Count Panel Data

A. Colin Cameron and Pravin K. Trivedi
Department of Economics, University of California - Davis.
School of Economics, University of Queensland.

July 5, 2013

Abstract

This paper surveys panel data methods for count dependent variable that takes nonnegative integer values, such as number of doctor visits. The focus is on short panels, as the literature has concentrated on this case. The survey covers both static and dynamic models with random and fixed effects. The paper surveys quasi-ML methods based on the Poisson, as well as richer more parametric models - negative binomial models, finite mixture models, hurdle models and with-zeros models.

Keywords: count; Poisson; negative binomial; panel; fixed effects; random effects; finite mixtures; latent class; hurdle; with-zeros; dynamic models; GMM.

JEL Classification: C13, C23, C25.

This paper is prepared for the Badi H. Baltagi ed., Oxford Handbook of Panel Data Econometrics, Oxford University Press. It has benefitted considerably from the comments of two referees.

Contents

1 Introduction

This paper surveys panel data methods for count dependent variable that takes nonnegative integer values, such as number of doctor visits. The focus is on short panels, with T fixed and $n \to \infty$, as the literature has concentrated on this case.

The simplest panel models specify the conditional mean to be of exponential form, and specify the conditional distribution to be Poisson or, in some settings, a particular variant of the negative binomial. Then it can be possible to consistently estimate slope parameters provided only that the conditional mean is correctly specified, and to obtain standard errors that are robust to possible misspecification of the distribution. This is directly analogous to panel linear regression under normality where consistent estimation and robust inference are possible under much weaker assumptions than normality. In particular, it possible to consistently estimate the slope parameters in a fixed effects version of the Poisson model, even in a short panel.

Richer models account for special features of count data. In particular, the Poisson is inadequate in modelling the conditional distribution as it is a one parameter distribution that imposes variance-mean equality. In most applications the conditional variance exceeds the conditional mean. Richer parametric models are negative binomial models and finite mixture models. Furthermore, even for a given parametric model there can be a bunching or excess of zeros, leading to modified count models – hurdle models and with-zeros models. These considerations are especially important for applications that need to model the conditional distribution, not just the conditional mean. For example, interest may lie in predicting the probability of an excessive number of doctor visits.

Section 2 briefly reviews standard cross-section models for count data, the building block for section 3 that presents standard static models for panel counts with focus on short panels. Section 4 presents extension to the dynamic case, where the current count depends on lagged realizations of the count. Again the emphasis is on short panels, and the Arellano-Bond estimator for linear dynamic models with fixed effects can be adapted to count data. Section 5 considers extensions that address more complicated features of count data.

2 Models for Cross-section Count Data

The main cross-section data models for counts are the Poisson and negative binomial models, hurdle and zero-inflated variants of these models, and latent class or finite mixture models.

2.1 Poisson Quasi-MLE

The Poisson regression model specifies that y_i given $\mathbf{x}_i$ is Poisson distributed with density

$$f(y_i|\mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, \ldots \tag{1}$$

and mean parameter

$$\mathrm{E}[y_i|\mathbf{x}_i] = \mu_i = \exp(\mathbf{x}_i'\beta). \tag{2}$$

The exponential form in (2) ensures that $\mu_i > 0$. It also permits β to be interpreted as a semi-elasticity, since $\beta = [\partial \mathrm{E}[y_i|\mathbf{x}_i]/\partial \mathbf{x}_i]/\mathrm{E}[y_i|\mathbf{x}_i]$. In the statistics literature the model is often called a log-linear model, since the logarithm of the conditional mean is linear in the parameters: $\ln \mathsf{E}[y_i|\mathbf{x}_i] = \mathbf{x}_i'\beta$.

Given independent observations, the log-likelihood is $\ln \mathsf{L}(\beta) = \sum_{i=1}^n \{y_i \mathbf{x}_i'\beta - \exp(\mathbf{x}_i'\beta) - \ln y_i!\}$. The Poisson MLE $\widehat{\beta}_{\mathsf{P}}$ solves the first-order conditions

$$\sum_{i=1}^n (y_i - \exp(\mathbf{x}_i'\beta))\mathbf{x}_i = \mathbf{0}. \tag{3}$$

These first-order conditions imply that the essential condition for consistency of the Poisson MLE is that $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta)$, i.e., that the conditional mean is correctly specified – the data need not be Poisson distributed.

The Poisson quasi-MLE is then asymptotically normally distributed with mean β and variance-covariance matrix

$$\mathrm{V}[\widehat{\beta}_{\mathsf{P}}] = \left(\sum_{i=1}^n \mu_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \left(\sum_{i=1}^n \mathrm{V}[y_i|\mathbf{x}_i]\mathbf{x}_i\mathbf{x}_i'\right) \left(\sum_{i=1}^n \mu_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}, \tag{4}$$

where $\mu_i = \exp(\mathbf{x}_i'\beta)$. This can be consistently estimated using a heteroskedasticity-robust estimate

$$\widehat{\mathrm{V}}[\widehat{\beta}_{\mathsf{P}}] = \left(\sum_{i=1}^n \widehat{\mu}_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \left(\sum_{i=1}^n (y_i - \widehat{\mu}_i)^2\mathbf{x}_i\mathbf{x}_i'\right) \left(\sum_{i=1}^n \widehat{\mu}_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}. \tag{5}$$

A property of the Poisson distribution is that the variance equals the mean. Then $\mathrm{V}[y_i|\mathbf{x}_i] = \mu_i$, so (4) simplifies to $\mathrm{V}[\widehat{\beta}_{\mathsf{P}}] = (\sum_{i=1}^n \mu_i \mathbf{x}_i\mathbf{x}_i')^{-1}$. In practice for most count data, the conditional variance exceeds the conditional mean, a feature called overdispersion. Then using standard errors based on $\mathrm{V}[\widehat{\beta}_{\mathsf{P}}] = (\sum_{i=1}^n \mu_i \mathbf{x}_i\mathbf{x}_i')^{-1}$, the default in most Poisson regression packages, can greatly understate the true standard errors; one should use (5).

The robustness of the Poisson quasi-MLE to distributional misspecification, provided the conditional mean is correctly specified, means that Poisson regression can also be applied to continuous nonnegative data. In particular, OLS regression of $\ln y$ on $\mathbf{x}$ cannot be performed if $y = 0$ and leads to a retransformation problem if we wish to predict y. Poisson regression of y on $\mathbf{x}$ (with exponential conditional mean) does not have these problems.

The robustness to distributional misspecification is shared with the linear regression model with assumed normal errors. More generally, this holds for models with specified density in the linear exponential family, i.e., $f(y|\mu) = \exp\{a(\mu) + b(y) + c(\mu)y\}$. In the statistics literature this class is known as generalized linear models (GLM). It includes the normal, Poisson, geometric, gamma, Bernoulli and binomial.

2.2 Parametric Models

In practice, the Poisson distribution is too limited as it is a one-parameter distribution, depending on only the mean μ. In particular, the distribution restricts the variance to equal the mean, but count data used in economic applications generally are overdispersed.

The standard generalization of the Poisson is the negative binomial (NB) model, most often the NB2 variant that specifies the variance to equal $\mu + \alpha\mu^2$. Then

$$f(y_i|\mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{y_i}, \quad \alpha > 0,\ y_i = 0, 1, 2, ... \quad (6)$$

This reduces to the Poisson for $\alpha \to 0$. Specifying $\mu_i = \exp(\mathbf{x}_i'\beta)$, the MLE solves for β and α the first-order conditions

$$\begin{aligned} \sum_{i=1}^{n} \frac{y_i - \mu_i}{1 + \alpha\mu_i} \mathbf{x}_i = \mathbf{0} \\ \sum_{i=1}^{n} \left\{ \frac{1}{\alpha^2} \left(\ln(1 + \alpha\mu_i) - \textstyle\sum_{j=0}^{y_i - 1} \frac{1}{(j + \alpha^{-1})} \right) + \frac{y_i - \mu_i}{\alpha\,(1 + \alpha\mu_i)} \right\} = 0. \end{aligned} \quad (7)$$

As for the Poisson, the NB2 MLE for β is consistent provided $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta)$.

A range of alternative NB models can be generated by specifying $\mathrm{V}[y|\mathbf{x}] = \mu + \alpha\mu^p$ where p is specified or is an additional parameter to be estimated. The most common alternative model is the NB1 that sets $p = 1$, so the conditional variance is a multiple of the mean. For these variants the quasi-MLE is no longer consistent – the distribution needs to be correctly specified. Yet another variation parameterizes α to depend on regressors.

The NB models are parameterized to have the same conditional mean as the Poisson. In theory the NB MLE is more efficient than the Poisson QMLE if the NB model is correctly specified, though in practice the efficiency gains are often small. The main reason for using the NB is in settings where the desire is to fit the distribution, not just the conditional mean. For example, interest may lie in predicting the probability of ten or more doctor visits. And a fully parametric model such as the NB may be necessary if the count is incompletely observed, due to truncation, censoring and interval-recording (e.g. counts recorded as 0, 1, 2, 3-5, more than 5).

Both the Poisson and NB models are inadequate if zero counts do not come from the same process as positive counts. Then there are two commonly-used modified count models, based on different behavioral models. Let $f_2(y)$ denote the latent count density. A hurdle or two-part model specifies that positive counts are observed only after a threshold is crossed, with probability $1 - f_1(0)$. Then we observe $f(0) = f_1(0)$ and, for $y > 0$, $f(y) = f_2(y)(1 - f_1(0))/(1 - f_2(0))$. A zero-inflated or with-zeros model treats some zero counts as coming from a distinct process due to, for example, never participating in the activity or mismeasurement. In that case $f_2(0)$, the probability of zero counts from the baseline density, is inflated by adding a probability of, say, π. Then we have $f(0) = \pi + (1 - \pi)f_2(0)$ and, for $y > 0$, $f(y) = (1 - \pi)f_2(y)$.

A final standard adaptation of cross-section count models is a latent class or finite mixtures model. Then y is a draw from an additive mixture of C distinct populations with component (subpopulation) densities $f_1(y), ..., f_C(y)$, in proportions $\pi_1, ..., \pi_C$, where $\pi_j \geqslant 0$, $j = 1, ..., C$, and $\sum_{j=1}^{C} \pi_j = 1$. The mixture density is then $f(y) = \sum_{j=1}^{C} \pi_j f_j(y)$. Usually the π_j are not parameterized to depend on regressors, the $f_j(y)$ are Poisson or NB models with regressors, and often $C = 2$ is adequate.

3 Static Panel Count Models

The standard methods for linear regression with data from short panels – pooled OLS and FGLS, random effects and fixed effects – extend to Poisson regression and, to a lesser extent, to NB regression. Discussion of other panel count models is deferred to section 5.

3.1 Individual Effects in Count Models

Fixed and random effects models for short panels introduce an individual-specific effect. For count models, with conditional mean restricted to be positive, the effect is multiplicative in the conditional mean, rather than additive. Then

$$\mu_{it} \equiv \mathrm{E}[y_{it}|\mathbf{x}_{it}, \alpha_i] = \alpha_i \lambda_{it} = \alpha_i \exp(\mathbf{x}_{it}'\beta), \;\; i = 1, ..., n, \;\; t = 1, ..., T, \tag{8}$$

where the last equality specifies an exponential functional form. Note that the intercept is merged into α_i, so that now the regressors $\mathbf{x}_{it}$ do not include an intercept.

In this case the model can also be expressed as

$$\mu_{it} \equiv \exp(\delta_i + \mathbf{x}_{it}'\beta), \tag{9}$$

where $\delta_i = \ln \alpha_i$. For the usual case of an exponential conditional mean, the individual-specific effect can be interpreted as either a multiplicative effect or as an intercept shifter. If there is reason to specify a conditional mean that is not of exponential then a multiplicative effects model may be specified, with $\mu_{it} \equiv \alpha_i g(\mathbf{x}_{it}'\beta)$, or an intercept shift model may be used, with $\mu_{it} \equiv g(\delta_i + \mathbf{x}_{it}'\beta)$.

Unlike the linear model, consistent estimation of β here does not identify the marginal effect. The marginal effect given (8) is

$$\mathrm{ME}_{itj} \equiv \frac{\partial \mathrm{E}[y_{it}|\mathbf{x}_{it}, \alpha_i]}{\partial \mathbf{x}_{itj}} = \alpha_i \exp(\mathbf{x}_{it}'\beta)\beta_j = \beta_j \mathrm{E}[y_{it}|\mathbf{x}_{it}, \alpha_i], \tag{10}$$

which depends on the unknown α_i. Instead, the slope coefficient β_j can be interpreted as a semi-elasticity, giving the proportionate increase in $\mathrm{E}[y_{it}|\mathbf{x}_{it}, \alpha_i]$ associated with a one-unit change in x_{itj}. For example, if $\beta_j = .06$ then a one-unit change in x_j is associated with a 6% increase in y_{it}, after controlling for both regressors and the unobserved individual effect α_i.

3.2 Pooled or Population-Averaged Models

Before estimating models with individual-specific effects, namely fixed and random effects models, we consider pooled regression. A pooled Poisson model bases estimation on the marginal distributions of the individual counts y_{it}, rather than on the joint distribution of the counts $y_{i1}, ..., y_{iT}$ for the i^{th} individual.

The pooled Poisson QMLE is obtained by standard Poisson regression of y_{it} on an intercept and $\mathbf{x}_{it}$. Define $\mathbf{z}_{it}' = [1\ \mathbf{x}_{it}']$ and $\gamma' = [\delta\ \beta']$, so $\exp(\mathbf{z}_{it}'\gamma) = \exp(\delta + \mathbf{x}_{it}'\beta)$. Then the first-order conditions are

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \exp(\mathbf{z}_{it}'\gamma))\mathbf{z}_{it} = \mathbf{0}. \tag{11}$$

The estimator is consistent if

$$\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \exp(\delta + \mathbf{x}_{it}'\beta) = \alpha \exp(\mathbf{x}_{it}'\beta) \tag{12}$$

i.e., if the conditional mean is correctly specified. Default standard errors are likely to be incorrect, however, as they assume that y_{it} is equidispersed and is uncorrelated over time for individual i. Instead it is standard in short panels to use cluster-robust standard errors, with clustering on the individual, based on the variance matrix estimate

$$\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{\mu}_{it}\mathbf{z}_{it}\mathbf{z}_{it}'\right]^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{u}_{it}\widehat{u}_{is}\mathbf{z}_{it}\mathbf{z}_{is}'\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{\mu}_{it}\mathbf{z}_{it}\mathbf{z}_{it}'\right]^{-1}, \tag{13}$$

where $\widehat{\mu}_{it} = \exp(\mathbf{z}_{it}'\widehat{\gamma})$, and $\widehat{u}_{it} = y_{it} - \exp(\mathbf{z}_{it}'\widehat{\gamma})$.

The multiplicative effects model (8) for $\mathrm{E}[y_{it}|\mathbf{x}_{it}, \alpha_i]$ leads to condition (12) for $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$ if α_i is independent of $\mathbf{x}_{it}$ and $\alpha = \mathrm{E}_{\alpha_i}[\alpha_i]$. This condition holds in a random effects model, see below, but not in a fixed effects model. The statistics literature refers to the pooled estimator as the population-averaged estimator, since (12) is assumed to hold after averaging out any individual-specific effects. The term marginal analysis, meaning marginal with respect to α_i, is also used.

More efficient pooled estimation is possible by making assumptions about the correlation between y_{it} and y_{is}, $s \neq t$, conditional on regressors $\mathbf{X}_i = [\mathbf{x}_{i1}' \cdots \mathbf{x}_{iT}']'$. Let $\mu_i(\gamma) = [\mu_{i1} \cdots \mu_{iT}]'$ where $\mu_{it} = \exp(\mathbf{z}_{it}'\gamma)$ and let $\mathbf{\Sigma}_i$ be a model for $\mathrm{V}[\mathbf{y}_i|\mathbf{X}_i]$ with ts^{th} entry $\mathrm{Cov}[y_{it}, y_{is}|\mathbf{X}_i]$. For example, if we assume data are equicorrelated, so $\mathrm{Cor}[y_{it}, y_{is}|\mathbf{X}_i] = \rho$ for all $s \neq t$, and that data are overdispersed with variance σ_{it}^2, then $\mathbf{\Sigma}_{i,ts} \equiv \mathrm{Cov}[y_{it}, y_{is}|\mathbf{X}_i] = \rho\sigma_{it}\sigma_{is}$. An alternative model permits more flexible correlation for the first m lags, with $\mathrm{Cor}[y_{it}, y_{i,t-k}|\mathbf{X}_i] = \rho_k$ where $\rho_k = 0$ for $|k| > m$. Such assumptions enable estimation by more efficient feasible nonlinear generalized least squares. The first-order conditions for γ are

$$\sum_{i=1}^{n}\frac{\partial\mu_i'(\gamma)}{\partial\gamma}\widehat{\mathbf{\Sigma}}_i^{-1}(\mathbf{y}_i - \mu_i(\gamma)) = \mathbf{0}, \tag{14}$$

where $\widehat{\mathbf{\Sigma}}_i$ is obtained from initial first-stage pooled Poisson estimation of β and consistent estimation of any other parameters that determine $\mathbf{\Sigma}_i$.

The statistics literature calls this estimator the Poisson generalized estimating equations (GEE) estimator. The variance model $\mathbf{\Sigma}_i$ is called a working matrix, as it is possible to obtain a cluster-robust estimate of the asymptotic variance matrix robust to misspecification of $\mathbf{\Sigma}_i$, provided $n \to \infty$. Key references are Zeger and Liang (1986) and Liang and Zeger (1986). Liang, Zeger and Qaqish (1992) consider generalized GEE estimators that jointly estimate the regression and correlation parameters. Brännäs and Johansson (1996) allow for time-varying random effects α_{it} and estimation by generalized method of moments (GMM).

The preceding pooled estimators rely on correct specification of the conditional mean $\mathsf{E}[y_{it}|\mathbf{x}_{it}]$. In richer parametric models, such as a hurdle model or an NB model other than NB2, stronger assumptions are needed for estimator consistency. The log-likelihood for pooled ML estimation is based for individual i on $\prod_{t=1}^{T} f(y_{it}|\mathbf{x}_{it})$, the product of the marginal densities, rather than the joint density $f(\mathbf{y}_i|\mathbf{X}_i)$. Consistent estimation generally requires that the marginal density $f(y_{it}|\mathbf{x}_{it})$ be correctly specified. Since y_{it} is in fact correlated over t, there is an efficiency loss. Furthermore, inference should be based on cluster-robust standard errors, possible given $n \to \infty$.

3.3 Random Effects Models and Extensions

A random effects (RE) model is an individual effects model with the individual effect α_i (or δ_i) assumed to be distributed independently of the regressors. Let $f(y_{it}|\mathbf{x}_{it}, \alpha_i)$ denote the density for the it^{th} observation, conditional on both α_i and the regressors. Then the joint density for the i^{th} observation, conditional on the regressors, is

$$f(\mathbf{y}_i|\mathbf{X}_i) = \int_0^{\infty} \left[\prod_{t=1}^{T} f(y_{it}|\alpha_i, \mathbf{x}_{it})\right] g(\alpha_i|\eta) d\alpha_i, \qquad (15)$$

where $g(\alpha_i|\eta)$ is the specified density of α_i. In some special cases there is an explicit solution for the integral (15). Even if there is no explicit solution, Gaussian quadrature numerical methods work well since the integral is only one-dimensional, or estimation can be by simulated maximum likelihood.

The Poisson random effects model is obtained by supposing y_{it} is Poisson distributed, conditional on $\mathbf{x}_{it}$ and α_i, with mean $\alpha_i \lambda_{it}$, and additionally that α_i is gamma distributed with mean 1, a normalization, and variance $1/\gamma$. Then, integrating out α_i, the conditional mean $\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \lambda_{it}$, the conditional variance $\mathrm{V}[y_{it}|\mathbf{x}_{it}] = \lambda_{it} + \lambda_{it}^2/\gamma$, and there is a closed form solution to (15), with

$$f(\mathbf{y}_i|\mathbf{X}_i) = \left[\prod_t \frac{\lambda_{it}{}^{y_{it}}}{y_{it}!}\right] \times \left(\frac{\gamma}{\sum_t \lambda_{it} + \gamma}\right)^{\gamma} \times \left(\sum_t \lambda_{it} + \gamma\right)^{-\sum_t y_{it}} \frac{\Gamma\left(\sum_t y_{it} + \gamma\right)}{\Gamma(\gamma)}. \qquad (16)$$

For exponential conditional mean the ML first-order conditions for $\widehat{\beta}$ are

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it}-\lambda_{it}\frac{\overline{y}_i+\gamma/T}{\overline{\lambda}_i+\gamma/T}\right)=\mathbf{0}, \tag{17}$$

where $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\overline{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T}\lambda_{it}$. A sufficient condition for consistency is that $\mathrm{E}[y_{it}|\mathbf{X}_i] = \lambda_{it}$. The model has the same conditional mean as the pooled Poisson, but leads to more efficient estimation if in fact overdispersion is of the NB2 form.

The Poisson random effects model was proposed by Hausman, Hall and Griliches (1984). They also presented a random effects version of the NB2 model, with y_{it} specified to be i.i.d. NB2 with parameters $\alpha_i\lambda_{it}$ and ϕ_i, where $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta)$. Conditional on λ_{it}, α_i, and ϕ_i, y_{it} has mean $\alpha_i\lambda_{it}/\phi_i$ and variance $(\alpha_i\lambda_{it}/\phi_i)\times(1+\alpha_i/\phi_i)$. A closed form solution to (15) is obtained by assuming that $(1+\alpha_i/\phi_i)^{-1}$ is a beta-distributed random variable with parameters (a, b).

The preceding examples specify a distribution for α_i that leads to a closed-form solution to (15). This is analogous to specifying a natural conjugate prior in a Bayesian setting. Such examples are few, and in general there is no closed form solution to (15). Furthermore, the most obvious choice of distribution for the multiplicative effect α_i is the lognormal, equivalent to assuming that δ_i in $\exp(\delta_i+\mathbf{x}_{it}'\beta)$ is normally distributed. Since there is then no closed form solution to (15), Gaussian quadrature or maximum simulated likelihood methods are used. If $\delta_i \sim \mathrm{N}[\delta, \sigma_\delta^2]$ then $\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \exp(\delta+\sigma_\delta^2/2)\lambda_{it}$, a rescaling of the conditional mean in the Poisson-gamma random effects model. This is absorbed in the intercept if $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta)$.

More generally, slope coefficients in addition to the intercept may vary across individuals. A random coefficients model with exponential conditional mean specifies $\mathrm{E}[y_{it}|\mathbf{x}_{it}, \delta_i, \beta_i] = \exp(\delta_i + \mathbf{x}_{it}'\beta_i)$. Assuming $\delta_i \sim \mathrm{N}[\delta, \sigma_\delta^2]$ and $\beta_i \sim \mathrm{N}[\beta, \mathbf{\Sigma}_\beta]$ implies $\delta_i + \mathbf{x}_{it}'\beta_i \sim \mathrm{N}[\delta + \mathbf{x}_{it}'\beta, \sigma_\delta^2+\mathbf{x}_{it}'\mathbf{\Sigma}_\beta\mathbf{x}_{it}]$. The conditional mean is considerably more complicated, with $\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \exp\{\delta + \mathbf{x}_{it}'\beta + (\sigma_\delta^2 + \mathbf{x}_{it}'\mathbf{\Sigma}_\beta\mathbf{x}_{it})/2\}$. This model falls in the class of generalized linear latent and mixed models; see Skrondal and Rabe-Hesketh (2004). Numerical integration methods are more challenging as the likelihood now involves multi-dimensional integrals.

One approach is to use Bayesian Markov chain Monte Carlo (MCMC) methods. Chib, Greenberg and Winkelmann (1998) consider the following model. Assume $y_{it}|\gamma_{it}$ is Poisson distributed with mean $\exp(\gamma_{it})$, where $\gamma_{it} = \mathbf{x}_{it}'\beta + \mathbf{w}_{it}'\alpha_i$ and $\alpha_i \sim \mathrm{N}[\alpha, \mathbf{\Sigma}_\alpha]$. The RE model is the specialization $\mathbf{w}_{it}'\alpha_i = \alpha_i$, and the random coefficients model sets $\mathbf{w}_{it} = \mathbf{x}_{it}$, though Chib et al. (1998) argue that $\mathbf{x}_{it}$ and $\mathbf{w}_{it}$ should share no common variables to avoid identification and computational problems. Data augmentation is used to add γ_{it} as parameters leading to augmented posterior $p(\beta, \eta, \mathbf{\Sigma}, \gamma|\mathbf{y}, \mathbf{X})$. A Gibbs sampler is used where draws from $p(\gamma|\beta, \eta, \mathbf{\Sigma}, \mathbf{y}, \mathbf{X})$ use the Metropolis-Hastings algorithm, while draws from the other full conditionals $p(\beta|\eta, \mathbf{\Sigma}, \gamma, \mathbf{y}, \mathbf{X})$, $p(\eta|\beta, \mathbf{\Sigma}, \gamma, \mathbf{y}, \mathbf{X})$, and $p(\mathbf{\Sigma}|\beta, \eta, \gamma, \mathbf{y}, \mathbf{X})$ are straightforward if independent normal priors for β and η and a Wishart prior for $\mathbf{\Sigma}^{-1}$ are specified.

Another generalization of the RE model is to model the time-invariant individual effect to depend on the average of individual effects, an approach proposed for linear regression by Mundlak (1978) and Chamberlain (1982). The conditionally correlated random (CCR) effects model specifies that α_i in (8) can be modelled as

$$\alpha_i = \exp(\bar{\mathbf{x}}_i'\lambda + \varepsilon_i), \tag{18}$$

where $\bar{\mathbf{x}}_i$ denotes the time-average of the time-varying exogenous variables and ε_i may be interpreted as unobserved heterogeneity that is uncorrelated with the regressors. Substituting into (8) yields

$$\mathrm{E}[y_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \alpha_i] = \exp(\mathbf{x}_{it}'\beta + \bar{\mathbf{x}}_i'\lambda + \varepsilon_i). \tag{19}$$

This can be estimated as an RE model, with $\bar{\mathbf{x}}_i$ as an additional regressor.

3.4 Fixed Effects Models

Fixed effects (FE) models treat the individual effect α_i in (8) as being random and potentially correlated with the regressors $\mathbf{X}_i$. In the linear regression model with additive errors the individual effect can be eliminated by mean-differencing or by first-differencing. In the nonlinear model (8), α_i can be eliminated by quasi-differencing as follows.

Assume that the regressors $\mathbf{x}_{it}$ are strictly exogenous, after conditioning on α_i, so that

$$\mathrm{E}[y_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \alpha_i] \equiv \mathsf{E}[y_{it}|\mathbf{X}_i, \alpha_i] = \alpha_i\lambda_{it}. \tag{20}$$

This is a stronger condition than (8) which conditions only on $\mathbf{x}_{it}$ and α_i. Averaging over time for individual i, it follows that $\mathsf{E}[\bar{y}_i|\mathbf{X}_i, \alpha_i] = \alpha_i\bar{\lambda}_i$, where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^T y_{it}$ and $\bar{\lambda}_i = \frac{1}{T}\sum_{t=1}^T \lambda_{it}$. Subtracting from (20) yields

$$\mathrm{E}\left[\left(y_{it} - (\lambda_{it}/\bar{\lambda}_i)\bar{y}_i\right)|\alpha_i, \mathbf{X}_i\right] = 0, \tag{21}$$

and hence by the law of iterated expectations

$$\mathrm{E}\left[\mathbf{x}_{it}\left(y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right)\right] = \mathbf{0}. \tag{22}$$

Given assumption (20), β can be consistently estimated by the method of moments estimator that solves the sample moment conditions corresponding to (22):

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it} - \frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\right) = \mathbf{0}. \tag{23}$$

For short panels a panel-robust estimate of the variance matrix of $\widehat{\beta}$ can be obtained using standard GMM results.

Wooldridge (1990) covers this moment-based approach in more detail and gives more efficient GMM estimators when additionally the variance is specified to be of the form $\psi_i\alpha_i\lambda_{it}$. Chamberlain (1992a) gives semi-parametric efficiency bounds for models using only specified first moment of form (8). Attainment of these bounds is theoretically possible but practically difficult, as it requires high-dimensional nonparametric regressions.

Remarkably, the method of moments estimator defined in (23) coincides with the Poisson fixed effects estimator in the special case that $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta)$. This estimator in turn can be derived in two ways.

First, suppose we assume that $y_{it}|\mathbf{x}_{it}, \alpha_i$ is independently distributed over i and t as Poisson with mean $\alpha_i \lambda_{it}$. Then maximizing the log likelihood function $\sum_{i=1}^{n}\sum_{t=1}^{T} \ln f(y_{it}|\mathbf{x}_{it}, \alpha_i)$ with respect to both β and $\alpha_1, ..., \alpha_n$ leads to first-order conditions for β that after some algebra can be expressed as (23), while $\widehat{\alpha}_i = \overline{y}_i/\widehat{\overline{\lambda}}_i$. This result, given in Blundell, Griffith, and Windmeijer (1997, 2002) and Lancaster (1997) is analogous to that for MLE in the linear regression model under normality with fixed effects – in principle the α_i introduce an incidental parameters problem, but in these specific models this does not lead to inconsistent estimation of β, even if T is small.

Second, consider the Poisson conditional MLE that additionally conditions on $T\overline{y}_i = \sum_{t=1}^{T} y_{it}$. Then some algebra reveals that α_i drops out of the conditional log-likelihood function $\sum_{i=1}^{n}\sum_{t=1}^{T} \ln f(y_{it}|\mathbf{x}_{it}, \alpha_i, T\overline{y}_i)$, and that maximization with respect to β leads to the first-order conditions (23). This is the original derivation of the Poisson FE estimator due to Palmgren (1981) and Hausman, Hall, and Griliches (1984). Again, a similar result holds for the linear regression model under normality.

In general it is not possible to obtain consistent estimates of β in a fixed effects model with data from a short panel, due to too many incidental parameters $\alpha_i, ..., \alpha_n$. The three leading exceptions are regression with additive errors, regression with multiplicative errors, including the Poisson, and the logit model. Hausman, Hall, and Griliches (1984) additionally proposed a fixed effects estimator for the NB1 model, but Guimarães (2008) shows that this model places a very strong restriction on the relationship between α_i and the NB1 overdispersion parameter. One consequence, pointed out by Allison and Waterman (2002), is that the coefficients of time-invariant regressors are identified in this model.

One alternative is to estimate a regular NB model, such as NB2, with a full set of individual dummies. While this leads to inconsistent estimation of β in short panels due to the incidental parameters problem, Allison and Waterman (2002) and Greene (2004) present simulations that suggest that this inconsistency may not be too large for moderately small T; see also Fernández-Val (2009) who provides theory for the probit model. A second alternative is to use the conditionally correlated random effects model presented in (18).

The distinction between fixed and random effects is fundamentally important, as pooled and random effects estimators are inconsistent if in fact the data are generated by the individual-specific effects model (8) with α_i correlated with $\mathbf{x}_{it}$. Let β_1 denote the subcomponent of β that is identified in the fixed effects model (i.e. the coefficient of time-varying regressors), or a subset of this, and let $\widehat{\beta}_{1,\mathrm{RE}}$ and $\widetilde{\beta}_{1,\mathrm{FE}}$ denote, respectively, the corresponding RE and FE estimators. The Hausman test statistic is

$$\mathrm{H} = (\widehat{\beta}_{1,\mathrm{RE}} - \widetilde{\beta}_{1,\mathrm{FE}})' \left[\widehat{\mathsf{V}}[\widetilde{\beta}_{1,\mathrm{FE}} - \widehat{\beta}_{1,\mathrm{RE}}]\right]^{-1} (\widehat{\beta}_{1,\mathrm{RE}} - \widetilde{\beta}_{1,\mathrm{FE}}). \tag{24}$$

If H< $\chi^2_\alpha(\dim(\beta_1))$ then at significance level α we do not reject the null hypothesis that the individual specific effects are uncorrelated with regressors. In that case there is no need for fixed effects estimation.

This test requires an estimate of $\mathsf{V}[\widetilde{\beta}_{1,\text{FE}} - \widehat{\beta}_{1,\text{RE}}]$. This reduces to $\mathsf{V}[\widetilde{\beta}_{1,\text{FE}}]-\text{V}[\widehat{\beta}_{1,\text{RE}}]$, greatly simplifying analysis, under the assumption that the RE estimator is fully efficient under the null hypothesis. But it is very unlikely that this additional restriction is met. Instead in short panels one can do a panel bootstrap that resamples over individuals. In the b^{th} resample compute $\widetilde{\beta}^{(b)}_{1,\text{FE}} - \widehat{\beta}^{(b)}_{1,\text{RE}}$ and, given B bootstraps, compute the variance of these B differences.

4 Dynamic Panel Count Models

An individual-specific effect α_i induces dependence over time in y_{it}. An alternative way to introduce dependence over time is a dynamic model that specifies the distribution of y_{it} to depend directly on lagged values of y_{it}.

4.1 Specifications for Dynamic Models

We begin by considering the Poisson model in the time series case, with y_t Poisson distributed with mean that is a function of y_{t-1} and regressors $\mathbf{x}_t$. Then a much wider range of specifications for a dynamic model have been proposed than in the linear case; Cameron and Trivedi (2013, chapter 7) provide a survey.

One obvious time series model, called exponential feedback, is that y_t is Poisson with mean $\exp(\rho y_{t-1} + \mathbf{x}_t'\beta)$, but this model is explosive if $\rho > 0$. An alternative is to specify the mean to be $\exp(\rho \ln y_{t-1} + \mathbf{x}_t'\beta)$, but this model implies that if $y_{t-1} = 0$ then y_t necessarily equals zero. A linear feedback model specifies the mean to equal $\rho y_{t-1} + \exp(\mathbf{x}_t'\beta)$. This model arises from a Poisson integer-valued autoregressive model of order 1 (INAR(1)), a special case of the more general class of INARMA models.

For panel data we allow for both dynamics and the presence of an individual specific effect. Define the conditional mean to be

$$\mu_{it} = \mathsf{E}[y_{it}|\mathbf{X}_i^{(t)}, \mathbf{Y}_i^{(t-1)}, \alpha_i], \tag{25}$$

where $\mathbf{X}_i^{(t)} = \{\mathbf{x}_{it}, \mathbf{x}_{i,t-1}, ...\}$ and $\mathbf{Y}_i^{(t-1)} = \{y_{i,t-1}, y_{i,t-2}, ...\}$. Blundell, Griffith and Windmeijer (1997, 2002) discuss various forms for μ_{it} and emphasize the linear feedback model

$$\mu_{it} = \rho y_{i,t-1} + \alpha_i \exp(\mathbf{x}_{it}'\beta), \tag{26}$$

where for simplicity we consider models where μ_{it} depends on just the first lag of y_{it}. The exponential feedback model instead specifies

$$\mu_{it} = \alpha_i \exp(\rho y_{i,t-1} + \mathbf{x}_{it}'\beta). \tag{27}$$

Yet another model, proposed by Crepon and Duguet (1997), is that

$$\mu_{it} = h(y_{i,t-1}, \gamma)\alpha_i \exp(\mathbf{x}_{it}'\beta), \tag{28}$$

where the function $h(y_{it-1}, \gamma)$ parameterizes the dependence of μ_{it} on lagged values of y_{it}. A simple example is $h(y_{i,t-1}, \gamma) = \exp(\gamma \mathbf{1}[y_{i,t-1} > 0])$ where $\mathbf{1}[\cdot]$ is the indicator function. More generally a spline-type specification in which a set of dummies determined by ranges taken by y_{it-1} might be specified.

4.2 Pooled Dynamic Models

Pooled dynamic models assume that all regression coefficients are the same across individuals, so that there are no individual-specific fixed or random effects. Then one can directly apply the wide range of methods suggested for time series data, even for small T provided $n \to \infty$. This approach is given in Diggle, Heagarty, Liang and Zeger (2002, chapter 10), who use autoregressive models that directly include $y_{i,t-k}$ as regressors. Brännäs (1995) briefly discusses a generalization of the INAR(1) time series model to longitudinal data.

Under weak exogeneity of regressors, which requires that there is no serial correlation in $(y_{it} - \mu_{it})$, the models can be estimated by nonlinear least squares, GEE, method of moments, or GMM based on the sample moment condition $\sum_i \sum_t \mathbf{z}_{it}(y_{it} - \mu_{it})$ where $\mathbf{z}_{it}$ can include $y_{i,t-1}$ and $\mathbf{x}_{it}$ and, if desired, additional lags in these variables.

This approach leads to inconsistent estimation if fixed effects are present. But inclusion of lagged values of y_{it} as a regressor may be sufficient to control for correlation between y_{it} and lagged y_{it}, so that there is no need to additionally include individual-specific effects.

4.3 Random Effects Dynamic Models

A random effects dynamic model is an extension of the static RE model that includes lagged y_{it} as regressors. However, the log-likelihood will depend on initial condition y_{i0}, this condition will not disappear asymptotically in a short panel, and most importantly it will be correlated with the random effect α_i (even if α_i is uncorrelated with $\mathbf{x}_{it}$). So it is important to control for the initial condition.

Heckman (1981) writes the joint distribution of $y_{i0}, y_{i1}, ..., y_{iT}, \alpha_i | \mathbf{x}_{it}$ as

$$f(y_{i0}, y_{i1}, \ldots, y_{iT}, \alpha_i | \mathbf{X}_i) = f(y_{i1}, \ldots, y_{iT} | \mathbf{X}_i, y_{i0}, \alpha_i) f(y_{i0} | \mathbf{X}_i, \alpha_i) f(\alpha_i | \mathbf{X}_i). \tag{29}$$

Implementation requires specification of the functional forms $f(y_{i0} | \mathbf{X}_i, \alpha_i)$ and $f(\alpha_i | \mathbf{X}_i)$ and, most likely, numerical integration; see Stewart (2007).

Wooldridge (2005) instead proposed a conditional approach, for a class of nonlinear dynamic panel models that includes the Poisson model, based on the decomposition

$$f(y_{i1}, \ldots, y_{iT}, \alpha_i | \mathbf{X}_i, y_{i0}) = f(y_{i1}, \ldots, y_{iT} | \mathbf{X}_i, y_{i0}, \alpha_i) f(\alpha_i | y_{i0}, \mathbf{X}_i). \tag{30}$$

This simpler approach conditions on y_{i0} rather than modelling the distribution of y_{i0}. Then the standard random effects conditional ML approach identifies the parameters of interest. One possible model for $f(\alpha_i|y_{i0,}\mathbf{X}_i)$ is the CCR model in (18) with y_{i0} added as a regressor, so

$$\alpha_i = \exp(\delta_0 y_{i0} + \bar{\mathbf{x}}_i'\lambda + \varepsilon_i), \tag{31}$$

where $\bar{\mathbf{x}}_i$ denotes the time-average of the time-varying exogenous variables, and ε_i is an i.i.d. random variable. Then the model (30)-(31) can be estimated using RE model software commands. Note that in a model with just one lag of $y_{i,t-1}$ as a regressor, identification in the CCR model requires three periods of data (y_{i0}, y_{i1}, y_{i2}).

4.4 Fixed Effects Dynamic Models

The Poisson FE estimator eliminates fixed effects under the assumption that $\mathsf{E}[y_{it}|\mathbf{X}_i] = \alpha_i\lambda_{it}$; see (20). This assumption rules out predetermined regressors. To allow for predetermined regressors that may be correlated with past shocks, we make the weaker assumption that regressors are weakly exogenous, so

$$\mathrm{E}[y_{it}|\mathbf{X}_i^{(t)}] = \mathrm{E}[y_{it}|\mathbf{x}_{it}, \ldots, \mathbf{x}_{i1}] = \alpha_i\lambda_{it}, \tag{32}$$

where now conditioning is only on current and past regressors. Then, defining $u_{it} = y_{it} - \alpha_i\lambda_{it}$, $\mathrm{E}[u_{it}|\mathbf{x}_{is}] = 0$ for $s \leq t$, so future shocks are indeed uncorrelated with current $\mathbf{x}$, but there is no restriction that $\mathrm{E}[u_{it}|\mathbf{x}_{is}] = 0$ for $s > t$.

For dynamic models, lagged dependent variables also appear as regressors, and we assume

$$\mathrm{E}[y_{it}|\mathbf{X}_i^{(t)}, \mathbf{Y}_i^{(t-1)}] = \mathrm{E}[y_{it}|\mathbf{x}_{it}, \ldots, \mathbf{x}_{i1}, y_{i,t-1}, ..., y_{i1}] = \alpha_i\lambda_{it}, \tag{33}$$

where conditioning is now also on past values of y_{it}. (For the linear feedback model defined in (26), $\alpha_i\lambda_{it}$ in (33) is replaced by $\rho y_{i,t-1} + \alpha_i\lambda_{it}$.)

If regressors are predetermined then the Poisson FE estimator is inconsistent, since quasi-differencing subtracts $(\lambda_{it}/\bar{\lambda}_i)\bar{y}_i$ from y_{it}, see (21), but $\bar{y}_i$ includes future values y_{is}, $s > t$. This problem is analogous to the inconsistency (or Nickell bias) of the within or mean-differenced fixed effects estimator in the linear model.

Instead GMM estimation is based on alternative differencing procedures that eliminate α_i under the weaker assumption (33). These generalize the use of first differences in linear dynamic models with fixed effects. Chamberlain (1992b) proposed eliminating the fixed effects α_i by the transformation

$$q_{it}(\theta) = \frac{\lambda_{i,t-1}}{\lambda_{it}}y_{it} - y_{i,t-1}, \tag{34}$$

where $\lambda_{it} = \lambda_{it}(\theta)$. Wooldridge (1997) instead proposed eliminating the fixed effects using

$$q_{it}(\theta) = \frac{y_{i,t-1}}{\lambda_{i,t-1}} - \frac{y_{it}}{\lambda_{it}}. \tag{35}$$

For either specification of $q_{it}(\theta)$ it can be shown that, given assumption (33),

$$\mathrm{E}[q_{it}(\theta)|\mathbf{z}_{it}] = 0, \tag{36}$$

where $\mathbf{z}_{it}$ can be drawn from $\mathbf{x}_{i,t-1}, \mathbf{x}_{i,t-2}, ...$ and, if lags up to $y_{i,t-p}$ appear as regressors, $\mathbf{z}_{it}$ can also be drawn from $\mathbf{y}_{i,t-p-1}, \mathbf{y}_{i,t-p-2}, ...$ Often $p = 1$, so $y_{i,t-2}, y_{i,t-3}, ...$ are available as instruments.

In the just-identified case in which there are as many instruments as parameters, the method of moments estimator solves

$$\sum_{i=1}^{n}\sum_{t=1}^{T} \mathbf{z}_{it} q_{it}(\theta) = \mathbf{0}. \tag{37}$$

In general there are more instruments $\mathbf{z}_{it}$ than regressors and the GMM estimator of β minimizes

$$\left(\sum_{i=1}^{n}\sum_{t=1}^{T} \mathbf{z}_{it} q_{it}(\theta)\right)' \mathbf{W}_n \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \mathbf{z}_{it} q_{it}(\theta)\right). \tag{38}$$

Given two-step GMM estimation, model adequacy can be tested using an over-identifying restrictions test.

It is also important to test for serial correlation in $q_{it}(\theta)$, using $q_{it}(\widehat{\theta})$, as correct model specification requires that $\mathrm{Cor}[q_{it}(\theta), q_{is}(\theta)] = 0$ for $|t-s| > 1$. Blundell, Griffith and Windmeijer (1997) adapt serial correlation tests proposed by Arellano and Bond (1991) for the linear model. Crepon and Duguet (1997) and Brännäs and Johansson (1996) apply serial correlation tests in the GMM framework.

Windmeijer (2008) provides a broad survey of GMM methods for the Poisson panel model, including the current setting. Two-step GMM estimated coefficients and standard errors can be biased in finite samples. Windmeijer (2008) proposes an extension of the variance matrix estimate of Windmeijer (2005) to nonlinear models. In a Monte Carlo exercise with predetermined regressor he shows that this leads to improved finite sample inference, as does the Newey and Windmeijer (2009) method applied to the continuous updating estimator variant of GMM.

Blundell, Griffith, and Windmeijer (2002) proposed an alternative transformation, the mean-scaling transformation

$$q_{it}(\theta) = y_{it} - \frac{\overline{y}_{i0}}{\lambda_{i0}}\lambda_{it}, \tag{39}$$

where $\overline{y}_{i0}$ is the presample mean value of y_i and the instruments are $(\mathbf{x}_{it} - \mathbf{x}_{i0})$. This estimator is especially useful if data on the dependent variable are available farther back in time than data on the explanatory variables. The transformation leads to inconsistent estimation, but in a simulation this inconsistency is shown to be small and efficiency is considerably improved.

The GMM methods of this section can be adapted to estimate FE models with endogenous regressors. Suppose the conditional mean of y_{it} with exogenous regressors is

$$\mu_{it} = \alpha_i \exp(\mathbf{x}_{it}'\beta). \tag{40}$$

Due to endogeneity of regressor(s), however, $\mathrm{E}[y_{it} - \mu_{it}|\mathbf{x}_{it}] \neq 0$, so the standard Poisson FE estimator is inconsistent. Windmeijer (2000) shows that in the panel case, the individual-specific fixed effects α_i can only be eliminated if a multiplicative errors specification is assumed and if the Wooldridge transformation is used. Then nonlinear IV or GMM estimation is based on $q_{it}(\theta)$ defined in (35), where the instruments $\mathbf{z}_{it}$ satisfy $\mathrm{E}[(y_{it} - \mu_{it})/\mu_{it}|\mathbf{z}_{it}] = 0$ and $\mathbf{z}_{it}$ can be drawn from $\mathbf{x}_{i,t-2}, \mathbf{x}_{i,t-3}, ...$

5 Extensions

In this section we survey recent developments that extend to the panel setting complications for counts that were introduced briefly in section 2.2 on cross-section data models. We consider panel versions of hurdle models, latent class models, and dynamic latent class models.

5.1 Hurdle Models

The panel count models covered in previous sections specify the same stochastic process for zero counts and for positive counts. Both the hurdle model and the zero-inflated model relax this restriction. Here we focus on panel versions of the hurdle or two-part model; similar issues arise for the zero-inflated model.

We specify a two-part data generating process. The split between zeros and positives is determined by a Bernoulli distribution with probabilities of, respectively, $f_1(0|\mathbf{z}_{it})$ and $1 - f_1(0|\mathbf{z}_{it})$. The distribution of positives is determined by a truncated-at-zero variant of the count distribution $f_2(y_{it}|\mathbf{x}_{it})$. Then

$$f(y_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}) = \begin{cases} f_1(0|\mathbf{z}_{it}) & \text{if } y_{it} = 0 \\ (1 - f_1(0|\mathbf{z}_{it}))\dfrac{f_2(y_{it}|\mathbf{x}_{it})}{1 - f_2(0|\mathbf{x}_{it})} & \text{if } y_{it} \geq 1, \end{cases} \tag{41}$$

which specializes to the standard model only if $f_1(0|\mathbf{z}_{it}) = f_2(0|\mathbf{x}_{it})$, and $\mathbf{z}_{it} = \mathbf{x}_{it}$. In principle, $\mathbf{z}_{it}$ and $\mathbf{x}_{it}$ may have distinct and/or overlapping elements, though in practice they are often the same. This model can handle both excess zeros in the count distribution $f_2(y_{it}|\mathbf{x}_{it})$, if $f_1(0) > f_2(0)$, and too few zeros if $f_2(0) > f_1(0)$.

This model is simply a pooled version of the standard cross-section hurdle model. Its implementation involves no new principles if the cross section assumptions are maintained, though cluster-robust standard errors analogous to those in (13) for pooled Poisson should be used. Because the two parts of the model are functionally independent, maximum likelihood estimation can be implemented by separately maximizing the two terms in the likelihood. A binary logit specification is usually used to model the positive outcome, and a Poisson or negative binomial specification is used for $f_2(y_{it}|\mathbf{x}_{it})$.

A random effects variant of this model introduces individual-specific effects, so

$$f(y_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}, \alpha_{1i}, \alpha_{2i}) = \begin{cases} f_1(y_{it}|\mathbf{z}_{it}, \alpha_{1i}) & \text{if } y_{it} = 0 \\ (1 - f_1(0|\mathbf{z}_{it}, \alpha_{1i}))\dfrac{f_2(y_{it}|\mathbf{x}_{it}, \alpha_{2i})}{1 - f_2(0|\mathbf{x}_{it}, \alpha_{2i})} & \text{if } y_{it} \geq 1, \end{cases} \quad (42)$$

where α_{1i} and α_{2i} are individual-specific effects for the first and second part of the model, respectively. Under the assumption of exogeneity of $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$, and given the bivariate density of $(\alpha_{1i}, \alpha_{2i})$ denoted by $h(\alpha_{1i}, \alpha_{2i})$, the marginal distribution of y_{it} is given by

$$\int\int f(y_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}, \alpha_{1i}, \alpha_{2i})h(\alpha_{1i}, \alpha_{2i})d\alpha_{1i}d\alpha_{2i}. \quad (43)$$

This calculation can be expected to be awkward to implement numerically. First, the likelihood no longer splits into two pieces that can be maximized individually. Second, it seems plausible that the individual-specific effects in the two distributions should not be independent. In some cases the assumption of a bivariate normal distribution is appropriate, perhaps after transformation such as for $\ln \alpha_{2i}$ rather than α_{2i}. Experience with even simpler problems of the same type suggest that more work is needed on the computational aspects of this problem; see Olsen and Schafer (2001).

Consistent estimation of a fixed effects variant of this model in a short panel is not possible. Conditional likelihood estimation is potentially feasible for some special choices of $f_1(\cdot)$, but a sufficient statistic for α_{2i} in a zero-truncated model is not available. Given T sufficiently large, individual-specific dummy variables may be added to the model. Then the profile likelihood approach (Dhaene and Jochmans, 2011) is potentially appealing, but there is no clear guidance from the literature. Yet another approach is to specify a conditionally correlated random effects model, introduced in (18).

5.2 Latent Class Models

Latent class models, or finite mixture models (FMM), have been used effectively in cross-section analysis of count data. They are generally appealing because they offer additional flexibility within a parametric framework. In this section we consider their extension to panel counts.

The key idea underlying latent class modeling is that an unknown distribution may be parsimoniously approximated by a mixture of parametric distributions with a finite and small number of mixture components. For example, a mixture of Poissons may be used to approximate an unknown distribution of event counts. Such models can provide an effective way of handling both excess zeros and overdispersion in count models (Deb and Trivedi, 2002).

The general expression for a panel latent class model in which all parameters are assumed to vary across latent classes is

$$f(y_{it}|\mathbf{x}_{it}, \theta_1, .., \theta_C, \pi_1, ..., \pi_{C-1}) = \sum_{j=1}^{C} \pi_j f_j(y_{it}|\mathbf{x}_{it}, \theta_j), \quad (44)$$

where $0 \leq \pi_j \leq 1$, $\pi_1 > \pi_2.... > \pi_C$, $\sum_j \pi_j = 1$, $\mathbf{x}_{it}$ is a vector of K exogenous variables, and θ_j denotes the vector of unknown parameters in the j^{th} component. For simplicity the component probabilities π_j in (44) are time-invariant and individual-invariant, an assumption that is relaxed below.

The estimation objective is to obtain consistent estimates of (π_j, θ_j), $j = 1, ..., C$, where C also should be determined from the data. For simplicity the analysis below concentrates on modeling issues that are specific to panel models, avoiding some general issues that arise in identification and estimation of all latent class models, and just a two-component mixture is considered, so we assume $C = 2$ is adequate.

We begin by considering a pooled panel latent class model. Introduce an unobserved variable d_{it} that equals j if individual i in period t is in the j^{th} latent class, and let

$$f(y_{it}|d_{it} = j, \mathbf{x}_{it}) = \mathsf{P}(\mu_{it}^{(j)}), \ j = 1, 2,$$

where $\mathsf{P}(\mu_{it}^{(j)})$ is the density of a Poisson distribution with mean $\mu_{it}^{(j)} = \exp(\mathbf{x}_{it}'\beta_j)$. For the case $C = 2$, d_{it} is a Bernoulli random variable. Let $\Pr[d_{it} = 1] = \pi$, for the moment constant over i and t. Then the joint density of (d_{it}, y_{it}) is

$$f(y_{it}, d_{it}|\mathbf{x}_{it}, \beta_1, \beta_2, \pi) = \left[\pi \mathsf{P}(\mu_{it}^{(1)})\right]^{\mathbf{1}[d_{it}=1]} \left[(1-\pi)\mathsf{P}(\mu_{it}^{(2)})\right]^{\mathbf{1}[d_{it}=2]}, \tag{45}$$

where $\mathbf{1}[A] = 1$ if event A occurs and equals 0 otherwise. The marginal density of y_{it} is

$$f(y_{it}|\mathbf{x}_{it}, \beta_1, \beta_2, \pi) = \pi \mathsf{P}(\mu_{it}^{(1)}) + (1-\pi)\mathsf{P}(\mu_{it}^{(2)}). \tag{46}$$

Let $\theta = (\beta_1, \beta_2)$. Under the assumption that the observations are independent across individuals and over time, the complete-data (joint) likelihood, conditioning on both y_{it} and d_{it} for all i and t, is

$$L^c(\theta, \pi) = \prod_{i=1}^{n}\prod_{t=1}^{T} \left[\pi \mathsf{P}(\mu_{it}^{(1)})\right]^{\mathbf{1}[d_{it}=1]} \left[(1-\pi)\mathsf{P}(\mu_{it}^{(2)})\right]^{\mathbf{1}[d_{it}=2]}, \tag{47}$$

and the marginal likelihood, conditioning on y_{it} and d_{it} for all i and t, is

$$L^m(\theta, \pi) = \prod_{i=1}^{n}\prod_{t=1}^{T} \left[\left\{\pi \mathsf{P}(\mu_{it}^{(1)}) + (1-\pi)\mathsf{P}(\mu_{it}^{(2)})\right\}\right]. \tag{48}$$

ML estimation may be based on an EM algorithm applied to (47) or, more directly, a gradient-based algorithm applied to (48). These expressions, especially the marginal likelihood, have been used to estimate a pooled panel model; see Bago d'Uva (2005). When following this pooled approach, the modeling issues involved are essentially the same as those for the cross-section latent class models.

In a panel with sufficiently long time series dimension there is some motivation for considering transitions between classes (states). One way to do so is to allow the mixture proportion π, assumed constant in the above exposition, to vary over time (and possibly individuals). This can be done by specifying π as a function of some time-varying regressors. Let $\Pr[d_{it} = 1|\mathbf{z}_{it}] = F(\mathbf{z}_{it}'\gamma)$ where F denotes a suitable c.d.f. such as logit or probit, and $\theta = (\gamma, \beta_1, \beta_2)$. Then the complete-data likelihood is

$$L^c(\theta, \gamma) = \prod_{i=1}^{n} \prod_{t=1}^{T} \left[F(\mathbf{z}_{it}'\gamma) \mathsf{P}(\mu_{it}^{(1)}) \right]^{\mathbf{1}[d_{it}=1]} \left[(1 - F(\mathbf{z}_{it}'\gamma)) \mathsf{P}(\mu_{it}^{(2)}) \right]^{\mathbf{1}[d_{it}=2]}. \tag{49}$$

This specification was used by Hyppolite and Trivedi (2012).

If, in the interests of parsimonious specification C is kept low, then some latent class components may still show substantial within-class heterogeneity. This provides motivation for adding individual-specific effects to improve the fit of the model. A multiplicative random effects model variant of (47), with individual-specific effects α_i and Poisson component means $\alpha_i \lambda_{it}^{(j)}$ where $\lambda_{it}^{(j)} = \exp(\mathbf{x}_{it}'\beta_j)$, has the following form:

$$L^c(\theta, \pi|\alpha_1, ..., \alpha_n) = \prod_{i=1}^{n} \prod_{t=1}^{T} \left[\pi \mathsf{P}(\alpha_i \lambda_{it}^{(1)}) \right]^{\mathbf{1}[d_{it}=1]} \left[(1 - \pi) \mathsf{P}(\alpha_i \lambda_{it}^{(2)}) \right]^{\mathbf{1}[d_{it}=2]}. \tag{50}$$

Assuming that the parametric distribution $g(\alpha_i v)$ for the individual-specific effects is the same for both latent classes, the individual-specific effects can be integrated out, analytically or numerically, yielding the likelihood function

$$L^c(\theta, \pi, \eta) = \prod_{i=1}^{n} \int \left\{ \prod_{t=1}^{T} \left[\pi \mathsf{P}(\alpha_i \lambda_{it}^{(1)}) \right]^{\mathbf{1}[d_{it}=1]} \left[(1 - \pi) \mathsf{P}(\alpha_i \lambda_{it}^{(2)}) \right]^{\mathbf{1}[d_{it}=2]} \right\} g(\alpha_i|\eta) d\alpha_i. \tag{51}$$

For a popular specification of $g(\alpha_i|\eta)$ such as the gamma, the integral will be a mixture of two negative binomial distributions (Deb and Trivedi, 2002); for the log-normal specification there is no closed form but a suitable numerical approximation can be used. Under the more flexible assumption that the two classes have different distributions for the α_i, likelihood estimation is potentially more complicated. More generally the slope coefficients may also vary across individuals; Greene and Hensher (2013) estimate a latent class model with random coefficients for cross-section multinomial data.

The fixed effects model is very popular in econometric studies as it allows the individual specific effects to be correlated with the regressors. Until recently, however, there has been no attempt to combine finite mixtures and fixed effects. In a recent paper, Deb and Trivedi (2013) take the first steps in this direction. They use the conditional likelihood approach to eliminate the incidental parameters α_i from the likelihood. The resulting likelihood is a complete-data form likelihood which is maximized using an EM algorithm.

For the one-component Poisson panel model with $y_{it} \sim \mathrm{P}(\alpha_i \lambda_{it})$ and $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta)$, the incidental parameters can be concentrated out of the likelihood using the first-order conditions with respect to α_i. Then $\widehat{\alpha}_i = \sum_t y_{it} / \sum_t \lambda_{it}$, leading to the following concentrated likelihood function, ignoring terms not involving β:

$$\ln L_{\text{conc}}(\beta) \propto \sum_{i=1}^{n} \sum_{t=1}^{T} \left[y_{it} \ln \lambda_{it} - y_{it} \ln \left(\sum\nolimits_{s=1}^{T} \lambda_{is} \right) \right]. \tag{52}$$

We wish to extend this conditional maximum likelihood approach to Poisson finite mixture models. The above conditioning approach will not work for the mixture of Poissons because in this case a sufficient statistic for the α_i is not available. However, the approach can work if the incidental parameters α_i are first concentrated out of the mixture components and the mixture is expressed in terms of the concentrated components. Denote by s_i the sufficient statistic for α_i. Then the mixture representation after conditioning on s_i is

$$f(y_{it}|\mathbf{x}_{it}, s_i, \theta_1, .., \theta_C, \pi_1, ..., \pi_{C-1}) = \sum_{j=1}^{C} \pi_j f(y_{it}|\mathbf{x}_{it}, s_i, \theta_j). \tag{53}$$

Specializing to the Poisson mixture, the complete-data concentrated likelihood is

$$L_{conc}(\theta, \pi_1, ..., \pi_C) = \prod_{i=1}^{n} \prod_{t=1}^{T} \sum_{j=1}^{C} \left[\pi_j \mathsf{P} \left(\frac{\sum_{s=1}^{T} y_{is}}{\sum_{s=1}^{T} \lambda_{is}^{(j)}} \times \lambda_{it}^{(j)} \right) \right]^{\mathbf{1}[d_{it}=j]}. \tag{54}$$

Because in this case the sufficient statistic $s_i = \sum_t y_{it} / \sum_t \lambda_{it}^{(j)}$ depends on model parameters and not just on data, the EM algorithm needs to be applied to the full-data likelihood. For a Monte Carlo evaluation and an empirical application, see Deb and Trivedi (2013).

5.3 Dynamic Latent Class Models

Dynamics can be introduced into the general model (44) in several ways. One way is to introduce dynamics into the component densities, such as using $f_j(y_{it}|\mathbf{x}_{it}, y_{i,t-1}, \theta_j)$. Böckenholt (1999) does this using a pooled version of the Poisson with conditional mean $\rho_j y_{i,t-1} + \exp(\mathbf{x}_{it}'\beta_j)$ for the j^{th} component.

Alternatively dynamics can be introduced through the latent class membership probabilities. Specifically, the class to which an individual belongs may evolve over time, depending on class membership in the previous period. If we assume that all unobserved past information useful in predicting class membership is contained in the most recent class membership, the process that determines the class of an individual can be characterized by a first-order Markov Chain.

First assume that the chains for each individual are characterized by the same time-homogeneous transition matrix and the same initial probability vector. Let p_{kl} denote the probability that an individual in state k switches to state l in the next period, where there are two possible states in a two-component model. Then

$$p_{kl} = \Pr[d_{it} = l | d_{i,t-1} = k], \quad k, l = 1, 2, \tag{55}$$

and, since $\sum_{l=1}^{2} p_{kl} = 1$, there are two free parameters, say p_{11} and p_{21}. The corresponding transition matrix is

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1 - p_{12} \\ p_{21} & 1 - p_{22} \end{bmatrix}. \tag{56}$$

We additionally need to specify the probabilities at initial time $t = 1$, and let

$$\pi = \Pr[d_{i1} = 1], \tag{57}$$

and $1 - \pi = \Pr[d_{i1} = 2]$. The bivariate discrete-time process (d_{it}, y_{it}), where d_{it} is an unobserved Markov chain and $y_{it}|(d_{it}, \mu_{it}^{(1)}, \mu_{it}^{(2)})$ is independent, is known as a hidden Markov model. The hidden Markov model is a mixture whose mixing distribution is a Markov chain. The joint density for (d_{it}, y_{it}) is given by

$$f(y_{it}, d_{it}|\mathbf{x}_{it}, \theta, \mathcal{I}_{i,t-1}) = \left(\Pr[d_{it} = 1|\mathcal{I}_{i,t-1}] \times \mathsf{P}(\mu_{it}^{(1)})\right)^{\mathbf{1}[d_{it}=1]} \left(\Pr[d_{it} = 2|\mathcal{I}_{i,t-1}] \times \mathsf{P}(\mu_{it}^{(2)})\right)^{\mathbf{1}[d_{it}=2]}, \tag{58}$$

where $\theta = (\beta_1, \beta_2, p_{11}, p_{12}, \pi)$ and $\mathcal{I}_{it}$ denotes information about individual i available up to time t. Because time dependence is modeled as a first-order Markov chain, the probability of being in a given state at a given point in time now depends on the previous history of the bivariate process. The difference between (58) and (45) is that in (58) the whole history of the process matters.

The corresponding marginal density is then

$$f(y_{it}|\mathbf{x}_{it}, \theta, \mathcal{I}_{i,t-1}) = \sum_{j=1}^{2} \Pr(d_{it} = j|\mathcal{I}_{i,t-1})\mathsf{P}(\mu_{it}^{(j)}). \tag{59}$$

Again the difference between (46) and (59) is that the history of the process enters the marginal density of the hidden Markov model.

The exact expression for the full-data likelihood depends upon whether or not the path of each individual chain is observed. For a detailed discussion of different assumptions, as well as estimation algorithms, see Hyppolite and Trivedi (2012).

The restriction that the transition matrix is time invariant can be relaxed. A more flexible model results if we parametrize the transition matrix. One such model specifies

$$\mathbf{P}_{it} = \left[\begin{array}{ll} F(\mathbf{z}_{it}'\gamma) & 1 - F(\mathbf{z}_{it}'\gamma) \\ F(\mathbf{z}_{it}'\gamma + \lambda) & 1 - F(\mathbf{z}_{it}'\gamma + \lambda) \end{array} \right],$$

where F is a suitable c.d.f. such as the logit and the probit. The complete-data likelihood and the marginal likelihood for this model are obtained the same way as for previous models. For details see Hyppolite and Trivedi (2012).

6 Conclusion

The methods of sections 2 to 4 are well established, and many of the methods have been integrated into the leading econometrics packages. Many econometrics textbooks provide discussion of count data models, though the treatment of panel counts is generally brief. The specialized monograph of Cameron and Trivedi (2013) provides a more comprehensive presentation. The richer parametric models of section 5 seek to model features of the data not well captured in some applications by the simpler panel models. These richer models are computationally more demanding, and eliminating fixed effects in these models is challenging.

7 References

Allison P.D., and R.P. Waterman (2002), "Fixed–Effects Negative Binomial Regression Models," *Sociological Methodology*, 32, 247-265.

Arellano, M., and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-298.

Bago d'Uva, T. (2005), "Latent Class Models for Use of Primary Care: Evidence from a British Panel," *Health Economics*, 14, 873–892.

Blundell, R., R. Griffith, and F. Windmeijer (1997), "Individual Effects and Dynamics in Count Data," manuscript, University College London.

Blundell, R., R. Griffith, and F. Windmeijer (2002), "Individual Effects and Dynamics in Count Data," *Journal of Econometrics*, 108, 113-131.

Böckenholt, U. (1999), "Mixed INAR(1) Poisson Regression Models: Analyzing Heterogeneity and Serial Dependencies in Longitudinal Count Data," *Journal of Econometrics*, 89, 317-338.

Brännäs, K. (1995), "Explanatory Variables in the AR(1) Model," Umea Economic Studies No. 381, University of Umea.

Brännäs, K., and P. Johansson (1996), "Panel Data Regression for Counts," *Statistical Papers*, 37, 191-213.

Cameron, A.C., and P.K. Trivedi (2013), *Regression Analysis of Count Data*, Second Edition, Econometric Society Monograph No. 53, Cambridge, Cambridge University Press.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5-46.

Chamberlain, G. (1992a), "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567-596.

Chamberlain, G. (1992b), "Comment: Sequential Moment Restrictions in Panel Data," *Journal of Business and Economic Statistics*, 10, 20-26.

Chib, S., E. Greenberg, and R. Winkelmann (1998), "Posterior Simulation and Bayes Factors in Panel Count Data Models," *Journal of Econometrics*, 86, 33-54.

Crepon, B., and E. Duguet (1997), "Estimating the Innovation Function from Patent Numbers: GMM on Count Data," *Journal of Applied Econometrics*, 12, 243-264.

Deb, P., and P.K. Trivedi (2002), "The Structure of Demand for Health Care: Latent Class versus Two-part Models," *Journal of Health Economics,* 21, 601-625.

Deb, P., and P.K. Trivedi (2013), "Finite Mixture for Panels with Fixed Effects," *Journal of Econometric Methods,* forthcoming.

Dhaene, G. and K. Jochmans (2011), " Profile-score Adjustments for Nonlinear Fixed-effect Models," Paper presented at the 17th International Panel Data Conference, Montreal.

Diggle, P.J., P. Heagerty, K.-Y. Liang, and S.L. Zeger (2002), *Analysis of Longitudinal Data*, Second Edition, Oxford, Oxford University Press.

Fernández-Val, I. (2009), "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models," *Journal of Econometrics*, 150, 71-85.

Greene, W.H. (2004), "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects," *Econometrics Journal*, 7, 98–119.

Greene, W.H., and D.A. Hensher (2013), "Revealing Additional Dimensions of Preference Heterogeneity in a Latent Class Mixed Multinomial Logit Model," Applied Economics, 45, 1897-1902.

Guimarães, P. (2008), "The Fixed Effects Negative Binomial Model Revisited," *Economics Letters*, 99, 63-66.

Hausman, J.A., B.H. Hall, and Z. Griliches (1984), "Econometric Models For Count Data With an Application to the Patents-R and D Relationship," *Econometrica*, 52, 909-938.

Heckman, J.J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge (MA), MIT Press.

Hyppolite, J., and P.K. Trivedi (2012), "Alternative Approaches for Econometric Analysis of Panel Count Data using Dynamic Latent Class Models (with Application to Doctor Visits Data)," *Health Economics*, 21, S101-128.

Lancaster, T. (1997), "Orthogonal Parameters and Panel Data," Working Paper No. 1997-32, Department of Economics, Brown University.

Liang, K.-Y., and S.L. Zeger (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13-22.

Liang K.-Y., S.L. Zeger, and B. Qaqish (1992), "Multivariate Regression Analyses for Categorical Data," *Journal of the Royal Statistical Society (B)*, 54, 3-40.

Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69-85.

Newey, W.K., and F. Windmeijer (2009), "GMM with Many Weak Moment Conditions," *Econometrica*, 77, 687-719.

Olsen, M.K., and J.L. Schafer (2001), "The Two-Part Random Effects Model for Semicontinuous Longitudinal Data," *Journal of American Statistical Association*, 96, 730-745.

Palmgren, J. (1981), "The Fisher Information Matrix for Log-Linear Models Arguing Conditionally in the Observed Explanatory Variables," *Biometrika*, 68, 563-566.

Skrondal, A., and S. Rabe-Hesketh (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, London, Chapman and Hall.

Stewart, M. (2007), "The Interrelated Dynamics of Unemployment and Low-wage Employment," *Journal of Applied Econometrics*, 22, 511–531.

Windmeijer F. (2000), "Moment Conditions for Fixed Effects Count Data Models with Endogenous Regressors," *Economics Letters*, 68, 21–24.

Windmeijer F. (2005), "A Finite Sample Correction for the Variance of Linear Efficient Two-step GMM Estimators, *Journal of Econometrics*, 126, 25–517.

Windmeijer, F. (2008), "GMM for Panel Count Data Models," *Advanced Studies in Theoretical and Applied Econometrics*, 46, 603-624.

Wooldridge, J.M. (1990), "Distribution-Free Estimation of Some Nonlinear Panel Data Models," Working Paper No. 564, Department of Economics, Massachusetts Institute of Technology.

Wooldridge, J.M. (1997), "Multiplicative Panel Data Models without the Strict Exogeneity Assumption," *Econometric Theory*, 13, 667-679.

Wooldridge, J.M. (2005), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, 20, 39-54.

Zeger, S.L., and K.-Y. Liang (1986), "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42, 121-130.