
		Poisson	Log gamma	Lognormal	Poisson	Log gamma	Lognormal
θ			0.5707 (59.96)			0.8289 (64.44)	
κ			1.7522 (59.96)			1.2064 (64.44)	
$\sigma(\varepsilon)$			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		

Estimated coefficients for year dummy variables, excluding year 1984, are not reported. θ = the estimated parameter for the log gamma (NB) model, $\kappa = 1/\theta =$ 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. σ(h) = $\sqrt{\kappa}$ for the log gamma model, $\sqrt{\exp(\sigma^2)[\exp(\sigma^2) - 1]}$ for the lognormal model 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 4. 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 versus 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 (1989) statistic suggested in (2.4–6) equals 2.329 in favor of the lognormal model.

4.3 Two part models

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

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

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

Table 4 Estimated sample selection model, males

Variable	$\rho = 0$, 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		−0.06454 (−5.67)		−0.06691 (−5.73)
AGESQ		0.8689 (6.62)		0.8974 (6.65)
HSAT		−0.2548 (−44.51)		−0.2565 (−45.15)
HANDDUM	0.07328 (1.06)	0.006177 (0.12)	0.07362 (1.06)	0.008152 (0.16)
HANDPER	−0.0000 (−0.001)	0.007824 (9.28)	0.0000 (−0.001)	0.007792 (9.24)
MARRIED	−0.01347 (−0.31)	0.06478 (1.80)	−0.01353 (−0.33)	0.06610 (1.83)
EDUC	−0.1808 (−7.39)	−0.01501 (1.91)	−0.1813 (−7.09)	−0.01681 (−1.58)
HHNINC	−10.024 (−11.94)	−0.08076 (−0.85)	−1.025 (−14.41)	−0.08917 (−0.85)
HHKIDS	0.06995 (1.85)	−0.02862 (−0.89)	0.07007 (1.88)	−0.03045 (−0.95)
SELF		−0.3116 (−5.54)		−0.3264 (−5.11)
CIVIL		0.02953 (0.38)		0.01633 (0.19)
BLUEC		0.006345 (0.18)		−0.005491 (−0.13)
WORKING		0.03976 (0.90)		0.05594 (1.15)
YEAR1985	0.04480 (0.75)	0.07375 (1.53)	0.04418 (0.73)	0.07153 (1.49)
YEAR1986	−0.01027 (−0.17)	0.1704 (3.52)	−0.01094 (−0.18)	0.1694 (3.51)
YEAR1987	−0.08772 (−1.14)	0.07857 (1.38)	−0.08815 (−1.14)	0.07343 (1.30)
YEAR1988	−0.06434 (−1.15)	0.09794 (2.02)	−0.06476 (−1.15)	0.09290 (1.93)
YEAR1991	−0.03130 (−0.55)	0.1077 (2.06)	−0.03161 (−0.56)	0.1015 (1.95)
YEAR1994	0.05901 (0.98)	0.3428 (6.52)	0.05850 (0.96)	0.3397 (6.48)
AGE3034	−0.2211 (−4.06)		−0.2208 (−4.00)	
AGE3539	−0.3089 (−6.02)		−0.3089 (−6.0)	
AGE4044	0.5164 (9.96)		0.5160 (10.08)	
AGE4549	−0.1473 (−2.73)		−0.1468 (−2.78)	
AGE5054	0.2120 (3.55)		0.2113 (3.43)	
AGE5559	0.4462 (6.53)		0.4463 (6.23)	
AGE60UP	0.5531 (7.25)		0.5532 (7.04)	
HAUPTS	0.3631 (3.36)		0.3635 (3.51)	
REALS	−0.3657 (−3.21)		−0.3646 (−3.24)	
FACHHS	0.1456 (0.95)		0.1476 (0.94)	
ABITUR	0.06202 (0.40)		0.06490 (0.40)	
UNIV	0.03097 (0.31)		0.03274 (0.31)	
WHITEC	1.1305 (27.83)		1.1307 (30.23)	
σ		1.2377 (99.92)		1.2394 (97.08)
ρ	0.0000 (fixed)		0.02246 (0.23)	
ln L	−4294.89	−24044.20	−28339.01	
n	14243	12264	14243	12264

model.[21] 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 4 and 5 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.

Figure 1 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 6 and 7 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 6, 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 2 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 7 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 7, this correlation is assumed to be zero. In the second model, the correlation is unrestricted; the estimated value is 0.154. However,

[21] 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.

Table 5 Estimated sample selection model, females (t ratios in parentheses)

Variable	$\rho = 0$, 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		−0.04117 (−4.37)		−0.03866 (−4.07)
AGESQ		0.5218 (4.93)		0.4857 (4.56)
HSAT		−0.2215 (−45.51)		−0.2197 (−43.31)
HANDDUM	0.1104 (1.19)	0.1001 (2.48)	0.07195 (0.85)	0.1205 (2.43)
HANDPER	0.003088 (1.97)	0.004725 (6.64)	0.0032745 (2.32)	0.004808 (5.93)
MARRIED	0.003124 (0.95)	0.05571 (1.91)	−0.0027795 (−0.06)	0.03836 (1.28)
EDUC	−0.1678 (−7.38)	0.02748 (3.91)	−0.1593 (−7.02)	0.06077 (7.47)
HHNINC	−1.183 (−12.9)	−0.1607 (−2.08)	−1.1620(−15.00)	0.08070 (0.97)
HHKIDS	0.08852 (1.95)	−0.1306 (−4.44)	0.08804 (1.89)	−0.1255 (−4.23)
SELF		−0.3444 (−4.78)		−0.2340 (−3.14)
CIVIL		0.2667 (1.58)		0.4428 (2.80)
BLUEC		−0.09816 (−2.73)		0.03352 (0.90)
WORKING		0.04561 (1.61)		−0.05897 (−1.88)
YEAR1985	−0.00536 (−0.07)	−0.0279 (2.19)	0.007422 (0.11)	−0.03663 (−0.86)
YEAR1986	−0.00267 (−0.04)	0.09287 (0.028)	−0.005941 (−0.09)	0.1289 (3.08)
YEAR1987	−0.1215 (−1.21)	−0.03763 (−0.79)	−0.08333 (−0.89)	−0.08684 (−1.54)
YEAR1988	−0.0909 (−1.42)	−0.1467 (−3.46)	−0.09197 (−1.43)	−0.1391 (−3.18)
YEAR1991	0.03407 (0.51)	−0.06931 (−1.54)	0.01679 (0.26)	−0.06926 (−1.51)
YEAR1994	0.1665 (2.28)	0.2642 (5.86)	0.1555 (2.16)	0.2202 (4.76)
AGE3034	−0.1565 (−2.27)		−0.1604 (−2.35)	
AGE3539	−0.1581 (−2.49)		−0.1642 (−2.57)	
AGE4044	0.1901 (3.00)		0.1809 (2.81)	
AGE4549	−0.1017 (−1.56)		−0.08129 (−1.24)	
AGE5054	0.06098 (0.89)		0.07611 (1.14)	
AGE5559	0.1323 (1.83)		0.1231 (1.72)	
AGE60UP	0.1221 (1.68)		0.1095 (1.47)	
HAUPTS	0.4737 (3.89)		0.3792 (3.21)	
REALS	0.2016 (1.56)		0.1203 (0.98)	
FACHHS	0.4686 (2.66)		0.3346 (1.97)	
ABITUR	0.4092 (2.38)		0.2705 (1.64)	
UNIV	−0.1886 (1.70)		−0.2206 (2.00)	
WHITEC	0.9482 (18.99)		0.9087 (18.98)	
σ		1.0560 (109.96)		1.0993 (84.92)
ρ	0.0000 (fixed)		−0.5339 (−8.87)	
ln L	−3099.5	−27833.5	−30927.7	
n	13083	11939	13083	11939

Table 6 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)	−0.01432 (−0.07)	2.5722 (57.10)
AGE		−0.02387 (−5.44)		−0.005731 (−3.21)
AGESQ		0.3693 (7.45)		0.1285 (6.34)
HSAT		−0.2253 (−104.1)		−0.1564 (−186.94)
HANDDUM		0.06899 (4.09)	0.1274 (3.58)	0.07641 (12.53)
HANDPER		0.00286 (10.04)	−0.01374 (−15.65)	0.00081 (8.34)
MARRIED		0.05831 (3.89)	−0.1228 (−3.79)	−0.01578 (−2.71)
EDUC		−0.02348 (−8.43)	−0.01112 (−0.65)	−0.01896 (−14.04)
HHNINC		−0.2220 (−5.93)	−0.06780 (−0.92)	−0.2240 (−14.07)
HHKIDS		−0.07598 (−5.75)	0.07732 (2.67)	−0.03190 (−6.19)
SELF		−0.2110 (−8.98)		−0.1084 (−10.49)
CIVIL		0.09144 (3.78)		0.1111 (10.16)
BLUEC		0.01779 (1.24)		0.04336 (7.21)
WORKING		−0.05539 (−3.17)		−0.05992 (−9.02)
PUBLIC		0.1001 (4.27)		0.07612 (7.07)
ADDON		0.06655 (1.63)		−0.07040 (−3.67)
YEAR1985		2.7719 (28.85)		0.09084 (11.58)
YEAR1986		−0.02387 (−5.44)		0.1840 (23.93)
YEAR1987		0.3693 (7.45)		0.1150 (14.64)
YEAR1988		−0.2253 (−104.1)		0.00065 (0.08)
YEAR1991		0.06899 (4.09)		−0.1058 (−12.99)
YEAR1994		0.00286 (10.04)		0.1810 (22.07)
AGE3034			0.02866 (0.69)	
AGE3539			0.06424 (1.58)	
AGE4044			−0.1434 (−3.45)	
AGE4549			0.1521 (3.47)	
AGE5054			−0.1562 (−3.47)	
AGE5559			−0.1864 (−3.93)	
AGE60UP			−0.3261 (−5.89)	
HAUPTS			0.06593 (0.79)	
REALS			0.07693 (0.85)	
FACHHS			0.1143 (0.95)	
ABITUR			0.2539 (2.10)	
UNIV			0.01076 (0.13)	
WHITEC			−0.005596 (−0.21)	
σ				0.0000 (fixed)
ρ			(0.0000) (fixed)	
ln L	−42774.7		−35757.0	
n	14243		14243	
Voung stat.	0.00		28.83	

Table 7 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	−0.3218 (−0.95)	2.4564 (8.83)	−0.3015 (−0.93)	2.5220 (8.74)
AGE		−0.02405 (−2.02)		−0.02436 (−1.96)
AGESQ		0.3650 (2.67)		0.36660 (2.56)
HSAT		−0.2310 (−41.48)		−0.2312 (−39.85)
HANDDUM	0.3784 (5.02)	0.03262 (0.72)	0.3933 (5.12)	0.03832 (0.82)
HANDPER	−0.02667 (−6.36)	0.002292 (2.98)	−0.02830 (−6.36)	0.001453 (1.43)
MARRIED	−0.2088 (−3.29)	0.005705 (0.14)	−0.1916 (−3.03)	0.007373 (0.17)
EDUC	−0.02912 (−0.96)	−0.01201 (−1.52)	−0.02994 (−1.03)	−0.01053 (−1.26)
HHNINC	−0.2411 (−1.51)	−0.2310 (−2.39)	−0.2410 (−1.53)	−0.2498 (−2.46)
HHKIDS	0.1193 (2.05)	−0.02546 (−0.72)	0.1105 (1.92)	−0.02517 (−0.65)
SELF		−0.2350 (−4.22)		−0.2474 (−4.33)
CIVIL		0.07712 (1.23)		0.06009 (0.94)
BLUEC		0.04352 (1.13)		0.04617 (1.15)
WORKING		−0.05616 (−1.21)		−0.04573 (−0.94)
PUBLIC		0.06828 (1.21)		0.04932 (0.88)
ADDON		0.05530 (0.52)		0.05892 (0.56)
YEAR1985		0.1125 (2.41)		0.1237 (2.67)
YEAR1986		0.2171 (4.67)		0.2198 (4.73)
YEAR1987		0.2142 (4.06)		0.2299 (4.26)
YEAR1988		0.05805 (1.26)		0.06311 (1.37)
YEAR1991		0.02592 (0.53)		0.02908 (0.59)
YEAR1994		0.3226 (6.43)		0.3189 (6.21)
AGE3034	0.05143 (0.71)		0.05491 (0.80)	
AGE3539	0.1164 (1.70)		0.1168 (1.76)	
AGE4044	−0.2525 (−3.57)		−0.2409 (−3.43)	
AGE4549	0.1770 (2.47)		0.1735 (2.51)	
AGE5054	−0.2060 (−2.49)		−0.1892 (−2.41)	
AGE5559	−0.1555 (−1.79)		−0.1464 (−1.75)	
AGE60UP	−0.4165 (−3.29)		−0.3874 (−3.19)	
HAUPTS	0.1363 (0.83)		0.1045 (0.68)	
REALS	0.1287 (0.73)		0.1034 (0.63)	
FACHHS	0.2027 (0.91)		0.1796 (0.86)	
ABITUR	0.4061 (1.84)		0.3680 (1.76)	
UNIV	0.07217 (0.51)		0.09214 (0.69)	
WHITEC	−0.03377 (−0.63)		−0.02770 (−0.52)	
σ		0.9875 (70.08)		0.9902 (66.31)
ρ	0.0000 (fixed)		0.1540 (0.14)	
ln$L$	−27183.9		−27183.1	
n	14243		14243	
Vuong stat.	24.16		24.17	

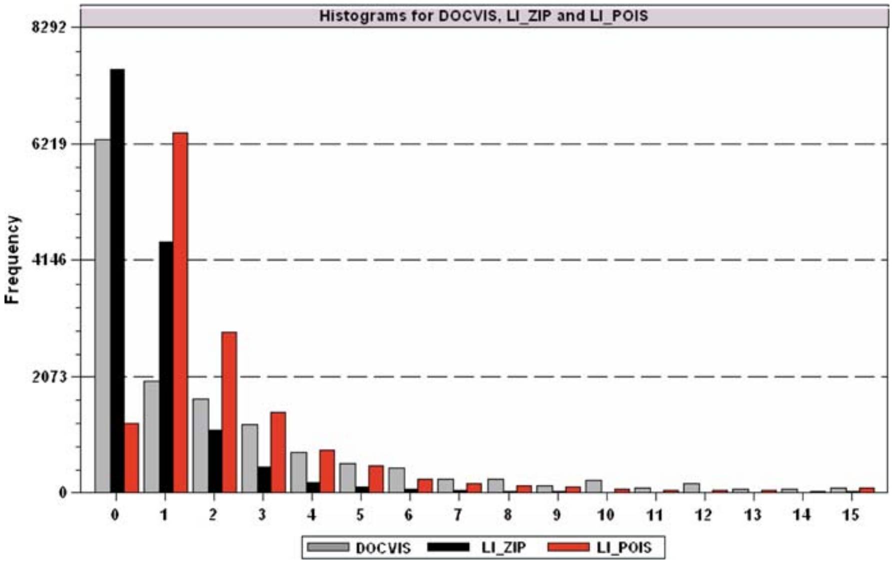

Fig. 2 Predictions from ZIP and Poisson models and actual DocVis

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. 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.

5 Conclusions

This study has examined several two part extensions to some familiar models for count data that allow for endogeneity of the participation decision in the first equation; We have also applied the techniques in an analysis of a large sample of German households.

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), Riphahn et al. (2003), Greene

(2008)a, have suggested that the normal distribution would be a preferable platform on which to build the model.[22] 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 Terza (1995, 1998), Greene (1997), Munkin and Trivedi (1999) and Riphahn et al. (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.[23]

Finally, the methods developed here were applied to the data set used in Riphahn et al. (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: Derivatives of partial effects in the poisson model with sample selection

The conditional mean function is

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

The partial effects are

$$\frac{\partial E[y_i|\mathbf{x}_i, \mathbf{w}_i]}{\partial \mathbf{x}_i} = \lambda_i \left[\frac{\exp((\rho\sigma)^2/2)\boldsymbol{\Phi}(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})}{\boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta})}\right]\boldsymbol{\beta} = \mathbf{g_x}$$

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

For the variables in the count model,

$$\frac{\partial \mathbf{g_x}}{\partial(\alpha, \boldsymbol{\beta}')} = \lambda_i \left[\frac{\exp((\rho\sigma)^2/2)\boldsymbol{\Phi}(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})}{\boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta})}\right]\left[\underline{\mathbf{0}}\ \mathbf{I}\right] + \mathbf{g_x}(1, \mathbf{x}_i')$$

$$\frac{\partial \mathbf{g_x}}{\partial \rho} = \mathbf{g_x}\left[\rho\sigma^2 + \sigma\frac{\phi(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})}{\boldsymbol{\Phi}(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})}\right]$$

[22] Winkelmann (2003) reports that the AIC for the lognormal model is higher than that for the log gamma model in some applications.

[23] We concede the complexity of some of the analytic results in the appendices that might invoke some attraction to the simpler Poisson model as a robust pseudo MLE. On the other hand, once results such as these are incorporated into widely used software such as Stata, SAS and LIMDEP, the complexity becomes a moot point for th practitioner.

$$\frac{\partial \mathbf{g_x}}{\partial \sigma} = \frac{\partial \mathbf{g_x}}{\partial \rho}\frac{\rho}{\sigma}$$
$$\frac{\partial \mathbf{g_x}}{\partial \boldsymbol{\delta}'} = \lambda_i \exp(\rho\sigma)^2 \left[\frac{\phi(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})}{\boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta})} - \frac{\boldsymbol{\Phi}(\rho\sigma + \mathbf{w}_i'\boldsymbol{\delta})\phi(\mathbf{w}_i'\boldsymbol{\delta})}{\left[\boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta})\right]^2}\right]\boldsymbol{\beta}\mathbf{w}'.$$

For the variables in the selection equation,

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

Appendix B: Derivatives of partial effects of zip model with endogenous zero inflation

Let $\mathrm{A}(\varepsilon_i) = \frac{(\mathbf{w}_i'\boldsymbol{\delta} + \rho\varepsilon_i)}{\sqrt{1-\rho^2}}$ and $\Delta = \frac{1}{\sqrt{1-\rho^2}}$.

The partial effects are

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

The derivatives are

$$\frac{\partial \mathbf{g_x}}{\partial(\alpha, \boldsymbol{\beta}')} = \mathbf{g_x}(1, \mathbf{x}_i') + \lambda_i \left[\underline{\mathbf{0}}\ \mathbf{I}\right] \int_{-\infty}^{\infty} \boldsymbol{\Phi}\left[A(\varepsilon_i)\right] \exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i$$

$$\frac{\partial \mathbf{g_x}}{\partial \sigma} = \lambda_i \boldsymbol{\beta} \int_{-\infty}^{\infty} \boldsymbol{\Phi}\left[A(\varepsilon_i)\right] \varepsilon_i \exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i$$

$$\frac{\partial \mathbf{g_x}}{\partial \boldsymbol{\delta}'} = \lambda_i \boldsymbol{\beta} \mathbf{w}_i' \Delta \int_{-\infty}^{\infty} \phi\left[A(\varepsilon_i)\right] \exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i$$

$$\frac{\partial \mathbf{g_x}}{\partial \rho} = \lambda_i \boldsymbol{\beta} \Delta \int_{-\infty}^{\infty} \phi\left[A(\varepsilon_i)\right] \exp(\sigma\varepsilon_i) \left[\varepsilon_i - \rho\Delta^2 A(\varepsilon_i)\right] \phi(\varepsilon_i)d\varepsilon_i$$

$$\frac{\partial \mathbf{g_w}}{\partial(\alpha, \boldsymbol{\beta}')} = \mathbf{g_w}(1, \mathbf{x}_i')$$

$$\frac{\partial \mathbf{g_w}}{\partial \sigma} = \lambda_i \Delta \left\{ \int_{-\infty}^{\infty} \phi\left[A(\varepsilon_i)\right] \varepsilon_i \exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i \right\}$$

$$\frac{\partial \mathbf{g_w}}{\partial \boldsymbol{\delta}'} = \lambda_i \boldsymbol{\delta} \mathbf{w}' \Delta_i^2 \left\{ \int_{-\infty}^{\infty} -A(\varepsilon_i)\phi\left[A(\varepsilon_i)\right] \exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i \right\}$$

$$\frac{\partial \mathbf{g_w}}{\partial \sigma} = \mathbf{g_w}\rho\Delta^2 + \lambda_i \boldsymbol{\delta} \Delta^2 \left\{ \int_{-\infty}^{\infty} -A(\varepsilon_i)\phi\left[A(\varepsilon_i)\right] \left[\varepsilon_i - \rho\Delta^2 A(\varepsilon_i)\right] \exp(\sigma\varepsilon_i)\phi(\varepsilon_i)d\varepsilon_i \right\}.$$

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

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

$$\frac{\partial \mathbf{g_x}}{\partial(\alpha, \boldsymbol{\beta}', \sigma)} = \mathbf{g_x}(1, \mathbf{x}_i', \sigma),$$

$$\frac{\partial \mathbf{g_x}}{\partial \boldsymbol{\delta}'} = \lambda_i \boldsymbol{\beta} \mathbf{w}_i'[\phi(\mathbf{w}_i'\boldsymbol{\delta}) \exp(\sigma^2/2)]$$

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

$$\frac{\partial \mathbf{g_w}}{\partial(\alpha, \boldsymbol{\beta}', \sigma)} = \mathbf{g_w}(1, \mathbf{x}_i', \sigma),$$

$$\frac{\partial \mathbf{g_w}}{\partial \boldsymbol{\delta}'} = -\mathbf{g_w}\mathbf{w}_i'(\mathbf{w}_i'\boldsymbol{\delta}) + \lambda_i \phi(\mathbf{w}_i'\boldsymbol{\delta}) \exp(\sigma^2/2)\mathbf{I}.$$

Appendix C: 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 $\Lambda(\mathbf{w}_i'\boldsymbol{\delta})$ for $\boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\delta})$, $\{\Lambda(\mathbf{w}_i'\boldsymbol{\delta})[1-\Lambda(\mathbf{w}_i'\boldsymbol{\delta})]\}$ for $\phi(\mathbf{w}_i'\boldsymbol{\delta})$ and $[1-2\Lambda(\mathbf{w}_i'\boldsymbol{\delta})]\{\Lambda(\mathbf{w}_i'\boldsymbol{\delta})[1-\Lambda(\mathbf{w}_i'\boldsymbol{\delta})]\}$ for—$(\mathbf{w}_i'\boldsymbol{\delta})\phi(\mathbf{w}_i'\boldsymbol{\delta})]$ in what follows. The conditional mean is

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

The partial effects are

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

The derivatives are cumbersome. We proceed as follows: Write

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

Then

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

Thus,

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

Then,

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

$$\frac{\partial \mathbf{g_w}}{\partial(\alpha, \boldsymbol{\beta}')} = \phi(\mathbf{w}_i'\boldsymbol{\delta})c'(\lambda_i)\lambda_i\boldsymbol{\delta}(1, \mathbf{x}_i'),$$
$$\frac{\partial \mathbf{g_w}}{\partial \boldsymbol{\delta}'} = -(\mathbf{w}_i'\boldsymbol{\delta})\phi(\mathbf{w}_i'\boldsymbol{\delta})c(\lambda_i)\lambda_i\boldsymbol{\delta}\mathbf{w}_i'.$$

Appendix D: Partial effects and derivatives of partial effects in hurdle models with endogenous participation

The conditional mean is

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

For convenience, let

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

so that

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

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

Then

$$a_i'(u_i) = \frac{da(u_i)}{du_i} = \frac{a_i}{u_i}(1 - a_i\exp(-u_i)),$$

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

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

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

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

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

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

The conditional mean function is

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

and the partial effects are

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

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

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

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

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

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

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

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

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

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

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

References

Abramovitz M, Stegun I (1971) Handbook of mathematical functions. Dover Press, New York

Asplund M, Sandin R (1999) The number of firms and production capacity in relation to market size. J Ind Econ 47:69–85

Butler J, Moffitt R (1982) A computationally efficient quadrature procedure for the one factor multinomial probit model. Econometrica 50:761–764

Cameron A, Trivedi P (1986) Econometric models based on count data: comparisons and applications of some estimators and tests. J Appl Econ 1:29–54

Cameron A, Trivedi P (1990) Regression based tests for overdispersion in the poisson model. J Econ 46:347–364

Cameron A, Trivedi P (1998) Regression analysis of count data. Cambridge University Press, New York

Cameron A, Trivedi P (2005) Microeconometrics: methods and applications. Cambridge University Press, New York

Contoyannis C, Jones A, Rice N (2004) The dynamics of health in the british household panel survey. J Appl Econ 19:473–503

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

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

Freedman D (2006) On the so-called "Huber sandwich estimator" and "robust standard errors". Am Stat 60(4):299–302

Gourieroux C, Monfort A (1996) Simulation-based econometric methods. Oxford University Press, Oxford

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

Greene W (1994) Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working paper EC-94-10, Department of Economics, Stern School of Business, New York University

Greene W (1997) FIML estimation of sample selection models for count data. Working paper EC-97-02, Department of Economics, Stern School of Business, New York University

Greene W (2006) A general approach to incorporating 'selectivity' in a nonlinear model. Working paper EC-06-10, Department of Economics, Stern School of Business, New York University

Greene W (2007a) Functional form and heterogeneity in models for count data. Found Trends Econ 1(2):113–218

Greene W (2007b) Correlation in the bivariate Poisson model. Working paper 07-15, Department of Economics, Stern School of Business, New York University

Greene W (2007c) LIMDEP 9.0 Reference guide. Econometric Software, Inc., Plainview, New York

Greene W (2008) Econometric analysis. 6. Prentice Hall, Englewood Cliffs

Hall D (2000) Zero inflated Poisson and negative binomial regression with random effects: a case study. Biometrics 56:1030–1039

Harris M, Zhao X (2007) A zero inflated ordered probit model, with an application to modeling tobacco consumption. J Econo 141:1044–1072

Hausman J, Hall B, Griliches Z (1984) Economic models for count data with an application to the patents—R&D relationship. Econometrica 52:909–938

Heckman J (1979) Sample selection bias as a specification error. Econometrica 47:153–161

Heilbron D (1994) Zero altered and other regression models for count data with added zeros. Biom J 36:531–547

Hilbe J (2007) Negative binomial regression. Cambridge University Press, Cambridge

Hur K (1998) A random effects zero inflated Poisson regression model for clustered extra—zero counts. Ph.D. Dissertation, Dartmouth Medical School

Krinsky I, Robb L (1986) On approximating the statistical properties of elasticities. Rev Econ Stat 68:715–719

Lambert D (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34:1–14

Million A (1998) Models for correlated count data. Unpublished dissertation, University of Munich

Mullahy J (1986) Specification and testing of some modified count data models. J Econ 33:341–365

Munkin M, Trivedi P (1999) Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. Econ J 2:29–49

Murphy K, Topel R (2002) Estimation and inference in two step econometric models. J Bus Econ Stat 20:88–97

Riphahn R, Wambach A, Million A (2003) Incentive effects in the demand for health care: a bivariate panel count data estimation. J Appl Econ 18(4):387–405

Rose C, Martin W, Wannemuehler K, Plikaytis B (2006) On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. J Biopharm Stat 16(4):463–481

Smith M, Moffatt P (1999) Fisher's information on the correlation coefficient in bivariate logistic models. Aust N Z J Stat 41:315–330

Stata (2006) Stata user's guide, version 9. Stata Press, College Station

Terza J (1985) A tobit type estimator for the censored Poisson regression model. Econ Lett 18:361–365

Terza J (1995) Estimating count data models with endogenous switching and sample selection. Working Paper IPRE-95-14, Department of Economics, Penn State University

Terza J (1998) Estimating count data models with endogenous switching: sample selection and endogenous treatment effects. J Econ 84(1):129–154

Train K (2003) Discrete choice models with simulation. Cambridge University Press, Cambridge

Vuong Q (1989) Likelihood ratio tests for model selection and non-nested models. Econometrica 57:307–334

Wang P, Cockburn I, Puterman L (1998) Analysis of patent data—a mixed Poisson regression model approach. J Bus Econ Stat 16(1):27–41

Winkelmann R (2003) Econometric analysis of count data. 4. Springer, Heidelberg

Winkelmann R (2004) Health reform and the number of doctor visits—an econometric analysis. J Appl Econ 19(4):455–472

Xie H, Hur K, McHugo G (2006) 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

Yen S, Adamowicz V (1994) Participation, trip frequency and site choice: a multinomial Poisson hurdle model of recreation demand. Can J Agric Econ 42(1):65–76

Zorn C (1998) An analytic and empirical examination of zero-inflated and hurdle Poisson specifications. Sociol Methods Res 26(3):368–400