Topics in Panel Data Econometrics

Supplement to Lecture 1

July 3, 2015

Asymptotic bias of the autoregressive coefficient in the linear dynamic panel data model
Last time, we derived the asymptotic bias of $\hat{\beta}$ for the case when $T = 2$ and

$$y_{it} = \alpha_i + \beta y_{it-1} + \varepsilon_{it} \tag{1}$$

Recall that the LSDV/within estimator $\hat{\beta}$ has the following asymptotic behavior when $n \to \infty$:

$$\hat{\beta} \xrightarrow{p} \beta - \frac{\mathbb{E}\left[(\varepsilon_{i2} - \varepsilon_{i1})(y_{i1} - y_{i0})\right]}{\mathbb{E}\left[(y_{i1} - y_{i0})^2\right]}$$

Note that $Cov(\Delta\varepsilon_{i2}, \Delta y_{i1}) \neq 0$ because y_{i1} and ε_{i1} are correlated.

To examine the magnitude of the asymptotic bias, I assumed that y_{it} is stationary for every i and that $\varepsilon_{it} \sim iid\left(0, \sigma_\varepsilon^2\right)$ for every i and t. Recall that $\mathbb{E}\left[(\varepsilon_{i2} - \varepsilon_{i1})(y_{i1} - y_{i0})\right] = -\sigma_\varepsilon^2$. I may have made a small mistake in the calculation of $\mathbb{E}\left[(y_{i1} - y_{i0})^2\right]$. Here is the corrected version:

$$\begin{aligned}
\mathbb{E}\left[(y_{i1} - y_{i0})^2\right] &= \mathbb{E}\left[(\alpha_i + (\beta - 1) y_{i0} + \varepsilon_{i1})^2\right] \\
&= \mathbb{E}\left(\alpha_i^2\right) + (\beta - 1)^2 \mathbb{E}\left(y_{i0}^2\right) + \mathbb{E}\left(\varepsilon_{i1}^2\right) + 2(\beta - 1)\mathbb{E}(\alpha_i y_{i0}) \\
&\quad +2\mathbb{E}(\alpha_i \varepsilon_{i1}) + 2(\beta - 1)\mathbb{E}(y_{i0}\varepsilon_{i1})
\end{aligned}$$

If we interpret the α_i's as fixed constants, then

$$\begin{aligned}
\mathbb{E}\left[(y_{i1} - y_{i0})^2\right] &= \alpha_i^2 + (\beta - 1)^2 \mathbb{E}\left(y_{i0}^2\right) + \mathbb{E}\left(\varepsilon_{i1}^2\right) + 2\alpha_i(\beta - 1)\mathbb{E}(y_{i0}) \\
&\quad +2\alpha_i \mathbb{E}(\varepsilon_{i1}) + 2(\beta - 1)\mathbb{E}(y_{i0}\varepsilon_{i1}) \\
&= \alpha_i^2 + (\beta - 1)^2 \left[Var(y_{i0}) + (\mathbb{E}(y_{i0}))^2\right] + 2\alpha_i(\beta - 1)\mathbb{E}(y_{i0}) \\
&\quad +2\alpha_i \mathbb{E}(\varepsilon_{i1}) + 2(\beta - 1)\left[Cov(y_{i0}, \varepsilon_{i1}) + \mathbb{E}(\varepsilon_{i1})\mathbb{E}(y_{i0})\right] \qquad (2)
\end{aligned}$$

Under stationarity of y_{it}, we need to have the same mean and variance for all t. This means that

$$\begin{aligned}\mathbb{E}(y_{it}) &= \alpha_i + \beta\mathbb{E}(y_{it-1}) + \mathbb{E}(\varepsilon_{it}) \\ \mathbb{E}(y_{it}) &= \alpha_i + \beta\mathbb{E}(y_{it}) \\ \mathbb{E}(y_{it}) &= \frac{\alpha_i}{1-\beta} \qquad (3)\end{aligned}$$

Similarly, the variance is given by

$$\begin{aligned}Var(y_{it}) &= \beta^2 Var(y_{it-1}) + Var(\varepsilon_{it}) + 2Cov(y_{it-1}, \varepsilon_{it}) \\ Var(y_{it}) &= \beta^2 Var(y_{it}) + \sigma_\varepsilon^2 \\ Var(y_{it}) &= \frac{\sigma_\varepsilon^2}{1-\beta^2} \qquad (4)\end{aligned}$$

Substituting (3), (4), and the assumption on ε_{it} into (2), we have

$$\mathbb{E}\left[(y_{i1} - y_{i0})^2\right] = \sigma_\varepsilon^2 \frac{1-\beta}{1+\beta} + \sigma_\varepsilon^2 = \frac{2\sigma_\varepsilon^2}{1+\beta}$$

If we interpret the α_i's as random variables, then the statements above can be recomputed by conditioning on α_i and then apply the law of iterated expectations. For example, $\mathbb{E}(y_{it}|\alpha_i) = \alpha_i/(1-\beta)$. The problem is that we have to modify the iid assumption $\varepsilon_{it} \sim iid\left(0, \sigma_\varepsilon^2\right)$ slightly. Some would assume $\mathbb{E}\left(\varepsilon_{it}|y_i^{t-1}, \alpha_i\right) = \mathbb{E}(\varepsilon_{it}|y_{it-1}, y_{it-2}, \ldots, y_{i0}, \alpha_i) = 0$ and sometimes $Var\left(\varepsilon_{it}|y_i^{t-1}, \alpha_i\right) = \sigma_\varepsilon^2$. Some would just directly assume that $\mathbb{E}(y_{i0}\varepsilon_{it}) = 0$ for $t > 0$ and $\mathbb{E}(\alpha_i\varepsilon_{it}) = 0$.

Nickell (1981) bias Nickell (1981) derived the asymptotic bias of the LSDV/within estimator of the autoregressive coefficient in a linear AR(1) dynamic panel data moel with individual-specific fixed effects under the assumptions that $n \to \infty$, T fixed, stationarity of y_{it}, and that the process started in the infinite past. From (1), we have

$$y_{it} = \frac{\alpha_i}{1-\beta} + \sum_{l=0}^{\infty} \beta^l \varepsilon_{it-l}$$

Assuming $\mathbb{E}(\alpha_i\varepsilon_{it}) = 0$ for all t and $\varepsilon_{it} \sim iid\left(0, \sigma_\varepsilon^2\right)$, we have

$$y_{it}|\alpha_i \sim iid\left(\frac{\alpha_i}{1-\beta}, \frac{\sigma_\varepsilon^2}{1-\beta^2}\right)$$

Stacking (1) across all i and t, we have $y = D\alpha + y_{-1}\beta + \varepsilon$. In matrix form, the LSDV estimator for β (which is a scalar) is given by

$$\hat{\beta} = \frac{y_{-1}' M_D y}{y_{-1}' M_D y_{-1}} = \beta + \frac{y_{-1}' M_D \varepsilon}{y_{-1}' M_D y_{-1}}$$

As always, we have to compute the large sample limit of the second term of the previous expression. If you go through the derivations, you will find that as $n \to \infty$,

$$\hat{\beta} - \beta \quad \xrightarrow{p} \quad -\frac{1+\beta}{T-1} \frac{1 - \frac{1}{T}\frac{1-\beta^T}{1-\beta}}{1 - \frac{2\beta}{(1-\beta)(T-1)}\left[1 - \frac{1}{T}\frac{1-\beta^T}{1-\beta}\right]}$$

This is not really a neat expression but if we multiply both the nnumerator and denominator of the above expression by $T-1$, we will have

$$\begin{aligned}
\hat{\beta} - \beta \quad &\xrightarrow{p} \quad -\frac{1+\beta}{1} \frac{1 - \frac{1}{T}\frac{1-\beta^T}{1-\beta}}{T - 1 - \frac{2\beta}{1-\beta}\left[1 - \frac{1}{T}\frac{1-\beta^T}{1-\beta}\right]} \\
&= \quad -\frac{1+\beta}{T} \frac{1 - \frac{1}{T}\frac{1-\beta^T}{1-\beta}}{1 - \frac{1}{T} - \frac{2\beta}{T(1-\beta)}\left[1 - \frac{1}{T}\frac{1-\beta^T}{1-\beta}\right]} \\
&= \quad -\frac{1+\beta}{T} \frac{1 - \frac{1}{T}\frac{1-\beta^T}{1-\beta}}{1 - \frac{1}{T}\frac{1+\beta}{1-\beta} + \frac{2\beta\left(1-\beta^T\right)}{T^2(1-\beta)^2}}
\end{aligned}$$

Note that

$$\left[1 - \frac{1}{T}\frac{1+\beta}{1-\beta} + \frac{2\beta\left(1-\beta^T\right)}{T^2(1-\beta)^2}\right]^{-1} = 1 + \frac{1}{T}\frac{1+\beta}{1-\beta} + O\left(T^{-2}\right)$$

As a result,

$$\hat{\beta} - \beta \xrightarrow{p} -\frac{1+\beta}{T} + O\left(T^{-2}\right)$$

The idea can be extended to allow for strictly exogenous regressors but the closed form is much harder to obtain. Kiviet (1995, JoE) develops an asymptotic expansion to approximate the asymptotic bias. He shows that the biggest part of the asymptotic bias is really the $O\left(T^{-1}\right)$ portion. More of this type of work can be found in Kiviet (1999) and Bun and Kiviet (2003).

Asymptotic bias of MLE in static logit model I introduced the static logit model where

$$\begin{aligned}\Pr\left(y_{i1}=1|x_{i1}=0,x_{i2}=1,\alpha_i\right) &= \frac{\exp\left(\alpha_i\right)}{1+\exp\left(\alpha_i\right)}\\ \Pr\left(y_{i2}=1|x_{i1}=0,x_{i2}=1,\alpha_i\right) &= \frac{\exp\left(\alpha_i+\beta\right)}{1+\exp\left(\alpha_i+\beta\right)}\end{aligned}$$

I assumed that y_{i1} and y_{i2} are independent conditional on $\left(x_{i1},x_{i2},\alpha_i\right)$. We derived the log-likelihood function and calculated an estimator for the α_i which we called $\widehat{\alpha_i}\left(\beta\right)$. We substituted the estimator to the score equation $\partial\log L\left(\alpha_1,\ldots,\alpha_n,\beta\right)/\partial\beta=0$ and solved for β. The MLE for β is given by

$$\begin{aligned}\hat{\beta} &= 2\log\left[\frac{\frac{1}{n}\sum_{\{i:y_{i1}+y_{i2}=1\}}y_{i2}}{1-\frac{1}{n}\sum_{\{i:y_{i1}+y_{i2}=1\}}y_{i2}}\right]\\ &\xrightarrow{p} 2\log\left[\frac{\mathbb{E}\left(y_{i2}|y_{i1}+y_{i2}=1\right)}{1-\mathbb{E}\left(y_{i2}|y_{i1}+y_{i2}=1\right)}\right]\end{aligned}$$

By the law of iterated expectations,

$$\begin{aligned}\mathbb{E}\left(y_{i2}|y_{i1}+y_{i2}=1,x_{i1}=0,x_{i2}=1\right) &= \mathbb{E}\left[\mathbb{E}\left(y_{i2}|y_{i1}+y_{i2}=1,x_{i1}=0,x_{i2}=1,\alpha_i\right)\right]\\ &= \mathbb{E}\left[\Pr\left(y_{i2}|y_{i1}+y_{i2}=1,x_{i1}=0,x_{i2}=1,\alpha_i\right)\right]\\ &= \mathbb{E}\left[\frac{\exp\beta}{1+\exp\beta}\right]\\ &= \frac{\exp\beta}{1+\exp\beta}\end{aligned}$$

As a result, $\hat{\beta}\xrightarrow{p}2\beta$. When you substitute $\widehat{\alpha_i}\left(\beta\right)$ into the score equation for the parameter of interest β, the resulting score is called a profile score. For the MLE to be consistent, it must be the case that the score has expectation zero. Try writing down the profile score again and calculate its expectation with respect to the DGP. You will find that the expectation is not zero. Alternatively, you can substitute $\widehat{\alpha_i}\left(\beta\right)$ into the log-likelihood to obtain a log profile likelihood. Once again, the log profile likelihood will have a maximizer that is not consistent for β because the maximizer of the expectation of the log profile likelihood is not the true value β.