Regression Analysis of Count Data

A. Colin Cameron

Pravin K. Trivedi

CAMBRIDGE
UNIVERSITY PRESS

CHAPTER 9

Longitudinal Data

9.1 Introduction

Longitudinal data or *panel data* are observations on a cross-section of individual units such as persons, households, firms, and regions that are observed over several time periods. The data structure is similar to that of multivariate data considered in Chapter 8. Analysis is simpler than for multivariate data because for each individual unit the same outcome variable is observed, rather than several different outcome variables. Analysis is more complex because this same outcome variable is observed at different points in time, introducing time series data considerations presented in Chapter 7.

In this chapter we consider longitudinal data analysis if the dependent variable is a count variable. Remarkably, many count regression applications are to longitudinal data rather than simpler cross-section data. Econometrics examples include the number of patents awarded to each of many individual firms over several years, the number of accidents in each of several regions, and the number of days of absence for each of many persons over several years. A political science example is the number of protests in each of several different countries over many years. A biological and health science example is the number of occurrences of a specific health event, such as seizure, for each of many patients in each of several time periods.

A key advantage of longitudinal data over cross-section data is that they permit more general types of individual heterogeneity. Excellent motivation was provided by Neyman (1965), who pointed out that panel data enable one to control for heterogeneity and thereby distinguish between true and apparent contagion. For example, consider estimating the impact of research and development expenditures on the number of patent applications by a firm, controlling for individual firm-specific propensity to patent. For a single cross-section these controls can only depend on observed firm-specific attributes such as industry, and estimates may be inconsistent if there is additionally an unobserved component to individual firm-specific propensity to patent. With longitudinal data one can additionally include a firm-specific term for unobserved firm-specific propensity to patent.

The simplest longitudinal count data regression models are standard count models, with the addition of an individual specific term reflecting individual heterogeneity. In a *fixed effects* model this is a separate parameter for each individual. Creative estimation methods are needed if there are many individuals and hence parameters in the sample. In a *random effects* model this individual specific term is instead drawn from a specified distribution. Then creativity is required either in choosing a distribution that leads to tractable analytical results or in obtaining estimates if results are not tractable.

Asymptotic theory requires that the number of observations, here the number of individual units times the number of time periods, goes to infinity. We focus on the most common case of a *short panel*, in which only a few time periods are observed and the number of cross-sectional units goes to infinity. We also consider briefly the case in which the number of cross-sectional units is small but is observed for a large number of periods, as can be the case for cross-country studies. Then the earlier discussion for handling individual specific terms is mirrored in a similar discussion for time-specific effects. It is important to realize that the distribution of estimators, and which estimators are preferred, varies according to the type of sampling scheme.

In longitudinal data analysis the data are assumed to be independent over individual units for a given year but are permitted to be correlated over time for a given individual unit. In the simplest models this correlation over time is assumed to be adequately controlled for by individual-specific effects. In more general models *correlation* over time is additionally introduced in ways similar to those used in time series analysis. Finally, as in time series models, one can consider *dynamic models* or *transition models* that add a dynamic component to the regression function, allowing the dependent variable this year to depend on its own value in previous years.

A review of the standard linear models for longitudinal data, with fixed effects and random effects, is given in section 9.2, along with a statement of the analogous models for count data. In section 9.3 fixed effects models for count data are presented, along with application to data on the number of patents awarded to each of 346 firms in each of the years 1975 through 1979. Random effects models are studied in section 9.4. In sections 9.3 and 9.4 both MLEs and moment-based estimators are detailed. A discussion of applications and of the relative merits of fixed effects and random effects approaches is given in section 9.5. Model specification tests are presented in section 9.6. Dynamic models, in which the regressors include lagged dependent variables, are studied in section 9.7.

9.2 Models for Longitudinal Data

In this chapter we consider almost exclusively models that include fixed or random individual-specific effects. Even simpler models ignore such effects, assuming that the variation in regressors across individuals is sufficient to capture the differences in the dependent variable across individuals. We give little attention to these simpler models, as they do not exploit the advantage of longitudinal data over cross-section data.

9.2.1 *Linear Models*

Standard references for linear models for longitudinal data include Hsiao (1986), Diggle, Liang, and Zeger (1994), and Baltagi (1995). We give a brief review.

A quite general linear model for longitudinal data is

$$y_{it} = \alpha_{it} + \mathbf{x}'_{it}\beta_{it} + u_{it}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{9.1}$$

where y_{it} is a scalar dependent variable, $\mathbf{x}_{it}$ is a $k \times 1$ vector of independent variables and u_{it} is a scalar disturbance term. The subscript i indexes an individual person, firm, or country in a cross-section, and the subscript t indexes time. The distinguishing feature of longitudinal data models is that the intercept α_{it} and regressor coefficients β_{it} may differ across individuals or time. Such variation in coefficients reflects individual and time-specific effects. But the model (9.1) is too general and is not estimable. Further restrictions need to be placed on the extent to which α_{it} and β_{it} vary with i and t, and on the behavior of the error u_{it}.

The simplest linear model is the *one-way individual-specific effect* model

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\beta + u_{it}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{9.2}$$

where u_{it} is iid with mean 0 and variance σ_u^2. This is the standard linear regression model, except that rather than one intercept α there are n individual specific intercepts $\alpha_1, \ldots, \alpha_n$. The two standard models based on (9.2) are the *fixed effects linear* model, which treats α_i as a parameter to be estimated and excludes an intercept from $\mathbf{x}_{it}$, and the *random effects linear* model, which treats α_i as an iid random variable with mean 0 and variance σ_α^2 and includes an intercept in $\mathbf{x}_{it}$.

For the fixed effects linear model the estimator of the slope coefficients is

$$\hat{\beta}_{\mathsf{LFE}} = \left[\sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\right]^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i), \tag{9.3}$$

where $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}$ and $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ are individual-specific averages over time. The individual-specific fixed effects can be estimated by $\hat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}'_i\hat{\beta}_{\mathsf{LFE}}$. For a short panel, that is, $n \to \infty$ and T is fixed, $\hat{\beta}_{\mathsf{FE}}$ is consistent for β, while $\hat{\alpha}_i$ is not consistent for α_i as only T observations are used in estimating each α_i.

The linear fixed effects estimator $\hat{\beta}_{\mathsf{LFE}}$ can be motivated in several ways. First, joint estimation of α and β in (9.2) by OLS yields (9.3) for β. Second, if ε_{it} is assumed to be normally distributed, then (9.3) is obtained by maximizing with respect to β the conditional likelihood function given $\sum_{t=1}^{T} y_{it}$,

$i = 1, \ldots, n$, where $\sum_{t=1}^{T} y_{it}$ can be shown to be the sufficient statistic for α_i. Third, (9.2) implies

$$(y_{it} - \bar{y}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\beta + (u_{it} - \bar{u}_i), \tag{9.4}$$

meaning that differencing around the mean eliminates α_i. The GLS estimator of this equation can be shown to be OLS, and OLS of $(y_{it} - \bar{x}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ yields (9.3). Using this interpretation $\hat{\beta}_{\mathsf{LFE}}$ is called the *within estimator* as it explains variation in y_{it} around $\bar{y}_i$ by variation in $\mathbf{x}_{it}$ around $\bar{\mathbf{x}}_i$ – only variation within each individual is used.

For the random effects linear model the estimator of the slope coefficient estimator $\hat{\beta}_{\mathsf{LRE}}$ can be shown to be a matrix-weighted average of $\hat{\beta}_{\mathsf{LFE}}$, defined in (9.3), and $\hat{\beta}_{\mathsf{LB}}$ obtained from the OLS regression

$$(\bar{y}_i - \bar{y}) = (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'\beta + (\bar{u}_i - \bar{u}),$$

where $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \bar{\mathbf{x}}_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_i$. The estimator $\hat{\beta}_{\mathsf{LB}}$ is called the *between estimator* as it uses only variation between individuals, essentially ignoring the additional information available in a panel compared with a single cross-section. The weights used to form $\hat{\beta}_{\mathsf{LRE}}$ from $\hat{\beta}_{\mathsf{LFE}}$ and $\hat{\beta}_{\mathsf{LB}}$ depend on the variances σ_u^2 and σ_α^2. See basic treatments of this model.

The random effects estimator can be obtained in several ways. First, given the assumptions on the means and variances of u_{it} and α_i, it is the GLS estimator from estimation of (9.2). Second, it is asymptotically equivalent to the MLE, which additionally assumes that u_{it} and α_i are normally distributed. The MLE in practice leads in small samples to different estimates of the variances σ_u^2 and σ_α^2, and hence a different estimator of β, and is more difficult to estimate, as the log-likelihood is nonlinear in the parameters. A third method, which provides better small-sample estimates of σ_u^2 and σ_α^2, is the restricted MLE method of Patterson and Thompson (1971) and Harville (1977), reviewed in Diggle, Liang, and Zeger (1994, pp. 65–68).

The random and fixed effects linear models are compared by, for example, Hsiao (1986, pp. 41–47). The models are conceptually different, with the fixed effects analysis being conditional on the effects for individuals in the sample; random effects is an unconditional or marginal analysis with respect to the population. A major practical difference is that the fixed effects analysis provides only estimates of time-varying regressors. Thus, for example, it does not allow estimation of an indicator variable for whether or not a patient in a clinical trial was taking the drug under investigation (rather than a placebo). Another major difference is that the random effects model assumption that individual effects are iid implies that individual effects are uncorrelated with the regressors. If, instead, unobserved individual effects are correlated with observed effects, the random effects estimator is inconsistent. Many econometrics studies in particular prefer fixed effects estimators because of this potential problem.

Standard extensions to the linear model (9.2) are serial correlation in the error, for example, $u_{it} = \rho u_{it-1} + \varepsilon_{it}$; dynamic models, for example, $\mathbf{x}_{it}$ including

$y_{i,t-1}$; and more general random effects models with random slope coefficients in addition to random intercepts, for example, $\mathbf{x}_{it}'\beta$ is replaced by $\mathbf{x}_{1it}'\beta_1 + \mathbf{x}_{2it}'\beta_{2i}$ where β_{2i} is iid with mean β_2 and variance Σ_{β_2}.

9.2.2 *Count Models*

For count models for longitudinal data, the starting point is the Poisson regression model with exponential mean function and multiplicative individual specific term

$$\begin{aligned} y_{it} &\sim \mathsf{P}[\mu_{it} = \alpha_i \lambda_{it}] \\ \lambda_{it} &= \exp(\mathbf{x}_{it}'\beta), \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T. \end{aligned} \tag{9.5}$$

Note that α used here refers to the individual effect and is not used in the same way as in previous chapters, where it was an overdispersion parameter.

In the fixed effects model the α_i are unknown parameters. Like the linear model, estimation is possible by eliminating α_i, either by conditioning on $\sum_{t=1}^{T} y_{it}$, which requires fully parametric assumptions, or by using a quasi-differencing procedure that requires only first-moment assumptions.

In the random effects model the α_i are instead iid random variables. As in the linear model, estimation is possible either by assuming a distribution for α_i or by making second-moment assumptions, although unlike in the linear model under normality these can lead to quite different estimators.

A key departure from the linear model is that the individual specific effects in (9.5) are multiplicative, rather than additive as in the linear model (9.2). Given the exponential form for λ_{it}, the multiplicative effects can still be interpreted as a shift in the intercept because

$$\begin{aligned} \mathsf{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i] &= \mu_{it} \\ &= \alpha_i \exp(\mathbf{x}_{it}'\beta) \\ &= \exp(\delta_i + \mathbf{x}_{it}'\beta), \end{aligned} \tag{9.6}$$

where $\delta_i = \ln \alpha_i$.

Note that this equality between multiplicative effects and intercept shift does not hold in some count data models, nor does it hold in noncount models such as binary models to which similar longitudinal methods might be applied. Suppose the starting point is a more general conditional mean function $g(\mathbf{x}_{it}'\beta)$. Then some models and estimation methods continue with multiplicative effects, so

$$\mathsf{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i] = \mu_{it} = \alpha_i g(\mathbf{x}_{it}'\beta), \tag{9.7}$$

while other methods use a shift in the intercept

$$\mathsf{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i] = \mu_{it} = g(\delta_i + \mathbf{x}_{it}'\beta). \tag{9.8}$$

Results are most easily obtained for the Poisson. Extensions to the negative binomial do not always work, and when they do work they do so for some methods for the NB1 model and in other cases for the NB2 model. It should be kept in mind, however, that a common reason for such extensions in using cross-section data is to control for unobserved heterogeneity. The longitudinal data methods already control for heterogeneity, and Poisson longitudinal models may be sufficient.

The following sections begin with fixed effects and random effects models, with no consideration to either serial correlation and dynamics. In particular, for multiplicative effects models the regressors $\mathbf{x}_{it}$ are initially assumed to be *strictly exogenous*, so that

$$\mathsf{E}[y_{it} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \alpha_i] = \alpha_i \lambda_{it}. \tag{9.9}$$

This is a stronger condition than $\mathsf{E}[y_{it} \mid \mathbf{x}_{it}, \alpha_i] = \alpha_i \lambda_{it}$. This condition is relaxed when time series models are presented in section 9.7.

9.3 Fixed Effects Models

We consider three approaches to estimation of fixed effect count data models. First, we consider direct estimation by maximum likelihood, which may not necessarily lead to consistent estimates for the common case in which T is fixed and $n \to \infty$. Second, we present conditional maximum likelihood, which does analysis conditional on sufficient statistics for the individual effects. This works for NB1 models in addition to Poisson. Third, we consider a moment-based approach that bases estimation on a differencing transformation, which differs from that in the linear model, as here the effects are multiplicative, not additive.

9.3.1 *Maximum Likelihood*

The simplest fixed effects model for count data is the *Poisson fixed effects* model (9.5) where, conditional on λ_{it} and parameters α_i, y_{it} is iid $\mathsf{P}[\mu_{it} = \alpha_i \lambda_{it}]$, λ_{it} is a specified function of $\mathbf{x}_{it}$ and β, and $\mathbf{x}_{it}$ excludes an intercept. At times we specialize to the exponential form (9.6).

If n is small this model is easily estimated. In particular, the exponential mean specification (9.6) can be rewritten as $\exp(\sum_{j=1}^{n} \delta_j d_{jit} + \mathbf{x}_{it}'\beta)$, where d_{jit} is an indicator variable equal to one if the it^{th} observation is for individual j and zero otherwise. Thus we can use standard Poisson software to regress y_{it} on $d_{1it}, d_{2it}, \ldots, d_{nit}$ and $\mathbf{x}_{it}$.

This is impractical, however, if n is so large that $(n + \dim(\beta))$ exceeds software restrictions on the maximum number of regressors. In this chapter we focus on the case in which n is large and T is small, in which case this barrier is likely to be encountered. Then analytical expressions for estimators of β and the α_i are needed, analogous to those obtained for the linear model by partitioning of the OLS estimator.

A potentially more serious problem is possible inconsistency of parameter estimates if T is small and $n \to \infty$. This possibility arises because as $n \to \infty$ the number of parameters, $n + \dim(\beta)$, to be estimated goes to infinity, possibly negating the benefit of a larger sample size, nT. The individual fixed effects can be viewed as *incidental parameters*, because real interest lies in the slope coefficients. For some fixed effects panel data models, too many incidental parameters lead to inconsistent parameter estimates of β, in addition to α_i. A leading example is the logit model with fixed effects, with

$$\Pr[y_{it}=1]=\left[\alpha_i+\exp(\mathbf{x}_{it}'\beta)\right]/\left[1+\alpha_i+\exp(\mathbf{x}_{it}'\beta)\right].$$

Hsiao (1986, section 7.3.1) demonstrates the inconsistency of the MLE for β in this case, for fixed T and $n \to \infty$. This inconsistency disappears, of course, if $T \to \infty$. In the case of the linear model, however, there is no such incidental parameters problem.

An interesting question therefore is whether there is an incidental parameters problem for the Poisson fixed effects model. The literature has generally not directly addressed this issue, although it has suggested that there is a problem.* For y_{it} iid $\mathsf{P}[\alpha_i\lambda_{it}]$, the conditional joint density for the i^{th} observation is

$$\begin{aligned}
\Pr[y_{i1},\ldots,y_{iT}\mid\alpha_i,\beta]
&=\prod_t\left[\exp(-\alpha_i\lambda_{it})\,(\alpha_i\lambda_{it})^{y_{it}}/y_{it}!\right]\\
&=\exp\left(-\alpha_i\sum_t\lambda_{it}\right)\prod_t\alpha_i^{y_{it}}\prod_t\lambda_{it}^{y_{it}}\Big/\prod_t y_{it}!.
\end{aligned}\tag{9.10}$$

The corresponding log-density is

$$\begin{aligned}
\ln\Pr[y_{i1},\ldots,y_{iT}\mid\alpha_i,\beta]=-\alpha_i\sum_t\lambda_{it}+\ln\alpha_i\sum_t y_{it}\\
+\sum_t y_{it}\ln\lambda_{it}-\sum_t\ln y_{it}!.
\end{aligned}$$

Differentiating with respect to α_i and setting to zero yields

$$\hat{\alpha}_i=\frac{\sum_t y_{it}}{\sum_t\lambda_{it}}.\tag{9.11}$$

Substituting this back into (9.10), simplifying and considering all n observations yields the concentrated likelihood function,

$$\begin{aligned}
\mathsf{L}_{\text{conc}}(\beta)&=\prod_i\left[\exp\left(-\sum_t y_{it}\right)\prod_t\left(\frac{\sum_t y_{it}}{\sum_t\lambda_{it}}\right)^{y_{it}}\prod_t\lambda_{it}^{y_{it}}\Big/\prod_t y_{it}!\right]\\
&\propto\prod_i\left[\prod_t\left(\frac{\lambda_{it}}{\sum_s\lambda_{is}}\right)^{y_{it}}\right].
\end{aligned}\tag{9.12}$$

* We thank Frank Windmeijer and Tony Lancaster for pointing out that there is no incidental parameters problem here. The proof given here is due to Tony Lancaster.

This is the likelihood for n independent observations on a T-dimensional multinomial variable with cell probabilities

$$p_{it} = \frac{\lambda_{it}}{\sum_s \lambda_{is}} = \frac{\exp(\mathbf{x}_{it}'\beta)}{\sum_s \exp(\mathbf{x}_{is}'\beta)}.$$

It follows that for the Poisson fixed effects model there is no incidental parameters problem. Estimates of β that are consistent for fixed T and $n \to \infty$ can be obtained by maximization of $\ln \mathsf{L}_{\text{conc}}(\beta)$ in (9.12). The first-order conditions for this estimator are given in section 9.3.2, and its distribution is given in section 9.3.3 under much weaker conditions than those assumed here. Estimates of α_i can then be obtained from (9.11) and are consistent if in fact $T \to \infty$.

This consistency of the MLE for β despite the presence of incidental parameters is a special result that holds for the Poisson multiplicative fixed effects and, for continuous data, linear additive fixed effects. It holds in few other models, if any, in which case we need to transform the model into one in which the individual effects do not appear. The next two subsections present different ways to do this. We begin with the simplest case, the Poisson, even though as already noted there is no incidental parameters problem in this case.

9.3.2 *Conditional Maximum Likelihood*

The *conditional maximum likelihood* approach of Andersen (1970) performs inference conditional on the sufficient statistics for $\alpha_1, \ldots, \alpha_n$, which for LEF densities such as the Poisson are the individual-specific totals $T\bar{y}_i = \sum_{t=1}^{T} y_{it}$. In section 9.8.1 it is shown that for y_{it} iid $\mathsf{P}[\mu_{it}]$, the conditional joint density for the i^{th} observation is

$$\Pr\left[y_{i1}, \ldots, y_{iT} \,\middle|\, \sum_{t=1}^{T} y_{it}\right] = \frac{(\sum_t y_{it})!}{\prod_t y_{it}!} \times \prod_t \left(\frac{\mu_{it}}{\sum_s \mu_{is}}\right)^{y_{it}}. \tag{9.13}$$

This is a multinomial distribution, with probabilities $p_{it} = \mu_{it}/\sum_t \mu_{it}$, $t = 1, \ldots, T$, which has already been used in section 8.6.

Models with multiplicative effects set $\mu_{it} = \alpha_i \lambda_{it}$. This has the advantage that simplification occurs as α_i cancels in the ratio $\mu_{it}/\sum_s \mu_{is}$. Then (9.13) becomes

$$\Pr\left[y_{i1}, \ldots, y_{iT} \,\middle|\, \sum_{t=1}^{T} y_{it}\right] = \frac{(\sum_t y_{it})!}{\prod_t y_{it}!} \times \prod_t \left(\frac{\lambda_{it}}{\sum_s \lambda_{is}}\right)^{y_{it}}. \tag{9.14}$$

Because $y_{i1}, \ldots, y_{iT} \mid \sum_t y_{it}$ is multinomial distributed with probabilities $p_{i1}, \ldots, p_{iT}$, where $p_{it} = \lambda_{it}/\sum_s \lambda_{is}$, it follows that y_{it} has mean $p_{it}\sum_s y_{is}$. Given (9.6) this implies that we are essentially estimating the fixed effects α_i by $\sum_s y_{is}/\sum_s \lambda_{is}$.

In the special case $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta)$ this becomes

$$\Pr\left[y_{i1},\ldots,y_{iT}\,\middle|\,\sum_{t=1}^{T} y_{it}\right] = \frac{(\sum_t y_{it})!}{\prod_t y_{it}!} \times \prod_t \left(\frac{\exp(\mathbf{x}_{it}'\beta)}{\sum_s \exp(\mathbf{x}_{is}'\beta)}\right)^{y_{it}}. \tag{9.15}$$

The *conditional* MLE of the *Poisson fixed effects model* $\hat{\beta}_{\mathsf{PFE}}$ therefore maximizes the conditional log-likelihood function

$$\begin{aligned}\mathcal{L}_c(\beta) = \sum_{i=1}^{n}\Bigg[\ln\left(\sum_{t=1}^{T} y_{it}\right)! - \sum_{t=1}^{T}\ln(y_{it}!) \\ + \sum_{t=1}^{T} y_{it}\ln\left(\frac{\exp(\mathbf{x}_{it}'\beta)}{\sum_{s=1}^{T}\exp(\mathbf{x}_{is}'\beta)}\right)\Bigg].\end{aligned} \tag{9.16}$$

Note that this is proportional to the natural logarithm of $\mathsf{L}_{\mathrm{conc}}(\beta)$ given in (9.12), and therefore here the concentrated MLE equals the MLE.

Differentiation of (9.16), or equivalently (9.12), with respect to β yields first-order conditions for $\hat{\beta}_{\mathsf{PFE}}$ that can be reexpressed as

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it} - \lambda_{it}\frac{\bar{y}_i}{\bar{\lambda}_i}\right) = \mathbf{0}, \tag{9.17}$$

where $\bar{y}_i = \frac{1}{T}\sum_t y_{it}$ and $\bar{\lambda}_i = \frac{1}{T}\sum_t \lambda_{it}$ and $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta)$; see Blundell, Griffith, and Windmeijer (1995). The distribution of the resulting estimator can be obtained using standard maximum likelihood theory. In practice it is better to use results, given in the next subsection, obtained under weaker assumptions than y_{it} iid $\mathsf{P}[\alpha_i\lambda_{it}]$.

The log-likelihood function (9.16) is similar to that of the multinomial logit model, except that y_{it} is not restricted to taking only values zero or one and to sum over t to unity. Also the most standard form of a multinomial logit model with T outcomes has regressors fixed and parameters varying over the choices: $p_{it} = \exp(\mathbf{x}_i'\beta_t)/\sum_{s=1}^{T}\exp(\mathbf{x}_i'\beta_s)$. Here instead the parameters β are constant and the regressors $\mathbf{x}_{it}$ are time-varying.

The Poisson fixed effects model was proposed by Palmgren (1981) and Hausman, Hall, and Griliches (1984). The latter authors additionally presented a *negative binomial fixed effects* model. Then y_{it} is iid NB1 with parameters $\alpha_i\lambda_{it}$ and ϕ_i, where $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta)$, so y_{it} has mean $\alpha_i\lambda_{it}/\phi_i$ and variance $(\alpha_i\lambda_{it}/\phi_i)\times(1+\alpha_i/\phi_i)$.

This negative binomial model is of the less common NB1 form, with the variance a multiple of the mean. The parameter α_i is the individual-specific fixed effect; the parameter ϕ_i is the negative binomial overdispersion parameter, which is permitted to vary across individuals. Clearly α_i and ϕ_i can only be identified to the ratio α_i/ϕ_i, and even this ratio drops out for conditional maximum likelihood.

Some considerable algebra yields the conditional joint density for the i^{th} observation

$$\Pr\left[y_{i1},\ldots,y_{iT}\,\middle|\,\sum_{t=1}^{T}y_{it}\right]=\left(\prod_t\frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)}\right)\times\frac{\Gamma\left(\sum_t\lambda_{it}\right)\Gamma\left(\sum_t y_{it}+1\right)}{\Gamma\left(\sum_t\lambda_{it}+\sum_t y_{it}\right)},\qquad(9.18)$$

which involves β through λ_{it} but does not involve α_i and ϕ_i. This distribution for integer λ_{it} is the negative hypergeometric distribution. The log-likelihood function follows from this density and the MLE $\hat{\beta}_{\mathrm{NB1FE}}$ is obtained in the usual way.

McCullagh and Nelder (1989, section 7.2) consider the conditional maximum likelihood method in a quite general setting. Diggle, Liang, and Zeger (1994, section 9.2) specialize to GLMs with canonical link function (see section 2.4.4), in which case we again obtain the multinomial form (9.14). They also consider more general fixed effects in which the conditional mean function is of the form $g(\mathbf{x}_{it}'\beta+\mathbf{d}_{it}'\alpha_i)$ where $\mathbf{d}_{it}$ takes a finite number of values and α_i is now a vector. Hsiao (1986) specializes to binary models and finds that the conditional maximum likelihood approach is tractable for the logit model but not the probit model; that is, the method is tractable for individual intercepts if the canonical link function is used.

9.3.3 *Moment-Based Methods*

In the linear model (9.2) with additive fixed effects, there are several ways to transform the model to eliminate the fixed effects and hence obtain a consistent estimator of β. Examples are subtraction from y_{it} of the observation in another time period, say y_{i2}, or subtraction from y_{it} of the average over all time periods $\bar{y}_i$. The latter transformation, given in (9.4), yields the fixed effects estimator $\hat{\beta}_{\mathrm{LFE}}$.

Similarly in the Poisson model (9.5) with multiplicative effects, there are several ways to transform the model to eliminate the multiplicative effect. One example is subtraction from y_{it} of the observation in another time period, say y_{i2}, where y_{i2} is scaled to have the same mean as y_{it}. Thus we consider $(y_{it}-(\lambda_{it}/\lambda_{i2})y_{i2})$. Alternatively we could subtract the average over all time periods, appropriately rescaled, and consider $(y_{it}-(\lambda_{it}/\bar{\lambda}_i)\bar{y}_i)$. Then given (9.9) it follows that

$$\mathsf{E}[(y_{it}-(\lambda_{it}/\bar{\lambda}_i)\bar{y}_i)\,|\,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}]=0,$$

and hence

$$\mathsf{E}\left[\mathbf{x}_{it}\left(y_{it}-\frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right)\right]=\mathbf{0}.\qquad(9.19)$$

This suggests method of moments estimation of β by solving the corresponding sample moment conditions

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it}-\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\right)=\mathbf{0}. \tag{9.20}$$

These are the first-order conditions (9.17) of both the Poisson fixed effects conditional MLE $\hat{\beta}_{\text{PFE}}$ and the Poisson fixed effects MLE from section 9.3.1. Thus, the essential requirement for consistency of $\hat{\beta}_{\text{PFE}}$ is that (9.9) is the correct specification for the conditional mean. For example, $\hat{\beta}_{\text{PFE}}$ is also a consistent estimate of β in the negative binomial fixed effects model. Furthermore, the distribution of $\hat{\beta}_{\text{PFE}}$ can be obtained under weaker second-moment assumptions than variance–mean equality for y_{it}, or equivalently weaker than those imposed by the multinomial conditional distribution (9.15) for $y_{i1},\ldots,y_{iT}\mid\sum_t y_{it}$. The discussion is similar to that in section 3.2 for the cross-section Poisson model.

The first-order conditions (9.20) have first-derivative matrix with respect to β

$$\mathbf{A}_n=\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}'\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}-\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_{it}\mathbf{x}_{is}'\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\lambda_{is}\right], \tag{9.21}$$

for $\lambda_{it}=\exp(\mathbf{x}_{it}'\beta)$, while the outer product on taking expectations and eliminating cross-products in i and $j\neq i$ due to independence is

$$\mathbf{B}_n=\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{x}_{it}\mathbf{x}_{is}'\left(y_{it}-\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\right)\left(y_{is}-\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{is}\right)'. \tag{9.22}$$

Using the general result in section 2.5.1, an estimator of $\mathsf{V}[\hat{\beta}_{\text{PFE}}]$ that requires only first-moment assumptions, that is, the robust sandwich estimate, is

$$\mathsf{V}_{\text{RS}}[\hat{\beta}_{\text{PFE}}]=\hat{\mathbf{A}}_n^{-1}\hat{\mathbf{B}}_n\hat{\mathbf{A}}_n^{-1}, \tag{9.23}$$

where $\hat{\mathbf{A}}_n$ and $\hat{\mathbf{B}}_n$ are $\mathbf{A}_n$ and $\mathbf{B}_n$ evaluated at $\hat{\beta}_{\text{PFE}}$. By contrast, usual maximum likelihood estimates of the standard error are $\hat{\mathbf{A}}_n^{-1}$, using minus the inverse of the second derivatives of the log-likelihood function, or $\hat{\mathbf{B}}_n^{-1}$ using the BHHH estimate. These MLEs of $\mathsf{V}[\hat{\beta}_{\text{PFE}}]$ are inconsistent if the conditional variance does not equal the conditional mean $\alpha_i\mu_{it}$. If the conditional variance equals a constant γ times $\alpha_i\mu_{it}$, however, then a consistent estimate of $\mathsf{V}[\hat{\beta}_{\text{PFE}}]$ is γ times $\hat{\mathbf{A}}_n^{-1}$ or $\hat{\mathbf{B}}_n^{-1}$.

A quite general treatment of the distribution of the multinomial conditional MLE is given by Wooldridge (1990c), who considers a multiplicative fixed effect for general specifications of $\lambda_{it}=g(\mathbf{x}_{it}'\beta)$. In addition to giving robust variance matrix estimates, he gives more efficient GMM estimators if the conditional mean is specified to be of form $\alpha_i\lambda_{it}$ with other moments not specified, and when additionally the variance is specified to be of the form $\psi_i\alpha_i\lambda_{it}$. Chamberlain (1992a) gives semiparametric efficiency bounds for models using only specified first moment of form (9.6). Attainment of these bounds is theoretically

Table 9.1. *Patents: Poisson PMLE with NB1 standard errors*

	Poisson PMLE		Poisson fixed effects CMLE		NB1 fixed effects CMLE	
Variable	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
$\ln R_0$	.19	.16	.32	.07	.32	.07
$\ln R_{-1}$	−.07	.10	−.09	.10	−.08	.09
$\ln R_{-2}$	.07	.10	.08	.09	.06	.09
$\ln R_{-3}$	.06	.09	−.01	.08	−.01	.01
$\ln R_{-4}$	.16	.08	−.01	.08	.04	.07
$\ln R_{-5}$	.17	.12	−.03	.06	.01	.05
ln SIZE	.59	.07				
DSCI	.30	.13				
Sum ln R	.58		.32		.33	

Note: Poisson fixed effects conditional MLE with NB1 standard errors. NB1 fixed effects conditional MLE with MLE standard errors. All models include four time dummies for years 1976 to 1979.

possible but practically difficult, as it requires high-dimensional nonparametric regressions.

9.3.4 *Example: Patents*

Many longitudinal count-data studies, beginning with Hausman, Hall, and Griliches (1984), consider the relationship between past research and development (R&D) expenditures and the number of patents y_{it} awarded to the i^{th} firm in the t^{th} year, using data in a short panel. Here we consider data used by Hall, Griliches, and Hausman (1986) on 346 firms for 5 years' 1975 through 1979. Regression results are given in Table 9.1.

The Poisson PMLE estimates treat the data as one long cross-section, with y_{it} having conditional mean $\exp(\mathbf{x}_{it}'\beta)$. The reported standard errors are corrected for the considerable overdispersion in the data. The regressors of interest are $\ln R_0, \ldots, \ln R_{-5}$, the logarithm of current and up to 5 past years' research and development expenditures. Given the logarithmic transformation and the exponential conditional mean, the coefficient of $\ln R_{-j}$ is an elasticity, so that the coefficients of $\ln R_0, \ldots, \ln R_{-5}$ should sum to unity if a doubling of R&D expenditures leads to a doubling of patents. To control for firm-specific effects, the estimated model includes two time-invariant regressors, *SIZE*, the logarithm of firm book value in 1972 which is a measure of firm size, and *DSCI*, an indicator variable equal to one if the firm is in the science sector. If firm size doubles the number of patents increases by 59%.

The key empirical result for the Poisson PMLE estimates is that the coefficients of current and lagged R&D expenditures, $\ln R_{-j}$, sum to 0.58, which is considerably less than one, statistically so at conventional levels of significance. One possible explanation is that this is an artifact of failure to control adequately for firm-specific effects. However, the Poisson and NB1 fixed effects estimators, also given in Table 9.1, are even further away from one. (Estimated coefficients for ln *SIZE* and *DSCI* are not given, because the coefficients of time-invariant regressors are not identified in a fixed effects model.) These longitudinal estimators imply that in the long run a doubling of R&D expenditures leads to only a 33% increase in the number of patents. Qualitatively similar results have been found with other data sets and estimators, leading to a large literature on alternative estimators that may lead to results closer to a priori beliefs.

9.4 Random Effects Models

The simplest random effects model for count data is the *Poisson random effects* model. This model is given by (9.5), that is, y_{it} conditional on α_i and λ_{it} is iid Poisson ($\mu_{it} = \alpha_i \lambda_{it}$) and λ_{it} is a function of $\mathbf{x}_{it}$ and parameters $\boldsymbol{\beta}$. But in a departure from the fixed effects model, the α_i are iid random variables.

One approach is to specify the density $f(\alpha_i)$ of α_i and then integrate out α_i to obtain the joint density of $y_{i1}, \ldots, y_{iT}$ conditional on just $\lambda_{i1}, \ldots, \lambda_{iT}$. Then

$$\begin{aligned}\Pr[y_{i1}, \ldots, y_{iT}] &= \int_0^\infty \Pr[y_{i1}, \ldots, y_{iT} \mid \alpha_i] f(\alpha_i)\, d\alpha_i \\ &= \int_0^\infty \left[\prod_t \Pr[y_{it} \mid \alpha_i]\right] f(\alpha_i)\, d\alpha_i, \qquad (9.24)\end{aligned}$$

where for notational simplicity dependence on $\lambda_{i1}, \ldots, \lambda_{iT}$ is suppressed as in the fixed effects case. This integral appears similar to those in Chapter 4, except that here there is only one draw of α_i for the T random variables $y_{i1}, \ldots, y_{iT}$, so that this integral does not equal the product $\prod_t [\int_0^\infty \Pr[y_{it} \mid \alpha_i] f(\alpha_i)\, d\alpha_i]$ of mixtures considered in Chapter 4.

Different distributions for α_i lead to different distributions for $y_{i1}, \ldots, y_{iT}$. Analytical results can be obtained as they would be obtained in a similar Bayesian setting: by choosing $f(\alpha_i)$ to be conjugate to $\prod_t \Pr[y_{it} \mid \alpha_i]$. Conjugate densities exist for Poisson and NB2. In these standard count models the conjugate density is not the normal. Nonetheless there is considerable interest in results if $f(\alpha_i)$ is the normal density, because if results can be obtained for scalar α_i then they can be extended to random effects in slope coefficients. A number of methods have been proposed. Another solution if analytical results for the distribution are not available is to use moment methods if at least an analytical expression for the mean is available.

9.4.1 *Conjugate-Distributed Random Effects*

The gamma density is conjugate to the Poisson. In the pure cross-section case a Poisson–gamma mixture leads to the negative binomial; see section 4.2.2.

A similar result is obtained in the longitudinal setting. In section 9.8.2 it is shown that for y_{it} iid $\mathsf{P}[\alpha_i\lambda_{it}]$, where α_i is iid gamma(δ, δ) so that $\mathsf{E}[\alpha_i]=1$ and $\mathsf{V}[\alpha_i]=1/\delta$, integration with respect to α_i leads to the joint density for the i^{th} individual

$$\Pr[y_{i1},\ldots,y_{iT}] = \left[\prod_t \frac{\lambda_{it}{}^{y_{it}}}{y_{it}!}\right] \times \left(\frac{\delta}{\sum_t \lambda_{it}+\delta}\right)^{\delta} \times \left(\sum_t \lambda_{it}+\delta\right)^{-\sum_t y_{it}} \frac{\Gamma\left(\sum_t y_{it}+\delta\right)}{\Gamma(\delta)}. \tag{9.25}$$

This is the density of the Poisson random effects model (with gamma-distributed random effects). For this distribution $\mathsf{E}[y_{it}]=\lambda_{it}$ and $\mathsf{V}[y_{it}]=\lambda_{it}+\lambda_{it}^2/\delta$ so that overdispersion is of the NB2 form. Maximum likelihood estimation of $\boldsymbol{\beta}$ and δ is straightforward. For $\lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$, the first-order conditions for $\hat{\boldsymbol{\beta}}_{\mathsf{PRE}}$ can be expressed as

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it}-\lambda_{it}\frac{\bar{y}_i+\delta/T}{\bar{\lambda}_i+\delta/T}\right)=\mathbf{0}; \tag{9.26}$$

see exercise 9.3. Thus this estimator, like the Poisson fixed effects estimator, can be interpreted as being based on a transformation of y_{it} to eliminate the individual effects, and consistency essentially requires correct specification of the conditional mean of y_{it}. As for NB2 in the cross-section case the information matrix is block-diagonal and the first-order conditions for δ are quite complicated.

Hausman, Hall, and Griliches (1984) proposed this model and additionally considered the negative binomial case. Then y_{it} is iid NB2 with parameters $\alpha_i\lambda_{it}$ and ϕ_i, where $\lambda_{it}=\exp(\mathbf{x}_{it}'\boldsymbol{\beta})$, and hence y_{it} has mean $\alpha_i\lambda_{it}/\phi_i$ and variance $(\alpha_i\lambda_{it}/\phi_i)\times(1+\alpha_i/\phi_i)$. It is assumed that $(1+\alpha_i/\phi_i)^{-1}$ is a beta-distributed random variable with parameters (a, b).

Hausman, Hall, and Griliches show after considerable algebra that the *negative binomial random effects* model (with beta-distributed random effects) has joint density for the i^{th} individual

$$\Pr[y_{i1},\ldots,y_{iT}] = \left(\prod_t \frac{\Gamma(\lambda_{it}+y_{it})!}{\Gamma(\lambda_{it})!\Gamma(y_{it}+1)!}\right) \times \frac{\Gamma(a+b)\,\Gamma\left(a+\sum_t\lambda_{it}\right)\Gamma\left(b+\sum_t y_{it}\right)}{\Gamma(a)\,\Gamma(b)\,\Gamma\left(a+b+\sum_t\lambda_{it}+\sum_t y_{it}\right)}. \tag{9.27}$$

This is the basis for maximum likelihood estimation of $\boldsymbol{\beta}$, a, and b.

9.4.2 *Gaussian Random Effects*

An alternative random effects model is to allow the random effects to be normally distributed. In these models it is standard to assume an exponential mean function. Thus for the Poisson model the data y_{it} are assumed to be iid $\mathsf{P}[\exp(\delta_i + \mathbf{x}_{it}'\beta)]$, where the random effect δ_i is iid $\mathsf{N}[0, \sigma_\delta^2]$. From (9.6) this model can be rewritten as $y_{it} \sim \mathsf{P}[\alpha_i \exp(\mathbf{x}_{it}'\beta)]$, where $\alpha_i = \exp \delta_i$, and is therefore the preceding model where the random effects are log-normally distributed.

Unfortunately there is no analytical expression for the unconditional density (9.24) in this case. Development of estimation methods for such problems is an active area of research in generalized linear models. One solution (Schall, 1991; McGilchrist, 1994) is to linearize the model and use linear model techniques. An alternative is to directly compute the unconditional density by numerical integration or using simulation methods (Fahrmeir and Tutz, 1994, chapter 7). A recent example, using a Markov-chain Monte Carlo scheme to simulate, is Chib, Greenberg, and Winkelmann (1998). They apply their methods to epilepsy data from Diggle et al. (1994), patent data from Hall et al. (1986), and German work absence data.

9.4.3 *Moment-Based Methods*

In the linear random effects model (9.2) the OLS estimator from regression of y_{it} on $\mathbf{x}_{it}$ is still consistent. This is because if α_i is iid with zero mean the marginal mean of y_{it}, i.e., the mean conditional on $\mathbf{x}_{it}$ but marginal with respect to α_i, is $\mathbf{x}_{it}'\beta$. The OLS standard errors need to be corrected for the correlation induced by the random effects α_i, however, and it is more efficient to use the GLS estimator discussed in section 9.2.1.

Zeger and Liang (1986) carried this idea over to random effects in GLMs. Ideally, one would estimate by nonlinear feasible GLS, but this is not practical because unlike the linear case there is no simple analytical way to invert the covariance matrix of $\mathbf{y}$ conditional on $\mathbf{x}$. Instead, following a similar approach to that of Zeger (1988) for serially correlated error time series models presented in section 7.6, Zeger and Liang proposed estimation by nonlinear WLS, with corrections made to standard errors to ensure that they are consistently estimated.

For count data it is assumed that the marginal distribution of y_{it}, that is conditional on $\mathbf{x}_{it}$ but marginal on α_i, has first two moments

$$\begin{aligned} \mu_{it} = \mathsf{E}[y_{it} \mid \mathbf{x}_{it}] &= \exp(\mathbf{x}_{it}'\beta) \\ \sigma_{it}^2 = \mathsf{Var}[y_{it} \mid \mathbf{x}_{it}] &= \phi \exp(\mathbf{x}_{it}'\beta), \end{aligned} \tag{9.28}$$

where the multiplicative scalar ϕ implies that the random effects induce heteroskedasticity of NB1 form. The random effects additionally induce correlation between y_{it} and y_{is}, but this correlation is ignored. A consistent estimator for β is the generalized estimating equations estimator or nonlinear WLS estimator

$\hat{\beta}_{\mathsf{WLS}}$, with first-order conditions

$$\sum_{i=1}^{n} \mathbf{D}_i' \hat{\mathbf{V}}_i (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{9.29}$$

where $\mathbf{D}_i$ is the $T \times k$ matrix with tj^{th} element $\partial\mu_{it}/\partial\beta_j$, $\mathbf{V}_i$ is a $T \times T$ diagonal weighting matrix with t^{th} entry $[1/\sigma_{it}^2]$ or equivalently for this model $[1/\mu_{it}]$, $\mathbf{y}_i$ is the $T \times 1$ vector with t^{th} entry y_{it} and $\boldsymbol{\mu}_i$ is the $T \times 1$ vector with t^{th} entry μ_{it}. This is similar to the linear WLS estimator given in section 2.4.1 (see also section 7.6), and $\hat{\beta}_{\mathsf{WLS}}$ is asymptotically normal with mean β and variance

$$\begin{aligned} \mathsf{V}[\hat{\beta}_{\mathsf{WLS}}] = & \left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i \mathbf{D}_i\right]^{-1} \times \sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i \boldsymbol{\Omega}_i \mathbf{V}_i \mathbf{D}_i \\ & \times \left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i \mathbf{D}_i\right]^{-1}, \end{aligned} \tag{9.30}$$

where $\boldsymbol{\Omega}_i$ is the covariance matrix of $\mathbf{y}_i$. In practice $\boldsymbol{\Omega}_i$ is left unspecified and $\mathsf{V}[\hat{\beta}_{\mathsf{WLS}}]$ is consistently estimated by (9.30) with $\boldsymbol{\Omega}_i$ replaced by $(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'$.

Zeger and Liang (1986) and Liang and Zeger (1986) call this approach *marginal analysis*, as estimation is based on moments of the distribution marginal to the random effects. Zeger, Liang, and Albert (1988) consider mixed GLMs in which the random effects may interact with regressors. They call the approach *population-averaged*, as the random effects are averaged out, and contrast this with subject-specific models that explicitly model the individual effects. These papers present estimating equations of the form (9.29) with little explicit discussion of the random effects and the precise form of the correlation and modified heteroskedasticity that they induce. More formal treatment of random effects using this approach is given in Thall and Vail (1990).

Brännäs and Johansson (1996) consider a more tightly specified model than Zeger and Liang (1986), the Poisson model with multiplicative random effects in which the random effects are also time varying, that is, α_{it} replaces α_i in (9.6). They generalize (9.29) so that the weighting matrix has nonzero off-diagonal entries, reflecting correlation induced by the random effects. In addition they consider estimation by GMM, which exploits more of the moment conditions implied by the model. They apply these estimation methods to data on number of days absent from work for 895 Swedish workers in each of 11 years.

9.5 Discussion

The fixed effects model can be generalized from linear models to count data models. The conditional maximum likelihood approach leads to tractable results for some count models – for example, for the Poisson (9.13) simplifies to

(9.14) – but not for all count data models. Moment-based methods can more generally be used for all models with multiplicative individual effects as in (9.7).

The random effects model can also be used in a wide range of settings. The maximum likelihood approach is generally computationally difficult, unless a model with conjugate density for the random effects, such as Poisson-gamma, is used. Moment-based methods can again be used in a much wider range of settings.

The strengths and weaknesses of fixed effects versus random effects models in the linear case carry over to nonlinear models. For the linear model a considerable literature exists on the difference between fixed and random effects, see especially Mundlak (1978) and a summary by Hsiao (1986). The random effects model is appropriate if the sample is drawn from a population and one wants to do inference on the population; the fixed effects model is appropriate if one wishes to confine oneself to explaining the sample. The random effects model more easily accommodates random slope parameters as well as random intercepts. For the fixed effects model, coefficients of time-invariant regressors are absorbed into the individual-specific effect α_i and are not identified. For the random effects model, coefficient estimates may be inconsistent if the random effects are correlated with regressors. A test of correlation with regressors is presented in the next subsection.

We have focused on individual fixed effects in a short panel. Time-specific effects can additionally be included to form a two-way fixed effects error-component model. This can be estimated using conditional maximum likelihood as outlined previously, where the regressors $\mathbf{x}_{it}$ include time dummies. The results can clearly be modified to apply to a long panel with few individuals. Conditional maximum likelihood would then condition on $\sum_i y_{it}$ where, for example, y_{it} is iid $\mathsf{P}[\alpha_t \lambda_{it}]$.

In the linear model the total sample variability is split into between-group variability and within-group variability, where variability is measured by sums of squares. Hausman, Hall, and Griliches (1984) attempt a similar decomposition for count data models, where sample variability is measured by the log-likelihood function. For the Poisson model with gamma-distributed random effects, the log-likelihood of the iid $\mathsf{P}[\mu_{it}]$ can be decomposed as the sum of the conditional (on $\sum_t y_{it}$) log-likelihood and a marginal (for $\sum_t y_{it}$) log-likelihood. The conditional log-likelihood is naturally interpreted as measuring within variation; the marginal log-likelihood can be interpreted as between variation, although it depends on $\sum_t \lambda_{it}$, which depends on $\boldsymbol{\beta}$, rather than $\bar{\mathbf{x}}_i$ alone. A similar decomposition for negative binomial is not as neat.

References to count applications outside economics are given in Diggle, Liang, and Zeger (1994) and Fahrmeir and Tutz (1994). Many of these applications use random effects models with Gaussian effects or use the generalized estimating equations approach of Liang and Zeger (1986). Here we focus on economics applications, which generally use the fixed effects models or random effects models with conjugate density for the random effects.

The paper by Hausman, Hall, and Griliches (1984) includes a substantial application to number of patents for 121 U.S. firms observed from 1968 through 1975. This paper estimates Poisson and NB models with both fixed and random effects. Other studies using patent data are discussed in section 9.7.

Ruser (1991) studies the number of workdays lost at 2788 manufacturing establishments from 1979 through 1984. He uses the NB fixed effects estimator and finds that workdays lost increase with higher workers' compensation benefits, with most of the effect occurring in smaller establishments whose workers' compensation insurance premiums are less experience-rated.

Blonigen (1997) applies the NB2 random effects model to data on the number of Japanese acquisitions in the United States across 365 three-digit Standard Industry Classification industries from 1975 through 1992. The paper finds that if the U.S. dollar is weak relative to the Japanese yen, Japanese acquisitions increase in industries more likely to involve firm-specific assets, notably high R&D manufacturing sectors, which can generate a return in yen without involving a currency transaction.

In a novel application, Page (1995) applies the Poisson fixed effects model to data on the number of housing units shown by housing agents to each of two paired auditors, where the two auditors are as much as possible identical except that one auditor is from a minority group and the other is not. Specifically black/white pairs and Hispanic/Anglo pairs are considered. Here the subscript i refers to a specific auditor pair, $i = 1, \ldots, n$; subscript $t = 1, 2$ refers to whether the auditor is minority (say $t = 1$) or nonminority (say $t = 2$). A simple model without covariates is that $\mathsf{E}[y_{it}] = \alpha_i \exp(\beta d_{it})$, where $d_{it} = 1$ if minority and equals 0 otherwise. Then $\exp(\beta)$ equals the ratio of population-mean housing units shown to minority auditors to those shown to nonminority, and $\exp(\beta) < 1$ indicates discrimination is present. Page shows that in this case the Poisson fixed effects conditional MLE has explicit solution $\exp(\hat{\beta}) = \bar{y}_1/\bar{y}_2$. For the data studied by Page (1995) $\exp(\hat{\beta})$ lies between 0.82 and 0.91, with robust standard errors using (9.23) of between 0.022 and 0.028. Thus discrimination is present. Further analysis includes regressors that might explain the aggregate difference in number of housing units shown.

Van Duijn and Böckenholt (1995) analyze the number of spelling errors by 721 first-grade pupils on each of four dictation tests. They consider a Poisson–gamma mixture model that leads to a conditional multinomial distribution. This does not adequately model overdispersion, so they consider a finite mixtures version of this model using the methods of section 4.8. On the basis of chi-square goodness-of-fit tests they prefer a model with two classes, essentially good spellers and poor spellers.

Pinquet (1997) uses estimates of individual effects from longitudinal models of the number and severity of insurance claims to determine "bonus-malus" coefficients used in experience-rated insurance. In addition to an application to an unbalanced panel of over 100,000 policyholders, the paper gives considerable discussion of discrimination between true and apparent contagion. A range of models, including the random effects model of section 9.4, is considered.

9.6 Specification Tests

9.6.1 *Fixed Versus Random Effects*

The random effects estimator assumes that α_i is iid distributed, which in particular implies that the random effects are uncorrelated with the regressors. Thus it is assumed that individual specific unobservables are uncorrelated with individual specific observables, a strong assumption. The fixed effects model makes no such assumption – α_i could be determined by individual-specific time-invariant regressors.

If the random effects model is correctly specified, then both fixed- and random effects models are consistent, while if the random effects are correlated with regressors the random effects estimator loses its consistency. The difference between the two estimators can therefore be used as the basis for a Hausman test, introduced in section 5.6.6. This test is easily implemented because the random effects estimator is fully efficient, so the covariance matrix of the difference between estimators equals the difference in covariance matrices.

Thus form

$$\mathsf{T_H} = (\hat{\beta}_{\mathsf{RE}} - \tilde{\beta}_{\mathsf{FE}})'[\mathsf{V}[\tilde{\beta}_{\mathsf{FE}}] - \mathsf{V_{ML}}[\hat{\beta}_{\mathsf{RE}}]]^{-1}(\hat{\beta}_{\mathsf{RE}} - \tilde{\beta}_{\mathsf{FE}}). \tag{9.31}$$

If $\mathsf{T_H} < \chi^2_\alpha(\dim(\beta))$ then at significance level α we do not reject the null hypothesis that the individual specific effects are uncorrelated with regressors. This test is used in Hausman, Hall, and Griliches (1984, pp. 921 and 928) and leads to rejection of the random effects model in their application.

9.6.2 *Tests for Serial Correlation*

Tests for serial correlation are considered by Hausman, Hall, and Griliches (1984). If individual effects are present, then models that ignore such effects will have residuals that are serially correlated. If this serial correlation disappears after controlling for individual effects, then time series methods introduced in section 9.7 are not needed. We consider in turn tests for these two situations.

The natural model for initial analysis of count longitudinal data is Poisson regression of y_{it} on λ_{it} where independence is assumed over both i and t. Residuals from this regression are serially correlated if in fact individual effects α_i are present. Furthermore, the serial correlation between residuals from periods t and s is approximately constant in $(t - s)$, because it is induced by α_i, which is constant over time. It is natural to base tests on standardized residuals such as the Pearson residual $\varepsilon_{it} = (y_{it} - \lambda_{it})/\sqrt{\lambda_{it}}$. Then we expect the correlation coefficient between ε_{it} and ε_{is}, estimated as $\sum_i \varepsilon_{it}\varepsilon_{is}/\sqrt{\sum_i \varepsilon^2_{it}}\sqrt{\sum_i \varepsilon^2_{is}}$, to equal zero, $t \neq s$, if individual effects are not present. In practice these correlations are often sufficiently large that a formal test is unnecessary.

If models with individual effects are estimated, the methods yield consistent estimates of β but not α_i. Thus residuals $y_{it} - \alpha_i\lambda_{it}$ cannot be readily computed and tested for lack of serial correlation. For the fixed effects Poisson,

$y_{i1}, \ldots, y_{iT} \mid \sum_t y_{it}$ is multinomial-distributed with probability $p_{it} = \lambda_{it} / \sum_s \lambda_{is}$. It follows that y_{it} has mean $p_{it} \sum_s y_{is}$ and variance $p_{it}(1-p_{it}) \sum_s y_{is}$, and the covariance between y_{it} and y_{is} is $-p_{it} p_{is} \sum_s y_{it}$. The residual $u_{it} = (y_{it} - p_{it} \sum_s y_{is}) / \sqrt{\sum_s y_{is}}$ therefore satisfies $\mathsf{E}[u_{it}^2] = (1-p_{it})p_{it}$ and $\mathsf{E}[u_{it}u_{is}] = -p_{it}p_{is}$, $t \neq s$. Hausman, Hall, and Griliches (1984) propose a conditional moment test based on these moment conditions, where one of the residuals is dropped because predicted probabilities sum to one.

The dynamic longitudinal model applications discussed in section 9.7 generally implement tests of serial correlation. Blundell, Griffith, and Windmeijer (1995) adapt serial correlation tests proposed by Arellano and Bond (1991) for the linear model. Crepon and Duguet (1997a) and Brännäs and Johansson (1996) apply serial correlation tests in the GMM framework.

9.7 Dynamic and Transition Models

9.7.1 *Some Approaches*

Dynamic or transition longitudinal models allow current realizations of the count y_{it} to depend on past realizations $y_{i,t-k}$, $k > 0$, where $y_{i,t-k}$ defines individual i in period $t-k$.

One approach is to ignore the panel nature of the data. Simply 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 time series methods presented in Chapter 7, even for small T provided $n \to \infty$. This approach is given in Diggle, Liang, and Zeger (1994, chapter 10), who use autoregressive models that directly include $y_{i,t-k}$ as regressors. Also Brännäs (1995a) briefly discusses a generalization of the INAR(1) time series model to longitudinal data.

This approach may be adequate if there is considerable serial correlation in the data, because then lagged values of the dependent variable might be an excellent control for an individual effect. There may be no need to additionally include fixed or random effects. For example, firm-specific propensity to patent might be adequately controlled for simply by including patents last year as a regressor. A refinement is to consider a finite mixtures model with, say, two or three different types of firm, constant parameters for all firms of the same type, and firm type determined by the methods presented in Chapter 4.

Analysis becomes considerably more complicated if individual specific effects are introduced. In this case many of the preceding methods for panel count data are no longer appropriate, especially for short panels where $n \to \infty$ but T is fixed.

A similar complication arises for linear models, and is discussed for example in Nickell (1981), Hsiao (1986), and Baltagi (1995). In the simplest case of a fixed effects linear model with $y_{i,t-1}$ the only regressor, that is, $y_{it} = \beta y_{i,t-1} + u_{it}$, the differenced model (9.4) is

$$(y_{it} - \bar{y}_i) = \beta(y_{i,t-1} - \bar{y}_{i,-1}) + (u_{it} - \bar{u}_i), \qquad t = 2, \ldots, T,$$

where $\bar{y}_{i,-1} = \frac{1}{T-1}\sum_{t=2}^{T} y_{i,t-1}$. OLS estimation for finite T leads to an inconsistent estimate of β because the regressor $(y_{i,t-1} - \bar{y}_{i,-1})$ is correlated with $\bar{u}_i$; to see this, lag the above equation by one period – hence, the regressor is correlated with the error term.

For linear models, one solution is to restrict attention to the case $T \to \infty$. Then the problem disappears because $\bar{u}_i$ is then a small component of $u_{it} - \bar{u}_i$. A second solution, for finite T, is to use an alternative differenced model that subtracts the lagged value of y_{it}, so

$$(y_{it} - y_{i,t-1}) = \beta(y_{i,t-1} - y_{i,t-2}) + (u_{it} - u_{it-1}), \qquad t = 2, \ldots, T.$$

A consistent estimate of β can be obtained by instrumental variables methods, using for example $(y_{i,t-2} - y_{i,t-3})$ as an instrument. A considerable literature has developed on increasing the efficiency of such moment-based estimators. A third solution is to use MLEs of random effects models, in which case consistency depends crucially on assumptions regarding starting values.

For dynamic count models with individual-specific effects, qualitatively similar solutions to the above for linear models can be used.

An example of the first solution is Hill, Rothchild, and Cameron (1998), who model the monthly incidence of protests using data from 17 western countries for 35 years. To control for overdispersion and dynamics they use a negative binomial model with lagged y_{it} appearing as $\ln(y_{i,t-1} + c)$, where c is a constant whose role was explained in section 7.5. Country-specific effects are additionally controlled for by inclusion of country-specific indicator variables, which poses no consistency problems because in this example $T \to \infty$ while $n = 17$ is small.

In this section we concentrate on applying the second solution to dynamic count panel data models with fixed effects. Moment-based methods have already been presented for nondynamic models with multiplicative fixed effects in section 9.3.2. Here we present extension of these moment methods to the dynamic case. This is an active area of research, with most applications being to count data on patents.

9.7.2 *Fixed Effects Models*

The methods in preceding sections have implicitly assumed that regressors are strictly exogenous, that is,

$$\mathsf{E}[y_{it} \mid \mathbf{x}_{it}] = \mathsf{E}[y_{it} \mid \mathbf{x}_{iT}, \ldots, \mathbf{x}_{i1}] = \alpha_i \lambda_{it}. \tag{9.32}$$

This rules out cases in which regressors are *weakly exogenous*, or

$$\mathsf{E}[y_{it} \mid \mathbf{x}_{it}] = \mathsf{E}[y_{it} \mid \mathbf{x}_{it}, \ldots, \mathbf{x}_{i1}] = \alpha_i \lambda_{it}, \tag{9.33}$$

as in dynamic models in which lagged dependent variables appear as regressors. In this section we present results to estimate dynamic longitudinal data models using first-moment conditions.

We begin by considering the Poisson fixed effects estimator introduced in section 9.3. Given independence over i, the first-order conditions for $\boldsymbol{\beta}$ given in (9.17) have expected value zero if

$$\mathsf{E}\left[\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it}-\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\right)\right]=\mathbf{0}. \tag{9.34}$$

The presence of the average $\bar{y}_i$, introduced to eliminate the fixed effects, in these moment conditions limits application of this estimator to strictly exogenous regressors. To see this, consider the t^{th} term in the sum and assume $\mathsf{E}[y_{it}\,|\,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}]=\alpha_i\lambda_{it}$ as in (9.32). Then $\mathsf{E}[\bar{y}_i\,|\,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}]=\alpha_i\bar{\lambda}_i$ and

$$\begin{aligned}\mathsf{E}\left[\mathbf{x}_{it}\left(y_{it}-\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\right)\right]&=\mathsf{E}_{\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}}\left[\mathbf{x}_{it}\mathsf{E}\left[y_{it}-\frac{\bar{y}_i}{\bar{\lambda}_i}\lambda_{it}\,|\,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}\right]\right]\\&=\mathsf{E}_{\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}}\left[\mathbf{x}_{it}\left(\alpha_i\lambda_{it}-\frac{\alpha_i\bar{\lambda}_i}{\bar{\lambda}_i}\lambda_{it}\right)\right]\\&=\mathbf{0}.\end{aligned}$$

Note that it is not enough to assume $\mathsf{E}[y_{it}\,|\,\mathbf{x}_{it}]=\alpha_i\lambda_{it}$, because this does not necessarily imply $\mathsf{E}[\bar{y}_i\,|\,\mathbf{x}_{it}]=\alpha_i\bar{\lambda}_i$. For example, suppose $\mathbf{x}_{it}=y_{it-1}$ and $\mathsf{E}[y_{it}\,|\,y_{it-1}]=\alpha_i\rho y_{it-1}$, or $\lambda_{it}=\rho y_{it-1}$. Then

$$\mathsf{E}\left[\bar{y}_i\,|\,y_{it-1}\right]=\mathsf{E}\left[\frac{1}{T}(y_{i1}+\cdots+y_{iT})\,|\,y_{it-1}\right]$$

while

$$\alpha_i\bar{\lambda}_i=\frac{1}{T}\alpha_i\rho(y_{i0}+\cdots+y_{iT-1}).$$

Equality of the two requires $\mathsf{E}[y_{is}\,|\,y_{it-1}]=\alpha_i\rho y_{is-1}$ for $s\neq t$, which is clearly not the case. Similar problems arise if we assume $\mathsf{E}[y_{it}\,|\,\mathbf{x}_{it},\ldots,\mathbf{x}_{i1}]=\alpha_i\lambda_{it}$, because again this does not imply $\mathsf{E}[\bar{y}_i\,|\,\mathbf{x}_{it},\ldots,\mathbf{x}_{i1}]=\alpha_i\bar{\lambda}_i$.

One could instead eliminate fixed effects by quasidifferencing, as noted at the beginning of section 9.3.2. For weakly exogenous regressors, Chamberlain (1992b) proposes eliminating the fixed effects by the transformation

$$q_{it}=y_{it}-\frac{\lambda_{it}}{\lambda_{it+1}}y_{it+1}. \tag{9.35}$$

Suppose instruments $\mathbf{z}_{it}$ exist such that

$$\mathsf{E}[y_{it}-\alpha_i\lambda_{it}\,|\,\mathbf{z}_{it},\ldots,\mathbf{z}_{i1}]=0. \tag{9.36}$$

Then

$$\begin{aligned}&\mathsf{E}[y_{it+1}-\alpha_i\lambda_{it+1}\,|\,\mathbf{z}_{it},\ldots,\mathbf{z}_{i1}]\\&\quad=\mathsf{E}_{\mathbf{z}_{it+1}}[\mathsf{E}[y_{it+1}-\alpha_i\lambda_{it+1}\,|\,\mathbf{z}_{it+1},\mathbf{z}_{it},\ldots,\mathbf{z}_{i1}]]\\&\quad=\mathsf{E}_{\mathbf{z}_{it+1}}[0]\\&\quad=0.\end{aligned}$$

It follows that

$$\mathsf{E}\left[y_{it} - \frac{\lambda_{it}}{\lambda_{it+1}} y_{it+1} \mid \mathbf{z}_{it}, \ldots, \mathbf{z}_{i1}\right] = \alpha_i \lambda_{it} - \frac{\lambda_{it}}{\lambda_{it+1}} \alpha_i \lambda_{it+1}$$
$$= 0.$$

In the case in which there are as many instruments as parameters one solves

$$\sum_{t=1}^{T} \mathbf{z}_{it} \left(y_{it} - \frac{\lambda_{it}}{\lambda_{is}} y_{is} \right) = \mathbf{0}. \tag{9.37}$$

As an example, suppose

$$\mathsf{E}[y_{it} \mid y_{it-1}, \ldots, \mathbf{x}_{it}, \ldots] = \alpha_i \lambda_{it} = \alpha_i \left(\rho y_{it-1} + \exp\left(\mathbf{x}_{it}'\beta\right)\right).$$

Then the natural choice of instruments is $\mathbf{z}_{it} = (y_{it-1}, \mathbf{x}_{it})$.

If there are more instruments $\mathbf{z}_{it}$ than regressors, such as through adding additional lags of regressors into the instrument set, one can consistently estimate $\boldsymbol{\beta}$ by the GMM estimator, which minimizes

$$\frac{1}{n} \left(\sum_{i=1}^{n} \mathbf{q}_i(\beta)' \mathbf{z}_i \right) \mathbf{W}_n^{-1} \left(\sum_{i=1}^{n} \mathbf{z}_i' \mathbf{q}_i(\beta) \right), \tag{9.38}$$

where $\mathbf{q}_i(\beta) = (q_{i1} \cdots q_{iT})'$, $\mathbf{z}_i = \left(\mathbf{z}_{i1}' \cdots \mathbf{z}_{iT}'\right)'$, $\mathbf{W}_n^{-1}$ is a weighting matrix, and, given specification of $\mathbf{q}_i$ and $\mathbf{z}_i$, the optimal choice of $\mathbf{W}_n$ is $\mathbf{W}_n = \sum_{i=1}^{n} \mathbf{z}_i' \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_i' \mathbf{z}_i$ where $\tilde{\mathbf{q}}_i = \mathbf{q}_i(\tilde{\beta})$ and $\tilde{\beta}$ is an initial consistent estimate obtained for example by minimizing (9.38) with $\mathbf{W}_n = \mathbf{I}_n$.

An alternative transformation to eliminate the fixed effects is

$$q_{it} = \frac{y_{it}}{\lambda_{it}} - \frac{y_{it+1}}{\lambda_{it+1}}, \tag{9.39}$$

proposed by Wooldridge (1997), which is simply the earlier choice divided by λ_{it}. Yet another possibility is the mean scaling transformation

$$q_{it} = y_{it} - \frac{\bar{y}_{i0}}{\lambda_{i0}} \lambda_{it}, \tag{9.40}$$

proposed by Blundell, Griffith, and Windmeijer (1995), where $\bar{y}_{i0}$ is the presample mean value of y_i and the instruments are $(\mathbf{x}_{it} - \mathbf{x}_{i0})$. The latter estimator leads to estimates that are inconsistent, but in a simulation this inconsistency is shown to be small, and efficiency is considerably improved. This estimator is especially useful if data on the dependent variable go back farther in time than data on the explanatory variables.

These methods are applicable to quite general models with multiplicative fixed effects. Several studies, beginning with Montalvo (1997), have refined and applied these methods, mostly to count data on patents. Application to patents is of particular interest for several reasons. There are few ways to measure innovation aside from patents, which are intrinsically a count. R&D expenditures affect patents with a considerable lag, so there is potentially parsimony and elimination of multicollinearity in having patents depend on lagged patents rather than a long-distributed lag in R&D expenditures. And, as noted in the example earlier, most studies using distributed lags on R&D expenditure find the R&D expenditure elasticity of patents to be much less than unity.

Blundell, Griffith, and Windmeijer (1995) model the U.S. patents data of Hall, Griliches, and Hausman (1986). They pay particular attention to the functional form for dynamics and the time series implications of various functional forms. The lagged dependent variable is introduced in either multiplicative fashion as

$$\mu_{it} = \alpha_i \exp\left(\rho \ln y^*_{it-1} + \mathbf{x}'_{it}\beta\right)$$

where $y^*_{it-1} = y_{it-1}$ unless $y_{it-1} = 0$ in which case $y^*_{it-1} = c$, or additive fashion as

$$\mu_{it} = \rho y_{it-1} + \alpha_i \exp\left(\mathbf{x}'_{it}\beta\right),$$

where $\rho > 0$. Another variant of the additive model, not considered, is

$$\mu_{it} = \alpha_i\left(\rho y_{it-1} + \exp\left(\mathbf{x}'_{it}\beta\right)\right).$$

In their application up to two lags of patents and three lags of R&D expenditures appear as regressors. The estimates indicate long lags in the response of patents to R&D expenditures.

Related studies by Blundell, Griffith, and Van Reenen (1995a, b) model the number of "technologically significant and commercially important" innovations commercialized by British firms. Dynamics are introduced more simply by including the lagged value of the knowledge stock, an exponentially weighted sum of past innovations.

Montalvo (1997) uses the Chamberlain (1992b) transformation to model the number of licensing agreements by individual Japanese firms and the Hall et al. (1986) data. Lagged dependent variables do not appear as regressors. Instead Montalvo argues that current R&D expenditures cannot be assumed to be strictly exogenous because patents depend on additional R&D expenditures for their full development. So there is still a need for quasidifferenced estimators.

Crepon and Duguet (1997a) apply GMM methods to French patents data. They also use a relatively simple functional form for dynamics. First, as regressor they use a measure of R&D capital. This capital measure is calculated as the weighted sum of current and past depreciated R&D expenditure and can be viewed as imposing constraints on R&D coefficients in a distributed lag model. Dynamics in patents are introduced by including indicator variables for whether y_{it-1} is in the ranges 1 to 5, 6 to 10, or 11 or more. Particular attention is paid to model specification testing and the impact of increasing the size of the instrument set $\mathbf{z}_i$ in (9.38).

In a more applied study, Cincera (1997) includes not only a distributed lag in firm R&D expenditures but also a distributed lag in R&D expenditures by other firms in the same sector to capture spillover effects. Application is to a panel of 181 manufacturing firms from six countries.

9.8 Derivations

9.8.1 *Conditional Density for Poisson Fixed Effects*

Consider the conditional joint density for observations in all time periods for a given individual, where for simplicity the individual subscript i is dropped. In general the density of $y_1, \ldots, y_T$ given $\sum_t y_t$ is

$$\Pr\left[y_1, \ldots, y_T \,\Big|\, \sum_t y_t\right] = \Pr\left[y_1, \ldots, y_T, \sum_t y_t\right] \Big/ \Pr\left[\sum_t y_t\right]$$
$$= \Pr[y_1, \ldots, y_T] \Big/ \Pr\left[\sum_t y_t\right],$$

where the last equality arises because knowledge of $\sum_t y_t$ adds nothing given knowledge of $y_1, \ldots, y_T$.

Now specialize to y_t iid Poisson (μ_t). Then $\Pr[y_1, \ldots, y_T]$ is the product of T Poisson densities, and $\sum_t y_t$ is Poisson $(\sum_t \mu_t)$. It follows that

$$\Pr\left[y_1, \ldots, y_T \,\Big|\, \sum_t y_t\right] = \frac{\prod_t \left(\exp(-\mu_t)\mu_t^{y_t}/y_t!\right)}{\exp(-\sum_t \mu_t)(\sum_t \mu_t)^{\sum_t y_t}/(\sum_t y_t)!}$$
$$= \frac{\exp(-\sum_t \mu_t)\prod_t \mu_t^{y_t}/\prod_t y_t!}{\exp(-\sum_t \mu_t)\prod_t (\sum_s \mu_s)^{y_t}/(\sum_t y_t)!}$$
$$= \frac{(\sum_t y_t)!}{\prod_t y_t!} \times \prod_t \left(\frac{\mu_t}{\sum_s \mu_s}\right)^{y_t}.$$

Introducing the subscript i yields (9.13) for $\Pr[y_{i1}, \ldots, y_{iT} \mid \sum_t y_{it}]$.

9.8.2 *Density for Poisson with Gamma Random Effects*

Consider the joint density for observations in all time periods for a given individual, where for simplicity the individual subscript i is dropped. From (9.13) the joint density of $y_1, \ldots, y_T$ if $y_t \mid \alpha$ is $\mathsf{P}[\alpha\lambda_t]$ is

$$\Pr[y_1, \ldots, y_T] = \int_0^\infty \left[\prod_t \left(e^{-\alpha\lambda_t}(\alpha\lambda_t)^{y_t}/y_t!\right)\right] f(\alpha)\, d\alpha$$
$$= \int_0^\infty \left[\prod_t \lambda_t^{y_t}/y_t!\right]\left(e^{-\alpha\sum_t \lambda_t} \cdot \alpha^{\sum_t y_t}\right) f(\alpha)\, d\alpha$$
$$= \left[\prod_t \lambda_t^{y_t}/y_t!\right] \times \int_0^\infty \left(e^{-\alpha\sum_t \lambda_t} \cdot \alpha^{\sum_t y_t}\right) f(\alpha)\, d\alpha.$$

Now let $f(\alpha)$ be the gamma density with parameters density. Similar algebra to that in section 4.2.2 yields the Poisson random effects density given in (9.25).

9.9 Bibliographic Notes

Longitudinal data models fall in the class of multilevel models, surveyed by Goldstein (1995), who includes a brief treatment of Poisson. Standard references for linear models for longitudinal data include Hsiao (1986), Diggle, Liang, and Zeger (1994), and Baltagi (1995). Diggle et al. (1994) and Fahrmeir and Tutz (1994) consider generalized linear models in detail. A useful reference for general nonlinear longitudinal data models is Mátyás and Sevestre (1995).

There are remarkably many different approaches to nonlinear models, and many complications including serial correlation, dynamics and unbalanced panels. The treatment here is comprehensive for models used in econometrics and covers many of the approaches used in other areas of statistics. Additional statistical references can be found in Diggle et al. (1994) and Fahrmeir and Tutz (1994). Lawless (1995) considers both duration and count models for longitudinal data for recurrent events. For dynamic models the GMM fixed effects approach is particularly promising. In addition to the count references given in section 9.7, it is useful to refer to earlier work for the linear model by Arellano and Bond (1991) and Keane and Runkle (1992).

9.10 Exercises

9.1 Show that the Poisson fixed effects conditional MLE of β that maximizes the log-likelihood function given in (9.16) is the solution to the first-order conditions (9.17).

9.2 Find the first-order conditions for the negative binomial fixed effects conditional MLE of β that maximizes the log-likelihood function based on the density (9.18). (Hint: Use the gamma recursion as in section 3.3.) Do these first-order conditions have a simple interpretation, like those for the Poisson fixed effect conditional MLE?

9.3 Verify that the first-order conditions for the Poisson random effects MLE for β can be expressed as (9.26).

9.4 Show that the Poisson fixed effects conditional MLE that solves (9.17) reduces to $\exp(\hat{\beta}) = \bar{y}_1/\bar{y}_2$ in the application by Page (1995) discussed at the end of section 9.5.