# **15** Panel Data Models for Discrete Choice

# William Greene, Department of Economics, Stern School of Business, New York University

- I. Introduction
  - A. Analytical Frameworks for Panel Data Models for Discrete Choice
  - B. Panel Data
- II Discrete Outcome Models
- III Individual Heterogeneity
  - A. Random Effects
    - A.1. Partial Effects
    - A.2. Alternative Models for Random Effects
    - A.3. Specification Tests
    - A.4. Choice Models
  - B. Fixed Effects Models
  - C. Correlated Random Effects
  - D. Attrition
- IV Dynamic Models
- V. Spatial Correlation

# I. Introduction

We survey the intersection of two large areas of research in applied and theoretical econometrics. Panel data modeling broadly encompasses nearly all of modern microeconometrics and some of macroeconometrics as well. Discrete choice is the gateway to and usually the default framework in discussions of nonlinear models in econometrics. We will select a few specific topics of interest: essentially modeling cross sectional heterogeneity in the four foundational discrete choice settings: binary, ordered multinomial, unordered multinomial and count data frameworks.

We will examine some of the empirical models used in recent applications, mostly in parametric forms of panel data models. There are many discussions elsewhere in this volume that discuss discrete choice models. The development here can contribute a departure point to the more specialized treatments such as Keane's (2013, this volume) study of panel data discrete choice models of consumer demand or more theoretical discussions, such as Lee's (2013, this volume) extensive development of dynamic multinomial logit models. Toolkits for practical application of most of the models noted here are built into familiar modern software such as Stata, SAS, R, NLOGIT, MatLab, etc. We will not develop detailed descriptions of 'how to' for specific kinds of applications. Space considerations also preclude extended numerical applications.

Formal development of the discrete outcome models described above can be found in numerous sources, such as Greene (2012) and Cameron and Trivedi (2005). We will focus on extensions of the models to panel data applications. The common element of the discussions that necessitates a separate treatment is the nonlinearity of the models. Familiar treatments such as models of fixed and random effects and dynamic specifications in linear regression models provide only scant guidance in extensions to nonlinear models such as binary choice.

#### A. Analytical Frameworks for Panel Data Models for Discrete Choice

There are two basic threads of development of discrete choice models. *Random utility* based models emphasize the *choice* aspect of discrete choice. Discrete choices are the observable revelations of underlying preferences. For example, McFadden (1974) develops the random utility approach to multinomial qualitative choice. A second group of models is quantitative in nature – regression models for counts of events. For our purposes, it is useful to consider these as discrete *choices* as well. The fundamental building block is the binary choice model, which we associate with an agent's revelation of their preference for one specific outcome over another. Ordered and unordered choice models build on this basic platform. Regression models for counts of events fit into this study because of the style of model building typically used, which has much in common with the counterparts in the random utility framework. Though counts are not typically modeled as revelations of preferences, some analysts have

done so, including Schmidheiny and Brülhart's (2011) model of location choice and Bhat, Paleti and Castro's (2013) analysis of out-of-home non-work episodes. The familiar estimation platforms, univariate probit and logit, ordered choice (see Greene and Hensher (2010)), and multinomial logit for the former type and Poisson and negative binomial regressions for counts have been developed and extended in a vast literature. The extension of panel data models for heterogeneity and dynamic effects, that have been developed for linear regression in an equally vast literature, into these nonlinear settings is a bit narrower, and is the subject of this essay.

Panel data models, beginning with discussions of the linear regression model, are documented in almost fifty years of literature beginning with Balestra and Nerlove's (1966) canonical study of the U.S. natural gas market. Landmark treatments have built on this framework, including Arellano and Bond (1991) and Arellano and Bover (1995) and a generation of results on dynamic linear models. (Some of that research is continued elsewhere in this handbook.) The early extension of panel data methods to nonlinear models, specifically discrete choice models, is relatively more limited. The treatment of binary choice begins (superficially) with Rasch's (1960) and Chamberlain's (1980, 1984) development of a fixed effects binary choice model and, for practical applications, Butler and Moffitt's (1982) development of an algorithm for random effects choice models. We will focus largely on these models and modern extensions that have appeared in the recent literature.

# **B.** Panel Data

The second dimension of the treatment here is panel data modeling. The modern development of large, rich *longitudinal survey data sets* such as the German Socioeconomic Panel (GSOEP), Household Income and Labor Dynamics in Australia (HILDA), Survey of Income and Program Participation (SIPP, US), British Household Panel Survey (BHPS), Medical Expenditure Panel Survey (MEPS, US) and European Community Household Panel Survey (ECHP) to name a few, has supported an ongoing interest in analysis of individual outcomes across households and within households through time. The BHPS, for example, now in its 18<sup>th</sup> wave, is long enough to have recorded a significant fraction of the life cycle of many family members. The National Longitudinal Survey (NLS, US) was begun in the 1960s, and notes that for some purposes, they have entered their second generation. Each of these surveys includes questions on discrete outcomes such as labor force participation, banking behavior, self assessed health, subjective well being, health care decisions, insurance purchase, and many others. The discrete choice models already noted are the natural platforms for analyzing these variables. For present purposes, a specific treatment of 'panel data models' is motivated by interesting features of the population that can be studied in the context of longitudinal data, such as cross sectional heterogeneity and dynamics in behavior and on estimation methods that differ from cross section linear regression counterparts. We will narrow

our focus to individual data. The analysis of market level data on aggregates such as pioneered in Berry, Levinsohn and Pakes (1995) and Goldberg (1995), do belong in the class of discrete choice analyses – though usually not in discussions of panel data applications. Nonetheless, given our limited ambition and space constraints, we will confine attention to the sorts of discrete decisions analyzed using individual data. Contemporary applications include many examples in health economics: such as in Riphahn, Wambach and Million's (2003) study of insurance takeup and health care utilization using the GSOEP and Contoyannis, Rice and Jones's (2004) analysis of self assessed health in the BHPS.

# **II. Discrete Outcome Models**

We will denote the models of interest here as *discrete outcome models*. The data generating process takes two specific forms, *random utility models* and *nonlinear regression models* for counts of events. In some applications, there is a bit of fuzziness of the boundary between these. Bhat and Pulugurta (1998) treat the number of vehicles owned, naturally a count, as a revelation of preferences for transport services, i.e., in a utility based framework. For random utility, the departure point is the existence of an individual preference structure that implies a utility index defined over states, or alternatives, that is,

$$U_{it,j} = U(\mathbf{x}_{it,j}, \mathbf{z}_i, A_i, \varepsilon_{it,j}).$$

Preferences are assumed to obey the familiar axioms – completeness, transitivity, etc. – we take the underlying microeconomic theory as given. In the econometric specification, 'j' indexes the alternative, 'i' indexes the individual and 't' may index the particular choice situation in a set of  $T_i$  situations. In the cross section case,  $T_i = 1$ ... In panel data applications, the case  $T_i > 1$  will be of interest. The index 't' is intended to provide for possible sequence of choices, such as consecutive observations in a longitudinal data setting or a stated choice experiment. The number of alternatives, J, may vary across both i and t - consider a stated choice experiment over travel mode or consumer brand choices in which individuals choose from possibly different available choice sets as the experiment progresses through time. Analysis of brand choices for, e.g., ketchup, yogurt and other consumer products based on the scanner data is a prominent example from marketing research. (See Allenby Garrett and Rossi (2010).) With possibly some small loss of generality, we will assume that J is fixed throughout the discussion.

The number of choice situations, T, may vary across i. Most received theoretical treatments assume fixed (*balanced*) T largely for mathematical convenience, although many actual longitudinal data sets are *unbalanced*, that is, have variation in  $T_i$  across i. At some points this is a minor mathematical inconvenience – variation in  $T_i$  across i mandates a much more cumbersome notation than fixed T in most treatments. But, the variation in  $T_i$  can be substantive. If 'unbalancedness' of the panel is the result of *endogenous attrition* in the context of the outcome model being studied, then a relative to the problem of *sample selection* becomes pertinent. (See Heckman (1979) and a vast literature.) The application to self assessed health in the BHPS by Contoyannis, Jones and Rice (2004) described below is an example. Wooldridge (2002) and Semykina and Wooldridge (2013) suggests procedures for modeling nonrandom attrition in binary choice and linear regression settings.

The data,  $\mathbf{x}_{it,j}$ , will include observable attributes of the outcomes, time varying characteristics of the chooser, such as age, and, possibly, previous outcomes;  $\mathbf{z}_i$  are time and choice invariant characteristics of the chooser, typically demographics such as gender;  $\varepsilon_{it,j}$  is time varying and/or time invariant, unobserved and random characteristics of the chooser. We will assume away at this point any need to consider the time series properties of  $\mathbf{x}_{it}$  – nonstationarity for example. These are typically of no interest in longitudinal data applications. We do note that as the length of some panels such as the NLS, GSOEP and the BHPS grow, the structural stability of the relationship under study might at least be questionable. Variables such as age and experience will appear nonstationary and mandate some consideration of the nature of cross period correlations. This consideration has also motivated a broader treatment of macroeconomic panel data such as the Penn World Tables. But, interest here is in individual, discrete outcomes for which these considerations are tangential or moot.) The remaining element of the model is  $A_i$  which will be used to indicate the presence of choice and time invariant, *unobservable heterogeneity*. As is common in other settings, the unobserved heterogeneity could be viewed as unobservable elements of  $\mathbf{z}_i$ , but it is more illuminating to isolate  $A_i$ .

We note the distinctions between *fully parametric models*, such as the multinomial logit model or loglinear Poisson regression, and *semiparametric* approaches to binary choice modeling such as Manski's *maximum score* (1975, 1985, 1986, 1987), Klein and Spady (1993) and Horowitz's (1992, 1993) smoothed maximum score estimator. Completely nonparametric approaches have been applied as well, such as Hoderlein at al.'s (2011) examination of life cycle income and retirement and Bontemp et al.'s (2009) comparison of parametric and nonparametric models of water demand. In the latter study, the authors argue that patterns in the data that cannot be discerned using parametric models are revealed with the kernel based methods. There are numerous applications of nonparametric methods for binary choice in cross sections, but relatively little extension to panel applications and to the other models of interest here. (See, for example, Racine's (2008) survey, which devotes but a single paragraph to the idea.) The discussion to follow will include some description of non- and semiparametric methods, but, like the received empirical literature, will focus largely on parametric models.

The observation mechanism defined over the alternatives can be interpreted as a revelation of preferences;

 $y_{it} = G(U_{it,1}, U_{it,2}, \dots, U_{it,J})$ 

The translation mechanism that maps underlying preferences to observed outcomes is part of the model. The most familiar (by far) application is the discrete choice over two alternatives, in which

(3)  $y_{it} = G(U_{it,1}, U_{it,2}) = \mathbf{1}(U_{it,2} - U_{it,1} > 0).$ 

Another common case is the *unordered multinomial choice* case in which G(.) indexes the alternative with maximum utility.

$$y_{it} = G(U_{it,1}, U_{it,2}, \dots, U_{it,J}) = j$$
 such that  $U_{it,j} > U_{it,k} \forall j \neq k; j,k = 1,\dots,J$ .

(See, e.g., McFadden (1974).) The convenience of the single outcome model comes with some loss of generality. For example, van Dijk, Fok and Paap (2007) examine a *rank ordered logit model* in which the observed outcome is the subject's vector of ranks (in their case, of six video games), as opposed to only the single most preferred choice. Multiple outcomes at each choice situation, such as this one, are somewhat unusual. Not much generality lost by maintaining the assumption of a scalar outcome – modification of the treatment to accommodate multiple outcomes will generally be straightforward. We can also consider a multivariate outcome in which more than one outcome is observed in each choice situation. (See, e.g., Chakir and Parent (2009.) The multivariate case is easily accommodated, as well. Finally, the *ordered multinomial choice* model is not one that describes utility maximization as such, but rather, a feature of the preference structure itself; G(.) is defined over a single outcome, such that

 $y_{it} = G(U_{it,1}) = j$  such that  $U_{it,1} \in$  the  $j^{\text{th}}$  interval of a partition of the real line,  $[-\infty, \mu_0, \mu_1, \dots, \mu_J, \infty]$ .

The preceding has focused on random utility as an organizing principle. A second thread of analysis is models for counts. These are generally defined by the observed outcome and a discrete probability distribution

 $y_{it}$  = #(events individual *i* at time *t*).

Note the inherently dynamic nature of the statement; in this context, 't' means observed in the interval from the beginning to the end of a time period denoted t. Applications are typically normalized on the length of the observation window, such as the number of traffic incidents per day at given locations, or the number of messages that arrive at a switch per unit of time, or a physical dimension of the observation mechanism, such as the incidence of diabetes per thousand individuals. The 'model' consists, again, of the observed data mechanism and a characterization of an underlying probability distribution ascribed to the rate of occurrence of events. The core model in this setting is a discrete process described by a distribution such as the Poisson or negative binomial distribution. A broader view might also count the number of events until some absorbing state is reached – for example, the number of periods that elapses until bankruptcy occurs, etc. The model may also define treatments of sources of random variation, such as the negative binomial model or normal mixture models for counts which add a layer of unobservable heterogeneity into the Poisson platform. There is an intersection of the two types of models we have described. A *hurdle model* (see Mullahy (1987) and, e.g., Harris and Zhao's (2007) analysis of smoking

behavior) consists of a binary (utility based) choice of whether to participate in an activity followed by an intensity equation or model that describes a count of events. Bago d'Uva (2006) for example, models health care usage using a latent class hurdle model and the BHPS data.

For purposes of developing the methodology of discrete outcome modeling in panel data settings, it is sufficient to work through the binary choice outcome in detail. Extensions to other choice models from this departure point are generally straightforward. However, we do note one important point at which this is decidedly not the case. A great deal has been written about semiparametric and nonparametric approaches to choice modeling. However, nearly all of this analysis has focused on binary choice models. The extension of these methods to multinomial choice, for example, is nearly nonexistent. Partly for this reason, and with respect to space limitations, with only an occasional exception, our attention will focus on parametric models. It also follows naturally that nearly all of the estimation machinery, both classical and Bayesian is grounded in likelihood based methods.

# **III. Individual Heterogeneity in a Panel Data Model of Binary Choice**

After conventional estimation, in some cases, a so called '*cluster correction*' (see Wooldridge (2003)) is often used to adjust the estimated standard errors for effects that would correspond to common unmeasured elements. But, the correction takes no account of heterogeneity in the estimation step. If the presence of unmeasured and unaccounted for heterogeneity taints the estimator, then correcting the standard errors for 'clustering' (or any other failure of the model assumptions) may be a moot point. This discussion will focus on accommodating heterogeneity in discrete choice modeling.

The binary choice model is the natural starting point in the analysis of 'nonlinear panel data models.' Once some useful results are established, extensions to ordered choice models are generally straightforward and uncomplicated. There are only relatively narrow received treatments in unordered choice – we consider a few below. This leaves count data models which are treated conveniently later in discussions of nonlinear regression.

The base case is

$$y_{it} = \mathbf{1}(U_{it,2} - U_{it,1} > 0)$$

$$U_{it,j} = U(\mathbf{x}_{it,j}, \mathbf{z}_i, A_i, \varepsilon_{it,j}), j = 1,2$$

A linear utility specification (e.g., McFadden (1974)) would be

 $U_{it,j} = U(\mathbf{x}_{it,j}, \mathbf{z}_i, A_i, \varepsilon_{it,j}) = \alpha_j + \beta_j' \mathbf{x}_{it,j} + \gamma' \mathbf{z}_i + \delta A_i + \varepsilon_{it,j}$ 

where  $\varepsilon_{it,j}$  are independent and identically distributed across alternatives *j*. McFadden also assumed a specific distribution (type I extreme value) for  $\varepsilon_{it,j}$ . Subsequent researchers, including Manski (1975,

1985), Horowitz (1992) and Klein and Spady (1993) weakened the distribution assumptions. Matzkin (1991) suggested an alternative formulation, in which

$$U_{it,j} = U(\mathbf{x}_{it,j}, \mathbf{z}_i, A_i, \varepsilon_{it,j}) = V(\mathbf{x}_{it,j}, \mathbf{z}_i, A_i) + \varepsilon_{it,j}$$

with  $\varepsilon_{it,i}$  specified nonparametrically. In each of these cases, the question of what can be identified from observed data is central to the analysis. For McFadden's model, for example, absent the complication of the unobserved  $A_i$ , all of the parameters shown are point identified, and probabilities and average partial effects can be estimated. Of course, the issue here is  $A_i$ , which is unobserved. Further fully parametric treatments, e.g., Train (2009), show how all parameters are identifiable. Under partially parametric approaches such as Horowitz (1992) or Klein and Spady (1993), parameters are identified up to scale (and location,  $\alpha$ ). This hampers computation of useful secondary results, such as probabilities and partial effects. Chesher and Smolinsky (2012) and Chesher and Rosen (2012a,b) and Chesher (2010, 2013) examine yet less parameterized cases in which point identification of interesting results such as marginal effects will be difficult. They consider specifications that lead only to set identification of aspects of preferences such as partial effects. (See also Hahn (2010).) Chernuzhukov, Fernandez-Val, Hahn and Newey (2013) also show that without some restrictions, average partial effects are not point identified in nonlinear models; they do indicate estimable sets for discrete covariates. As Wooldridge (2010) notes, what these authors demonstrate is the large payoff to the palatable restrictions that we do impose in order to identify useful quantities in the parametric models that we estimate. Altonji and Matzkin (2005) develop the common case of exchangeability, for example. (Other semiparametric specifications have been suggested, including Honoré and Kyriazidou (2000a,b) that are in some sense immune to variation in functional form and heteroscedasticity. These often require very narrow assumptions about the support of  $\mathbf{x}_{it}$ , for example, 2 periods, or 3 with same  $\mathbf{x}_{it}$  in two of them, etc. Some results have been obtained for nonparametric treatment of both V and ε. See, for example, Honoré (2002), Honoré and Kyriazidou (2000) and Altonji and Matzkin (2005). For purposes of non- and semiparametric estimation, a significant virtue of these huge data sets is that the less than root n consistency of kernel based estimators becomes less of a problem when sample sizes are in the tens of thousands. However, the necessary limits on the support of the data themselves continue to pose limitations. It is difficult to find useful guidance for analyzing long and richly textured longitudinal data sets such as HILDA, MEPS or the BHPS.

Parametric models such as McFadden's have the virtue of strong point identification. As a consequence, however, they are fragile with respect to robustness to violations of assumptions. But, those violations often involve untestable assumptions such as the distribution of random terms (logistic vs. normal) or the existence of higher moments of the independent variables. Heteroscedasticity is less opaque, however. Given the discrete nature of the outcome variable, it can be difficult to distinguish heteroscedasticity from nonlinearity of the utility index. Moreover, in the presence of heteroscedasticity,

it is necessary to redefine the quantities of interest in estimation of the model. There is some ambiguity as to how heteroscedasticity should enter the partial effects. (See Chen and Khan (2003) and Wooldridge (2010) for discussion.)

The generic model specializes in the binary case to

$$y_{it,j} = \mathbf{1}[V(\mathbf{x}_{it,j}, \mathbf{z}_i, A_i, \varepsilon_{it,j}) > 0].$$

The objective of estimation is to learn about features of the preferences, such as partial effects and probabilities attached to the outcomes as well as the superficial features of the model, which in the usual case would be a parameter vector. In the case of a probit model, for example, an overwhelming majority of treatment devoted to estimation of  $\beta$  when actual target is some measure of partial effect. This has been emphasized in some recent treatments, such as Wooldridge (2010), Fernandez-Val (2009).

Combine the  $T_i$  observations on  $(\mathbf{x}_{i1}, \dots, \mathbf{x}_{iTi})$  in data matrix  $\mathbf{X}_i$ . The joint conditional density of  $y_{it}$  and  $A_i$  is

$$f(y_{i1}, y_{i2}, \dots, y_{it}, A_i | \mathbf{X}_i) = f(y_{i1}, y_{i2}, \dots, y_{it} | \mathbf{X}_i, A_i) f(A_i | \mathbf{X}_i)$$

A crucial ingredient of the estimation methodology is:

• Conditional independence: Conditioned on the observed data and the heterogeneity, the Observed outcomes are independent. The joint density of the observed outcomes and the heterogeneity,  $A_i$ , can thus be written

$$f_{y_{1,\ldots,y_{T}}}(y_{i1},y_{i2},\ldots,y_{it}|\mathbf{X}_{i},A_{i})f_{A}(A_{i}|\mathbf{X}_{i}) = \left[\prod_{t=1}^{T_{i}}f_{y}(y_{it}|\mathbf{X}_{i},A_{i})\right]f_{A}(A_{i}|\mathbf{X}_{i}).$$

Models of spatial interaction would violate this assumption. (See Lee (2008) and Greene (2011a).) The assumption will also be difficult to sustain when  $\mathbf{x}_{ii}$  contains lagged values of  $y_{ii}$ .) The conditional log likelihood for a sample of *n* observations based on this assumption is

$$\log L = \sum_{i=1}^{n} \left\{ \left[ \sum_{t=1}^{T_i} \log f_y(y_{it} \mid A_i, \mathbf{X}_i) \right] + \log f_A(A_i \mid \mathbf{X}_i) \right\}$$

If  $f_A(A_i|\mathbf{X}_i)$  actually involves  $\mathbf{X}_i$  then this assumption is only a partial solution to setting up the estimation problem. It is difficult to construct a substantial application without this assumption. The challenge of developing models that include spatial correlation is the leading application. (See Section V below.)

The two leading cases are random and fixed effects. We will specialize to a linear utility function at this point,

 $U_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\gamma}' \mathbf{z}_i + A_i + \varepsilon_{it}$ 

and the usual observation mechanism

$$y_{it} = \mathbf{1}[V_{it} > 0].$$

We (semi) parameterize the data generating process by assuming that there is a continuous probability distribution governing the random part of the model,  $\varepsilon_{it}$ , with distribution function  $F(\varepsilon_{it})$ . At least implicitly, we are directing our focus to cross sectional variation. However, it is important to note

possible unsystematic time variation in the process. The most general approach might be to loosen the specification of the model to  $F_t(\varepsilon_{it})$ . This would still require some statement of what would change over time and what would not – the heterogeneity carries across periods for example. Time variation is usually not the main interest of the study. A common accommodation (again, see Wooldridge (2010)) is a set of time dummy variables, so that

 $U_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\gamma}' \mathbf{z}_i + \boldsymbol{\Sigma}_t \delta_t d_{it} + A_i + \boldsymbol{\varepsilon}_{it}.$ 

Our interest is in estimating characteristics of the data generating process for  $y_{it}$ . Prediction of the outcome variable is considered elsewhere – e.g., Elliot and Leili (2005). We have also restricted our attention to features of the mean of the index function and mentioned scaling, or heteroscedasticity only in passing. (There has been recent research on less parametric estimators that are immune to heteroscedasticity. See, for example, Chen and Khan (2009).) The semiparametric estimators suggested by Honoré and Kyriazidou (2002) likewise consider explicitly the issue of heteroscedasticity. In the interest of brevity, we will leave this discussion for more detailed treatments of modeling discrete choices.

Two additional assumptions needed to continue are:

- *Random Sampling of the observation units*: All observation units *i* and *l* are generated and observed independently (within the overall framework of the data generating process).
- *Independence of the random terms in the utility functions*: Conditioned on  $\mathbf{x}_{it}, \mathbf{z}_i, A_i$ , the unique random terms,  $\varepsilon_{it}$ , are statistically independent for all *i*, *t*.

The random sampling assumption is formed on the basis of all of the information that enters the analysis. Conceivably, the assumption could be violated, for example in modeling choices made by participants in a social network or in models of spatial interaction. However, the apparatus described so far is wholly inadequate to deal with a modeling setting at that level of generality. (See, e.g., Durlauf and Brock (2001a,b, 2002), Durlauf et al. (2010).) Some progress has been made in modeling spatial correlation in discrete choices. However, the random effects framework has provided the only path to forward progress in this setting. The conditional independence assumption is crucial to the analysis.

# A. Random Effects in a Static Model

The binary choice model with a common effect is

$$U_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\gamma}' \mathbf{z}_i + \Sigma_t \delta_t d_{it} + A_i + \varepsilon_{it},$$
  
$$f_{At}(A_i | \mathbf{X}_i, \mathbf{z}_i) = f_A(A_i),$$
  
$$y_{it} = \mathbf{1}[U_{it} > 0].$$

Definitions of what constitutes a *random effects model* hinge on assumptions of the form of  $f_A(A_i|\mathbf{X}_i,\mathbf{z}_i)$ . For simplicity, we have made the broadest assumption, that the DGP of  $A_i$  is time invariant and independent of  $\mathbf{X}_{i}, \mathbf{z}_{i}$ . This implies that the conditional mean is free of the observed data;  $E[A_{i}|\mathbf{X}_{i}, \mathbf{z}_{i}] = E(A_{i})$ . If there is a constant term in  $\mathbf{x}_{it}$ , then no generality is lost if we make the specific assumption  $E[A_{i}] = 0$  for all *t*. Whether the mean equals zero given all  $(\mathbf{X}_{i}, \mathbf{z}_{i})$ , or equals zero given only the current (period *t*) realization of  $\mathbf{x}_{it}$ , or specifically given only the past or only the future values of  $\mathbf{x}_{it}$  (none of which are testable) may have an influence on the estimation method employed. (See, e.g., Wooldridge (2010, chapter 15).) We also assume that  $\varepsilon_{it}$  are mutually independent and normally distributed for all *i* and *t*, which makes this a *random effects probit model*. Given the ubiquity of the logit model in cross section settings, we will return below to the possibility of a *random effects logit* specification. The remaining question concerns the marginal (and, by assumption, conditional) distribution of  $A_{i}$ . For the present, motivated by the central limit theorem, we assume that  $A_{i} \sim N[0, \sigma_{A}^{2}]$ .

The log likelihood function for the parameters of interest is

$$\log L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta} \mid A_1, \dots, A_n) = \sum_{i=1}^n \log \left\{ \prod_{t=1}^{T_i} f_y(y_{it} \mid \mathbf{x}_{it}, t, A_i) \right\}$$

The obstacle to estimation is the unobserved heterogeneity. The unconditional log likelihood is

$$\log L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \sum_{i=1}^{n} \log \left\{ E_A \left[ \prod_{t=1}^{T_i} f_y(y_{it} \mid \mathbf{x}_{it}, A_i) \right] \right\}$$
$$= \sum_{i=1}^{n} \log \left\{ \int_{-\infty}^{\infty} \left[ \prod_{t=1}^{T_i} f_y(y_{it} \mid \mathbf{x}_{it}, A_i) f_A(A_i) dA_i \right] \right\}.$$

It will be convenient to specialize this to the random effects probit model. Write  $A_i = \sigma u_i$  where  $u_i \sim N[0,1]$ . The log likelihood becomes

$$\log L(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\sigma}) = \sum_{i=1}^{n} \log \left\{ \int_{-\infty}^{\infty} \left[ \prod_{t=1}^{T_{i}} \Phi[(2y_{it}-1)(\boldsymbol{\alpha}+\boldsymbol{\beta}'\mathbf{x}_{it}+\boldsymbol{\gamma}'\mathbf{z}_{i}+\boldsymbol{\Sigma}_{t}\boldsymbol{\delta}_{t}d_{it}+\boldsymbol{\sigma}u_{i})]\phi(u_{i})du_{i} \right] \right\}.$$

(Note that we have exploited the symmetry of the normal distribution to combine the  $y_{it} = 0$  and  $y_{it} = 1$  terms.) To save some notation, for the present we will absorb the constant, time invariant variables and time dummy variables in  $\mathbf{x}_{it}$  and the corresponding parameters in  $\boldsymbol{\beta}$  to obtain

$$\log L(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \sum_{i=1}^{n} \log \left\{ \int_{-\infty}^{\infty} \left[ \prod_{t=1}^{T_i} \Phi[(2y_{it} - 1)(\boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\sigma} u_i)] \phi(u_i) du_i \right] \right\}.$$

Two methods can be used in practice to obtain the maximum likelihood estimates of the parameters, Gauss-Hermite quadrature as developed by Butler and Moffitt (1982) and maximum simulated likelihood as analyzed in detail in Train (2009) and Greene (2012). The approximations to the log likelihood are

$$\log L_H(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \sum_{i=1}^n \log \left\{ \sum_{h=1}^H W_h \left[ \prod_{t=1}^{T_i} \Phi[(2y_{it} - 1)(\boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\sigma} W_h)] \right] \right\}$$

for the Butler and Moffitt approach, where  $(w,W)_h$ , h=1,...,H are the weights and nodes for an H point Hermite quadrature, and

$$\log L_{\mathcal{S}}(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \sum_{i=1}^{n} \log \left\{ \frac{1}{R} \sum_{r=1}^{R} \left[ \prod_{i=1}^{T_{i}} \Phi[(2y_{ii} - 1)(\boldsymbol{\beta}' \mathbf{x}_{ii} + \boldsymbol{\sigma}u_{ir})] \right] \right\},\$$

for the maximum simulated likelihood approach, where  $u_{ir}$ , r = 1,...,R are R pseudo-random draws from the standard normal population. Assuming that the data are well behaved and the approximations are sufficiently accurate, the likelihood satisfies the usual regularity conditions, and the MLE (or MSLE) is root-n consistent, asymptotically normally distributed and invariant to one to one transformations of the parameters. (See Train (2009) for discussion of the additional assumptions needed to accommodate the use of the approximations to the log likelihood. Bhat (1999) discusses the use of Halton sequences and other nonrandom methods of computing logL<sub>s</sub>. The quadrature method is widely used in contemporary software such as Stata - see Rebe-Hesketh, Skrondal and Pickles (2005) - SAS and NLOGIT.) Inference can be based on the usual trinity of procedures.

A random effects logit model would build off the same underlying utility function,

$$U_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + u_i + \varepsilon_{it},$$
  

$$f_u(u_i) = \mathbf{N}[0,1], f_{\varepsilon}(\varepsilon_{it}) = \frac{\exp(\varepsilon_{it})}{\left[1 + \exp(\varepsilon_{it})\right]^2}$$
  

$$y_{it} = \mathbf{1}[U_{it} > 0].$$

The change in the earlier log likelihood is trivial – the normal CDF is replaced by the logistic (change ' $\Phi$ ' to ' $\Lambda$ ' in the theory). It is more difficult to motivate the mixture of distributions in the model. The logistic model is usually specified in the interest of convenience of the functional form, while the random effect is the aggregate of all relevant omitted time invariant effects – hence the appeal to the central limit theorem. As noted, the modification of either of the practical approaches to estimation is trivial. A more orthodox approach would retain the logistic assumption for  $u_i$  as well as  $\varepsilon_{it}$ . It is not possible to adapt the quadrature method to this case as the Hermite polynomials are based on the normal distribution. But, it is trivial to modify the simulation estimator. In computing the simulated log likelihood function and any derivative functions, pseudo random normal draws are obtained by using  $u_{ir} = \Phi^{-1}(U_{ir})$  where  $U_{ir}$  is either a pseudorandom U[0,1] draw, a Halton draw or some other intelligent draw. To adapt the estimator to a logistic simulation, it would only be necessary to replace  $\Phi^{-1}(U_{ir})$  with  $\Lambda^{-1}(U_{ir}) = \log[U_{ir}/(1-U_{ir})]$ . (I.e., replace one line of computer code.) The logit model becomes less natural as the model is extended in, e.g., multiple equation directions and gives way to the probit model in nearly all recent applications.

The preceding is generic. The log likelihood function suggested above needs only to be changed to the appropriate density for the variable to adapt it to, e.g., an ordered choice model or one of the models for count data. We will return briefly to this issue below.

# A.1 Partial Effects

Partial effects in the presence of the heterogeneity are

$$\Delta(\mathbf{x}) = \frac{\partial B(\boldsymbol{\beta}'\mathbf{x} + \sigma u)}{\partial \mathbf{x}} = \boldsymbol{\beta}B'(\boldsymbol{\beta}'\mathbf{x} + \sigma u)$$

where B(.) is the function of interest, such as the probability, odds ratio, willingness to pay, or some other function of the latent index,  $\beta' \mathbf{x} + \sigma u$ . The particular element of  $\mathbf{x}$  might be a binary variable, D, in which case, the effect would be computed as  $B(\beta' \mathbf{x} + \beta_D + \sigma u) - B(\beta' \mathbf{x} + \sigma u)$ . If the index function includes a categorical variable such as education coded in levels such as  $ED_{low}$ ,  $ED_{hs}$ ,  $ED_{college}$ ,  $ED_{post}$ , the partial effects might be computed in the form of a transition matrix of effects,  $\mathbf{T}$ , in which the  $ij^{th}$  element is

$$T_{from,to} = B(\boldsymbol{\beta}' \mathbf{x} + \beta_{to} + \sigma u) - B(\boldsymbol{\beta}' \mathbf{x} + \beta_{from} + \sigma u)$$

(See Contoyannis, Jones and Rice (2004) for an application of this type of computation.) For convenience, we will assume that  $\Delta(\mathbf{x})$  is computed appropriately for the application. The coefficients,  $\boldsymbol{\beta}$  and  $\sigma$ , have been consistently estimated. The partial effect can be estimated directly at specific values of *u*, for example its mean of zero. An *average partial effect* can also be computed. This would be

$$\Delta_{\mathbf{x}}(\mathbf{x}) = E_{u} \left[ \frac{\partial B(\mathbf{x}, u)}{\partial \mathbf{x}} \right] = \left[ \frac{\partial E_{u}[B(\mathbf{x}, u)]}{\partial \mathbf{x}} \right] = \left[ \frac{\partial [B_{\mathbf{x}}(\mathbf{x})]}{\partial \mathbf{x}} \right]$$

where  $B_{\mathbf{x}}(\mathbf{x})$  is the expected value of the function of interest. The average partial effect will not equal the partial effect, as  $B_{\mathbf{x}}(.)$  need not equal B(..). Whether this average function is of interest is specific to the application. For the random effects probability model we would usually begin with  $\operatorname{Prob}(Y=1|\mathbf{x},u)$ . In this case, we can find  $B(\mathbf{x},u) = \Phi(\beta'\mathbf{x} + \sigma u)$  while  $B_{\mathbf{x}}(\mathbf{x}) = \Phi(\beta'\mathbf{x}/(1 + \sigma^2)^{1/2})$ . The average partial effect is then

$$\Delta_{\mathbf{x}}(\mathbf{x}) = \frac{\partial \Phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}}{\sqrt{1+\sigma^2}}\right)}{\partial \mathbf{x}} = \frac{1}{\sqrt{1+\sigma^2}}\boldsymbol{\beta}\boldsymbol{\phi}\left(\frac{\boldsymbol{\beta}'\mathbf{x}}{\sqrt{1+\sigma^2}}\right)$$

With estimates of  $\beta$  and  $\sigma$  in hand, it would be possible to compute the partial effects at specific values of  $u_i$ , such as zero. Whether this is an interesting value to use is questionable. However, it is also possible to obtain an estimate of the average partial effect, directly after estimation. Indeed, if at the outset, one simply ignores the presence of the heterogeneity, and uses maximum likelihood to estimate the parameters of the '*population averaged model*,'

$$\operatorname{Prob}(y=1|\mathbf{x}) = \Phi(\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}),$$

Then the estimator consistently estimates  $\beta_x = \beta' x/(1+\sigma^2)^{1/2}$ . Thus, while conventional analysis does not estimate the parameters of the structural model, it does estimate something of interest, namely the parameters and partial effects of the population averaged model.

#### A.2. Alternative Models for the Random Effects

The random effects may enter the model in different forms. The so called *GEE* approach to this analysis is difficult to motivate rigorously, but it is (loosely) generated by a seemingly unrelated regressions approach built around

 $y_{it} = \Phi(\boldsymbol{\beta}' \mathbf{x}_{it}) + v_{it},$ 

where the probability is also the regression function. A similar view is suggested by the *panel probit model* in Bertschuk and Lechner (1998),

$$U_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \varepsilon_{it},$$
  

$$Cov(\varepsilon_{it}, \varepsilon_{js}) = \mathbf{1}[i = j]\sigma_{ts}.$$
  

$$y_{it} = \mathbf{1}[U_{it} > 0].$$

Here, the SUR specification applies to the latent utilities, rather than the observed outcomes. The GEE estimator is estimated by a form of nonlinear generalized least squares. The terms in the log likelihood function for Bertschuk and Lechner's model are *T*-variate normal probabilities. This necessitates computation of higher order normal integrals. The authors devise a GMM estimator that avoids the burdensome calculations. Recent implementations of the GHK simulator and advances in computation capabilities do make the computations more reasonable. See Greene (2004a).

Heckman and Singer (1984) questioned the need for a full parametric specification of the distribution of  $u_i$ . (Their analysis was in the context of models for duration, but extends directly to this one.) A semiparametric, discrete specification based on their model would be

$$F(u_i) = \operatorname{Prob}(u_i = \alpha_q) = \pi_q, q = 1, \dots, Q.$$

This gives rise to a 'latent class' model, for which the log likelihood would be

$$\log L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi}) = \sum_{i=1}^{n} \log \left\{ \sum_{q=1}^{Q} \pi_q \left[ \prod_{t=1}^{T_i} \Phi[(2y_{it} - 1)(\boldsymbol{\alpha}_q + \boldsymbol{\beta}' \mathbf{x}_{it})] \right] \right\}.$$

This would be a partially semiparametric specification – it retains the fully parametric probit model as the platform. Note that this is a discrete counterpart to the continuous mixture model in (20).

The random effects model is, in broader terms, a *mixed model*. A more general statement of the *mixed model* would be

$$U_{it} = (\mathbf{\beta} + \mathbf{u}_i)'\mathbf{x}_{it} + \varepsilon_{it},$$
  

$$F(\mathbf{u}_i | \mathbf{X}_i, \mathbf{z}_i) = f(\mathbf{u}_i) = \mathbf{N}[\mathbf{0}, \mathbf{\Sigma}],$$
  

$$y_{it} = \mathbf{1}[U_{it} > 0].$$

The extension here is that the entire parameter vector, not just the constant term, is heterogeneous. The mixture model used in recent applications is either continuous. (See, e.g., Train (2009) and Rebe-Hesketh, Skrondal and Pickles (2005) or discrete in the fashion suggested by Heckman and Singer

(1984); see Greene and Hensher (2010). Altonji and Matzkin (2005) considered other semiparametric specifications.)

#### A.3. Specification Tests

It would be of interest to test for the presence of random effects against the null of the 'pooled' model. That is, ultimately, a test of  $\sigma = 0$ . In the random effects probit model, direct approaches based on the Wald or LR tests are available. The LM test has a peculiar feature; the score of the log likelihood is identically zero at  $\sigma = 0$ . Chesher (1984), Chesher and Lee (1986) and Cox and Hinkley (1974) suggest reparameterization of such models as a strategy for setting up the LM test. Greene and McKenzie (2012) derived the appropriate statistic for the random effects probit model. The phenomenon would reappear in an ordered probit or ordered logit model as well.

A second specification test of interest might be the distributional assumption. There is no natural residual based test such as the Bera and Jarque (1982) test for the linear regression. A test for the pooled (cross section) probit model based essentially on Chesher and Irish's (1987) generalized residuals is suggested by Bera, Jarque and Lee (1984). It is not clear how the test could be adapted to a random effects model, however, nor, in fact, whether it could be extended to other models such as ordered choice models.

#### A.4. Other Discrete Choice Models

Application of the random effects models described above to an ordered choice model requires only a minor change in the assumed density of the observed outcome. See Greene and Hensher (2010, pp. 275-278). All other considerations are the same. The ordered probit model does contain an additional source of heterogeneity, in the thresholds. Ongoing development of the ordered choice methodology includes specifications of the thresholds, which may respond to observed effects (Pudney and Shields (2000), Greene and Hensher (2010)) and to unobserved random effects (Harris, Hollingsworth and Greene (2012).

Random effects in count data models would build on a familiar specification in the cross section form. For a Poisson regression, we would have

$$\operatorname{Prob}(\mathbf{Y} = y_{it} | \mathbf{x}_{it}, u_i) = \frac{\exp(-\lambda_{it})\lambda_{it}^{y_{it}}}{y_{it}!}, \ \lambda_{it} = \exp(\boldsymbol{\beta}' \mathbf{x}_{it} + \sigma u_i).$$

Since  $\lambda_{it}$  is the conditional mean, at one level, this is simply a nonlinear random effects regression model. However, maximum likelihood is the preferred estimator. If  $u_i$  is assumed to have a log-gamma distribution (see Hausman, Hall and Griliches (HHG, 1984)), then the unconditional model becomes a negative binomial (NB) regression. Recent applications have used a normal mixture approach. See, for example, Riphahn, Wambach and Million (2003). The normal model would be estimated by maximum simulated likelihood or by quadrature based on Butler and Moffitt (1982). (See Greene (1995) for an application.) A random effects negative binomial model would be obtained by applying the same methodology to the NB probabilities. One could argue that the RENB model arises by having two layers of heterogeneity, a unique component,  $w_{it}$ , that transforms the base case Poisson and a second that embodies the common unobserved effect,  $u_i$ . HHG (1984) treat the NB model as a distinct specification rather than as the result of the mixed Poisson. The normal mixed NB model is discussed in Greene (2012).

There is an ambiguity in the mixed unordered multinomial choice model because it involves several utility functions. A fully specified random effects multinomial logit model would be

$$\operatorname{Prob}(y_{it} = j) = \frac{\exp(\alpha_j + \boldsymbol{\beta}' \mathbf{x}_{it,j} + u_{i,j})}{\sum_{j=1}^{J} \exp(\alpha_j + \boldsymbol{\beta}' \mathbf{x}_{it,j} + u_{i,j})}$$

A normalization is required since the probabilities sum to one – the constant and the random effect in the last utility function equal zero. An alternative specification would treat the random effect as a single choice invariant characteristic of the chooser, which would be constant across utility functions. It would seem that this would be easily testable using the likelihood ratio statistic. However, this specification involves more than a simple parametric restriction. In the first specification, (we assume) the random effects are uncorrelated. In the second, by construction, the utility functions are equicorrelated. This is a substantive change in the preference structure underlying the choices. (The intermediate case, of equal standard deviations on the *J*-1 random effects, seems difficult to interpret.) Finally, the counterpart to the fully random parameters model is the mixed logit model,

$$\operatorname{Prob}(\mathbf{y}_{it} = \mathbf{j}) = \frac{\exp(\alpha_{j,i} + (\boldsymbol{\beta} + \mathbf{u}_i)'\mathbf{x}_{it,j})}{\sum_{j=1}^{J} \exp(\alpha_{j,i} + (\boldsymbol{\beta} + \mathbf{u}_i)'\mathbf{x}_{it,j})}$$

See McFadden and Train (2000), Hensher, Rose and Greene (2005) and Hensher and Greene (2003).

# **B.** Fixed Effects in a Static Model

The single index model is

$$f(\mathbf{y}_{it}|\mathbf{x}_{it},\mathbf{z}_{i},\alpha_{i}) = f(\mathbf{y}_{it},\boldsymbol{\beta}'\mathbf{x}_{it}+\boldsymbol{\gamma}'\mathbf{z}_{i}+\alpha_{i}) = f(\mathbf{y}_{it},a_{it}).$$

For empirical purposes, the model is recast with the unobserved effects treated as parameters to be estimated;

$$a_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\gamma}' \mathbf{z}_i + \sum_{i=1}^n \alpha_i d_{it},$$

where  $d_{it}$  is a set of *n* group dummy variables. (Note, this is the estimation strategy. The model specification does not imply that the common effects are parameters in the same way that elements of  $\beta$  are. At this point,  $\mathbf{x}_{it}$  does not contain an overall constant term.) The leading cases in the received literature are the *fixed effects probit model*,

$$f(y_{it}, a_{it}) = \operatorname{Prob}(y_{it} = 1 | a_{it}) = \Phi[(2y_{it} - 1)a_{it}],$$

where  $\Phi(w)$  is the standard normal CDF, and *fixed effects logit model* 

$$f(y_{it},a_{it}) = \Lambda[(2y_{it}-1)a_{it}] = \exp[(2y_{it}-1)a_{it}]/\{1+\exp[(2y_{it}-1)a_{it}]\}.$$

The fixed effects model is distinguished from the random effects model by relaxing the assumption that  $f_A[A_i|\mathbf{X}_i,\mathbf{z}_i] = f_A(A_i)$ . In the fixed effects case, the conditional distribution is not specified and may depend on  $\mathbf{X}_i$ . Other cases of interest are the ordered choice models and the Poisson and negative binomial models for count data. We will examine the binary choice models first, then briefly consider the others. Fixed effects models have not provided an attractive framework for analysis of multinomial unordered choices. For most of the discussion, we can leave the model in generic form and specialize when appropriate.

No specific assumption is made about the relationship between  $\alpha_i$  and  $\mathbf{x}_{it}$ . The possibility that  $E[\alpha_i | \mathbf{x}_{i1,...,} \mathbf{x}_{iT}] = m(\mathbf{X}_i)$  is not ruled out. If no restrictions are placed on the joint distribution of the unobservable  $\alpha_i$  and the observed  $\mathbf{X}_i$ , then the random effects apparatus of the previous sections is unusable  $-\mathbf{x}_{it}$  becomes endogenous by dint of the omitted  $\alpha_i$ . Explicit treatment of  $\alpha_i$  is required for consistent estimation.

Any time invariant individual variables (TIVs),  $\mathbf{z}_i$ , will lie in the column space of the unobservable  $\alpha_i$ . The familiar identification (multicollinearity) issue arises in the linear regression case and in nonlinear models. Coefficients  $\gamma$  cannot be identified without further restrictions. (See Plumper and Troeger (2007, 2011), Greene (2011b), Breusch et al (2011) and Hahn and Meinecke (2005).) Consider a model with a single TIV,  $z_i$ . The log likelihood is

$$\log L = \sum_{i=1}^{n} \sum_{t=1}^{T} \log f(y_{it}, a_{it})$$

The likelihood equations for  $\alpha_i$  and  $\gamma$  are

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{t=1}^T \frac{\partial f(y_{it}, a_{it}) / \partial a_{it}}{f(y_{it}, a_{it})} \times 1 = \sum_{t=1}^T g_{a_{it}} = 0,$$
  
$$\frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^n \sum_{t=1}^T g_{a_{it}} z_i = \sum_{i=1}^n z_i \frac{\partial \log L}{\partial \alpha_i} = 0.$$

This produces the singularity in the second derivatives matrix for the full set of parameters that is a counterpart to multicollinearity in the linear case. Gradient based maximization methods will fail to converge because of the singularity of the weighting matrix, however formed. Bayesian methods

(Lancaster, 1999, 2000, 2001)) will be able to identify the model parameters on the strength of informative priors. (For an example of Bayesian identification of individual effects on the strength of informative priors, see Koop et al. (1997). For a comment on diffuse priors, see Hahn (2004).) The GMM approach suggested by Laisney and Lechner (2002) seems to provide a solution to the problem. The authors note, however,

Thus the coefficients of the time invariant regressors are identified provided there is at least one time varying regressor, .... However, since this identification hinges on the local misspecification introduced by the Taylor series approximation, it seems preferable not to attempt an estimation of the coefficients of the time invariant variables, and to subsume the impact of the latter in the individual effect.

This would be an extreme example of *identification by the functional form of the model*. The *fixed effects negative binomial model* proposed in Hausman, Hall and Griliches (HHG,1984) is a surprising exception to this broad generality. We defer that special case for the moment and assume that the model does not contain time invariant effects. It is worth noting that for purpose of analyzing modern longitudinal data sets, the inability to accommodate time invariant covariates is a vexing practical shortcoming of the fixed effects model. (See, again, Plumper and Troeger (2007).) The hybrid formulations based on Mundlak's (1978) formulation or on *correlated random effects* in the next section present a useful approach that appears in many recent applications.

Strategies for estimation of models with fixed effects generally begin by seeking a way to avoid estimation of *n* effects parameters in the fully specified model. (See, e.g., Hahn (2009).) This turns on the existence of a sufficient statistic,  $S_i$  for the fixed effect such that the joint density,  $f(y_{i1}, ..., y_{iT}|S_i, \mathbf{X}_i)$  does not involve  $\alpha_i$ . In the linear regression model,  $\Sigma_i y_{it}$  provides the statistic – the estimator based on the conditional distribution is the within groups linear least squares estimator. In all but a small few other cases (only two of any prominence in the contemporary literature), there is no sufficient statistic for  $\alpha_i$  in the log likelihood for the sample. In the Poisson regression, and in the binary logit model,  $\Sigma_i y_{it}$  provides the statistic. (See Lancaster (2000) for a few additional cases (that are not discrete outcome models). Chamberlain (1984) mentions a counterpart for a form of the multinomial logit model.) For the Poisson model, the marginal density is

$$f(\mathbf{y}_{it}, a_{it}) = \frac{\exp(-\lambda_{it})\lambda_{it}^{\mathbf{y}_{it}}}{\mathbf{y}_{it}!}, \lambda_{it} = \exp(\mathbf{\beta}'\mathbf{x}_{it} + \alpha_i) = \exp(\alpha_i)\exp(\mathbf{\beta}'\mathbf{x}_{it}).$$

The likelihood equation for  $\alpha_i$  is

$$\frac{\partial \log L}{\partial \alpha_i} = \left(\sum_{t=1}^{T_i} -\lambda_{it}\right) + \left(\sum_{t=1}^{T_i} y_{it}\right) = 0$$

which can be solved for

$$\boldsymbol{\alpha}_{i} = \log \begin{pmatrix} \boldsymbol{\Sigma}_{t=1}^{T_{i}} \boldsymbol{y}_{it} \\ \boldsymbol{\Sigma}_{t=1}^{T_{i}} \boldsymbol{\beta}' \mathbf{x}_{it} \end{pmatrix}.$$

Note that there is no solution when  $y_{it}$  equals zero for all *t*. There need not be within group variation; the only requirement is that the sum be positive. Such observation groups must be dropped from the sample. The result for  $\alpha_i$  can be inserted into the log likelihood to form a concentrated log likelihood. The remaining analysis appears in HHG (1984). (HHG did not consider the case in which  $\sum_{i,t} y_{it} = 0$ , as in their data,  $y_{it}$  was always positive.) A second case, perhaps not surprisingly given its relationship to the Poisson model, would be the exponential regression model,

$$G(y_{it},a_{it}) = \lambda_{it} \exp(-y_{it}\lambda_{it}), \ \lambda_{it} = \exp(\boldsymbol{\beta}' \mathbf{x}_{it} + \alpha_i).$$

Finally, for the binary logit model, the familiar result is

$$Prob(y_{i1}, y_{i2}, ..., y_{t,T_i}) = f(y_{i1}, y_{i2}, ..., y_{t,T_i} | X_i, \Sigma_{t=1}^{T_i})$$
$$= \frac{\exp(\Sigma_{t=1}^{T_i} y_{it}(\boldsymbol{\beta}' \mathbf{x}_{it}))}{\sum_{\Sigma_t d_{it} = \Sigma_t y_{it}} \exp(\Sigma_{t=1}^{T_i} d_{it}(\boldsymbol{\beta}' \mathbf{x}_{it}))}$$

which is free of the fixed effects. The denominator in the probability is the sum over all  $\begin{pmatrix} T_i \\ \Sigma_{t=1}^T y_{it} \end{pmatrix}$ 

configurations of the sequence of outcomes that sum to the same  $\Sigma_t y_{it}$ . This computation can, itself, be daunting – for example, if  $T_i = 20$  and  $\Sigma_t y_{it} = 10$ , there are  $20!/(10!)^2 = 184,756$  terms that all involve  $\beta$ . A recursive algorithm provided by Krailo and Pike (1984) greatly simplifies the calculations. (In an experiment with 500 individuals and T= 20, estimation of the model required about 0.25 seconds on an ordinary desktop computer.) Chamberlain (1980) details a counterpart of this method for a multinomial logit model. We are unaware of any applications of this estimator for the multinomial logit case, however.

In the probit model, which has attracted considerable interest, the practical implementation of the FEM requires estimation of the model with n dummy variables actually in the index function – there is no way to concentrate them out and no sufficient statistic. The complication of nonlinear models with possibly tens of thousands of coefficients to be estimated all at once has long been viewed as a substantive barrier to implementation of the model. See, e.g., Maddala, (1983). The algorithm given in Greene (2004b, 2012) presents a solution to this practical problem. Fernandez-Val (2009) reports that he used this method to fit an FE probit model with 500,000 dummy variables. Thus, the physical complication is not a substantive obstacle in any problem of realistic dimensions. (In practical terms, the complication of fitting a model with 500,000+K coefficients would be a covariance matrix that would occupy nearly a terabyte of memory. Greene's algorithm exploits the fact that nearly the entire matrix is zeros to reduce the matrix storage requirements to linear in n rather than quadratic.)

The impediment to application of the fixed effects probit model is a persistent bias labeled the incidental parameters problem. As has been widely documented in a long sequence of Monte Carlo studies and theoretical analyses, there is a persistent bias of O(1/T) in the maximum likelihood estimation of the parameters in many fixed effects model estimated by maximum likelihood. (Again, the Poisson regression is the well known exception.) The incidental parameters problem was first reported in Neyman and Scott (1948), where it is shown that the MLE of  $\sigma^2$  in a fixed effects linear regression model, e'e/nT, has plim  $s^2 = \sigma^2(T-1)/T$ . This is potentially far less than  $\sigma^2$  and does not improve as N increases. The obvious remedy, correcting for degrees of freedom, does not eliminate the vexing shortcoming of a perfectly well specified maximum likelihood estimator in other internally consistent model specifications. The problem persists in nonlinear settings where there is no counterpart 'degrees of freedom correction.' (See Lancaster (2000) for a detailed history.) The extension of this result to other, nonlinear models has entered the orthodoxy of the field, though a precise result has actually been formally derived for only one case, the binomial logit model when T = 2, where it is shown that  $plim\hat{\beta}_{ML} = 2\beta$ . (See, e.g., Abrevaya (1997) and Hsiao (2003).) Although the regularity seems to be equally firm for the probit model and can be demonstrated with singular ease with a random number generator with any modern software, it has not been proved formally. Nor has a counterpart been found for any other T, for the unbalanced panel case, or for any other model. Other specific cases such as the ordered probit and logit models have been persuasively demonstrated by Monte Carlo methods. (See, e.g., Katz (2001) and Greene (2004b). The persistent finding is that the MLE for discrete choice models is biased away from zero. (Greene (2004b) finds (again, experimentally) that this result seems not to be general. When the dependent variable is continuous, other outcomes can occur - lack of bias in the slope estimators in a tobit model and a downward bias in the MLE of  $\beta$  in a truncated regression model, for example. The result that does seem to persist is that when the incidental parameters problem arises, it does so with a proportional impact on some or all of the model parameters.) The bias does not appear to depend substantively on the nature of the data support - it appears in the same form regardless of the process assumed to underlie the independent variables in the model. Rather, it is due to the presence of *n* additional estimation equations. We do note, once again, the generality of the bias, away from zero, appears to be peculiar to discrete outcome models. Moreover, the effect appears not to be confined to variance parameters in continuous outcome models – it shows up in both  $\beta$  and  $\sigma^2$  in a truncated regression model, but only in the variance terms in Tobit and stochastic frontier models. (See Greene (2004b).)

Solutions to the incidental parameters problem in discrete choice cases – that is, consistent estimators of  $\beta$  - are of two forms. As discussed in Lancaster (2000), for a few specific cases, there exist sufficient statistics that will allow formation of a conditional density that is free of the fixed effects. The

binary logit and Poisson regression cases are noted earlier. Lancaster notes a generic solution based on orthogonalization of the log likelihood – a reparameterization that produces a partition of the log likelihood function into two terms, one of which involves only  $\beta$ . Orthogonalization has not proved to be a viable strategy in very many cases, however. Lancaster notes a duration model based on the Weibull distribution.

Several recent applications have suggested a '*bias reduction*' approach. The central result as shown, for example, in Hahn and Newey (1994) and Hahn and Kuersteiner (2011) largely (again) for binary choice models is

plim 
$$\hat{\boldsymbol{\beta}}_{ML} = \boldsymbol{\beta} + \mathbf{B}/T + O(1/T^2).$$

(See, as well, Arellano and Hahn. (2007).) That is, the unconditional MLE converges to a constant that is biased of O(1/*T*). Three approaches have been suggested for eliminating **B**/*T*, a *penalized criterion* (modified log likelihood), *modified estimation (likelihood) equations* and direct *bias correction* by estimating the bias, itself. In the first case, the direct log likelihood is augmented by a term in  $\beta$  whose maximizer is a good estimator of  $-\mathbf{B}/T$ . (See Carro and Traferri (2011).) In the second case, an estimator of  $-\mathbf{B}/T$  is added to the MLE. See, e.g., Fernandez-Val (2009). (The received theory has made some explicit use of the apparent proportionality result, that the bias in fixed effect discrete choice models, which are the only cases ever examined in detail, appears to be multiplicative, by a scalar of the form  $1+b/T + O(1/T^2)$ . The effect seems to attach itself to scale estimation, not location estimators. The regression case noted earlier is obvious by construction. The binary choice case, though less so, does seem to be consistent with this. Write the model as  $y = \mathbf{1}[\beta'\mathbf{x} + \alpha_i + \sigma w_{it} > 0]$ . The estimated parameters are  $\beta/\sigma$ , not  $\beta$ , where  $\sigma$  is typically normalized to 1 for identification. But, the multiplicative bias of the MLE does seem to affect the implicit 'estimate' of the scale factor. The same result appears to be present in the MLE of the FE tobit model. (See Greene (2004b).) Fernandez-Val (2009) discusses this result at some length.

There is a loose end in the received results. The bias corrected estimators begin from the unconditional, brute force estimator that also estimates the fixed effects. However, this estimator, regardless of the distribution assumed (that will typically be the probit model), is incomplete. The estimator of  $\alpha_i$  is not identified when there is no within group variation in  $\mathbf{y}_i$ . For the probit model, the likelihood equation for  $\alpha_i$  is

$$\frac{\partial \log \mathcal{L}}{\partial \alpha_{i}} = \sum_{t=1}^{T_{i}} \frac{(2y_{it}-1)\phi[(2y_{it}-1)(\boldsymbol{\beta}'\mathbf{x}_{it}+\alpha_{i})]}{\Phi[(2y_{it}-1)(\boldsymbol{\beta}'\mathbf{x}_{it}+\alpha_{i})]} = 0$$

If  $y_{it}$  equals one (zero) for all *t*, then the derivative is necessarily positive (negative) and cannot be equated to zero for any finite  $\alpha_i$ . In the 'Chamberlain' estimator, groups for which  $y_{it}$  is always one or zero fall

out of the estimation – they contribute log(1.0) = 0.0 to the log likelihood. Such groups must also be dropped for the unconditional estimator.

The starting point for consistent estimation of FE discrete choice models is the binary logit model. For the two period case, there are two obvious consistent estimators of  $\beta$ , the familiar textbook conditional estimator and  $\frac{1}{2}$  times the unconditional MLE. For more general (different T) cases, the well known estimator developed by Rasch (1960) and Chamberlain (1980), builds on the conditional joint distribution,  $Prob(y_{i1}, y_{i2}, \dots, y_{i,Ti} | \Sigma_t y_{it}, \mathbf{X}_i)$  which is free of the fixed effects. Two important shortcomings of the conditional approach are: (1) it does not provide estimators of any of the  $\alpha_i$  so it is not possible to compute probabilities or partial effects (see Wooldridge (2010, p. 622)) and (2) it does not extend to other distributions or models. It does seem that there could be a remedy for (1). With a consistent estimator of  $\beta$  in hand, one could estimate individual terms of  $\alpha_i$  by solving the likelihood equation noted earlier for the probit model (at least for groups that have within group variation). The counterpart for the logit model is  $\sum_{i} [y_{it} - \Lambda(\boldsymbol{\beta}' \mathbf{x}_{it} + \alpha_{i})] = 0$ . A solution exists for  $\alpha_{i}$  for groups with variation over t. Each individual estimator is inconsistent as it is based on fixed T observations. Its asymptotic variance is O(1/T). It remains to be established whether the estimators are systematically biased (upward or downward) when they are based on a consistent estimator of  $\beta$ . If not, it might pay to investigate whether the average over the useable groups provides useful information about  $E[\alpha_i]$ , which is what is needed to solve problem (1). The bias reduction estimators, to the extent that they solve the problem of estimation of  $\beta$ , may also help to solve this subsidiary problem. This was largely the finding of Hahn and Newey (2002). The conditional MLE in the binary logit model would appear to be a solution. This finding would be broadly consistent with Wooldridge's arguments for the random effects pooled, or 'population averaged' estimator.

The ordered choice cases are essentially the same as the binary cases as regards the conventional (brute force) estimator and the incidental parameters problem. There is no sufficient statistic for estimation of  $\beta$  in either case. However, the  $2\beta$  result for T = 2 appears to extend to the ordered choice models. The broad nature of the result for T > 2 would seem to carry over as well. [See Greene and Hensher (2010).] the ordered logit model provides an additional opportunity to manipulate the sample information. The base outcome probability for a fixed effects ordered logit model is

 $\operatorname{Prob}(y_{it} = j | \mathbf{x}_{it}) = \Lambda(\mu_j - \boldsymbol{\beta}' \mathbf{x}_{it} - \alpha_i) - \Lambda(\mu_{j-1} - \boldsymbol{\beta}' \mathbf{x}_{it} - \alpha_i).$ 

The implication is

 $\operatorname{Prob}(y_{it} \geq j | \mathbf{x}_{it}) = \Lambda(\boldsymbol{\beta}' \mathbf{x}_{it} + \alpha_i - \mu_j) = \Lambda(\boldsymbol{\beta}' \mathbf{x}_{it} + \delta_i(j)).$ 

Define the new variable  $D_{ii}(j) = \mathbf{1}[y_{ii} \ge j]$ , j = 1, ..., J. This defines J-1 binary fixed effects logit models, each with its own set fixed effects, though they are the same save for the displacement by  $\mu_j$ . The

Rasch/Chamberlain estimator can be used for each one. This does produce *J*-1 numerically different estimators of  $\beta$  that one might reconcile using a minimum distance estimator. The covariance matrices needed for the efficient weighting matrix are given in Brant (1990). An alternative estimator is based on the sums of outer products of the score vectors from the *J*-1 log likelihoods. Das and van Soest (2000) provide an application.

Large sample bias corrected applications of the ordered choice models have been developed in Bester and Hansen (2009) and in Carro and Traferri (2012). The methods employed limit attention to a three outcome case (low/medium/high). It is unclear if they can be extended to more general cases.

As has been documented elsewhere (e.g., Cameron and Trivedi (2005)), the conditional fixed effects estimator for the Poisson model is algebraically identical to the unconditional estimator. The upshot would be that for the Poisson model, there is no incidental parameters problem. The mathematics of the result is straightforward enough. The logic still seems elusive. We would surmise that in contrast to the binary choice cases, there is no implicit random variation around the mean – no disturbance variance defined in the model. The fixed effects negative binomial model is rather more involved. A form of the model was proposed in HHG (1984) and was the received standard until quite recently. Applied researchers would occasionally bump into a surprising result that in contrast to every other model considered thus far, a FENB model with time invariant variables z in the index function 'worked,' in that estimation of all parameters including those on z (and even an overall constant) were estimated routinely. Allison and Waterman (2002) examined the HHG model in detail (see also Greene (2012)) and demonstrated that unlike every other familiar case, this received FE model was not a single index model. In the HHG model, the time invariant heterogeneity appears in the scale parameter of the log-gamma heterogeneity that extends the NB model from the Poisson base. A more natural NB model – at least in terms of its relationship to other models, would take the usual form, as a conditional Poisson regression

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

where *u* has a log gamma( $\theta$ , $\theta$ ) distribution. The mixed Poisson produces an NB model with fixed effects. This model appears to be impacted by the IP problem. Recourse to a pseudo maximum likelihood approach – that is, to a Poisson regression, might be useable strategy. This remains an avenue for further research.

The preceding is focused on estimation of the parameters of fixed effects models. We also noted the possibility of conventional inference about parameters, and for estimation of partial effects. A remaining question is whether it is possible to test for the presence of fixed effects. The behavior of the MLE under the null hypothesis is the pooled estimator, which is easily established. Behavior under the alternative is less clear because of the incidental parameters problem. The MLE of the parameters converge to something (see Hahn and Newey (1994)) but not to the 'true' parameters of the model. The behavior of the likelihood ratio statistic remains to be settled. One practical approach based on Mundlak's approximation is considered in the next section. Finally, the force of the IP problem seems to be more pronounced when lagged values in model. However, relatively little is known about the behavior of the MLE in this case. (See Lee (2013, this volume).)

### **C.** Correlated Random Effects

Mundlak (1978) suggested an approach between the questionable orthogonality assumptions of the random effects model and the frustrating limitations of the fixed effects specification,

$$y_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \alpha_i + \varepsilon_{it}$$
  
$$\alpha_i = \alpha + \boldsymbol{\gamma}' \overline{\mathbf{x}}_i + w_i.$$

Chamberlain (1980) proposed a less restrictive formulation,

$$\alpha_i = \alpha + \Sigma_t \gamma_t' \mathbf{x}_{it} + w_i$$

This formulation is a bit cumbersome if the panel is not balanced – particularly if, as Wooldridge (2010) considers, the unbalancedness is due to endogenous attrition. The model examined by Plumper and Troeger (2007) is similar to Mundlak's;

$$\alpha_i = \alpha + \gamma' \mathbf{z}_i + w_i$$

(This is a 'hierarchical model,' or multi (two) level model – see Bryk and Raudenbush (2002).) In all of these cases, the assumption that  $E[w_i \mathbf{x}_{it}] = \mathbf{0}$  point identifies the parameters and the partial effects. The direct extension of this approach to nonlinear models such as the binary choice, ordered choice and count data models converts them to random effects specifications that can be analyzed by conventional techniques. Whether the auxiliary equation should be interpreted as the conditional mean function in a structure or as a projection that, it is hoped, provides a good approximation to the underlying structure is a minor consideration that nonetheless appears in the discussion. For example, Hahn, Ham and Moon (2011) assume Mundlak's formulation as part of the structure at the outset, while Chamberlain (1980) would view that as restriction on the more general model.

The correlated random effects specification has a number of virtues for nonlinear panel data models. The practical appeal of a random effects vs. a full fixed effects approach is considerable. There are a number of conclusive results that can be obtained for the linear model that cannot be established for nonlinear models, such as Hausman's (1978) specification test for fixed vs. random effects. In the correlated random effects case, although the conditions needed to motivate Hausman's test are not met – the fixed effects is not robust; it is not even consistent under either hypothesis – a variable addition test (Wu (1973)) is easily carried. In the Mundlak form, the difference between this version of the fixed effects model and the random effects model is the nonzero  $\gamma$ , which can be tested with a Wald test. Hahn, Ham and Moon (2011) explored this approach in the context of panels in which there is very little within

group variation and suggested an alternative statistic for the test. (The analysis of the data used in the World Health Report (WHO (2000)) by Gravelle et al. (2002) would be a notable example.)

# **D** Attrition and Unbalanced Panels

Unbalanced panels may be more complicated than just a mathematical inconvenience. If the unbalanced panel results from attrition from what would otherwise be a balanced panel, and if the attrition is connected to the outcome variable, then the sample configuration is endogenous, and may taint the estimation process. Contoyannis, Jones and Rice (2004) examine self assessed health (SAH) in eight waves of the British Household Panel Survey. Their results suggest that individuals left the panel during the observation window in ways connected to the sequence of values of SAH. A number of authors, beginning with Verbeek and Nijman (1992) and Verbeek (2000) have suggested methods of detecting and correcting for endogenous attrition in panel data. Wooldridge (2002) proposes an 'inverse probability weighting' procedure to weight observations in relation to their length of stay in the panel as a method of undoing the attrition bias. The method is refined in Wooldridge (2010) as part of an extension to a natural sample selection treatment.

#### **IV. Dynamic Models**

An important benefit of panel data is the ability to study dynamic aspects of behavior in the model. The dynamic linear panel data regression

$$y_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \delta y_{i,t-1} + \alpha_i + \varepsilon_i$$

has been intensively studied since the field originated with Balestra and Nerlove (1966). Analysis of dynamic effects in discrete choice modeling has focused largely on binary choice. An empirical exception is Contoyannis, Jones and Rice's (2004) ordered choice model for SAH. (Wooldridge (2005) also presents some more general theoretical results, e.g., for ordered choices.) For the binary case, the random effects treatment is untenable. The base case would be

 $y_{it} = \mathbf{1}[\boldsymbol{\beta}'\mathbf{x}_{it} + \delta y_{i,t-1} + \boldsymbol{\gamma}'\mathbf{z}_i + u_i + \varepsilon_{it} > 0].$ 

Since the common effect appears in every period,  $u_i$  cannot be treated as a random effect. A second complication is the 'initial conditions problem' (Heckman (1981)). The path of  $y_{it}$  will be determined at least partly (if not predominantly) by the value it took when the observation window opened. (The idea of initial conditions, itself, is confounded by the nature of the observation. It will rarely be the case that a process is observed from its beginning. Consider, for example, a model of insurance takeup or health status. Individuals have generally already participated in the process in periods before the observation begins. In order to proceed, it may be necessary to make some assumptions about the process, perhaps

that it has reached an equilibrium at time  $t_0$  when it is first observed. (See, e.g., Heckman (1981) and Wooldridge (2002).) Arellano and Honoré (2001) consider this in detail as well.

Analysis of binary choice with lagged dependent variables, such as Lee (2013, this volume) suggest that the incidental parameters problem is exacerbated by the lagged effects. See, e.g., Heckman (1981), Hahn and Kuersteiner (2002) and Fernandez Val (2009). Even under more restrictive assumptions, identification (and consistent estimation) of model parameters is complicated owing to the several sources of persistence in  $y_{it}$ , the heterogeneity itself and the state persistence induced by the lagged value. Analysis appears in Honoré and Kyriazidou (2000), Chamberlain (1992), Hahn (2001) and Hahn and Moon (2006).

Semiparametric approaches to dynamics in panel data discrete choice have provided fairly limited guidance. Arellano and Honoré (2001) examine two main cases, one in which the model contains only current and lagged dependent variables and a second, three period model that has one regressor for which the second and third periods are equal. Lee (2013) examines the multinomial logit model in similar terms. The results are suggestive, though perhaps more of methodological than practical interest. A practical approach is suggested by Heckman (1981), Hsiao (2003) and Wooldridge (2010) and Semikyna and Wooldridge (2010). In a model of the form

 $y_{it} = \mathbf{1}[\boldsymbol{\beta}'\mathbf{x}_{it} + \delta y_{i,t-1} + u_i + \varepsilon_{it} > 0],$ 

the starting point,  $y_{i0}$ , is likely to be crucially important to the subsequent sequence of outcomes, particularly if *T* is small. We condition explicitly on the history;

 $Prob(y_{it} = 1 | \mathbf{X}_{i}, u_{i}, y_{i,t-1}, \dots, y_{i1}, y_{i0}) = f[y_{it}, (\mathbf{\beta}' \mathbf{x}_{it} + \delta y_{i,t-1} + u_{i})].$ 

One might at this point take the initial outcome as exogenous and build up a likelihood,

$$f(y_{i1},...,y_{iT} | \mathbf{X}_{i},y_{i0},u_{i}) = \prod_{t=1}^{T} f[(2y_{it}-1)(\boldsymbol{\beta}'\mathbf{x}_{it}+\delta y_{i,t-1}+u_{i})],$$

then use the earlier methods to integrate  $u_i$  out of the function and proceed as in the familiar random effects fashion –  $y_{i0}$  appears in the first term. The complication is that it is implausible to assume the common effect out of the starting point and have it appear suddenly at t = 1, even if the process (for example, a labor force participation study that begins at graduation) begins at time 1. An approach suggested by Heckman (1981) and refined by Wooldridge (2005, 2010) is to form the joint distribution of the observed outcomes given ( $\mathbf{X}_{i}, y_{i0}$ ) and a plausible approximation to the marginal distribution  $f(u_i|y_{i0}, \mathbf{X}_i)$ . For example, if we depart from a probit model and use the Mundlak device to specify

$$u_i \mid y_{i0}, X_i \sim N[\eta + \theta' \overline{\mathbf{x}}_i + \lambda y_{i0}, \sigma_w^2]$$

then

$$y_{it} = \mathbf{I}[\boldsymbol{\beta}'\mathbf{x}_{it} + \delta y_{i,t-1} + \eta + \boldsymbol{\theta}'\mathbf{\overline{x}}_i + \lambda y_{i0} + w_i + \varepsilon_{it} > 0] .$$

(Some treatments, such as Chamberlain (1982), extend all of the rows of  $X_i$  individually rather than use the group means. This creates a problem for unbalanced panels and, for a large model with even moderately large *T* creates an uncomfortably long list of right hand side variables. Recent treatments have usually used the projection onto the means instead.) Wooldridge (2010, page 628) considers computation of average partial effects in this context. An application of these results to a dynamic random effects Poisson regression model appears in Wooldridge (2005). Contoyannis, Jones and Rice (2004) specified a random effects dynamic ordered probit model, as

$$h_{it}^* = \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\gamma}' \mathbf{h}_{i,t-1} + \alpha_i + \varepsilon_{it}$$
$$h_{it} = j \text{ if } \mu_{j-1} < h_{it}^* \le \mu_j$$
$$\alpha_i = \eta + \alpha_1' \mathbf{h}_{i0} + \alpha_2' \mathbf{x}_i + w_i$$

This is precisely the application suggested above (with the Mundlak device). One exception concerns the treatment of the lagged outcome. Here, since the outcome variable is the label of the interval in which  $h_{ii}$ \* falls,  $\mathbf{h}_{i,t}$  is a vector of *J* dummy variables for the *J*+1 possible outcomes (dropping one of them).

### V. Spatial Panels and Discrete Choice

The final class of models noted is spatial regression models. Spatial regression has been well developed for the linear regression model. The linear model with *spatial autoregression* is

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta} + \lambda \mathbf{W} \mathbf{y}_t + \boldsymbol{\varepsilon}_t$$

where the data indicated are a sample of *n* observations at time *t*. The panel data counterpart will consist of *T* such samples. The matrix **W** is the *spatial weight matrix*, or *contiguity matrix*. Nonzero elements  $w_{ij}$ define the two observations as neighbors. The relative magnitude of  $w_{ij}$  indicates how close the neighbors are. **W** is defined by the analyst. Rows of **W** are standardized to sum to one. The crucial parameter is the spatial autoregression coefficient,  $\lambda$ . The transformation to the spatial moving average form is

$$\mathbf{y}_t = (\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{X}_t \boldsymbol{\beta} + (\mathbf{I} - \lambda \mathbf{W})^{-1} \boldsymbol{\varepsilon}_t$$

This is a generalized regression with disturbance covariance matrix  $\Omega = \sigma^2 (\mathbf{I} - \lambda \mathbf{W})^{-1} (\mathbf{I} - \lambda \mathbf{W})^{-1}$ . Some discussion of the model formulation may be found, e.g., in Arbia (2006). An application to residential home sale prices is Bell and Bockstael (2006). Extension of this linear model to panel data is developed at length in Lee and Yu (2010). An application to UK mental health expenditures appears in Moscone, Knapp and Tosetti (2007).

Extensions of the spatial regression model to discrete choice are relatively scarce. A list of applications includes binary choice models Smirnov (2010), Pinske and Slade (1998), Bhat and Sener (2009), Klier and McMillen (2008) and Beron and Vijverberg (2004); a sample selection model applied to Alaskan trawlers by Flores Lagunes and Schnier (2012); an ordered probit analysis of accident severity by

Kockelman and Wang (2009); a spatial multinomial probit model in Chakir and Parent (2009) and, an environmental economics application to zero inflated counts by Rathbun and Fei (2006).

It is immediately apparent that if the spatial regression framework is applied to the underlying random utility specification in a discrete choice model that the density of the observable random vector,  $\mathbf{y}_t$  becomes intractable. In essence, the sample becomes one enormous fully autocorrelated observation. There is no transformation of the model that produces a tractable log likelihood. Each of the applications above develops a particular method of dealing with the issue. Smirnov, for example, separates the autocorrelation into 'public' and 'private' parts, and assumes that the public part is small enough to discard. There is no generally applicable methodology in this setting on the level of the general treatment of simple dynamics and latent heterogeneity that has connected the applications up to this point. We note, as well, that there are no received applications of spatial panel data to discrete choice models.

# References

Abrevaya, J., 1997. "The Equivalence of Two Estimators of the Fixed Effects Logit Model," *Economics Letters*, 55, 1, pp. 41-43.

Allenby, G, J. Garratt and P. Rossi, 2010. "A Model for Trade-Up and Change in Considered Brands," *Marketing Science*, 29, 1, pp. 40-56.

Allison, P. and R. Waterman, 2002. "Fixed Effects Negative Binomial Regression Models," *Sociological Methodology*, 32, pp. 247-256.

Altonji, J. and R. Matzkin, 2005. "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 3, pp. 1053-1102.

Arbia, G., 2006. Spatial Econometrics, Springer, Berlin.

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, pp. 277-297.

Arellano, M. And O. Bover, 1995."Another Look at the Instrumental-Variable Estimation of Error-Components Models", *Journal of Econometrics*, 68, pp. 29-51.

Arellano, M. and J. Hahn, 2007. "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," in R. Blundell, W. Newey, and T. Persson, eds.: *Advances in Economics and Econometrics, Ninth World Congress*, Volume III, Cambridge University Press, pp. 381-409.

Arellano, M. and B. Honoré, 2001. "Panel Data Models: Some Recent Developments," in: J. Heckman and E. Leamer, Eds): *Handbook of Econometrics*, Volume 5, Chapter 53, North-Holland, 2001, pp. 3229-3296.

Bago d'Uva T., 2006. "Latent Class Models for Utilization of Health Care," Health Economics 15, 4, pp. 329-343.

Balestra, P. and M. Nerlove, 1966. "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica*, 34, pp. 585-612.

Bell, K. and N. Bockstael, 2006. "Applying the Generalized Method of Moments Approach to Spatial Problems Involving Micro-Level Data," *Review of Economics and Statistics*, 82, 1, pp. 72-82.

Bera, A. and C. Jarque, 1982. "Model Specification Tests: A Simultaneous Approach," *Journal of Econometrics*, 20, pp. 59-82.

Bera, A. C. Jarque and L. Lee, 1984. 'Testing the Normality Assumption in Limited Dependent Variable Models," *International Economic Review*, 25, pp. 563-578.

Beron, K. and W. Vijverberg, 2004. "Probit in a Spatial Context: A Monte Carlo Analysis," in L. Anselin, R. Florax and S. Rey, eds. *Advances in Spatial Econometrics: Methodology, Tools and Applications*, , New York, Springer, pp. 169-195.

Berry, S., J. Levinsohn, and A. Pakes, 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63, 4, pp. 841–890.

Bertschuk, I., and M. Lechner, 1998. "Convenient Estimators for the Panel Probit Model." *Journal of Econometrics*, 87, 2, pp. 329–372.

Bester, C. and C. Hansen, 2009. "A Penalty Function Approach to Bias Reduction in Non-linear Panel Models with Fixed Effects," *Journal of Business and Economic Statistics*, 27, 2, pp. 131-148.

Bhat, C., 1999. "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," Manuscript, Department of Civil Engineering, University of Texas, Austin.

Bhat, C. and I. Sener, 2009 "A Copula Based Closed Form Binary Logit Choice Model for Accommodating Spatial Correlation Across Observational Units," *Journal of Geographical Systems*, 11, pp. 243–272.

Bhat, C., R. Paleti, and M. Castro, 2013. "A New Econometric Approach to Multivariate Count Data Modeling," Technical Paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Bhat, C. and V. Pulugurta, 1998. "A Comparison of Two Alternative Behavioral Mechanisms for Car Ownership Decisions", *Transportation Research Part B*, 32, 1, pp. 61-75.

Breusch, T., M. Ward, H. Nguyen, and T. Kompas, 2011, "On the Fixed-Effects Vector Decomposition," *Political Analysis*, 19, 2, pp. 123-134

Bryk, A. and S. Raudenbush, 2002. *Hierarchical Linear Models, Advanced Quantitative Techniques*, Sage, New York.

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

Bontemps, C., J. Racine and M. Simion, 2009. "Nonparametric vs. Parametric Binary Choice Models: An Empirical Investigation, in *Selected Papers at the Agricultural & Applied Economics Association AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin, July.* 

Brant, R., 1990. "Assessing Proportionality in the Proportional Odds Model for Ordered Logistic Regression." *Biometrics*, 46, pp. 1171–1178.

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

Chamberlain, G., 1980 "Analysis with Qualitative Data," Review of Economic Studies, 47, pp. 225-238.

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

Chamberlain, G., 1984, "Panel Data," in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Vol. 2, North Holland, pp. 4-46.

Chamberlain, G., 1992. "Binary Response Models for Panel Data: Identification and Information.," Unpublished Manuscript, Department of Economics, Harvard University.

Carro J. and A. Traferri, 2011. State Dependence and Heterogeneity in Health Using a Bias Corrected Fixed Effects Estimator," Journal of Applied Econometrics, 26, pp. 1-27.

Chakir, R. and O. Parent, 2009. "Determinants of Land Use Changes: A Spatial Multinomial Probit Approach," *Papers in Regional Science*, 88, 2, pp. 328-346.

Chen, S. and S. Khan, 2003. "Rates of Convergence for Estimating Regression Coefficients in Heteroscedastic Discrete Response Models," *Journal of Econometrics*, 117, pp. 245-278.

Chesher, A., 1984. "Testing for Neglected Heterogeneity," *Econometrica*, 52, 4, pp. 865-872.

Chesher, A., 2010 "Instrumental Variables Models for Discrete Outcomes", Econometrica, 78, pp. 575-601.

Chesher, A., 2013. "Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification", *Econometric Theory*, forthcoming.

Chesher, A. and M. Irish, 1987. "Residual Analysis in the Grouped Data and Censored Normal Linear Model," *Journal of Econometrics*, 34, pp. 33–62.

Chesher, A. and L. Lee, 1986. "Specification Testing When Score Test Statistics are Identically Zero," *Journal of Econometrics*, 31, 2, pp. 121-149.

Cox, D. and D. Hinkley, 1974. *Theoretical Statistics*, Chapman and Hall, London.

Chesher, A. and K. Smolinsky, 2012. "IV Models of Ordered Choice", Journal of Econometrics, 166, pp. 33-48.

Chesher, A. and A. Rosen, 2012a, "An Instrumental Variable Random Coefficients Model for Binary Outcomes," CeMMAP Working Paper CWP 34/12.

Chesher, A. and A. Rosen, 2012b. "Simultaneous Equations for Discrete Outcomes: Coherence, Completeness and Identification," CeMMAP Working Paper CWP 21/12.

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

Das, M., and A. van Soest. "A Panel Data Model for Subjective Information on Household Income Growth." *Journal of Economic Behavior and Organization*, 40, pp. 409–426.

Durlauf, S. and W. Brock, 2001a. "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68, 2, pp. 235-260.

Durlauf, S. and W. Brock, 2001b. "A Multinomial Choice Model with Neighborhood Effects," American Economic Review, 92, pp. 298-303.

Durlauf, S. and W. Brock, 2002. "Identification of Binary Choice Models with Social Interactions," *Journal of Econometrics*, 140, 1, pp. 52-75.

Durlauf, S., L. Blume, W. Brock and Y. Ioannides, 2010. "Identification of Social Interactions," in J. Benhabib, A. Bisin, and M. Jackson, eds., *Handbook of Social Economics*, Amsterdam: North Holland.

Elliott, G. and R. Leili, 2005. "Predicting Binary Outcomes," Unpublished Working paper, Department of Economics, UCSD.

Fernandez-Val, I., 2009. "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models," *Journal of Econometrics*, 150, 1, pp. 71-85.

Flores-Lagunes, A. and Schnier, K., 2012. "Sample Selection and Spatial Dependence," *Journal of Applied Econometrics*, 27, 2, pp. 173-204.

Goldberg, P., 1995. "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63, pp. 891-951.

Gravelle H., R. Jacobs, A. Jones, and A. Street, 2002. "Comparing the Efficiency of National Health Systems: A Sensitivity Approach," Manuscript, University of York, Health Economics Unit.

Greene, W., 1995. "Sample Selection in the Poisson Regression Model," Working Paper No. EC-95-6, Department of Economics, Stern School of Business, New York University.

Greene, W., 2004a. "Convenient Estimators for the Panel Probit Model." Empirical Economics, 29, 1, pp. 21-47.

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

Greene, W., 2011a. "Spatial Discrete Choice Models," Manuscript, Department of Economics, Stern School of Business, New York University, http://people.stern.nyu.edu/wgreene/SpatialDiscreteChoiceModels.pdf.

Greene, W., 2011b. "Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time Invariant Variables in Fixed Effects Models?" *Political Analysis*, 19, 2, pp. 135-146.

Greene, W., 2012. Econometric Analysis, 7th Ed., Prentice Hall, Upper Saddle River.

Greene, W. and D. Hensher, 2010. Modeling Ordered Choices, Cambridge University Press, Cambridge.

Greene, W. and C. McKenzie, 2012. "LM Tests for Random Effects," Working Paper EC-12-14, Department of Economics, Stern School of Business, New York University.

Hahn, J., 2001. "The Information Bound of a Dynamic Panel Logit Model with Fixed Effects," *Econometric Theory*, 17, pp. 913 - 932.

Hahn, J., 2004. "Does Jeffrey's Prior Alleviate the Incidental Parameters Problem?" *Economics Letters* 82, pp. 135-138.

Hahn, J., 2009. "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models," *Journal of Econometrics*, 150, 1, pp. 71-85.

Hahn, J., 2010, "Bounds on ATE with Discrete Outcomes," Economics Letters, 109, pp. 24-27.

Hahn, J., V. Chernozhukov, I. Fernandez-Val and W. Newey, 2013. "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, forthcoming.

Hahn, J., J. Ham and H. Moon, 2011. "Test of Random vs. Fixed Effects with Small Within Variation", *Economics Letters* 112, pp. 293-297.

Hahn, J., and G. Kuersteiner, 2002. "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica*, 70, pp. 1639-1657.

Hahn, J. and G. Kuersteiner, 2011. "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects", *Econometric Theory* 27, pp. 1152-1191.

Hahn, J. and J. Meinecke, 2005. "Time Invariant Regressor in Nonlinear Panel Model with Fixed Effects", *Econometric Theory*, 21, pp. 455-469.

Hahn, J. and H. Moon, 2006. "Reducing Bias of MLE in a Dynamic Panel Model", *Econometric Theory* 22, pp. 499-512.

Hahn, J. and W. Newey, 1994. "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", *Econometrica* 72, pp. 1295-1319.

Hahn, J., and W. Newey, 2002. "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Unpublished Manuscript, Department of Economics, UCLA.

Harris, M., B. Hollingsworth and W. Greene, 2012. "Inflated Measures of Self Assessed Health, Manuscript, School of Business, Curtin University.

Harris M. and Y. Zhao, 2007. "Modeling Tobacco Consumption with a Zero Inflated Ordered Probit Model," *Journal of Econometrics*, 141, pp.1073-99

Hausman, J., 1978. "Specification Tests in Econometrics." Econometrica, 46, pp. 1251–1271.

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

Heckman, J., 1979. "Sample Selection Bias as a Specification Error." Econometrica, 47, 1979, pp. 153–161.

Heckman, J. 1981 "Statistical Models for Discrete Panel Data." In C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.

Heckman, J., and B. Singer, 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, pp. 271–320.

Hensher, D. and W. Greene, 2003. "The Mixed Logit Model: The State of Practice," *Transportation Research, B*, 30, pp. 133-176;

Hensher, D., J. Rose, and W. Greene, 2006. Applied Choice Analysis, Cambridge University Press, Cambridge.

Hoderlein, S., E. Mammen and K. Yu, 2011. "Nonparametric Models in Binary Choice Fixed Effects Panel Data," *Econometrics Journal*, 14, 3, pp. 351-367.

Honoré, B. and E. Kyriazidou, 2000a. "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68, 4, pp. 839 - 874.

Honoré, B. and E. Kyriazidou, 2000b. "Estimation of Tobit-type Models with Individual Specific Effects," *Econometric Reviews* 19, pp. 341 - 366.

Honoré, B., 2002, "Nonlinear Models with Panel Data," Portuguese Economic Journal, 1, 2, pp. 163-179.

Horowitz, J., 1992. "A Smoothed Maximum Score Estimator for the Binary Response Model." *Econometrica*, 60, pp. 505–531.

Horowitz, J., 1993. "Semiparametric Estimation of a Work-Trip Mode Choice Model." *Journal of Econometrics*, 58, pp. 49–70.

Hsiao, C., 2003. Analysis of Panel Data, 2nd ed. New York: Cambridge University Press, 2003.

Katz E., 2001. "Bias in Conditional and Unconditional Fixed Effects Logit Estimation," *Political Analysis*, 9, 4, pp. 379-84.

Keane, M., 2013. "Discrete Choice Models of Consumer Demand for Panel Data," in B. Baltagi, ed., *Oxford Handbook of Panel Data*, Oxford University Press, Oxford (this volume).

Klein, R. and R. Spady, 1993. "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61, pp. 387-421.

Klier, T. and D. McMillen, 2008. "Clustering of Auto Supplier Plants in the United States: Generalized Method of Moments Spatial Logit for Large Samples," *Journal of Business and Economic Statistics*, 26, 4, pp. 460-471.

Kockelman, K and C. Wang, 2009. "Bayesian Inference for Ordered Response Data with a Dynamic Spatial Ordered Probit Model," Working Paper, Department of Civil and Environmental Engineering, Bucknell University.

Koop, G., J. Osiewalski, and M. Steel, 1997. "Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontiers," *Journal of Econometrics*, 76, pp. 77-106.

Krailo, M., and M. Pike, 1984. "Conditional Multivariate Logistic Analysis of Stratified Case-Control Studies." *Applied Statistics*, 44, 1, pp. 95–103.

Laisney, F. and M. Lechner, 2002. "Almost Consistent Estimation of Panel Probit Models with 'Small' Fixed Effects," ZEW Zentrum Discussion Paper No. 2002-64, ftp://ftp.zew.de/pub/zew-docs/dp/dp0264.pdf.

Lancaster, T., 1999. "Panel Binary Choice with Fixed Effects", unpublished discussion paper, Brown University.

Lancaster, T., 2000. "The Incidental Parameter Problem Since 1948", Journal of Econometrics, 95, pp. 391-413.

Lancaster, T., 2001. "Orthogonal Parameters and Panel Data", unpublished discussion paper, Brown University.

Lee, L. and J. Yu, 2010. "Estimation of Spatial Panels," Foundation and Trends in Econometrics, 4:1-2.

Lee, M., 2013. "Panel Conditional and Multinomial Logit," in B. Baltagi, ed., *Oxford Handbook of Panel Data*, Oxford University Press, Oxford (this volume).

Maddala, G., 1983. *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.

Manski, C., 1975. "The Maximum Score Estimator of the Stochastic Utility Model of Choice." *Journal of Econometrics*, 3, pp. 205–228.

Manski, C., 1985. "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator." *Journal of Econometrics*, 27, pp. 313–333.

Manski, C., 1986. "Operational Characteristics of the Maximum Score Estimator." *Journal of Econometrics*, 32, pp. 85–100.

Manski, C., 1987. "Semiparametric Analysis of the Random Effects Linear Model from Binary Response Data," *Econometrica*, 55, pp. 357–362.

Matzkin, R., 1991. "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, 59, 5, pp. 1315-1327.

Matzkin, R., 2005. "Identification of Consumers' Preferences when Individuals' Choices are Unobservable," *Economic Theory*, 26, 2, pp. 423-443.

McFadden, D., 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press, 1974.

McFadden, D. and K. Train, 2000. "Mixed MNL Models for Discrete Choice," *Journal of Applied Econometrics*, 15, 447-70.

Moscone, F., M. Knapp, and E. Tosetti, 2007. "Mental Health Expenditures in England: A Spatial Panel Approach." *Journal of Health Economics*, 26, 4, pp. 842-864.

Mullahy, J., 1987. "Specification and Testing of Some Modified Count Data Models." *Journal of Econometrics*, 33, pp. 341–365

Mundlak, Y. "On the Pooling of Time Series and Cross Sectional Data." Econometrica, 56, 1978, pp. 69-86.

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

Pinske, J. and M. Slade, 1998. "Contracting in Space: An Application of Spatial Statistics to Discrete Choice Models," *Journal of Econometrics*, 85, pp. 125-154.

Plümper, T. and V. Troeger, 2007. "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects," *Political Analysis*, 15, 2, pp. 124-139.

Plümper, T. and V. Troeger, 2011. "Fixed-Effects Vector Decomposition: Properties, Reliability, and Instruments," *Political Analysis*, 19, 2, pp. 147-164.

Pudney, S., and M. Shields, 2000. "Gender, Race, Pay and Promotion in the British Nursing Profession: Estimation of a Generalized Ordered Probit Model." *Journal of Applied Econometrics*, 15, 4, pp. 367–399.

Racine, J., 2008. "Nonparametric Econometrics: A Primer," Foundations and Trends in Econometrics, 3, 1.

Rasch, G., 1960. "Probabilistic Models for Some Intelligence and Attainment Tests." *Denmark Paedogiska*, Copenhagen.

Rathbun, S and L. Fei, 2006. "A Spatial Zero-Inflated Poisson Regression Model for Oak Regeneration," *Environmental Ecology Statistics*, 13, pp. 409-426.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A., 2005. "Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects," *Journal of Econometrics*, 128, pp. 301-323.

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

Schmidheiny, K. and M. Brülhart, 2011. "On the Equivalence of Location Choice Models: Conditional Logit, Nested Logit and Poisson." *Journal of Urban Economics*, 69, 2, pp. 214-222.

Semykina, A. and J. Wooldridge, J., 2013. "Estimation of Dynamic Panel Data Models with Sample Selection," Journal of Applied Econometrics," 28, 1, pp. 47-61.

Smirnov, A., 2010. "Modeling Spatial Discrete Choice," *Regional Science and Urban Economics*, 40, 5, pp. 292-298

Train, K., 2003. Discrete Choice Methods with Simulation, Cambridge: Cambridge University Press.

Train, K. 2010. Discrete Choice Methods with Simulation, 2<sup>nd</sup> edition. Cambridge: Cambridge University Press.

Van dijk R., D. Fok and R. Paap, 2007. "A Rank-Ordered Logit Model with Unobserved Heterogeneity in Ranking Capabilities," Econometric Institute, Erasmus University, Report 2007-07.

Verbeek, M., 2000. A Guide to Modern Econometrics, Wiley, Chichester.

Verbeek, M., and T. Nijman, 1992. "Testing for Selectivity Bias in Panel Data Models." *International Economic Review*, 33, 3, pp. 681–703.

World Health Organization, 2000. The World Health Report, 2000, Health Systems: Improving Performance. WHO, Geneva.

Wooldridge, J., 2002. "Inverse Probability Weighted M-Estimators for Sample Selection, Attrition, and Stratification," *Portuguese Economic Journal* 1, pp. 117-139.

Wooldridge, J., 2003. "Cluster-Sample Methods in Applied Econometrics," *American Economic Review* 93, pp. 133-138.

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

Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data, 2<sup>nd</sup> ed., MIT Press, Cambridge.

Wooldridge, J., 2013. "Estimation of Dynamic Panel Data Models with Sample Selection," *Journal of Applied Econometrics*, 28, 1, pp. 47-61.

Wu, D., 1973. "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, 41, pp. 733-750.