14.382 L10. NONLINEAR PANEL DATA

VICTOR CHERNOZHUKOV AND IVÁN FERNÁNDEZ-VAL

Abstract. We discuss bias corrections to deal with the incidental parameter problem in fixed effects estimation of nonlinear panel data models.

1. The Model

The model that we consider will be both nonlinear and dynamic, since once we allow for nonlinearities, we might as well consider dynamics as a form of nonlinearity. The model will be semiparametric, where we specify the distribution of the outcome $y_{k,t}$ for individual k at time t, given the information set I_t, as:

$$y_{k,t} \mid I_k \sim f(\cdot \mid x_{k,t}, \alpha, a_k, b_t), \;\; t = 1, \dots, T, \;\; k = 1, \dots, N, \tag{1.1}$$

where I_t represents the information available up to time t, including all past lags of the dependent variable as well as contemporaneous and past lags of the "controls".

The function $f(\cdot \mid x_{k,t}, \alpha, a_k, b_t)$ represents a parametric pdf or pmf of $y_{k,t}$ conditional on x_{kt}, which can include lagged outcomes $y_{k,t-1}, ..., y_{k,t-J}$ and the set of other observed controls $w_{k,t}$, the unobserved individual effect a_k, and the unobserved time effect b_t. The nonparametric part of the model is the distribution of the individual and time effects. This distribution is left completely unrestricted by treating the individual and time effects as parameters to be estimated in the so-called *fixed effects approach*. The main appeal of this approach is that it does not impose any restriction in the relationship between the unobserved effects and observed controls.

The target parameter will be α and a_k and b_t represent the individual "fixed effect" and the time "fixed effect," respectively. Thus, the overall parameter vector is

$$\beta := (\alpha', \gamma')' := (\alpha, {(a_k)_{k=1}^{N}}', {(b_t)_{t=1}^{T}}')',$$

with $\beta_0 := (\alpha_0', \gamma_0')' := (\alpha_0, {(a_{0k}')_{k=1}^{N}}', {(b_{0t}')_{t=1}^{T}}')'$ denoting the true value of the parameter. Since there are N observations depending on each time effect and T depending on each individual effect, and the model is nonlinear, we can not hope to estimate consistently the parameters without having both N and T large. Thus we must rely N and

T being large, formally

$$N \to \infty, \quad T \to \infty.$$

But in this case we have a large number of parameters, N individual effects and T time effects. The problem of having that many nuisance (non-target) parameters is called the *incidental parameter problem*.

Unlike in linear models, the requirements on data are substantial, as we can no longer afford arbitrary dependence of the observations across time. Thus, the statement of the model (1.1) contains the strong assumption that the dependence on the past information I_t for individual k can be captured by (x_{kt}, a_k, b_t), which includes J lags of the dependent variable, the contemporaneous observed controls and unobserved individual effect for individual k, and the contemporaneous time effect. This is an important assumption that yields simplifications in the limit distribution, since it makes the scores of the problem conditionally uncorrelated across t. This assumption also means that our modeling has also implicitly taken care of "clustering", namely that we have captured all of the relevant dependence for a given unit k by putting in enough lags of the dependent variable. Whether this is a good approximation to the real world or not is an empirical matter. Finally, we also assume that the data are independent across units k conditional on the unobserved time effects. The inclusion of the time effects is therefore a parsimonious way to capture cross sectional dependence induced for example by aggregate shocks.

As an example, consider a binary outcome $y_{k,t}$ taking values 0 and 1. The first-order dynamic binary response model specifies the conditional probability:

$$\mathrm{P}\{y_{k,t} = 1 \mid I_t\} = F(x_{k,t}'\alpha + a_k + b_t), \tag{1.2}$$

where F is the parametric link function, for example, the standard normal or logistic distributions, and $x_{k,t}$ is a vector of transformations of $y_{k,t-1}$ and $w_{k,t}$. To give some context, in the empirical application of Section 3, $y_{k,t}$ is an indicator of labor force participation by woman k in year t, $y_{k,t-1}$ is an indicator of labor force participation in the previous year $t-1$, and $w_{k,t}$ is the set of controls including age, husband's income and several indicators for number of kids of different ages. In this application, $N \approx 1,000$ and $T \approx 10$.

As in the binary response models of L6, in addition to the parameter α, we might be interested in the APE,

$$\theta_0 = \frac{1}{NT} \sum_{k=1}^{N} \sum_{t=1}^{T} \int \Delta_{k,t}(x, \beta_0) dM(x),$$

where $\Delta_{k,t}$ is the PE for individual k at time t, and M is the distribution of x in the population of interest, which we assume is stationary across k and t. For example, if $x_{k,t}^1$, the first component of $x_{k,t}$, is continuous, and F is differentiable

$$\Delta_{k,t}(x,\beta) = \partial F(x_{k,t}'\alpha + a_k + b_t)/\partial x_{k,t}^1.$$

2. Bias Corrections for Fixed Effects Estimators

The estimator here could be the "fixed effect" maximum likelihood estimator,

$$\hat{\beta} = \arg\max_{\beta} \prod_{k=1}^{N}\prod_{t=1}^{T} f(y_{k,t} \mid x_{k,t},\alpha,a_k,b_t) = \arg\max_{\beta} \sum_{k=1}^{N}\sum_{t=1}^{T} \ln f(y_{k,t} \mid x_{k,t},\alpha,a_k,b_t).$$

For instance, in the binary response model the log likelihood is

$$\ln f(y_{k,t} \mid x_{k,t},\alpha,a_k,b_t) = y_{k,t}\ln F(x_{k,t}'\alpha + a_k + b_t) + (1-y_{k,t})\ln\{1 - F(x_{k,t}'\alpha + a_k + b_t)\}.$$

The product form of the likelihood arises from the Markovian assumption across t underlying the dynamic model, and it also reflects the independence across observational units indexed by k conditional on I_t.

This estimator is a GMM estimator with the score functions:

$$g_{k,t}(Z_{k,t},\beta) = \frac{\partial}{\partial\beta}\ln f(y_{k,t} \mid w_{k,t},\alpha,a_i,b_t).$$

Here $Z_{k,t} = (y_{k,t}, x_{k,t}')'$, and the double index (k,t) corresponds to an observational unit i, where the total number of units is

$$n = NT.$$

We assume that the scores evaluated at the true value of the parameter

$$g_i(Z_i,\beta_0) = g_{k,t}(Z_{k,t},\beta_0)$$

are independent across i, but they are not identically distributed because they depend on the unobserved effects.

A fixed effects estimator of the APE can be constructed using the plug-in principle. For example, if M is the distribution of x in the entire panel,

$$\hat{\theta} = \frac{1}{NT}\sum_{k=1}^{N}\sum_{t=1}^{T}\Delta_{k,t}(x_{k,t},\hat{\beta}).$$

This estimator is also a GMM estimator with the score functions:

$$\tilde{g}_i(Z_i,\delta) = \begin{pmatrix} \theta - \Delta_i(Z_i,\beta) \\ g_i(Z_i,\beta) \end{pmatrix}, \quad \delta := (\theta,\beta')', \quad \Delta_i(Z_i,\beta) = \Delta_{k,t}(x_{k,t},\beta).$$

From the previous lecture, we expect that in this case the "fixed effect" estimators to have large bias, since

$$\frac{p^2}{n} = O\left(\frac{(N+T)^2}{NT}\right) \not\to 0.$$

Thus we should perform bias-correction either analytically or via sample splitting. With the bias correction, the condition could be made as weak as

$$\frac{p^{3/2}}{n} = O\left(\frac{(N+T)^{3/2}}{NT}\right) \to 0,$$

which seems plausible in many example. Of course, the only way to check if this condition makes sense in a particular example is through Monte-Carlo. For a concrete example of how to do this, see Section 3.

[4] shows that the first order bias of $\hat{\alpha}$ relative to α_0 has the form

$$\frac{D}{N} + \frac{B}{T},$$

and provides analytical expressions for D and B and estimators $\hat{B}$ and $\hat{D}$ in several leading cases.

The analytically bias-corrected estimator then takes the form

$$\check{\alpha} = \hat{\alpha} - \frac{\hat{B}}{T} - \frac{\hat{D}}{N},$$

and obeys

$$\check{\alpha} \overset{a}{\sim} N(\alpha_0, V_{11}/n),$$

where V_{11} is the block of the GMM asymptotic variance of $\hat{\theta}$ corresponding to the main parameter component.

The jackknife bias corrected estimator based on sample splitting takes the form:

$$\check{\alpha} = \hat{\alpha} - (\hat{\alpha}_{N,T/2} - \hat{\alpha}) - (\hat{\alpha}_{N/2,T} - \hat{\alpha}),$$

where

- $\bar{\alpha}_{N,T/2}$ is the average of the two fixed effects estimators computed over the subpanels resulting from partitioning the original panel in two halves along the time dimension;
- $\bar{\alpha}_{N/2,T}$ is the average of the two fixed effects estimators computed over the subpanels resulting from partitioning the original panel in two halves randomly selected along the cross section dimension. Here we can average over many partitions to reduce arbitrariness in the choice of the sample splitting.

The correction works because the first order biases of $\hat{\alpha}_{N,T/2} - \hat{\alpha}$ and $\hat{\alpha}_{N/2,T} - \hat{\alpha}$ are, respectively,

$$\frac{B}{T/2} - \frac{B}{T} = \frac{B}{T} \quad \text{and} \quad \frac{D}{N/2} - \frac{D}{N} = \frac{D}{N},$$

which is what we need to bias correction. This estimator has the same properties as the analytically bias-corrected estimator under the assumptions imposed, i.e.

$$\breve{\alpha} \overset{\text{a}}{\sim} N(\alpha_0, V_{11}/n).$$

The jackknife correction is more robust to misspecification of the model than the analytical correction, but is less robust to violations of stationarity.

Similar analytical and jackknife bias corrections can be constructed for the fixed effects estimator of the APE $\hat{\theta}$, see [4].[1]

3. Female Labor Force Participation

We illustrate the bias corrections with an application to the effect of fertility on female labor force participation (LFP) from [3].[2] The relationship between fertility and female labor force participation has been of longstanding interest in labor economics and demography. Research on the causal effect of fertility on labor force participation is complicated because both variables are jointly determined. Here, we adopt an empirical strategy that aims at solving this endogeneity problem by controlling for unobserved individual and time effects using panel data. The data comes from waves 13 to 22 of the Panel Study of Income Dynamics (PSID) and contains information for the ten calendar years 1979-1988. Only women aged 18-60 in 1985 who were continuously married with husbands in the labor force in each of the sample periods are included in the sample. The sample consists of 1,461 women, 664 of whom changed labor force participation status during the sample period. The first year of the sample is excluded for use as initial condition.[3]

We estimate the probit binary response model (1.2), where F is the normal cdf, $y_{k,t}$ is a LFP indicator for women k at time time; and $x_{k,t} = (y_{k,t-1}, w_{k,t})$ includes the LFP indicator of the previous period $y_{k,t-1}$, three fertility variables (the numbers of children aged 0-2, 3-5, and 6-17), the logarithm of the husband's earnings in 1995 thousands of dollars, and a quadratic function of age in years divided by 10. Descriptive statistics for the sample are given in Table 1. Roughly 72% of the women in the sample participated in the labor force

[1] The Stata commands `probitfe` and `logitfe` [2] implement the analytical and jackknife bias corrections for probit and logit models.

[2] See also [5] and [1].

[3] The code and data are available at http://sites.bu.edu/ivanf/research/.

at some period. The average numbers of children per woman were .2, .3, and 1.1 for the three categories 0-2 year-old, 3-5 year-old, and 6-17 year-old, respectively.

TABLE 1. Descriptive Statistics

Variable	Mean	Std. Dev.
lfp	0.724	0.447
laglfp	0.721	0.449
kids0_2	0.227	0.466
kids3_5	0.288	0.513
kids6_17	1.05	1.095
log husband income ($1995/1000)	10.431	0.690
age (years/10)	3.73	0.922
n=NT	13149	

Source: PSID 1979-1988

Table 2 reports uncorrected and bias corrected fixed effects estimates of the target parameter α. The results show that most of the uncorrected estimates are more than one standard error away from their bias-corrected counterparts. We report the same standard errors for all the estimators because the standard errors constructed from the uncorrected fixed effects estimators are consistent for $\sqrt{V_{11}/n}$ under the asymptotic approximation that we consider where $N \to \infty$ and $T \to \infty$. We verify if these standard errors approximate well the variability of the estimators with our sample size in a Monte Carlo example below. Table 3 reports the results for the APEs of the lagged LFP and fertility variables. Here, we only find significant differences between the uncorrected and bias corrected estimates of the lagged LFP, where the APE almost doubles. After correcting the bias, we estimate about 20% positive state dependence in the participation decision. Each child aged 0-2 and 3-5 reduces the probability of participation by 7-9 percent and 3-5 percent, respectively, while an additional child aged more than 6 years does not have a significant effect on the probability of participation (at the 5 percent level using the analytically corrected estimator).

To assess the properties of the estimators in our sample size, we conduct a Monte Carlo experiment calibrated to the empirical application. In particular, we draw 500 panels of size $N = 664$ and $T = 9$, with $w_{k,t}$ fixed to the values in the original data, and the values of $y_{k,t}$ generated sequentially as

$$y^*_{k,t} = 1(\hat{\alpha}' x^*_{k,t} + \hat{a}_i + \hat{\beta}_t \geq \varepsilon^*_{k,t}), \quad k = 1, \ldots, N, \quad t = 1, \ldots, T,$$

TABLE 2. Probit Model: Parameters

Variable	Fixed effects	Analytical BC	Jackknife BC	(Std. Err.)
lagged lfp	0.757	1.082	1.261	(0.043)
kids0_2	-0.553	-0.445	-0.605	(0.058)
kids3_5	-0.290	-0.201	-0.346	(0.053)
kids6_17	-0.074	-0.052	-0.129	(0.043)
log husband income	-0.252	-0.206	-0.329	(0.055)
age	2.333	1.854	1.828	(0.627)
age2	-0.244	-0.186	-0.188	(0.052)

Analytical uses 2 lags to estimate spectral expectations.

Split-Jackknife averages over 20 random partitions of the cross sectional dimension

TABLE 3. Probit Model: APEs

Variable	Fixed effects	Analytical BC	Jackknife BC	(Std. Err.)
lagged lfp	0.107	0.207	0.190	(0.007)
kids0_2	-0.068	-0.072	-0.085	(0.032)
kids3_5	-0.035	-0.033	-0.047	(0.017)
kids6_17	-0.009	-0.009	-0.015	(0.007)

Analytical uses 2 lags to estimate spectral expectations.

Split-Jackknife averages over 20 random partitions of the cross sectional dimension

where $x^*_{k,t} = (y^*_{k,t-1}, w_{k,t})$, $y^*_{k,0} = y_{k,0}$, the initial values in the data, and $\varepsilon^*_{k,t}$ are independent draws from the standard normal distribution. The parameters $(\hat{\alpha}, \hat{a}_i, \hat{b}_t)$ are calibrated to their uncorrected fixed effects estimates.[4]

Table 4 reports biases, standard deviations, root mean square errors, ratios of standard errors to standard deviations, and empirical coverage probabilities of confidence intervals with nominal level of 95% for the uncorrected and bias corrected estimators of the target parameter. To speed up computation, we use only one cross sectional partition in the jackknife following the order of the individuals in the original data. All the results, except for the coverage probabilities, are in percentage of the true value of the parameter. We find that the bias corrections drastically reduce bias and rmse, and have coverage probabilities close to their nominal level. The standard errors provide a good approximation to the variability of the analytical corrections, but they underestimate the variability of the jackknife correction because this correction increases finite-sample sample standard deviation. We can use the results of this table to correct the coverage probabilities of the confidence intervals in the empirical application. For example, the corrected coverage probability of the

[4]We do not need to redraw the women that do not change LFP status during the sample period because their values of $y_{k,t}$ can be perfectly predicted by setting $\hat{a}_i$ arbitrarily large or small. These observations are not informative about the target parameter α.

confidence interval for the state dependence based on jackknife, $1.261 \pm 1.96 \times 0.043$, is 81%, instead of 95%.

TABLE 4. Calibrated Monte Carlo, $N = 664$, $T = 9$

	Bias	Std. Dev.	RMSE	SE/SD	p;.95
		Coefficient of laglfp			
FE-Probit	-53.56	6.06	53.90	1.01	0.00
Analytical	-10.28	6.03	11.92	1.01	0.63
Jackknife	-5.05	7.56	9.09	0.81	0.81
		Coefficient of Kids0-2			
FE-Probit	34.34	12.67	36.60	0.91	0.17
Analytical	6.41	10.64	12.42	1.09	0.95
Jackknife	7.30	16.95	18.44	0.68	0.78
		Coefficient of Kids3-5			
FE-Probit	49.73	23.27	54.90	0.87	0.34
Analytical	11.77	19.36	22.64	1.04	0.92
Jackknife	22.13	30.97	38.04	0.65	0.70
		Coefficient of Kids6-17			
FE-Probit	58.10	75.73	95.39	0.83	0.81
Analytical	17.14	62.88	65.12	1.00	0.94
Jackknife	34.00	97.91	103.55	0.64	0.75
		Coefficient of log of husband income			
FE-Probit	26.16	28.74	38.84	0.85	0.79
Analytical	5.47	24.97	25.54	0.98	0.93
Jackknife	4.78	30.60	30.95	0.80	0.87
		Coefficient of age			
FE-Probit	34.89	34.69	49.18	0.84	0.74
Analytical	8.44	29.43	30.58	0.99	0.93
Jackknife	3.89	38.86	39.02	0.75	0.87
		Coefficient of age squared			
FE-Probit	36.06	27.67	45.44	0.84	0.64
Analytical	8.31	22.61	24.07	1.03	0.96
Jackknife	3.73	35.99	36.15	0.64	0.79

Notes: 500 replication of probit model calibrated to PSID.
All the entries are in percent of the true parameter value.
Analytical uses 2 lags to estimate spectral expectations.
Jackknife uses 1 partition of the cross sectional dimension.

REFERENCES

[1] Jesus M. Carro. Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, 140(2):503–528, October 2007.

[2] M. Cruz-Gonzalez, I. Fernandez-Val, and M. Weidner. probitfe and logitfe: Bias corrections for probit and logit models with two-way fixed effects. *ArXiv e-prints*, October 2016.

[3] Iván Fernández-Val. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71–85, May 2009.

[4] Iván Fernández-Val and Martin Weidner. Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics*, 192(1):291–312, 2016.

[5] Dean R. Hyslop. State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women. *Econometrica*, 67(6):1255–1294, November 1999.