Jackknife Bias Reduction for Nonlinear Dynamic Panel Data Models with Fixed Effects

Geert Dhaene* K.U.Leuven
Koen Jochmans K.U.Leuven
Bram Thuysbaert K.U.Leuven

November 2006

Abstract

Split-panel jackknife estimators are proposed for reducing the bias of the maximum likelihood estimator (MLE) of dynamic panel data models with fixed effects. The bias is reduced from $O(T^{-1})$ to $O(T^{-2})$ or smaller, where T is the number of periods observed. The split-panel jackknife combines the MLE computed from the full panel, $\widehat{\theta}$, with the MLEs computed from shorter subpanels. For example, the half-panel jackknife (defined in eq. (1) below) uses $\widehat{\theta}$ and the MLEs corresponding to two non-overlapping half-panels, each using $T/2$ observations and all N cross-sections units. The half-panel jackknife estimator has bias $O(T^{-2})$. The bias is further reduced to $O(T^{-3})$ (or smaller) if two (or more) partitions of the panel are used, for example two half-panels and three 1/3-panels, and the MLEs corresponding to the subpanels. The asymptotic distribution of the jackknife estimators is normal, correctly centered at the true value, has variance equal to that of the MLE, and allows T to grow only slowly with N. The split-panel jackknife can also be employed to correct the profile likelihood function to any order. Maximising the jackknife-corrected likelihood yields estimators with essentially the same properties as the jackknife-corrected MLE. The large N, fixed T asymptotic variance of the split-panel jackknife estimators can be estimated consistently by the bootstrap or by the delete-one jackknife in the cross-section dimension. Simulation results for the probit and logit binary AR(1) models and for the linear AR(1) model show that even in small, short panels such as $N = 25$ and $T = 9$, the split-panel jackknife is very effective in reducing the bias of the MLE, has smaller mean squared error, and yields confidence intervals with much better coverage.

Let the data be $z_{it} = (y_{it}, x_{it})$, where $i = 1, ..., N$ and $t = 1, ..., T$. Assume the conditional density of y_{it}, given x_{it} (which may contain lagged values of y_{it}),

*Corresponding author. Address: K.U.Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium. Tel. +32 16 326798; Fax +32 16 326796; Email: geert.dhaene@econ.kuleuven.be.

is $f_{it}(\theta_0, \alpha_{i0}) \equiv f(y_{it}|x_{it}; \theta_0, \alpha_{i0})$. The MLE of θ_0 is

$$\widehat{\theta} = \arg\max_{\theta} l(\theta), \qquad l(\theta) \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \log f_{it}(\theta, \widehat{\alpha}_i(\theta)),$$

where $\widehat{\alpha}_i(\theta) \equiv \arg\max_{\alpha_i} \frac{1}{T} \sum_{t=1}^{T} \log f_{it}(\theta, \alpha_i)$. In general, $p\lim_{N\to\infty} \widehat{\theta} \neq \theta_0$, but under suitable regularity conditions,

$$p \lim_{N\to\infty} \widehat{\theta} = \theta_0 + \frac{B_1}{T} + \frac{B_2}{T^2} + ... + \frac{B_k}{T^k} + o\left(T^{-k}\right)$$

where $B_1, ..., B_k$ are constants. Let $\widehat{\theta}^{(1)}$ and $\widehat{\theta}^{(2)}$ be the MLEs corresponding to the half-panels $\{1, ..., T/2\}$ and $\{T/2+1, ..., T\}$ (assuming T is even). Define the half-panel jackknife estimator as

$$\widehat{\theta}_{1/2} \equiv 2\widehat{\theta} - \overline{\theta}_{1/2}, \qquad \text{where} \qquad \overline{\theta}_{1/2} \equiv \frac{1}{2}(\widehat{\theta}^{(1)} + \widehat{\theta}^{(2)}). \tag{1}$$

Clearly, $p\lim_{N\to\infty} \widehat{\theta}_{1/2} = \theta_0 + O(T^{-2})$, so $\widehat{\theta}_{1/2}$ corrects the first-order bias of $\widehat{\theta}$. More generally, define the split-panel estimator

$$\widehat{\theta}_{1/g} \equiv \frac{g}{g-1}\widehat{\theta} - \frac{1}{g-1}\overline{\theta}_{1/g},$$

where $\overline{\theta}_{1/g}$ is the average of g MLEs corresponding to g non-overlapping subpanels, each using T/g observations (assuming g divides T). Then, $p\lim_{N\to\infty} \widehat{\theta}_{1/g} = \theta_0 + O(T^{-2})$ and, as $N, T \to \infty$ and $N/T^3 \to 0$,

$$\sqrt{NT}(\widehat{\theta}_{1/g} - \theta_0) \xrightarrow{d} N(0, \Omega), \qquad \Omega \equiv \text{Avar}_{N,T\to\infty}(\sqrt{NT}\widehat{\theta}).$$

Higher-order corrections of $\widehat{\theta}$ are obtained as follows. Let $G = (g_1, ..., g_k) \geq 2$ be k distinct integers dividing T, and define the multiple split-panel jackknife estimator

$$\widehat{\theta}_{1/G} \equiv \left(1 + \sum_{i=1}^{k} a_i\right)\widehat{\theta} - \sum_{i=1}^{k} a_i \overline{\theta}_{1/g_i}, \qquad a_i \equiv \frac{1}{g_i - 1} \prod_{j\neq i} \frac{g_j}{g_j - g_i}.$$

Then $p\lim_{N\to\infty} \widehat{\theta}_{1/G} = \theta_0 + O(T^{-k-1})$ and, as $N, T \to \infty$ and $N/T^{2k+1} \to 0$,

$$\sqrt{NT}(\widehat{\theta}_{1/G} - \theta_0) \xrightarrow{d} N(0, \Omega).$$

The definition of $\widehat{\theta}_{1/G}$ can be slightly generalised to accomodate the case where T is not divisible by one or more g_i, without affecting the properties of $\widehat{\theta}_{1/G}$.

The plit-panel jackknife can be employed to correct the profile likelihood. Assume, around θ_0,

$$p \lim_{N\to\infty} l(\theta) = \overline{E} \log f_{it}(\theta, \alpha_i(\theta)) + \frac{D_1}{T} + \frac{D_2}{T^2} + ... + \frac{D_k}{T^k} + o\left(T^{-k}\right),$$

for constants $D_1, ..., D_k$, where $\overline{E}(\cdot) \equiv \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(\cdot)$ and where $\alpha_i(\theta) \equiv \arg\max_{\alpha_i} E \log f(z_{it}; \theta, \alpha_i)$. Note that $\theta_0 = \arg\max_\theta \overline{E} \log f_{it}(\theta, \alpha_i(\theta))$. Let $\bar{l}_{1/2}(\theta)$ be the average of the two half-panel (profile) log-likelihood functions, and define the jackknife-corrected log-likelihood as

$$\dot{l}_{1/2}(\theta) \equiv 2l(\theta) - \bar{l}_{1/2}(\theta).$$

Clearly, $p\lim_{N\to\infty} \dot{l}_{1/2}(\theta) = \overline{E} \log f_{it}(\theta, \alpha_i(\theta)) + O(T^{-2})$. Define the maximum jackknife-corrected likelihood estimator as

$$\dot{\theta}_{1/2} \equiv \arg\max_\theta \dot{l}_{1/2}(\theta)$$

and define $\dot{\theta}_{1/g}$ and $\dot{\theta}_{1/G}$ by analogy to $\widehat{\theta}_{1/g}$ and $\widehat{\theta}_{1/G}$. Then $p\lim_{N\to\infty} \dot{\theta}_{1/G} = \theta_0 + O(T^{-k-1})$ and, as $N, T \to \infty$ and $N/T^{2k+1} \to 0$,

$$\sqrt{NT}(\dot{\theta}_{1/G} - \theta_0) \xrightarrow{d} N(0, \Omega)$$

and

$$\sqrt{NT}(\dot{\theta}_{1/G} - \widehat{\theta}_{1/G}) \xrightarrow{p} 0.$$

Thus, $\dot{\theta}_{1/G}$ and $\widehat{\theta}_{1/G}$ are asymptotically equivalent under the asymptotics considered. For fixed T, $\dot{\theta}_{1/G}$ and $\widehat{\theta}_{1/G}$ are not asymptotically equivalent. Remark that $\dot{\theta}_{1/G}$ is equivariant under one-to-one parameter transformations, whereas $\widehat{\theta}_{1/G}$ is not.

For any of the estimators proposed, the bootstrap (where the i's are resampled) and the standard jackknife (where each i is deleted, one at a time) yield consistent estimates of its large N, fixed T variance.