ADVANCED ECONOMETRICS

Martin Schumann

Technical University Dortmund

Winter Term 2018

This Version: December 04, 2018

Introduction

- Nonlinear models appear frequently when latent variables can only be partially observed.
- Moreover, they naturally arise when modeling economic decision-making.
  - An outcome variable that corresponds to a decision is typically modeled using binary or discrete variables.
- If panel data is available, one can allow to account for unobserved heterogeneity.
- As in linear panel data models, different assumptions imposed on the unobserved effects require different estimation techniques.

Nonlinear panel data models with fixed effects

- Consider for example the latent variable model

$$y_{it}^* = X_{it}'\theta_0 + c_{i0} + U_{it},\ i = 1, \dots, n\ t = 1, \dots, T.$$

  where X_{it} contains explanatory variables that vary across i and t and c_{i0} denotes an unobserved time-invariant effect.
- We observe

$$y_{it} = \begin{cases} 1, & y_{it}^* > 0 \\ 0, & \text{else} \end{cases}$$

- In fixed effects models, one imposes distributional assumptions on the model error U_{it} conditional on the time-series of explanatory variables and c_{i0}, but we do not make any further assumptions regarding c_{i0}!

The fixed effects model (1)

- For each individual i:
  - $\mathcal{Y}_{iT} := \{Y_{i1}, \dots, Y_{iT}\}$ is the time-series of outcomes
  - $\mathcal{X}_{iT} := \{X_{i1}, \dots, X_{iT}\}$ the time-series of explanatory variables
  - c_{i0} denotes the fixed effect.

(1) The outcomes $\mathcal{Y}_{iT}$ are drawn from the conditional density $f_{\mathcal{Y}_{iT}|\mathcal{X}_{iT}, c_{i0}; \theta_0}$, which is known up to θ_0.
  - The common parameter θ_0 is the object of interest.

- The joint density of $(\mathcal{X}_{iT}, c_{i0})$ is unknown. The relationship between c_{i0} and $\mathcal{X}_{iT}$ is unrestricted.

The fixed effects model (2)

- The likelihood for (θ, α_i), conditional on $(\mathcal{X}_{iT}, c_{i0})$, is

$$L_{iT}(\theta, c_i) := f_{\mathcal{Y}_{iT}|\mathcal{X}_{iT}, c_i; \theta}(\mathcal{Y}_{iT}).$$

The average (scaled) loglikelihood for the ith stratum is

$$\ell_{iT}(\theta, c_i) := T^{-1} \log L_{iT}(\theta, c_i).$$

- Additional to the distributional assumptions, we assume

(2) time-independence conditional on $\mathcal{X}_{iT}$ and c_i (can be relaxed)

(3) observations are independent across i.

(4) The likelihood is sufficiently well-behaved, so that it can be differentiated sufficiently many times. Further technical assumptions need to be made for the general theory.

Properties of Maximum Likelihood (1)

- We can compute the MLE as

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \ell_{iT}(\theta, \hat{c}_{iT}(\theta)),$$

where $\ell_{iT}(\theta, \hat{c}_{iT}(\theta))$ is called the "profile" or "concentrated likelihood" and

$$\hat{c}_{iT}(\theta) = \arg\max_{u} \ell_{iT}(\theta, u).$$

Properties of Maximum Likelihood (2)

- In general, it can be shown that

$$\hat{\theta} - \theta_0 = O_p(\frac{1}{\sqrt{nT}}) + O_p(\frac{1}{T}).$$

- The MLE is inconsistent if $n \to \infty$ while T is fixed.
- $\hat{\theta} \xrightarrow{p} \theta_0$ if both $n, T \to \infty$. However,

$$\sqrt{nT}(\hat{\theta} - \theta_0) = O_p(1) + O_p(\sqrt{\frac{n}{T}})$$

so that the asymptotic distribution is incorrectly centered if $n/T \not\to 0$.

- This is known as the "incidental parameters problem" of Maximum Likelihood, which has first been noted by Neyman and Scott (1948).

Example 1 (see Exercise 3 on Sheet 6)

Consider the simple linear regression

$$y_{it} = c_{i0} + U_{it}, \qquad U_{i1}, \ldots, U_{iT} | c_{i0} \sim \mathrm{NIID}(0, \sigma_0^2),$$

where individual specific heterogeneity arises only in the mean. The MLE of σ^2 is the average of the within-stratum sample variances, i.e.,

$$\tilde{\sigma}^2 := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} (y_{it} - \bar{y}_i)^2,$$

where $\bar{y}_i := \sum_{t=1}^{T} y_{it}/T$. But,

$$n \to \infty,\ T \text{ fixed} \Rightarrow \tilde{\sigma}^2 \xrightarrow{\mathrm{P}} \sigma_0^2(1 - \frac{1}{T}) \neq \sigma_0^2.$$

Example 2 (Fixed effects panel logit)

- Consider the model

$$y_{it} = \mathbb{1}(X_{it}'\theta_0 + \alpha_{i0} + U_{it} > 0),$$

where $U_{i1}, \ldots, U_{iT} | \mathcal{X}_{iT}, c_{i0} \sim \text{LogisticIID}$.

- The total loglikelihood can be derived as

$$\ell_{nT}(\theta, c_1, \ldots, c_n) = \sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}\log(\Lambda(X_{it}'\theta + c_i)) + (1-y_{it})\log(1-\Lambda(X_{it}'\theta + c_i))$$

- In general, there is no closed form for $\hat{c}_{iT}(\theta)$, since the FOC

$$\sum_{t=1}^{T} y_{it} - \Lambda(X_{it}'\theta + \hat{c}_{iT}(\theta)) = 0$$

cannot be solved for $\hat{c}_{iT}(\theta)$.

Panel logit (2)

- To find a closed form expression for $\hat{c}_{iT}(\theta)$, assume that $T = 2$, $X_{i1} = 0$ and $X_{i2} = 1$ for all i. Some algebra then shows that

$$\hat{c}_{iT}(\theta) = \begin{cases} -\infty, & y_{i1} + y_{i2} = 0 \\ -\frac{\theta}{2}, & y_{i1} + y_{i2} = 1 \\ \infty, & y_{i1} + y_{i2} = 2 \end{cases}$$

$\Rightarrow$ We cannot identify θ_0 from data of individuals whose outcome variable does not vary across t.

  - Intuitively, a time-constant outcome can be perfectly explained by a time-constant unobserved effect that is unrestricted. Due to the time-invariance of the outcome, we cannot distinguish the model $y_{it} = X_{it}'\theta_0 + c_{i0} + U_{it}$ from the model $y_{it} = c_{i0}^* + U_{it}^*$.

Panel logit (3)

- Using only the individuals whose outcome varies over time, we can show that

$$\hat{\theta} \xrightarrow{p} 2\theta_0$$

as $n \to \infty$ while T is fixed if $c_{i0} = k \in \mathbb{R} \forall i$.

$\Rightarrow$ The MLE is inconsistent as $n \to \infty$ while T fixed.

  - Given our discussion on the incidental parameters problem, this does not come as a surprise.

- However, the specific form of the logistic density allows us to consistently estimate θ_0 by conditioning on a *sufficient statistic*:

Panel logit (4)

- It can be shown that the distribution of

$$(y_{i1}, \ldots, y_{iT}) | \mathcal{X}_{iT}, c_{i0}, \sum_{t=1}^{T} y_{it}$$

does not depend on c_{i0}.

- This conditional density is still a density and can therefore be used to define a conditional likelihood function that only depends on θ.
- We can then apply standard maximum likelihood theory to show the consistency of the Conditional Maximum Likelihood Estimator as $n \to \infty$ while T is fixed.
  - Again, only individuals with time-varying outcomes can be used.

General situation

- In most models we cannot find a sufficient statistic.
  - E.g. static and dynamic panel probit, dynamic panel logit.
- The incidental parameters bias is often regarded as small sample bias that can be asymptotically corrected.
- Many bias corrections are available. Most have in common that they reduce the order of the bias from $O_{\mathrm{p}}(T^{-1})$ to $O_{\mathrm{p}}(T^{-2})$.
- Simulation studies suggest that this yields an improved small sample performance in panels with moderately large T.
- The general bias correction formulas for dynamic models are beyond the scope of this course.

Table 3. Design 1: Simulation results for the probit model ($n = 100$).

T	Estimator	Mean Bias	STD	MSE	MAE	Median Bias
2	$\hat{\theta}$	1.1280	0.4331	1.4597	1.1280	1.0781
	$\hat{\theta}_{avg}$	1.0695	0.4233	1.3228	1.0695	1.0194
	$\hat{\theta}_{exp}$	0.4607	0.2961	0.2998	0.4642	0.4250
	$\hat{\theta}$	0.1170	0.1514	0.0366	0.1499	0.1028
	$\hat{\theta}^*$	−0.0184	0.1600	0.0259	0.1281	−0.0325
	$\hat{\theta}^{(\infty)}$	−0.1509	0.2058	0.0651	0.2346	−0.2489
3	$\hat{\theta}$	0.7474	0.2972	0.6469	0.7474	0.7099
	$\hat{\theta}_{avg}$	0.6372	0.2917	0.4911	0.6372	0.6030
	$\hat{\theta}_{exp}$	0.2482	0.2003	0.1017	0.2561	0.2230
	$\hat{\theta}$	0.1163	0.1529	0.0369	0.1472	0.1020
	$\hat{\theta}^*$	−0.0139	0.1457	0.0214	0.1095	−0.0329
	$\hat{\theta}^{(\infty)}$	−0.1030	0.1554	0.0347	0.1682	−0.1497
4	$\hat{\theta}$	0.5333	0.2126	0.3296	0.5334	0.5115
	$\hat{\theta}_{avg}$	0.3920	0.2021	0.1945	0.3929	0.3702
	$\hat{\theta}_{exp}$	0.1544	0.1454	0.0450	0.1695	0.1394
	$\hat{\theta}$	0.0833	0.1308	0.0240	0.1194	0.0717
	$\hat{\theta}^*$	−0.0128	0.1068	0.0116	0.0843	−0.0227
	$\hat{\theta}^{(\infty)}$	−0.0702	0.0935	0.0137	0.0964	−0.0820
5	$\hat{\theta}$	0.3949	0.1662	0.1835	0.3951	0.3840
	$\hat{\theta}_{avg}$	0.2499	0.1517	0.0854	0.2533	0.2390
	$\hat{\theta}_{exp}$	0.0999	0.1187	0.0240	0.1237	0.0941
	$\hat{\theta}$	0.0537	0.1105	0.0151	0.0958	0.0479
	$\hat{\theta}^*$	−0.0087	0.0939	0.0089	0.0746	−0.0123
	$\hat{\theta}^{(\infty)}$	−0.0400	0.0820	0.0083	0.0731	−0.0409

Nonlinear random effects

- Besides modeling the distribution of the model errors conditional on the time series of explanatory variables and the unobserved effect, one now also imposes a distribution on the unobserved effect.
- In order to allow for correlation between the observed explanatory variables and the unobserved effect, one often permits the mean of the distribution of the unobserved effect to depend on the observed explanatory variables.
  - This approach is referred to as “correlated random effects”.
- For dynamic models, a popular approach is Wooldridge (2005).

Dynamic panel probit with random effects (1)

- We now include dynamics in the latent variable model:

$$y_{it}^* = \rho y_{it-1} + X_{it}'\theta_0 + c_{i0} + U_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T.$$

- The model error satisfies

$$U_{it}|y_{it-1}, \ldots, y_{i0}, \mathcal{X}_{it}, c_{i0} \sim \text{i.i.d.}\, \mathrm{N}(0, 1)$$

- This yields

$$\mathbb{P}(y_{it} = 1|y_{it-1}, \ldots, y_{i0}, \mathcal{X}_{iT}, c_{i0}) = \Phi(\rho y_{it-1} + X_{it}'\theta_0 + c_{i0}).$$

- The model contains several assumptions:
  - The dynamics are assumed to be of first order conditional on $\mathcal{X}_{iT}, c_{i0}$.
  - The explanatory variables are strictly exogenous.
  - The unobserved effect enters additively (as before).

Dynamic panel probit with random effects (2)

- In addition, assume

$$c_{i0} = \alpha_0 + \alpha_1 y_{i0} + \bar{X}_i' \alpha_3 + a_i$$

where $\bar{X}_1 = T^{-1} \sum_{t=1}^{T} X_{it}$ and

$$a_i \sim \text{i.i.d}\, \mathrm{N}(0, \sigma^2)$$

and a_i is independent of y_{i0} and the explanatory variables.

- This implies

$$c_{i0} | y_{i0}, \mathcal{X}_{it} \sim \mathrm{N}(\alpha_0 + \alpha_1 y_{i0} + \bar{X}_i' \alpha_2, \sigma^2),$$

i.e. the mean of the unobserved effect is allowed to depend on the observed characteristics and the initial condition y_{i0}.

Dynamic panel probit with random effects (3)

- We then obtain

$$\mathbb{P}(y_{it} = 1 | y_{it-1}, \ldots, y_{i0}, \mathcal{X}_i, \bar{X}_i, a_i) = \Phi(\rho y_{it-1} + X_{iT}' \theta_0 + \alpha_0 + \alpha_1 y_{i0} + \bar{X}_i' \alpha_3 + a_i)$$

- The conditional (on a_i) likelihood is

$$L_{iT}(\beta, a_i) = \prod_{t=1}^{T} \Phi(\ldots)_{it}^{y} (1 - \Phi(\ldots))^{1-y_{it}},$$

where β contains $\theta, \rho, \alpha_0, \alpha_1, \alpha_3$ and σ.

  - For a given value of a_i, the random effects model is a standard dynamic probit model enriched by the additional regressors $\bar{X}_i$ and y_{i0}.
  - Since a_i is unobserved, we need to integrate it out with respect to the assumed density.

Dynamic panel probit with random effects (4)

- Here, a_i is independent of the explanatory variables and the initial conditions, hence we compute the unconditional likelihood as

$$L_{iT}(\beta) = \int_{-\infty}^{\infty} L_{iT}(\beta, \alpha_i) \frac{1}{\sigma} \phi(\frac{a}{\sigma})\, da$$

- Writing $l_{iT}(\beta) = \log L_{iT}(\beta)$ and using the independence across i, we find the random effects estimator as

$$\hat{\beta} = \arg\max_{\beta} \sum_{i=1}^{n} l_{iT}(\beta).$$

- As usual, maximization has to be carried out numerically.

Estimating APEs (1)

- In nonlinear models, estimates of coefficients are informative about the direction of the effect of a variable, but not about the absolute magnitude of the effect.
- Therefore, the object of interest is often the Average Partial Effect (APE).
- In a fixed effects dynamic probit model, one would estimate the APE of y_{it-1} as

$$\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \Phi(\hat{\rho} + X_{it}' \hat{\theta} + \hat{\alpha}_{iT}(\theta)) - \Phi(X_{it}' \hat{\theta} + \hat{\alpha}_{iT}(\theta)).$$

- Due to the presence of $\hat{\alpha}_i(\theta)$, this estimator again suffers from the incidental parameters problem.
  - Small sample corrections exist.
  - The bias is less severe as compared to coefficient estimates.

Estimating APEs (2)

- In a random effects setting, one would need to integrate out the random effects again.
- In the dynamic random effects probit, the APE of y_{t-1} is

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Phi(\hat{\rho}_a + X_{it}'\hat{\theta}_a + \hat{\alpha}_{a0} + \hat{\alpha}_{a1}y_{i0} + \bar{X}_i'\hat{\alpha}_{a3}) \\ -\Phi(X_{it}'\hat{\theta}_a + \hat{\alpha}_{a0} + \hat{\alpha}_{a1}y_{i0} + \bar{X}_i'\hat{\alpha}_{a3}),$$

where the subscript a indicates multiplication of the original estimates with $(1+\sigma^2)^{-1/2}$.

Fixed versus random effects (1)

- In nonlinear models, fixed effects estimators typically suffer from the incidental parameters bias, whereas random effects estimators usually suffer from a misspecification bias.
- It has been shown by Arellano and Bonhomme (2009) that both the incidental parameters bias and the misspecification bias are of order $O_{\mathrm{p}}(T^{-1})$ for FE and RE estimators of the *coefficients*.
- For panels with long time series, one would therefore expect similar results with both methods.
- In panels with only moderately many time periods, one should use a bias corrected fixed effects estimator.

Fixed versus random effects (2)

- The situation for estimators of APEs is however different:
- The incidental parameters bias is of order $O_{\mathrm{p}}(T^{-1})$ again.
  - We can reduce the order to $O_{\mathrm{p}}(T^{-2})$ using bias corrections.
- Unlike the incidental parameters bias of FE, the misspecification bias of the RE estimator does not vanish as $T \to \infty$!

$\Rightarrow$ Even in fairly long panels, one should expect differences between FE and RE estimators of APEs, since the RE is likely to be inconsistent.