Fixed and Random Effects Models for Count Data

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

Abstract

The most familiar fixed effects (FE) and random effects (RE) panel data treatments for count data were proposed by Hausman, Hall and Griliches (HHG) (1984). The Poisson FE model is particularly simple and is one of a small few known models in which the incidental parameters problem is, in fact, not a problem. The same is not true of the negative binomial (NB) model. Researchers are sometimes surprised to find that the HHG formulation of the FENB model allows an overall constant – a quirk that has also been documented elsewhere. We resolve the source of the ambiguity, and consider the difference between the HHG FENB model and a 'true' FENB model that appears in the familiar index function form.

The familiar RE Poisson model using a log gamma heterogeneity term produces the NB model. The HHG RE NB model is also unlike what might seem the natural application in which the heterogeneity term appears as an additive common effect in the conditional mean. We consider the lognormal model as an alternative RENB model in which the common effect appears in a natural index function form.

JEL classification: C14; C23; C25

Keywords: Poisson regression; Negative binomial; Panel data; Heterogeneity; Lognormal; Fixed effects; Random effects

---

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

1 Introduction

The most familiar panel data treatments, fixed effects (FE) and random effects (RE), were proposed for count data models by Hausman, Hall and Griliches (HHG) (1984). The Poisson FE model is particularly simple to analyze, and has long been recognized as one of a small handful of models in which the incidental parameters problem [see Neyman and Scott (1948) and Lancaster (2000)] is, in fact, not a problem. The same is not true of the negative binomial (NB) model. Researchers are sometimes surprised to find, moreover, that the HHG formulation of the FENB model allows an overall constant – a quirk that has been documented elsewhere [see Allison (2000) and Allison and Waterman (2002), for example]. This note resolves the source of the ambiguity, and considers the difference between the HHG FENB model and a 'true' FENB model that appears in the familiar index function form that is used in other familiar settings. The true FENB model has not been used by applied researchers, in part because of the absence of a computational method. We have developed a method of computing the true FENB model that allows a comparison to the HHG formulation.

The familiar RE Poisson model using an additive log gamma heterogeneity term in the conditional mean produces an uncomplicated NB model. The HHG RENB model, however, is also unlike what might seem the natural application in which the heterogeneity term appears as an additive common effect in the conditional mean. Theirs was a practical solution to the problem. Here, we consider the lognormal model as an alternative and compare it to the HHG formulation.The lognormal model provides a means of specifying the RE NB model in a natural index function form. We will develop this model, and, once again, compare it to the HHG formulation. The various models discussed and developed below are applied in a large, rich panel data set that allows a detailed comparison.

2 Basic Functional Forms for Count Data Models

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

(2-1) $$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\Gamma(1+y_i)}, \lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta}), y_i = 0,1,..., i = 1,..., N,$$

where $\mathbf{x}_i$ is a vector of covariates and, $i = 1,\dots,N$, indexes the N observations in a random sample. [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 negative binomial model is the standard extension that is used to circumvent the equidispersion property of the Poisson model, $\text{Var}[y_i|\mathbf{x}_i] = E[y_i|\mathbf{x}_i] = \lambda_i$. The model can also be motivated by introducing latent heterogeneity into the Poisson model.[1] We write

(2-2) $$E[y_i|\mathbf{x}_i,\varepsilon_i] = \exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i) = h_i\lambda_i,$$

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

(2-3) $$f(h_i) = \frac{\theta^\theta \exp(-\theta h_i)h_i^{\theta-1}}{\Gamma(\theta)}, h_i \geq 0, \theta > 0.$$

Integrating h_i out of the conditional density produces the negative binomial marginal distribution,

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

(2-4) $$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\Gamma(y_i+\theta)r_i^{\theta}(1-r_i)^{y_i}}{\Gamma(y_i+1)\Gamma(\theta)}, y_i = 0,1,...;\ \theta > 0;\ r_i = \theta/(\theta+\lambda_i).$$

Both the Poisson and NB random variables have conditional mean functions

(2-5) $$E[y_i|\mathbf{x}_i] = \lambda_i.$$

The model in (2-4) is the "NB2" form of the model, in reference to the conditional variance,

(2-6) $$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i[1 + (1/\theta)\lambda_i.] = \lambda_i + \kappa\lambda_i^P,$$

where $P = 2$. [See CT (1986).] The NB1 form ($P = 1$), which has the same conditional mean function, λ_i, but conditional variance

(2-7) $$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i[1 + (1/\theta)] = \lambda_i[1 + \kappa],$$

is obtained by replacing θ with $\theta\lambda_i$ in the density (2-4). (This is not a simple reparameterization of the model. E.g., the log likelihood functions for the two models will differ, and the parameters of the NB1 model are not one to one functions of those in the NB2 model.)

Consider, instead, introducing the heterogeneity in (2-2) as a normally distributed variable with mean zero and standard deviation σ. Then, the conditional Poisson model is

(2-8) $$P(y_i|\mathbf{x}_i,\varepsilon_i) = \frac{\exp(-h_i\lambda_i)(h_i\lambda_i)^{y_i}}{\Gamma(1+y_i)},\ h_i\lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta} + \sigma\varepsilon_i),\ \varepsilon_i \sim \text{N}[0,1].$$

The unconditional density would be

(2-9) $$P(y_i|\mathbf{x}_i) = \int_{-\infty}^{\infty} \frac{\exp[-\exp(\sigma\varepsilon_i)\lambda_i][\exp(\sigma\varepsilon_i)\lambda_i]^{y_i}}{\Gamma(1+y_i)}\phi(\varepsilon_i)d\varepsilon_i,$$

where $\phi(\varepsilon_i)$ denotes the standard normal density. Maximum likelihood estimates of the model parameters are obtained by maximizing the unconditional log likelihood function with respect to the model parameters (α, $\boldsymbol{\beta}$, σ). Butler and Moffitt's (1982) Hermite quadrature based method may be used. [See, e.g., Greene (2007).] Simulation is another effective approach to maximizing the log likelihood function. [See Train (2003) and Greene (2007, 2008).]

3. Models for Panel Data

The Poisson fixed effects model,

(3-1) $$P(y_{it}|\mathbf{x}_{it}) = \frac{\exp(-\lambda_{it})\lambda_{it}^{y_{it}}}{\Gamma(y_{it}+1)}, \lambda_{it} = \exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}),$$

is one of only a few known cases in which maximization of the full log likelihood,

(3-2) $$\ln L = \sum_{i=1}^{N}\sum_{t=1}^{T} \ln P(y_{it}|\mathbf{x}_{it})$$

with respect to (α_i,i=1,...,N, $\boldsymbol{\beta}$) produces an estimate of $\boldsymbol{\beta}$ that is numerically identical to the maximizer of the conditional log likelihood based on

(3-3) $$P(y_{i1},...,y_{iT}|\,\Sigma_{t=1}^{T}y_{it},\mathbf{X}_i) \;=\; \frac{\Gamma\left[\left(\Sigma_{t=1}^{T}y_{it}\right)+1\right]}{\prod_{t=1}^{T}\Gamma(y_{it}+1)}\prod_{t=1}^{T}A_{it}^{y_{it}}\ ,$$

(3-4) $$A_{it} = \frac{\lambda_{it}}{\Sigma_{t=1}^{T}\lambda_{it}} = \frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}{\Sigma_{t=1}^{T}\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}\,.$$

[See Lancaster (2000).]

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

(3-5) $$p\left(y_{i1}, y_{i2}, \ldots, y_{iT_i} \mid \mathbf{X}_i, \Sigma_{t=1}^{T}y_{it}\right) \;=\; \frac{\Gamma\left[(\Sigma_{t=1}^{T}y_{it})+1\right]\Gamma\left(\Sigma_{t=1}^{T}\gamma_{it}\right)}{\Gamma\left(\Sigma_{t=1}^{T}y_{it}+\Sigma_{t=1}^{T}\gamma_{it}\right)}\prod_{t=1}^{T}\frac{\Gamma(y_{it}+\gamma_{it})}{\Gamma(y_{it}+1)\Gamma(\gamma_{it})},$$

$$\gamma_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta}),$$

$$\delta_i = \phi_i/\exp(\mu_i),$$

$$E[y_{it}|\mathbf{x}_{it}] \;=\; \gamma_{it}\,/\,\delta_i \;=\; \exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\mu_i)/\phi_i,$$

$$\mathrm{Var}[y_{it}|\mathbf{x}_{it}] \;=\; \gamma_{it}\,/\,\delta_i^2 = \exp(\mathbf{x}_{it}'\boldsymbol{\beta}+2\mu_i)/\phi_i^2.$$

The specification appears to allow fixed effects in both the mean (through μ_i) and the standard deviation (through ϕ_i). The conditional density in (3-5) is free of both fixed effects, which would seem to solve the heterogeneity problem in the familiar fashion. This is the default FENB formulation used in popular software packages such as Stata, SAS and LIMDEP. But, this leaves the conundrum: Researchers accustomed to the admonishment that fixed effects models cannot contain overall constants or time invariant covariates are sometimes surprised to find (perhaps accidentally) that this fixed effects model allows both. Why can this model coexist with an overall constant term or even an additional set of additive fixed effects?

To resolve the question, return to the HHG formulation of the conditional probability. Using their notation, the departure point is a Poisson model conditioned on an unobserved conditional mean,

$$\mathrm{Prob}[Y_{it} = y_{it}|\lambda_{it}] \;=\; \frac{\exp(-\lambda_{it})\lambda_{it}^{y_{it}}}{\Gamma(y_{it}+1)}\,.$$

Now, assume that the unobserved λ_{it} is distributed as Gamma(γ_{it},δ) where

$$\gamma_{it} \;=\; \exp(\mathrm{x}_{it}'\boldsymbol{\beta}).$$

Then $$f(\lambda_{it}|\mathbf{x}_{it}) \;=\; \frac{\gamma_{it}^{\delta}\exp(-\gamma_{it}\lambda_{it})\lambda_{it}^{\delta-1}}{\Gamma(\delta)}\,.$$

It follows, then $$E[\lambda_{it}|\mathbf{x}_{it}] \;=\; \gamma_{it}/\delta$$

and $$\mathrm{Var}(\lambda_{it}|\mathbf{x}_{it}) \;=\gamma_{it}/\delta^2.$$

By integrating λ_{it} out of the joint density for (y_{it},λ_{it}), we obtain the marginal density reported in HHG [their equation (3.1)]

$$\text{Prob}(Y_{it} = y_{it}|\mathbf{x}_{it}) \;=\; \frac{\Gamma(\gamma_{it}+y_{it})}{\Gamma(\gamma_{it})\Gamma(y_{it}+1)}\left(\frac{\delta}{1+\delta}\right)^{\gamma_{it}}\left(\frac{1}{1+\delta}\right)^{y_{it}}.$$

This is the NB1 model that is obtained by replacing λ_i in our (2.4) with $\lambda_{it} = \delta\gamma_{it}$ and θ with $1/\delta$. It follows, then, that the conditional mean function in the HHG model, in our notation, would be

$$E[y_{it}|\mathbf{x}_{it}] \;=\; \delta\lambda_{it}$$

If we now define

$$\theta_i = \exp(\alpha_i)/\phi_i,$$

then it appears that

$$E[y_{it}|\mathbf{x}_{it}] \;=\; \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i)\;/\;\phi_i$$

and

$$\text{Var}[y_{it}|\mathbf{x}_{it}] \;=\; \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + 2\alpha_i)\;/\;\phi_i^2.$$

and the HHG model follows.

The loose end in the derivation is that the interpretation of θ_i as displacing the mean and variance at the same time is incorrect. The firm specific scale factor θ_i is just that. It acts only on the variance of the random variable. It is a single parameter, not the product of two separately identified parameters. Indeed, θ_i could be written as the product of any number of individual specific parameters, and the group of them would still fall out of the conditional density. The apparent individual specific effect in the conditional mean is an artifact of the functional form chosen for θ_i. To see this clearly, note that α_i cannot vary independently of ϕ_i. Thus, HHG's statement that "both ϕ_i and μ_i are allowed to vary across firms" is incorrect. Only $\phi_i/\exp(\mu_i)$ is allowed to vary across firms.

In the two negative binomial models considered, the conditional mean functions are

(3-6) NB1(HHG): $E[y_{it} \mid \mathbf{x}_{it}] \;=\; (1/\theta_i)\phi_{it} = (1/\theta_i)\exp(\alpha + \mathbf{x}_{it}'\boldsymbol{\beta})$

NB2(FE): $E[y_{it} \mid \mathbf{x}_{it}] \;=\; \lambda_{it} \;=\; \exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})$,

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

(3-7) NB1(HHG): $\text{Var}\,[y_{it} \mid \mathbf{x}_{it}] = (1/\theta_i)\phi_{it}[1 + (1/\theta_i)]$,

NB2(FE): $\text{Var}\,[y_{it} \mid \mathbf{x}_{it}] = \lambda_{it}\,[1 + (1/\theta)\lambda_{it}]$.

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

Theory does not provide a reason to prefer the NB1 formulation over the more familiar NB2 model. The NB1 form does not share the usual interpretation of the fixed effect as carrying only the the time invariant heterogeneity into the conditional mean function. The HHG model being conditionally independent of the fixed effects, does finess the incidental parameters problem – the estimator of $\boldsymbol{\beta}$ *in this NB1 model* is consistent. This is not the case for the FENB2 form. But, it remains unclear what role the fixed effects play in the NB1 model, and how they relate to the fixed effects in other familiar treatments.

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

The random effects Poisson model can be formed by writing

$$\lambda_{it} = \exp(\alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + u_i) \tag{3-8}$$

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

$$P(y_{i1},y_{i2},\ldots,y_{iT}|\mathbf{X}_i) = \frac{\left[\Pi_{t=1}^{T}\lambda_{it}^{y_{it}}\right]\Gamma\left[\theta+\Sigma_{t=1}^{T}y_{it}\right]}{\Gamma(\theta)\left[\Pi_{t=1}^{T}\Gamma(y_{it}+1)\right]\left[\left(\Sigma_{t=1}^{T}\lambda_{it}\right)^{\Sigma_{t=1}^{T}y_{it}}\right]}Q_i^{\theta}(1-Q_i)^{\Sigma_{t=1}^{T}y_{it}} \tag{3-9}$$

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

This is a negative binomial, NB2 distribution for $Y_i = \Sigma_{t=1}^{T} y_{it}$ with mean $\Lambda_i = \Sigma_{t=1}^{T}\lambda_{it}$. The Poisson RE model could also be specified with lognormal heterogeneity. Analysis would follow precisely along the lines of Section 2. The joint probability would be computed from

$$\begin{aligned} P(y_{i1},\ldots,y_{iT}\,|\,\mathbf{X}_i) &= \int_{u_i}\prod_{t=1}^{T}\frac{\exp(-\exp(u_i)\lambda_{it})(\exp(u_i)\lambda_{it})^{y_{it}}}{\Gamma(1+y_{it})}f(u_i)du_i \\ &= \prod_{t=1}^{T}\left[\frac{\lambda_{it}^{y_{it}}}{\Gamma(1+y_{it})}\right]\int_{u_i}\exp\left[-\exp(u_i)\Sigma_{t=1}^{T}\lambda_{it}\right]\left[\exp(u_i)\right]^{\Sigma_{t=1}^{T}y_{it}}f(u_i)du_i. \end{aligned} \tag{3-10}$$

The implied log likelihood function and its derivatives can be approximated using either quadrature or simulation.

Like the fixed effects model, introducing random effects into the negative binomial model adds some additional complexity. An approach that would preserve the form of the model would be to begin with a Poisson model and write

$$\lambda_{it} = \exp(\alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + u_i), \tag{3-11}$$

where both ε_{it} and u_i are log gamma distributed with parameters θ and μ, respectively. This would correspond to a mixed negative binomial model. [The model used in Riphahn (2003).] If it is assumed that ε_{it} has the $G(\theta,\theta)$ distribution assumed in Section 2 and u_i has a normal distribution with mean zero and standard deviation σ, then we obtain a "true" random effects NB model that parallels the model developed earlier. The conditional negative binomial model will result from

$$P(y_{it}|\mathbf{x}_{it},u_i) = \int_{\varepsilon_{it}} P(y_{it}\,|\,x_{it},\varepsilon_i,u_i)\,f(\varepsilon_{it})d\varepsilon_{it}\,. \tag{3-12}$$

Changing the variable to $h_{it} = \exp(\varepsilon_{it})$ and integrating over h_{it} instead produces the negative binomial (NB2) model with conditional mean $E[y_{it}|\mathbf{x}_{it},u_i] = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + \sigma u_i)$ and dispersion parameter θ. The resulting conditional density is

$$P(y_{it}|\mathbf{x}_{it},u_i) = \frac{\Gamma(y_{it}+\theta)}{\Gamma(y_{it}+1)\Gamma(\theta)} r_{it}^{\theta}(1-r_{it})^{y_{it}},$$

(3-13)
$$\lambda_{it} = \exp(\alpha + \mathbf{x}_{it}'\boldsymbol{\beta}),$$

$$r_{it} = \theta / (\theta + \exp(\sigma u_i)\lambda_{it}).$$

We can then estimate the parameters by forming the conditional (on u_i) log likelihood and integrating u_i out either by quadrature or simulation. Note that u_i can be assumed to be either log gamma or normally distributed, but in either case, there will be no closed form for the integrals.

Hausman et al.'s (1984) random effects negative binomial model is a hierarchical model that derives from a heterogeneous Poisson model. The mean in the Poisson model is $\exp(u_i)\lambda_{it}$ where $\exp(u_i)$ has $G(\theta,\theta)$ density. This produces the NB kernel. The unconditional distribution is obtained by treating $p_{it} = [\exp(u_i)\lambda_{it}]/[\Sigma_t \exp(u_i)\lambda_{it}]$ as a random vector with Dirichlet mixing distribution. Each pair of means, $\mu_{it} = \exp(u_i)\lambda_{it}$ $\mu_{is} = \exp(u_i)\lambda_{is}$ is such that $\mu_{it}/(\mu_{it}+\mu_{is})$ has a beta distribution with parameters a and b. The resulting unconditional density is

(3-14)
$$p(y_{i1}, y_{i2},\ldots, y_{iT} \mid \mathbf{X}_i) = \frac{\Gamma(a+b)\Gamma\left(a+\Sigma_{t=1}^{T}\lambda_{it}\right)\Gamma\left(b+\Sigma_{t=1}^{T}y_{it}\right)}{\Gamma(a)\Gamma(b)\Gamma\left(a+\Sigma_{t=1}^{T}\lambda_{it}+b+\Sigma_{t=1}^{T}y_{it}\right)}.$$

This is the common form of the RENB model that is incorporated in several contemporary computer packages. As before, the relationship between the heterogeneity and the conditional mean function is unclear, and there is no obvious interpretation of the hyperparameters a and b – the distribution was chosen for mathematical convenience. The parameters can be directly interpreted in the effects model in (3-11), where the estimated standard deviation of u_i can be directly interpreted against the other parameters in the model. Moreover, the HHG model does not admit of a ready test of the homogeneous model. Estimates of the two forms of the fixed and random effects model are presented in Section 4 for a comparison.

4. Applications

In "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation," Riphahn, Wambach and Million (2003) employed a part of the German Socioeconomic Panel (GSOEP) data set to analyze two count variables, *DocVis*, the number of doctor visits in the last three months and *HospVis*, the number of hospital visits in the last year. The authors employed a bivariate panel data (random effects) Poisson model to study these two outcome variables. A central focus of the investigation was the role of the choice of private health insurance in the intensity of use of the health care system, i.e., whether the data contain evidence of moral hazard. We will use these data to illustrate the model extensions described above.[2]

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

The RWM data set is an unbalanced panel of 7,293 individual families observed from one to seven times. The number of observations per family varies from one to seven (1,525, 1,079, 825, 926, 1,051, 1000, 887) with a total number of observations of 27,326. The 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. The outcome variables of interest in the study were doctor visits in the last three months and number of hospital visits last year.

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

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

and a set of year effects,

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

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.

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

Estimates for the fixed and random effects negative binomial models appear in Table 4. The two sets of fixed effects estimates are quite different. The statistical significance and the signs of several of the coefficients change across the two specifications, including AGE, MARRIED, EDUC, CIVIL, and ADDON. The magnitudes of several of the coefficients change substantively, notably the coefficient on PUBLIC, which is five times larger in the "true" fixed effects estimates. The signs and statistical significance of the period effects reverse several times as well. The difference between the HHG and true FE models is that HHG builds the effects into the variance of the random variable, not the mean. Thus, we cannot conclude that the HHG estimator is a consistent estimator of a model that contains a heterogeneous mean. It is a consistent estimator in the context of a model with heterogeneous variance. We have convincing evidence from the Poisson model that there is substantial latent heterogeneity in the mean of the random variable. The log likelihood function for the "no effects" NB model falls to -27,480, which is thousands less than the log likelihood for either fixed effects specification. Thus, it is reasonable to conclude that the HHG estimator is at least potentially problematic. This finding does not weigh in favor of the true FE estimator, however. There is no minimally sufficient statistic for α_i in the NB2 model, so we are led to expect that the incidental parameters problem will surface in this setting. It remains to be investigated how substantial the biases (if there are any) will be, however. It seems unlikely that the simple proportional results widely known for the probit and logit models will carry over to this setting. The FE approach produces a bit of a Hobson's choice. The HHG model does not actually build the heterogeneity into the mean of the random variable, so we might suspect that it suffers from an "omitted variable" problem. The true fixed effects estimates differ enough from the HHG estimates in this very large sample that one might suspect the appearance of the incidental parameters problem.

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

5. Conclusions

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

9. References

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

Allison, P. and R. Waterman, “Fixed-Effects Negative Binomial Regression Models,” *Sociological Methodology*, 32,, 2002, pp. 247-265.

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

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

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

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

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

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

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

Greene, W., “Functional Form and Heterogeneity in Models for Count Data,” Working Paper Number 07-10, Department of Economics, Stern School of Business, New York University, 2007.

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

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

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

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

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

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

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

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

Table 1. Variables in German Health Care Data File

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

Data source: http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/.
From Riphahn, R., A. Wambach and A. Million "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation," *Journal of Applied Econometrics*, 18, 4, 2003, pp. 387-405.
Notes: * *NEWHSAT* = *HSAT*; 40 observations on HSAT recorded between 6 and 7 were changed to 7.
** Transformed variable not in raw data file.
*** Divided by 1,000 rather than 10,000 by RWM. We used this scale to ease comparison of coefficients.

Table 2. Descriptive Statistics by Gender

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

Table 3 Estimated Panel Data Poisson Models, Males (t ratios in parentheses)

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

Table 4 Estimated Panel Data Negative Binomial Models, Males (t ratios in parentheses)

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