EUSERS SUMMER SCHOOL

Performance and Governance of Services of General Interest.

Critical perspectives on Energy, Telecommunications, Transport and Water Reforms in the EU

***June, 27th – July 1st* 2016**

Some more tools for empirical analysis of SGI

Carlo Fiorio

Milan, June 29th, 2016

JEAN MONNET NETWORK

Co-funded by the European Union

Advantages of panel data

- Panel data sets are typically **larger** than cross-sectional or time series data sets, and **explanatory variables vary over two dimensions** (individuals and time) rather than one, estimators based on panel data are quite often **more accurate** than from other sources.
- Even with identical sample sizes, the use of a panel data set will often yield **more efficient estimators** than a series of independent cross-sections.
- If one is interested in changes from one period to another, a panel will yield more efficient estimators than a series of cross-sections.

Advantages of panel data

- Among the major advantages of panel data is the ability to **model individual dynamics**.
- Many economic models suggest that current behaviour depends upon **past behaviour** (persistence, habit formation, partial adjustment, etc.), so in many cases we would like to estimate a dynamic model on an individual level.
- The ability to do so is unique for panel data.

Autoregressive Panel Data Model (Dynamic)

- Consider the **linear dynamic model** with exogenous variables and a lagged dependent variable, that is
$$y_{it} = x_{it}'\beta + \gamma y_{i,t-1} + \alpha_i + \epsilon_{it}$$
where it is assumed that ε_{it} is $IID(0, \sigma_\epsilon)$.
- In the static model, we have seen arguments of consistency (robustness) and efficiency for choosing between a fixed or random effects treatment of the α_i.
- In a dynamic model the situation is substantially different, because $y_{i,t-1}$ will depend upon α_i , irrespective of the way we treat α_i .

Autoregressive Panel Data Model (Dynamic)

- To illustrate the problems that this causes, we first consider the case where there are **no exogenous variables** included and the model reads

$$y_{it} = x_{it}'\beta + \gamma y_{i,t-1} + \alpha_i + \epsilon_{it}, |\gamma| < 1$$

- Assume that we have observations on y_{it} for periods

  $t = 0,\ 1,. .., T$.

- The **fixed effects estimator** for γ is given by

$$\hat{\gamma}_{FE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_i)(y_{i,t-1} - \bar{y}_{i,-1})}{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{i,t-1} - \bar{y}_{i,-1}\right)^2}$$

  where $\bar{y}_i = (1/T) \sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i,-1} = (1/T) \sum_{t=1}^{T} y_{i,t-1}$.

Autoregressive Panel Data Model (Dynamic)

- The **fixed effects estimator** for γ can also be written as:

$$\hat{\gamma}_{FE} = \gamma + \frac{(1/(NT)) \sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{it} - \bar{\epsilon}_i)\,(y_{i,t-1} - \bar{y}_{i,-1})}{(1/(NT)) \sum_{i=1}^{N} \sum_{t=1}^{T} \quad \left(y_{i,t-1} - \bar{y}_{i,-1}\right)^2}$$

- This estimator, however, is biased and inconsistent for $N \to \infty$ and fixed T, as the last term in the right-hand side does not have expectation zero and does not converge to zero if N goes to infinity. In particular, it can be shown that (see Nickell, 1981):

$$plim_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{it} - \bar{\epsilon}_i)\,(y_{i,t-1} - \bar{y}_{i,-1}) \neq 0.$$

- Thus, for fixed T we have an **inconsistent estimator**.

Autoregressive Panel Data Model (Dynamic)

- Note that this inconsistency is not caused by anything we assumed about the α_is, as these are eliminated in estimation.
- The problem is that the within transformed lagged dependent variable is **correlated** with the within transformed error.
- If $T \to \infty$, it converges to 0 so that the fixed effects estimator is consistent for γ if both $T \to \infty$ and $N \to \infty$, because the former expression could be written as:

$$plim_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{it} - \bar{\epsilon}_i) \left(y_{i,t-1} - \bar{y}_{i,-1} \right) = -\frac{\sigma_\epsilon^2}{T^2} \frac{(T-1) - T\gamma + \gamma^2}{(1-\gamma)^2}$$

Is the asymptotic bias for fixed T large enough to be worrisome?

- Yes! For finite T the **bias can hardly be ignored**. For example, if the true value of γ equals 0.5, it can easily be computed that (for $N \to \infty$):
- plim $\hat{\gamma}_{FE} = -0.25$ if $T = 2$
- plim $\hat{\gamma}_{FE} = -0.04$ if $T = 3$
- plim $\hat{\gamma}_{FE} = 0.33$ if $T = 10$,

so even for moderate values of T the **bias is substantial**.

- Which ways out?

Ways out

- Fortunately, there are relatively easy ways to avoid these biases.
- Start with a **different transformation** to eliminate the individual effects α_i , in particular we take **first differences**.
- This gives:

$$y_{it} - y_{i,t-1} = \gamma\left(y_{i,t-1} - y_{i,t-2}\right) + \left(\epsilon_{it} - \epsilon_{i,t-1}\right),$$
$$t = 2, \ldots, T.$$

- If we estimate this by OLS we do not get a consistent estimator for γ because $y_{i,t-1}$ and $\epsilon_{i,t-1}$ are, by definition, correlated, even if $T \rightarrow \infty$.
- However, $y_{i,t-2}$ is correlated with $\left(y_{i,t-1} - y_{i,t-2}\right)$ but not with $\varepsilon_{i,t-1}$, unless ε_{it} exhibits autocorrelation (which we excluded by assumption).

Ways out

- This suggests an **instrumental variables estimator** for γ as
$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-2}\,(y_{i,t} - y_{i,t-1})}{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-2}\,(y_{i,t-1} - y_{i,t-2})}$$
- A necessary condition for consistency of this estimator is that
$$plim \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}(\epsilon_{it} - \epsilon_{i,t-1})\,y_{i,t-2} = 0$$

for either T, or N, or both going to infinity.

- This estimator is **one of the estimators** proposed by Anderson and Hsiao (1981). They also proposed an alternative, where $y_{i,t-2} - y_{i,t-3}$ is used as an instrument.
- A **method of moments** approach can unify the estimators and eliminate the disadvantages of reduced sample sizes.

Ways out

- It is well known that imposing **more moment conditions increases the efficiency** of the estimators (provided the additional conditions are valid, of course). Arellano and Bond (1991) suggest that the list of instruments **can be extended** by exploiting additional moment conditions and letting their number vary with t .
- The general **GMM approach does not impose** that ϵ_{it} is **i.i.d**. over individuals and time
- Under weak regularity conditions, the **GMM estimator for γ is asymptotically normal** for $N \to \infty$ and fixed T .
- Alvarez and Arellano (2003) show that, in general, the **GMM estimator is also consistent** when both N and T tend to infinity despite the fact that the number of moment conditions tends to infinity with the sample size. For **large *T***, however, the GMM estimator will be close to the fixed effects estimator..

Dynamic Models with Exogenous Variables

- If the model also contains **exogenous variables**, we have
$$y_{it} = x'_{it}\beta + \gamma y_{i,t-1} + \alpha_i + \epsilon_{it}$$
- which can also be estimated by the generalized instrumental variables or GMM approach. Depending upon the assumptions made about x_{it} , different sets of additional instruments can be constructed.
- Arellano and Bover (1995) provide a framework to integrate the above approach with the instrumental variables estimators of Hausman and Taylor (1981). Most importantly, they discuss how **information in levels** can also be exploited in estimation.
- That is, in addition to the moment conditions presented above, it is also possible to exploit the presence of valid instruments for the levels equation or its average over time (the between regression).

Limited dependent variable (LDV) models

- In practical applications one often has to cope with phenomena that are of a **discrete or mixed discrete continuous nature**.
  - E.g., one could be interested in explaining whether consumers are **satisfied for the water supply service** (yes or no), or whether consumers think that the **price of gas is very low, fairly accessible, fairly expensive or excessive** (ordered discrete).
- For this type of variables LDV models are often more appropriate. **Attention should be paid to the estimation and interpretation** of their parameters.
- Often the problems analyzed with this type of model are of a **micro-economic nature**, thus requiring data on individuals, households or firms.
- To stress this, we shall index all variables by i, running from 1 to sample size N

Binary Choice Models: linear regressions

- Suppose we want to explain whether a family have **access to gas supply**. Let the sole explanatory variable be the family income.
- We have data on *N* families *(i* = 1,.. .,*N)*, with observations on their income, x_{i2}, and whether or not they have access to gas services. This latter element is described by the **binary variable** y_i, defined as
  - $y_i = 1$ if family *i* has gas supply access
  - $y_i = 0$ if family *i* does not have gas supply access
- Suppose we would use a regression model to explain y_i from x_{i2} and an intercept term $(x_{i1} \equiv 1)$.
- This **linear model** would be given by:
  - $y_i = \beta_1 + \beta_2 x_{i2} + \epsilon_i = x_i' \beta + \epsilon_i$, where $x_i = (x_{i1}, x_{i2})'$

Binary Choice Models: linear regressions

- It seems reasonable to make the standard assumption that $E\{\varepsilon_i \,|x_i\} = 0$ such that $E\{y_i \,|x_i\} = x_i'\beta$.
- This implies that

$$E\{y_i \,|x_i\} = 1.P\,\{y_i = 1|x_i\} + 0.P\,\{y_i = 0|x_i\}$$
$$= P\,\{y_i = 1|x_i\} = x_i'\beta.$$

- The linear model implies that $x_i'\beta$ is a probability and should therefore lie between 0 and 1. This is only possible if the x_i values are bounded and if certain restrictions on β are satisfied.

Binary Choice Models: linear regressions

- In addition, the error term has a highly non-normal distribution and suffers from **heteroskedasticity**. Because y_i has only two possible outcomes (0 or 1), the error term, for a given value of x_i, has two possible outcomes as well.
- In particular, the distribution of ε_i can be summarized as

  $P\{\varepsilon_i = -x_i{}'\beta | x_i\} = P\{y_i = 0 | x_i\} = 1 - x_i'\beta$

  $P\{\varepsilon_i = 1 - x_i'\beta | x_i\} = P\{y_i = 1 | x_i\} = x_i'\beta$
- Hence, the **variance of the error term is not constant** but dependent upon the explanatory variables according to $V\{\varepsilon_i | x_i\} = x_i'\beta\,(1 - x_i'\beta)$.
  - Note that the error variance also depends upon the model parameters β.

Binary Choice Models: logit, probit

- To overcome the problems with the linear model, there exists a class of **binary choice models** (or univariate dichotomous models), designed to model the 'choice' between two discrete alternatives.
- These models essentially describe the probability that $y_i = 1$directly, although they are often derived from an underlying **latent variable model**.
- In general, we have

  $P\{y_i = 1|x_i\} = G(x_i, \beta)$ for some function $G(.)$.
- Clearly, the function $G(.)$ should take on values in the interval $[0, 1]$ only.
- Usually, one restricts attention to functions of the form $G(x_i, \beta) = F(x_i'\beta)$.
  - As $F(.)$ also has to be between 0 and 1, it seems natural to choose F to be some distribution function.

Binary Choice Models: logit, probit

- Common choices are the **standard normal distribution** function and the **standard logistic function**
  - $F(w) = \Phi(w) = \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt \rightarrow Probit\ model$
  - $F(w) = \mathrm{L}(\mathrm{w}) = \frac{e^w}{1+e^w} \quad \rightarrow Logit\ model$
- Both a standard normal and a standard logistic random variable have an expectation of zero, while the latter has a variance of $\frac{x^2}{3}$ instead of 1.
- A third choice is the **uniform** over the interval [0,1], i.e. with distribution:
  - F(w)=0, w<0;
  - $F(w) = w, 0 \leq z \leq 1$
  - $F(w) = 1, w > 1; \quad \rightarrow Linear\ probability\ model$

Binary Choice Models: logit, probit

- Apart from their **signs**, the coefficients in these binary choice models are not easy to **interpret** directly.
- One way to interpret the parameters (and to ease comparison across different models) is to consider the **partial derivative** (marginal effects) of the probability that y_i equals one with respect to a continuous explanatory variable, x_{ik} , say. For the three models above, we obtain:
- $\frac{\partial \Phi(x_i'\beta)}{\partial x_{ik}} = \phi(x_i'\beta)\beta_k$
- $\frac{\partial \mathrm{L}(x_i'\beta)}{\partial x_{ik}} = \frac{e^{x_i'\beta}}{\left(1+e^{x_i'\beta}\right)^2}\beta_k$
- $\frac{\partial x_i'\beta}{\partial x_{ik}} = \beta_k; (or\ 0)$

LDV Models: underlying latent model

- It is possible (but not necessary) to derive a binary choice model from underlying behavioural assumptions.
  - This leads to a **latent variable representation of the model**, which is in common use even when such behavioural assumptions are not made.
- Let us look at the **decision of a family to have access to gas supply**. The utility difference between having a gas supply and not having one depends upon the income earned but also on other personal and household characteristics, like the householder age and education, the location of the household, etc.
- Thus, for each person i we can write the utility difference between having gas supply and not having one as a function of observed characteristics, x_i say, and unobserved characteristics, ε_i.

LDV Models: underlying latent model

- Assuming a linear **additive relationship** we obtain for the **utility difference**, denoted y^*, which is referred to as a latent variable (indicated with an asterisk).
- **Our assumption** is that an individual chooses to have access to gas if the utility difference exceeds a certain threshold level, which can be set to zero without loss of generality.
- Consequently, we observe $y_i = 1$ (access) if and only if $y_i^* > 0$ and $y_i = 0$ (no access) otherwise.
- Thus we have:
- $P(y_i = 1) = P(y_i^* > 0) = P(x_i'\beta + \epsilon_i > 0) = P(-\epsilon_i \leq x_i'\beta) = F(x_i'\beta)$, where F denotes the distribution function of $-\epsilon_i$.
- As the scale of utility is not identified, a **normalization** on the distribution of ϵ_i is required. Usually this means that its variance is fixed at a given value.

Multi-response models

- In many applications, the number of alternatives that can be chosen is **larger than two**. For example, we can distinguish the choice between high satisfaction, mild satisfaction, mild dissatisfaction, high dissatisfaction of consumers as for the quality of gas supply services.
- Some **quantitative variables can only be observed to lie in certain ranges**. This may be because questionnaire respondents are unwilling to give precise answers, or are unable to do so, perhaps because of conceptual difficulties in answering the question.
- An important goal is to describe these probabilities with a **limited number of unknown parameters and in a logically consistent way**. For example, probabilities should lie between 0 and 1 and, over all alternatives, add up to one.

Multi-response models

- An important distinction exists between **ordered response models** and **unordered models**.
- An ordered response model is generally more **parsimonious** but can only be applied if there exists a logical ordering of the alternatives.
  - The reason is that there is assumed to exist one underlying latent variable that drives the choice between the alternatives. In other words, the **results will be sensitive to the ordering** of the alternatives, so this ordering should make sense.
- Unordered models are not sensitive to the way in which the alternatives are numbered. In many cases, they can be based upon the assumption that each alternative has a random utility level and that individuals choose the alternative that yields highest utility.

Multi-response models: ordered response models

- Let us consider the choice between *M* alternatives (*j*=1…*M*). If there is a logical ordering in these alternatives (for example, no car, 1 car, more than one car), a so-called **ordered response model** can be used. This model is also based on *one* underlying latent variable but with a different match from the latent variable, y_i^*, to the observed one $(y_i = 1, 2, \dots, M)$. Usually, one says that
  - $y_i^* = x_i'\beta + \epsilon_i$
  - $y_i = j$ if $\gamma_{j-1} < y_i^* \leq \gamma_j$ for unknown γ_js with $\gamma_0 = -\infty$, $\gamma_1 = 0$ and $\gamma_M = \infty$.
- Consequently, the probability that alternative j is chosen is the probability that the latent variable y^* is between two boundaries γ_{j-1} and γ_j. Assuming that ϵ_i is i.i.d. standard normal results in the **ordered probit model**. The logistic distribution gives the **ordered logit model**.
- For $M = 2$ we are back at the binary choice model.