New Developments in Econometrics
Lecture 6: Nonlinear Panel Data Models

Jeff Wooldridge

Cemmap Lectures, UCL, June 2009

1. Introduction
2. General Setup and Quantities of Interest
3. Exogeneity Assumptions
4. Conditional Independence
5. Assumptions about the Unobserved Heterogeneity
6. Dynamic Models
7. Estimating Popular Models

1. Introduction

- Consider a static, unobserved effects probit model for panel data:

$$P(y_{it} = 1|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i),\ t = 1,\ldots,T. \tag{1}$$

What are the quantities of interest for most purposes? Possibilities: (i) The element of $\boldsymbol{\beta}$, the β_j. These give the directions of the partial effects of the covariates on the response probability. For any two continuous covariates, the ratio of coefficients, β_j/β_h, is identical to the ratio of partial effects (and the ratio does not depend on the covariates or unobserved heterogeneity, c_i).

(ii) The magnitudes of the partial effects. These depend not only on the value of the covariates, say $\mathbf{x}_t$, but also on the value of the unobserved heterogeneity. In the continuous covariate case,

$$\frac{\partial P(y_t = 1|\mathbf{x}_t, c)}{\partial x_{tj}} = \beta_j \phi(\mathbf{x}_t \boldsymbol{\beta} + c). \tag{2}$$

• Questions: (i) Assuming we can estimate $\boldsymbol{\beta}$, what should we do about the unobservable c? (ii) If we can only estimate $\boldsymbol{\beta}$ up-to-scale, can we still learn something useful about magnitudes of partial effects? (iii) What kinds of assumptions do we need to estimate partial effects?

2. General Setup and Quantities of Interest

- Let $\{(\mathbf{x}_{it}, y_{it}) : t = 1, \ldots, T\}$ be a random draw from the cross section.

Suppose we are interested in

$$E(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i) = m_t(\mathbf{x}_{it}, \mathbf{c}_i). \tag{3}$$

$\mathbf{c}_i$ can be a vector of unobserved heterogeneity.

- Partial effects: if x_{tj} is continuous, then

$$\theta_j(\mathbf{x}_t, \mathbf{c}) \equiv \frac{\partial m_t(\mathbf{x}_t, \mathbf{c})}{\partial x_{tj}}, \tag{4}$$

or discrete changes.

- How do we account for unobserved $\mathbf{c}_i$? If we know enough about the distribution of $\mathbf{c}_i$ we can insert meaningful values for $\mathbf{c}$. For example, if $\boldsymbol{\mu}_{\mathbf{c}} = E(\mathbf{c}_i)$, then we can compute the *partial effect at the average (PEA)*,

$$PEA_j(\mathbf{x}_t) = \theta_j(\mathbf{x}_t, \boldsymbol{\mu}_{\mathbf{c}}). \tag{5}$$

Of course, we need to estimate the function m_t and $\boldsymbol{\mu}_{\mathbf{c}}$. If we can estimate the distribution of $\mathbf{c}_i$, or features in addition to its mean, we can insert different quantiles, or a certain number of standard deviations from the mean.

- Alternatively, we can obtain the *average partial effect* (APE) (or *population average effect*) by averaging across the distribution of $\mathbf{c}_i$:

$$APE(\mathbf{x}_t) = E_{\mathbf{c}_i}[\theta_j(\mathbf{x}_t, \mathbf{c}_i)]. \tag{6}$$

The difference between (5) and (6) can be nontrivial. In some leading cases, (6) is identified while (5) is not. (6) is closely related to the notion of the *average structural function* (ASF) (Blundell and Powell (2003)). The ASF is defined as

$$ASF(\mathbf{x}_t) = E_{\mathbf{c}_i}[m_t(\mathbf{x}_t, \mathbf{c}_i)]. \tag{7}$$

- Passing the derivative through the expectation in (7) gives the APE.

• How do APEs relate to parameters? Index model:

$$m_t(\mathbf{x}_t, c) = G(\mathbf{x}_t \boldsymbol{\beta} + c), \tag{8}$$

where $G(\cdot)$ is differentiable. Then

$$APE(\mathbf{x}_t) = \beta_j E_{c_i}[g(\mathbf{x}_t \boldsymbol{\beta} + c_i)], \tag{9}$$

where $g(\cdot)$ is the derivative of $G(\cdot)$. Even if $G(\cdot)$ is known, magnitude of effects cannot be estimated without making assumptions about the distribution of c_i

• Important: Definitions of partial effects do not depend on whether $\mathbf{x}_{it}$ is correlated with $\mathbf{c}_i$. Of course, whether and how we estimate them certainly does.

3. Exogeneity Assumptions

- As in linear case, cannot get by with just specifying a model for the contemporaneous conditional distribution, $D(y_{it}|\mathbf{x}_{it},\mathbf{c}_i)$.
- The most useful definition of strict exogeneity for nonlinear panel data models is

$$D(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\mathbf{c}_i) = D(y_{it}|\mathbf{x}_{it},\mathbf{c}_i). \tag{10}$$

Chamberlain (1984) labeled (10) *strict exogeneity conditional on the unobserved effects* $\mathbf{c}_i$. Conditional mean version:

$$E(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\mathbf{c}_i) = E(y_{it}|\mathbf{x}_{it},\mathbf{c}_i). \tag{11}$$

- The *sequential exogeneity* assumption is

$$D(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{it},\mathbf{c}_i) = D(y_{it}|\mathbf{x}_{it},\mathbf{c}_i). \tag{12}$$

Unfortunately, it is much more difficult to allow sequential exogeneity in in nonlinear models. (Most progress has been made for lagged dependent variables or specific functional forms, such as exponential.)

- Neither strict nor sequential exogeneity allows for contemporaneous endogeneity of one or more elements of $\mathbf{x}_{it}$, where, say, x_{itj} is correlated with unobserved, time-varying unobservables that affect y_{it}.

4. Conditional Independence

- In linear models, serial dependence of idiosyncratic shocks is easily dealt with, either by “cluster robust” inference or Generalized Least Squares extensions of Fixed Effects and First Differencing. With strictly exogenous covariates, serial correlation never results in inconsistent estimation, even if improperly modeled. The situation is different with most nonlinear models estimated by MLE.
- *Conditional independence* (CI) (under strict exogeneity):

$$D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i,\mathbf{c}_i) = \prod_{t=1}^{T} D(y_{it}|\mathbf{x}_{it},\mathbf{c}_i). \tag{13}$$

• In a parametric context, the CI assumption reduces our task to specifying a model for $D(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i)$, and then determining how to treat the unobserved heterogeneity, $\mathbf{c}_i$.

• In random effects and correlated random frameworks (next section), CI plays a critical role in being able to estimate the “structural” parameters and the parameters in the distribution of $\mathbf{c}_i$ (and therefore, in estimating PEAs). In a broad class of popular models, CI plays no essential role in estimating APEs.

5. Assumptions about the Unobserved Heterogeneity

Random Effects

- Generally stated, the key RE assumption is

$$D(\mathbf{c}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = D(\mathbf{c}_i). \tag{14}$$

Under (14), the APEs are actually nonparametrically identified from

$$r_t(\mathbf{x}_t) \equiv E(y_{it}|\mathbf{x}_{it} = \mathbf{x}_t). \tag{15}$$

- In some leading cases (RE probit and RE Tobit with heterogeneity normally distributed), if we want PEs for different values of **c**, we must assume more: strict exogeneity, conditional independence, and (14) with a parametric distribution for $D(\mathbf{c}_i)$.

Correlated Random Effects

A CRE framework allows dependence between $\mathbf{c}_i$ and $\mathbf{x}_i$, but restricted in some way. In a parametric setting, we specify a distribution for $D(\mathbf{c}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})$, as in Chamberlain (1980,1982), and much work since. Distributional assumptions that lead to simple estimation – homoskedastic normal with a linear conditional mean — can be restrictive.

- Possible to drop parametric assumptions and just assume

$$D(c_i|\mathbf{x}_i) = D(c_i|\bar{\mathbf{x}}_i), \tag{16}$$

without restricting $D(c_i|\bar{\mathbf{x}}_i)$. Altonji and Matzkin (2005, Econometrica).

- Other functions of $\{\mathbf{x}_{it} : t = 1,\ldots,T\}$ are possible.

• APEs are identified very generally. For example, under (16), a consistent estimate of the average structural function is

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1} \sum_{i=1}^{N} r_t(\mathbf{x}_t, \bar{\mathbf{x}}_i), \tag{17}$$

where $r_t(\bullet)$ is now the mean function $E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i)$.

• Need a random sample $\{\bar{\mathbf{x}}_i : i = 1, \ldots, N\}$ for the averaging out to work.

Fixed Effects

• The label "fixed effects" is used in different ways by different researchers. One view: $\mathbf{c}_i,\ i = 1,\ldots,N$ are parameters to be estimated. Usually leads to an "incidental parameters problem."

• Second meaning of "fixed effects": $D(\mathbf{c}_i|\mathbf{x}_i)$ is unrestricted and we look for objective functions that do not depend on $\mathbf{c}_i$ but still identify the population parameters. Leads to "conditional MLE" if we can find "sufficient statistics" $\mathbf{s}_i$ such that

$$D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i,\mathbf{c}_i,\mathbf{s}_i) = D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i,\mathbf{s}_i). \tag{18}$$

• Conditional Independence is usually maintained.

• Key point: PEAs and APEs are generally unidentified.

6. Dynamic Models

- Nonlinear models with only sequentially exogenous variables are difficult to deal with. More is known about models with lagged dependent variables and otherwise strictly exogenous variables:

$$D(\mathbf{y}_{it}|\mathbf{z}_{it},\mathbf{y}_{i,t-1},\ldots,\mathbf{z}_{i1},\mathbf{y}_{i0},\mathbf{c}_i),\ t=1,\ldots,T, \tag{19}$$

which we assume also is $D(\mathbf{y}_{it}|\mathbf{z}_i,\mathbf{y}_{i,t-1},\ldots,\mathbf{y}_{i1},\mathbf{y}_{i0},\mathbf{c}_i)$. Suppose this distribution depends only on $(\mathbf{z}_{it},\mathbf{y}_{i,t-1},\mathbf{c}_i)$ with density $f_t(\mathbf{y}_t|\mathbf{z}_t,\mathbf{y}_{t-1},\mathbf{c};\boldsymbol{\theta})$. The joint density of $(\mathbf{y}_{i1},\ldots,\mathbf{y}_{iT})$ given $(\mathbf{y}_{i0},\mathbf{z}_i,\mathbf{c}_i)$ is

$$\prod_{t=1}^{T} f_t(\mathbf{y}_t|\mathbf{z}_t,\mathbf{y}_{t-1},\mathbf{c};\boldsymbol{\theta}). \tag{20}$$

- How do we deal with $\mathbf{c}_i$ along with the initial condition, $\mathbf{y}_{i0}$? Various approaches have been suggested. One that meshes well with Stata's built-in commands (random effects probit, Tobit, count) was proposed by Wooldridge (2005, Journal of Applied Econometrics). Idea is to model $D(\mathbf{c}_i|\mathbf{y}_{i0}, \mathbf{z}_i)$ directly. Leads to $D(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iT}|\mathbf{y}_{i0}, \mathbf{z}_i)$ and MLE conditional on $(\mathbf{y}_{i0}, \mathbf{z}_i)$. This can be computationally simple for popular models, and can be made somewhat flexible in $D(\mathbf{c}_i|\mathbf{y}_{i0}, \mathbf{z}_i)$.
- The APEs for the conditional mean are easy to obtain.

7. **Estimating Popular Models**

- Pooled and random effects estimation commands in Stata (for probit, Tobit, Poisson, GLM, GEE) often can be used.
- Stata "egen" command for generating time averages. Need leads and lags of exogenous variables, and the initial condition, for dynamic models.
- For pooled methods, use the "panel bootstrap" feature in Stata to obtain standard errors or confidence intervals.
- Computational time is an issue for dynamic models because it uses full "random effects" with lots of covariates.

7.1 Binary and Fractional Response

• Unobserved effects (UE) "probit" model. For a binary or fractional y_{it},

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i),\ t = 1,\ldots,T. \tag{26}$$

Assume strict exogeneity (conditional on c_i) and Chamberlain-Mundlak device:

$$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i,\ a_i|\mathbf{x}_i \sim Normal(0,\sigma_a^2). \tag{27}$$

• In binary response case under serial independence, all parameters are identified and MLE (Stata: xtprobit) can be used. Just add the time averages $\bar{\mathbf{x}}_i$ as an additional set of regressors. Then $\hat{\mu}_c = \hat{\psi} + \bar{\mathbf{x}}\hat{\boldsymbol{\xi}}$ and $\hat{\sigma}_c^2 \equiv \hat{\boldsymbol{\xi}}'\left[N^{-1}\sum_{i=1}^{N}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\right]\hat{\boldsymbol{\xi}} + \hat{\sigma}_a^2$. Can evaluate PEs at, say, $\hat{\mu}_c \pm k\hat{\sigma}_c$.

• Only under restrictive assumptions does c_i have an unconditional normal distribution, although it becomes more reasonable as T gets large.

• Simple to test $H_0 : \boldsymbol{\xi} = \mathbf{0}$ as null that c_i, $\bar{\mathbf{x}}_i$ are independent.

- The APEs are identified from the ASF, estimated as

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{x}_t \hat{\boldsymbol{\beta}}_a + \hat{\psi}_a + \bar{\mathbf{x}}_i \hat{\boldsymbol{\xi}}_a) \tag{28}$$

where, for example, $\hat{\boldsymbol{\beta}}_a = \hat{\boldsymbol{\beta}}/(1 + \hat{\sigma}_a^2)^{1/2}$.

- For binary or fractional response, APEs are identified without the conditional serial independence assumption. Use pooled Bernoulli quasi-MLE (Stata: glm) or generalized estimating equations (Stata: xtgee) to estimate scaled coefficients based on

$$E(y_{it}|\mathbf{x}_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta}_a + \psi_a + \bar{\mathbf{x}}_i \boldsymbol{\xi}_a). \tag{29}$$

(Time dummies have been supressed for simplicity.)

- A more radical suggestion, but in the spirit of Altonji and Matzkin (2005), is to just use a flexible model for $E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i)$ directly, say,

$$
\begin{aligned}
E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = {} & \Phi[\theta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\gamma} + \\
& (\bar{\mathbf{x}}_i \otimes \bar{\mathbf{x}}_i)\boldsymbol{\delta} + (\mathbf{x}_{it} \otimes \bar{\mathbf{x}}_i)\boldsymbol{\eta}].
\end{aligned} \tag{30}
$$

Just average out over $\bar{\mathbf{x}}_i$ to get APEs.

- If we have a binary response, start with

$$
P(y_{it} = 1|\mathbf{x}_{it}, c_i) = \Lambda(\mathbf{x}_{it}\boldsymbol{\beta} + c_i), \tag{31}
$$

and assume CI, we can estimate $\boldsymbol{\beta}$ by FE logit without restricting $D(c_i|\mathbf{x}_i)$.

- In any nonlinear model using the Mundlak assumption $D(c_i|\mathbf{x}_i) = D(c_i|\bar{\mathbf{x}}_i)$, if $T \geq 3$ can include lead values, $\mathbf{w}_{i,t+1}$, to simply test strict exogeneity.

- Example: Married Women's Labor Force Participation: $N = 5,663$, $T = 5$ (four-month intervals).

- Following results include a full set of time period dummies (not reported).

- The APEs are directly comparable across models, and can be compared with the linear model coefficients.

LFP	(1)	(2)		(3)		(4)		(5)
Model	Linear	Probit		CRE Probit		CRE Probit		FE Logit
Est. Method	FE	Pooled MLE		Pooled MLE		MLE		MLE
	Coef.	Coef.	APE	Coef.	APE	Coef.	APE	Coef.
kids	−.0389	−.199	−.0660	−.117	−.0389	−.317	−.0403	−.644
	(.0092)	(.015)	(.0048)	(.027)	(.0085)	(.062)	(.0104)	(.125)
lhinc	−.0089	−.211	−.0701	−.029	−.0095	−.078	−.0099	−.184
	(.0046)	(.024)	(.0079)	(.014)	(.0048)	(.041)	(.0055)	(.083)
$\overline{kids}$	—	—	—	−.086	—	−.210	—	—
	—	—	—	(.031)	—	(.071)	—	—
$\overline{lhinc}$	—	—	—	−.250	—	−.646	—	—
	—	—	—	(.035)	—	(.079)	—	—

- Simple dynamic model (for binary only, not fractional response):

$$P(y_{it} = 1|\mathbf{z}_{it}, y_{i,t-1}, c_i) = \Phi(\mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + c_i). \tag{32}$$

A simple analysis is available if we specify

$$c_i|\mathbf{z}_i, y_{i0} \sim Normal(\psi + \xi_0 y_{i0} + \mathbf{z}_i\xi, \sigma_a^2) \tag{33}$$

Then

$$\begin{aligned} P(y_{it} = 1|\mathbf{z}_i, y_{i,t-1}, \ldots, y_{i0}, a_i) = \\ \Phi(\mathbf{z}_{it}\boldsymbol{\delta} + \rho y_{i,t-1} + \psi + \xi_0 y_{i0} + \mathbf{z}_i\boldsymbol{\xi} + a_i), \end{aligned} \tag{34}$$

where $a_i \equiv c_i - \psi - \xi_0 y_{i0} - \mathbf{z}_i\boldsymbol{\xi}$.

• Can use standard RE probit software (Stata: xtprobit), with explanatory variables $(1, \mathbf{z}_{it}, y_{i,t-1}, y_{i0}, \mathbf{z}_i)$ in time period t. Get APEs from ASF:

$$\widehat{ASF}(\mathbf{z}_t, y_{t-1}) = N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{z}_t \hat{\boldsymbol{\delta}}_a + \hat{\rho}_a y_{t-1} + \hat{\psi}_a + \hat{\xi}_{a0} y_{i0} + \mathbf{z}_i \hat{\boldsymbol{\xi}}_a), \tag{35}$$

with coefficients scaled by $(1 + \hat{\sigma}_a^2)^{-1/2}$.

• Labor force participation example with $N = 5,663$ and $T = 5$. The estimated APE is .260 (.026). It is .837 (.005) ignoring heterogeneity. Standard errors from 500 panel bootstrap replications.

- For estimating parameters, Honoré and Kyriazidou (2000) extend an idea of Chamberlain. With four time periods, $t = 0, 1, 2$, and 3, the conditioning that removes c_i requires $\mathbf{z}_{i2} = \mathbf{z}_{i3}$. HK show how to use a local version of this condition, that is, $\mathbf{z}_{i2} \approx \mathbf{z}_{i3}$, to consistenty estimate the parameters. The estimator is also asymptotically normal, but converges more slowly than the usual $\sqrt{N}$-rate.
- The condition that $\mathbf{z}_{i2} - \mathbf{z}_{i3}$ have a distribution with support around zero rules out aggregate year dummies. By design, cannot estimate magnitudes of effects.

Count and Other Multiplicative Models

- Conditional mean with multiplicative heterogeneity:

$$E(y_{it}|\mathbf{x}_{it}, c_i) = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta}) \tag{36}$$

where $c_i \geq 0$. Under strict exogeneity,

$$E(y_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i), \tag{37}$$

the "fixed effects" Poisson estimator is attractive: it does not restrict $D(y_{it}|\mathbf{x}_i, c_i)$, $D(c_i|\mathbf{x}_i)$, or serial dependence.

- The FE Poisson estimator is the conditional MLE derived under a Poisson and conditional independence assumptions. It is one of the rare cares where treating the c_i as parameters to estimate gives a consistent estimator of $\boldsymbol{\beta}$.
- The FE Poisson estimator is fully robust to any distributional failure (except (37)) and serial correlation. y_{it} does not even have to be is not a count variable! Fully robust inference is easy (though not currently built into Stata).

• Estimation under sequential exogeneity has been studied by Chamberlain (1992). Use moment conditions such as

$$E(y_{it}|\mathbf{x}_{it},\ldots,\mathbf{x}_{i1},c_i) = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta}). \tag{38}$$

Under this assumption, it can be shown that

$$E\{[y_{it} - y_{i,t+1}\exp((\mathbf{x}_{it} - \mathbf{x}_{i,t+1})\boldsymbol{\beta})|\mathbf{x}_{it},\ldots,\mathbf{x}_{i1}] = 0, \tag{39}$$

and, because these moment conditions depend only on observed data and the parameter vector $\boldsymbol{\beta}$, GMM can be used to estimate $\boldsymbol{\beta}$, and fully robust inference is straightforward.

- Wooldridge (2005, Journal of Applied Econometrics) shows how a dynamic Poisson model with conditional Gamma heterogeneity can be easily estimated in, say, Stata. In this case, one must take the distributional assumptions – for $D(y_{it}|\mathbf{z}_{it}, y_{it}, c_i)$ and $D(c_i|y_{i0}, \mathbf{z}_i)$ – seriously. (How seriously?)
- The same kinds of approaches – adding time averages in the strictly exogenous case, modeling $D(c_i|y_{i0}, \mathbf{z}_i)$ as Normally distributed in the dynamic cases – leads to very similar methods for Tobit models with corners (usually at zero).

- How can we handle heterogeneity and contemporaneously endogenous explanatory variables? There are GMM and control function approaches.
- A GMM approach – which extends Windmeijer (2002) – modifies the moment conditions under a sequential exogeneity assumption on instruments (rather than just explanatory variables):

$$y_{it} = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta})r_{it} \tag{40}$$

$$E(r_{it}|\mathbf{z}_{it},\ldots,\mathbf{z}_{i1},c_i) = 1, \tag{41}$$

which contains the sequentially exogenous case as a special case (with $\mathbf{z}_{it} = \mathbf{x}_{it}$).

- Now start with the transformation

$$\frac{y_{it}}{\exp(\mathbf{x}_{it}\boldsymbol{\beta})} - \frac{y_{i,t+1}}{\exp(\mathbf{x}_{i,t+1}\boldsymbol{\beta})} = c_i(r_{it} - r_{i,t+1}). \tag{42}$$

This is the starting point with sequential exogeneity but then with a multiplication by $\exp(\mathbf{x}_{it}\boldsymbol{\beta})$. In the sequential exogeneity case, $E(r_{it}|\mathbf{x}_{it},\ldots,\mathbf{x}_{i1},c_i) = E(r_{i,t+1}|\mathbf{x}_{it},\ldots,\mathbf{x}_{i1},c_i) = 1$, and so multiplying the moment conditions by any function of $\mathbf{x}_{it}$ is allowed. But not if $\mathbf{x}_{it}$ is contemporaneously endogenous (correlated with r_{it}).

- Using the moment conditions

$$E\left[\frac{y_{it}}{\exp(\mathbf{x}_{it}\boldsymbol{\beta})} - \frac{y_{i,t+1}}{\exp(\mathbf{x}_{i,t+1}\boldsymbol{\beta})}\middle|\mathbf{z}_{it},\ldots,\mathbf{z}_{i1}\right] = 0,\ t = 1,\ldots,T-1 \qquad (43)$$

generally causes computational problems. For example, if $x_{itj} \geq 0$ for some j and all i and t – for example, if x_{itj} is a time dummy – then the moment conditions can be made arbitarily close to zero by choosing β_j larger and larger.

- Windmeijer (2002, Economics Letters) suggested multiplying through by $\exp(\boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\beta})$ where $\boldsymbol{\mu}_{\mathbf{x}} = T^{-1}\sum_{r=1}^{T} E(\mathbf{x}_{ir})$.

- So, the modified moment conditions are

$$E\left[\frac{y_{it}}{\exp[(\mathbf{x}_{it}-\boldsymbol{\mu}_{\mathbf{x}})\boldsymbol{\beta}]}-\frac{y_{i,t+1}}{\exp[(\mathbf{x}_{i,t+1}-\boldsymbol{\mu}_{\mathbf{x}})\boldsymbol{\beta}]}\,\middle|\,\mathbf{z}_{it},\ldots,\mathbf{z}_{i1}\right]=0. \tag{44}$$

- As a practical matter, replace $\boldsymbol{\mu}_{\mathbf{x}}$ with the overall sample average,

$$\bar{\mathbf{x}}=(NT)^{-1}\sum_{i=1}^{N}\sum_{r=1}^{T}\mathbf{x}_{ir}. \tag{45}$$

- The deviated variables, $\mathbf{x}_{it}-\bar{\mathbf{x}}$, will always take on positive and negative values, and this seems to solve the GMM computational problem. (But more work could be done on this, especially in models with time dummies.)