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 and Université Laval

First version: 15 February 2006
Revised: 28 February 2006

Preliminary and incomplete

Abstract

We propose jackknife estimators for nonlinear dynamic panel data models with fixed effects that reduce the asymptotic bias of the maximum likelihood estimator (MLE) from $O(T^{-1})$ to $O(T^{-2})$ or smaller. The estimators are linear combinations of the MLE computed from the full panel and the MLE's computed from two or more shorter subpanels. The relative lengths of the subpanels determine the order of bias reduction that can be achieved. The jackknife can in a similar manner be applied to correct the score or the likelihood. Preliminary simulation results for the probit and logit binary AR(1) models are very encouraging. Even in small, short panels such as $N = 25$ and $T = 8$, the jackknife is very effective in reducing the bias of the MLE and has smaller mean squared error.

JEL: C13, C22
Keywords: Fixed effects, dynamic panel models, asymptotic bias, bias reduction, panel jackknife

*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. I am grateful to Manuel Arellano for stimulating discussion.

1 Introduction

Since Neyman and Scott (1948) it is known that the presence of fixed effects in panel models generally renders the MLE of the common parameters inconsistent as the number of cross-section units, N, tends to infinity while the number of observations along the time dimension, T, remains fixed. For certain nonlinear models there exist fixed-T consistent estimators, such as the conditional MLE for the static panel logit model and Honoré and Kyriazidou's (2000) estimator for the dynamic panel logit model. The latter estimator, however, has slow convergence and the conditional MLE approach requires the existence of orthogonal fixed effects relative to the common parameters (in the information sense) and the MLE of the orthogonal fixed effects to be free of the common parameters. These conditions are not satisfied in general, and fixed-T consistency may even be impossible due to lack of identification (Chamberlain, 1993), or the fixed-T information bound may be singular (Hahn, 2001).

The current approach, which has the advantage of greater generality, is to correct the MLE to reduce its bias from $O(T^{-1})$ down to at most $O(T^{-2})$, as in Hahn and Kuersteiner (2002), extending Kiviet (1995), for dynamic linear models; Woutersen (2002), Arellano (2003) and Hahn and Newey (2004) for nonlinear static models with iid data; and Hahn and Kuersteiner (2004) for nonlinear dynamic models. Arellano and Hahn (2005) survey the literature on bias corrections for nonlinear panel models and closely related work on likelihood corrections in the statistics literature. The corrections in Hahn and Newey (2004) and Hahn and Kuersteiner (2004) are based on an expansion of the large N bias of the MLE in powers of T^{-1}, from which the first term is analytically calculated (with the MLE as plug-in estimate) and subtracted from the MLE. Alternatively, as suggested by Hahn and Newey (2004) for nonlinear static panel models with iid data (including iid covariates), the bias can be reduced from $O(T^{-1})$ to $O(T^{-2})$ by a panel jackknife, thus yielding an automatic bias correction of the MLE.

In this paper we extend the panel jackknife approach of Hahn and Newey (2004) to nonlinear dynamic panel models. The jackknife exploits the fact that the bias of the MLE is roughly proportional to T^{-1} by comparing the MLE from the full sample with the MLE's computed from subsamples. This comparison yields an estimate of the bias term. In a panel setting with individual fixed effects, the subsamples are subpanels with fewer observations along the time dimension. While Hahn and Newey use the delete-one jackknife, we remove the $O(T^{-1})$ term of the bias by a half-panel jackknife. The idea is as follows: let $\hat{\theta}$ be MLE of the common parameter based on the full panel, and $\hat{\theta}_1$ and the $\hat{\theta}_2$ the MLE's based on the first and the second half-panel, respectively, where each half-panel consists of $T/2$ consecutive observations over time for all cross-section units. Then $\hat{\theta}^h = 2\hat{\theta} - \frac{1}{2}(\hat{\theta}_1 + \hat{\theta}_2)$ has bias $O(T^{-2})$ provided that the remainder of the first-order expansion of the bias of $\hat{\theta}$ is $O(T^{-2})$. Choosing the subpanels in this way ensures that the time series dependency structure is the same in the full panel and in the subpanels. Hence the full panel MLE and the subpanel MLE's share the same expansion of the bias in powers of T^{-1} and $(T/2)^{-1}$, respectively. Furthermore, by a standard argument, the large N, large T asymptotic variances of $\hat{\theta}^h$ and $\hat{\theta}$ are identical. Slight variations of the half-panel jackknife lead to higher-order bias reductions. By choosing two subpanels of different lengths, an appropriate linear combination of $\hat{\theta}$ and the subpanel MLE's has bias $O(T^{-3})$. Three subpanels of different lengths allow the bias to be reduced to $O(T^{-4})$, and so on. The bias reductions, especially those of higher order, come at the price of increased variance in finite samples. Nevertheless, it is worth considering higher order bias corrections in situations with relatively large N and small T, as often encountered in microeconometric applications. This point needs further investigation in a range of nonlinear panel models such as discrete choice models, Tobit models, truncated regression models, duration models, etc., all of which often have dynamics or at least dependent observations.

The paper is organised as follows. Section 2 considers the asymptotic bias of the MLE from a sequentially large N, large T perspective and discusses the relative merits of the jackknife in a fixed effects panel setting vs. a single time series setting. Section 3 presents the dynamic panel jackknife estimators, with an emphasis on the half-panel jackknife. In Section 4 we make the observation that the jackknife can also be applied to bias-correct the score function or, equivalently, the likelihood function. Section 5 presents some Monte Carlo results on Gaussian and binary fixed effects panel AR(1) models, where the panel jackknife is shown to be very effective in reducing the bias of the MLE. Section 6 concludes.

2 Asymptotic bias of the MLE

Suppose we observe $z_{it} \equiv (y_{it}, x_{it})$ for $i = 1, ..., N$ and $t = 0, ..., T$. Assume the time series processes z_{it} are stationary and ergodic for each i and independent across i but not necessarily identical. The heterogeneity across i is assumed to be captured by an unobserved fixed effect α_i which, together with a parameter vector θ that is common to all i, determines the (possibly discrete) conditional density $f_{it}(\theta, \alpha_i) \equiv f(y_{it}|x_{it}, z_{it-1}; \theta, \alpha_i)$, where f is known. We explicitly allow the covariates x_{it} to be dependent over time. This implies that, even if the model is static, i.e. $f(y_{it}|x_{it}, z_{it-1}; \theta, \alpha_i) = f(y_{it}|x_{it}; \theta, \alpha_i)$, z_{it} will generally be dependent over time. The restriction to first-order dynamics, rather than some other finite order, is without loss of generality. We take α_i to be scalar, but also this restriction is not essential to the argument. The true values of θ and α_i are (θ_0, α_{i0}), the unique maximisers of $E \log f_{it}(\theta, \alpha_i)$. We are interested in estimating θ_0 and consider the conditional MLE given z_{i0}, $i = 1, ..., N$,

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

where $\hat{\alpha}_i(\theta) \equiv \arg\max_{\alpha_i} \sum_{t=1}^{T} \log f_{it}(\theta, \alpha_i)$. For fixed T, the MLE is generally inconsistent for θ_0. That is, $p\lim_{N\to\infty} \hat{\theta} \neq \theta_0$, because $p\lim_{N\to\infty} \hat{\theta}$ maximises

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} E\left(\sum_{t=1}^{T} \log f_{it}(\theta, \hat{\alpha}_i(\theta))\right),$$

where the expectation and the limit are assumed to exist, while θ_0 is seen to maximise

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} E\left(\sum_{t=1}^{T} \log f_{it}(\theta, \alpha_i(\theta)\right),$$

where $\alpha_i(\theta) \equiv \arg\max_{\alpha_i} E \log f_{it}(\theta, \alpha_i)$. With $\hat{\alpha}_i(\theta) \neq \alpha_i(\theta)$, these two maximands are different, and so are their maximisers.

We make the high-level assumption that f and the processes z_{it} are sufficiently regular to allow the asymptotic bias of $\hat{\theta}$ to be expanded as

$$p \lim_{N\to\infty} \hat{\theta} - \theta_0 = \frac{B_1}{T} + \frac{B_2}{T^2} + ... + \frac{B_j}{T^j} + o\left(T^{-j}\right) \tag{1}$$

for some positive integer j. (We take a sequential limit here, $N \to \infty$ followed by $T \to \infty$.) Primitive conditions under wich this expansion holds for given j need to be established, but as in the iid case (where z_{it} is iid over time and thus $f(y_{it}|x_{it}, z_{it-1}; \theta, \alpha_i) \equiv f(y_{it}|x_{it}; \theta, \alpha_i)$ and x_{it} is iid) we anticipate the conditions to take the form of integrable Lipschitz bounds on $\log f_{it}$ and its derivatives with respect to θ and α_i up to the required order. Hahn and Kuersteiner (2004) give conditions ensuring the validity of (1) for $j = 1$ and derive an expression for B_1 in the more general case where $\hat{\theta}$ is an M-estimator. They suggest using plug-in estimates of the unknown quantities that appear in B_1 to form $\tilde{B}_1$, say, and then correcting $\hat{\theta}$ to yield the bias-corrected estimator $\hat{\theta} - \tilde{B}_1/T$.

The approach taken in this paper differs from Hahn and Kuersteiner (2004) in that, while also relying on the validity of (1), it employs the jackknife to bias-correct $\hat{\theta}$ in an automatic fashion, i.e. without any need to calculate likelihood-derived quantities. The idea of using the jackknife for bias reduction in panel models with fixed effects is due to Hahn and Newey (2004). They showed that, when z_{it} is iid, the "delete-one" panel jackknife estimator, $T\hat{\theta} - (T - 1)T^{-1}\sum_{t=1}^{T}\hat{\theta}_{(t)}$, where $\hat{\theta}_{(t)}$ is the MLE computed from the subpanel with the t-th observation removed for all i, has asymptotic bias $O(T^{-2})$, provided that (1) holds for $j = 2$. The application of the jackknife to fixed effects panel models with relatively large N and small T is actually much more powerful than the usual application of the jackknife, which may be thought of as taking place in a context where $N = 1$. To get an intuitive idea of the power of the jackknife in a panel context, it suffices to consider the reduction in mean-squared error (MSE) achieved by the jackknife: in a fixed effects panel model the MSE (equal to variance plus squared bias) is reduced from $O(N^{-1}T^{-1}) + O(T^{-2})$ to $O(N^{-1}T^{-1}) + O(T^{-4})$. Thus, when N is large compared to T, one expects the jackknife to reduce the MSE considerably, while at best only marginally so when $N = 1$.

Turning to the case with dependent z_{it} over time, as here, the delete-one panel jackknife is not expected to have the same bias reduction property. This is because, as noted by Hahn and Kuersteiner (2004) and Arellano and Hahn (2005), removing the t-th observation for all i from the panel yields a subpanel with a dependency structure different from that in the full panel, and therefore the expansion (1), when applied to $\hat{\theta}_{(t)}$ with T replaced with $T - 2$, which is the effective number of observations that $\hat{\theta}_{(t)}$ employs, is unlikely to have the same coefficients B_1 and B_2 . In the next section we show that, by designing the panel jackknife estimator differently, it is possible to generalise Hahn and Newey's (2004) approach to automatic bias correction to dependent data and to higher order bias correction.

3 Dynamic panel jackknife estimators

When (1) holds for some $j \geq 1$, the asymptotic bias of $\hat{\theta}$ can be reduced from $O(T^{-1})$ to $o(T^{-j})$. Before considering the bias correction in detail, it is useful to point out that, while much of the 1960's and 1970's literature on the jackknife emphasised the delete-one jackknife, some of the earliest contributions make more explicit mention of other possibilities (see Miller, 1974, for a review on the jackknife). Tukey (1958), referring to Quenouille (1956), in a famous abstract wrote "Let $y_{(\cdot)}$ be the estimate based on all the data, $y_{(i)}$ that based on all but the ith *piece* [our italics], $\bar{y}_{(i)}$ the average of the $y_{(i)}$" and then went on to summarise the key properties of the n quantities $ny_{(\cdot)} - (n-1)y_{(i)}$, where n is the number of data "pieces", taken to be of equal size. Quenouille (1956) already considered splitting the data in just two half-samples to form the bias-corrected estimator $2y_{(\cdot)} - \bar{y}_{(i)}$, in the above notation, and explicitly mentioned the possibility to use two half-series in estimation problems where the data are a single time series. This idea naturally extends to panel data with fixed effects.

3.1 First-order bias correction

Assume (1) holds for $j = 1$. Let A be a subset of $\mathcal{T} \equiv \{1, ..., T\}$ with at least two elements and consider the estimator

$$\hat{\theta}_A \equiv \arg\max_{\theta} \sum_{i=1}^{N} \sum_{t \in A} \log f_{it}(\theta, \hat{\alpha}_{iA}(\theta)),$$

where $\hat{\alpha}_{iA}(\theta) \equiv \arg\max_{\alpha_i} \sum_{t \in A} \log f_{it}(\theta, \alpha_i)$. Clearly, if (1) holds for $j = 1$ it must also hold for $\hat{\theta}_A$ if A is a set of consecutive integers. Now, take proper subsets $A_1 \equiv \{1, ..., T_1\}$ and $A_2 \equiv \{T - T_2 + 1, ..., T\}$ of $\mathcal{T}$ such that $T_1, T_2 \geq 2$, $A_1 \cup A_2 = \mathcal{T}$ and $\#(A_1 \cap A_2)$ is bounded as $T \to \infty$ while T_1/T and T_2/T are bounded away from zero. That is, A_1 and A_2 partition $\mathcal{T}$ apart from a bounded overlap, and A_1 and A_2 grow at the same rate as $\mathcal{T}$. Then

$$p\lim_{N\to\infty}\hat{\theta}_{A_k}-\theta_0=\frac{B_1}{T_k}+o\left(T_k^{-1}\right),\qquad k=1,2.$$

As a result,

$$p\lim_{N\to\infty}\frac{T_k}{T-T_k}(\hat{\theta}_{A_k}-\hat{\theta})=\frac{B_1}{T}+o(T^{-1}),\qquad k=1,2,$$

suggesting the subpanel jackknife estimator

$$\hat{\theta}^c\equiv\hat{\theta}-\frac{1}{2}\sum_{k=1,2}\frac{T_k}{T-T_k}\left(\hat{\theta}_{A_k}-\hat{\theta}\right),\tag{2}$$

which has asymptotic bias $o(T^{-1})$. The asymptotic bias reduces further to $O(T^{-2})$ if (1) also holds for $j=2$. By stationarity and ergodicity, $\hat{\theta}_{A_1}$ and $\hat{\theta}_{A_2}$ are almost independent for large T, and the covariances between $\hat{\theta}_{A_1}$ and $\hat{\theta}$, and between $\hat{\theta}_{A_2}$ and $\hat{\theta}$, are almost equal to the variance of $\hat{\theta}$. We may rewrite $\hat{\theta}^c$ as

$$\hat{\theta}^c=\frac{1}{2}\sum_{k=1,2}\left(\frac{T}{T-T_k}\hat{\theta}-\frac{T_k}{T-T_k}\hat{\theta}_{A_k}\right),$$

where the summand is easily shown to be almost independent over $k=1,2$ for large enough T. A further calculation shows that, for large N and large T, the variance of $\hat{\theta}^c$ is almost equal to the variance of $\hat{\theta}$ times $T^2/(4T_1T_2)$. Hence, because $T/(T_1+T_2)\to 1$, the asymptotic variance of $\hat{\theta}^c$ is minimised by choosing $T_1=T_2\cong T/2$. This suggests the half-panel jackknife estimator

$$\hat{\theta}^h\equiv\frac{T}{T-\lceil T/2\rceil}\hat{\theta}-\frac{\lceil T/2\rceil}{T-\lceil T/2\rceil}\frac{\hat{\theta}_{H_1}+\hat{\theta}_{H_2}}{2},$$

where $H_1\equiv\{1,...,\lceil T/2\rceil\}$ and $H_2\equiv\{\lfloor T/2\rfloor+1,...,T\}$ index two half-panels of equal length, with either one or no element in common. The choice of half-panels is further motivated by the fact that it minimises the leading term, which is $-\frac{1}{2}(T_1T_2)^{-1}B_2$, of the expansion of the asymptotic bias of $\hat{\theta}^c$, provided (1) holds with $j=2$. We return to this point below. In summary, while all subpanel jackknife estimators in the class described by $\hat{\theta}^c$ remove the first-order asymptotic bias from $\hat{\theta}$, $\hat{\theta}^h$ does so without increasing the asymptotic variance and at once has the smallest leading term in the expansion of the bias.

The half-panel jackknife estimator requires $T \geq 3$ to be applicable, because $T_1 = T_2 = 2$ is the minimum number of observations for which the subpanel estimators are defined. The minimum number of N is 1. Of course, for any value of N, if T is small, the variance of $\hat{\theta}^h$ is expected to be substantially larger than that of $\hat{\theta}$. One may also wonder whether, for small T, the large N asymptotic bias of $\hat{\theta}^h$ is smaller than that of $\hat{\theta}$. Large N, large T asymptotic arguments, like the one that motivated $\hat{\theta}^h$, are unlikely to shed any light on this question. Simulations with large N, however, give a reliable answer. To conclude, there is value in applying $\hat{\theta}^h$ even with very small T, provided the fixed-T asymptotic bias of $\hat{\theta}^h$ is smaller in absolute value than that of $\hat{\theta}$ and N is sufficiently large for the variance inflation of $\hat{\theta}^h$ to be outweighed by the bias reduction.

It is perhaps worth remarking that the subpanel jackknife estimators presented here are valid in an iid context under weaker conditions than the delete-one jackknife, because the latter requires the expansion of the asymptotic bias to be valid up to *and including* the term in T^{-2}.

3.2 Higher-order bias correction

The bias reduction can be carried further. Assume (1) holds for $j = 2$. To obtain a second order bias reduction for $\hat{\theta}$ we can apply a correction to $\hat{\theta}^h$ (see Quenouille, 1956, and Schucany, Gray and Owen, 1971, for a similar proposal) or directly to $\hat{\theta}$. If (1) holds with $j = 2$, then

$$p \lim_{N \to \infty} \hat{\theta}^h - \theta_0 = \frac{B_2'}{T^2} + o\left(T^{-2}\right),$$

where $B_2' = -4B_2$. For any subset A of $\mathcal{T}$ consisting of consecutive integers that has at least three elements, let $\hat{\theta}_A^h$ be $\hat{\theta}^h$ applied to A just as $\hat{\theta}_A$ is $\hat{\theta}$ applied to A. Then, with H_1 and H_2 the half-panel index sets defined earlier,

$$p\lim_{N\to\infty} \hat{\theta}^h_{H_k} - \theta_0 = \frac{B_2'}{\lceil T/2 \rceil^2} + o\left(T^{-2}\right), \qquad k = 1, 2,$$

leading to the nested half-panel jackknife estimator

$$\hat{\theta}^{hh} \equiv \frac{T^2}{T^2 - \lceil T/2 \rceil^2}\hat{\theta}^h - \frac{\lceil T/2 \rceil^2}{T^2 - \lceil T/2 \rceil^2}\frac{\hat{\theta}^h_{H_1} + \hat{\theta}^h_{H_2}}{2},$$

which has asymptotic bias $o(T^{-2})$ or, if (1) holds for $j = 3$, $O(T^{-3})$. The computational requirements for $\hat{\theta}^{hh}$ are very modest: only $1 + 2(1 + 2) = 7$ maximum likelihood estimates need to be computed. The minimal T for this procedure is 5. For large N and T, is can be shown that the variance of $\hat{\theta}^{hh}$ is almost equal to that of $\hat{\theta}^h$ and hence to that of $\hat{\theta}$. For small T, however, it is expected that the variance of $\hat{\theta}^{hh}$ will be considerably larger than that of $\hat{\theta}^h$.

To remove the second order bias directly from $\hat{\theta}$, we need to estimate $B_1/T + B_2/T^2$. This requires combining $\hat{\theta}$ with at least two estimators, $\hat{\theta}_{A_1}$ and $\hat{\theta}_{A_2}$, based on subpanels of different length. Now choose the subsets A_1 and A_2 of $\mathcal{T}$ according to the rules given earlier and, in addition, $T_1 \neq T_2$. Defining

$$a_1 \equiv \frac{T_1^2}{(T - T_1)(T_1 - T_2)}, \qquad a_2 \equiv \frac{T_2^2}{(T - T_2)(T_2 - T_1)},$$

a straightforward calculation shows that

$$p\lim_{N\to\infty} (a_1\hat{\theta}_{A_1} + a_2\hat{\theta}_{A_2} - (a_1 + a_2)\hat{\theta}) = \frac{B_1}{T} + \frac{B_2}{T^2} + o(T^{-2}).$$

Hence the subpanel jackknife estimator

$$\hat{\theta}^{cc} \equiv (1 + a_1 + a_2)\hat{\theta} - (a_1\hat{\theta}_{A_1} + a_2\hat{\theta}_{A_2})$$

has asymptotic bias $o(T^{-2})$, or even $O(T^{-3})$ if (1) holds for $j = 3$. This estimator requires only 3 maximum likelihood estimates to be computed, and is applicable when $T \geq 4$. It is not clear yet how one should best partition $\mathcal{T}$. It can be shown that, if (1) holds for $j = 3$, then

$$p\lim_{N\to\infty} \hat{\theta}^{cc} - \theta_0 = \left(\frac{T^2}{T_1 T_2}\right)\frac{B_3}{T^3} + o\left(T^{-3}\right).$$

Thus, the next term in the asymptotic bias is minimised if T_1 and T_2 are almost, but not exactly, equal. This is likely, however, to inflate the asymptotic variance, because with T_1 close to T_2 the weights a_1 and a_2 become very large. Further analysis is required on this point. Finally, the variance of $\hat{\theta}^{cc}$ is reduced by replacing $\hat{\theta}_{A_1}$ and $\hat{\theta}_{A_2}$ with $(\hat{\theta}_{A_1}+\hat{\theta}_{A_1'})/2$ and $(\hat{\theta}_{A_2}+\hat{\theta}_{A_2'})/2$, respectively, where (A_1', A_2') are obtained by 'mirroring' (A_1, A_2)—for example, if $A_1 = \{1, 2, 3\}$ and $A_2 = \{3, 4, 5, 6\}$, then $A_1' = \{4, 5, 6\}$ and $A_2' = \{1, 2, 3, 4\}$. This mirrored version of $\hat{\theta}^{cc}$ is implemented in the Monte Carlo experiments reported in Section 5.

The bias reduction can, in principle, be carried to any order provided the expansion of $p\lim_{N\to\infty}\hat{\theta}$ in powers of T^{-1} is valid up to that order. Suppose (1) is valid for a given j. Partition $\mathcal{T}$ roughly into proper subsets $A_1, ..., A_j$, consisting of $T_1, ..., T_j$ consecutive integers, respectively, such that $A_1 \cup ... \cup A_j = \mathcal{T}$ and, for all k and $l \neq k$, $T_k \geq 2$, $T_k \neq T_l$ and $\#(A_k \cap A_l)$ is bounded as $T \to \infty$ while T_k/T is bounded away from zero. Let

$$a_k \equiv \frac{T_k^j}{(T - T_k)\prod_{l=1;l\neq k}^{j}(T_k - T_l)}, \qquad k = 1, ..., j.$$

Then

$$\hat{\theta}^{c\ldots c} \equiv \left(1 + \sum_{k=1}^{j} a_k\right)\hat{\theta} - \sum_{k=1}^{j} a_k \hat{\theta}_{A_k}$$

has bias $o(T^{-j})$. If (1) is valid also for $j+1$, then some algebra shows that

$$p\lim_{N\to\infty}\hat{\theta}^{c\ldots c} - \theta_0 = (-1)^j\left(\frac{T^j}{\prod_{l=1}^{j} T_l}\right)\frac{B_{j+1}}{T^{j+1}} + o\left(T^{-j-1}\right).$$

Hence, at any order of bias reduction, the next term in the expansion of the bias is minimised by partitioning the full panel into subpanels with approximately, but not exactly, equal length. Finally, a multi-mirrored version of $\hat{\theta}^{c\ldots c}$ will have smaller variance than $\hat{\theta}^{c\ldots c}$.

4 Jackknife corrections of the score and the likelihood

Hahn and Arellano (2005) show, in the iid panel context, how the score function and the likelihood function can also be analytically bias-corrected, in fact more easily than the MLE itself. As might be expected, the subpanel jackknife can be turned into an automatic bias-correction of the score function and the likelihood function that is also valid with dependent data. For any $A \subset \mathcal{T}$, denote the normalised concentrated log-likelihood and score functions associated with A as

$$L_A(\theta) = \frac{1}{N \cdot \# A} \sum_{i=1}^{N} \sum_{t \in A} \log f_{it}(\theta, \hat{\alpha}_{iA}(\theta)),$$
$$s_A(\theta) = \partial L_A(\theta) / \partial \theta,$$

where, as before, $\hat{\alpha}_{iA}(\theta) \equiv \arg\max_{\alpha_i} \sum_{t \in A} \log f_{it}(\theta, \alpha_i)$. Let $L \equiv L_{\mathcal{T}}$ and $s \equiv s_{\mathcal{T}}$ be the full panel log-likelihood and score functions. Assume the bias of the score function at θ_0 admits an expansion

$$p \lim_{N \to \infty} s(\theta_0) = \frac{C_1}{T} + \frac{C_2}{T^2} + ... + \frac{C_j}{T^j} + o\left(T^{-j}\right),$$

for some $j \geq 1$. Then, for the half-panel score functions we get

$$p \lim_{N \to \infty} \frac{\lceil T/2 \rceil}{T} s_{H_k}(\theta_0) = \frac{C_1}{T} + o(T^{-1}), \qquad k = 1, 2.$$

Hence, $\tilde{\theta}^h$, the solution of

$$s(\theta) - \frac{1}{2} \sum_{k=1,2} \frac{\lceil T/2 \rceil}{T} s_{H_k}(\theta) = 0$$

has bias $o(T^{-1})$ or, if $j \geq 2$, $O(T^{-2})$. A correction of the likelihood function is obtained as follows. Assuming that, for some $j \geq 1$,

$$p \lim_{N \to \infty} L(\theta_0) - \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E \log f_{it}(\theta_0, \alpha_i) = \frac{D_1}{T} + \frac{D_2}{T^2} + ... + \frac{D_j}{T^j} + o\left(T^{-j}\right),$$

we have

$$p\lim_{N\to\infty}\left(L(\theta_0)-\frac{1}{2}\sum_{k=1,2}\frac{\lceil T/2\rceil}{T}L_{H_k}(\theta_0)\right)$$
$$=\left(1-\frac{1}{2}\sum_{k=1,2}\frac{\lceil T/2\rceil}{T}\right)\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\log f_{it}(\theta_0,\alpha_i)+o\left(T^{-1}\right).$$

Solving

$$\max_{\theta}\left(L(\theta)-\frac{1}{2}\sum_{k=1,2}\frac{\lceil T/2\rceil}{T}L_{H_k}(\theta)\right)$$

gives, again, $\tilde{\theta}^h$. As with $\hat{\theta}^h$, the correction to the score or likelihood does not increase the asymptotic variance of the ensuing estimator, $\tilde{\theta}^h$, and higher-order corrections are also possible. We conjecture that $p\lim_{N\to\infty}(\hat{\theta}^h-\tilde{\theta}^h)=O(T^{-3})$ for smooth enough likelihoods.

5 Monte Carlo results

We conducted some Monte Carlo experiments to compare the MLE, $\hat{\theta}$, with the first-order and second-order bias correcting panel jackknife, $\hat{\theta}^h$ and $\hat{\theta}^{cc}$, in the Gaussian AR(1) panel model and in binary AR(1) panel models, all with fixed effects. In all cases, $\hat{\theta}^h$ is based on two half-panels of exactly equal length, $\lceil T/2\rceil$, and $\hat{\theta}^{cc}$ is the mirrored version of the panel jackknife defined by two non-intersecting half-panels whose lengths, $\lceil (T+1)/2\rceil$ and $\lfloor (T-1)/2\rfloor$, differ by one or two. The Gaussian AR(1) model is linear and there exist fixed-T consistent estimators for it. At this stage we use this model to conduct quick and easy experiments with the jackknife estimators that we propose, because we see no a priori reason why the linearity assumption would drive the experimental results.

5.1 Gaussian AR(1) panel model

The Gaussian AR(1) panel model with fixed effects is

$$y_{it}=\alpha_i+\theta y_{it-1}+\epsilon_{it},\qquad \epsilon_{it}\sim \text{iid } N(0,1),\qquad t=1,...,T,\qquad i=1,...,N,$$

with $y_{i0} \sim$ iid $N(\alpha_i(1-\theta)^{-1}, (1-\theta)^2)$. Since the conditional MLE of θ, given the initial values y_{i0}, $i = 1, ..., N$, is pooled least-squares on the demeaned variables for each i and so sweeps out the fixed effects, the MLE is invariant with respect to the α_i's. The MLE is known to be severely biased towards zero, especially when θ is close to one. We run the experiment with $N \in \{5, 50, 500, 5000\}$, $T \in \{3, ..., 8, 10, 12\}$ and $\theta \in \{0.5, 0.9, 1\}$. Although the assumption of stationarity, of course, does not cover the unit root case, we included it in the simulation because it is well known that otherwise fixed-T consistent IV estimators that use lagged levels as instruments suffer from a weak instrument problem when θ is close to one, and are inconsistent, as $T \to \infty$, when $\theta = 1$. In the unit root case, the data are generated by $y_{i0} = 0$ (or any other set of values, in view of the invariance of the MLE) and $y_{it} = y_{it-1} + \epsilon_{it}$, or one may take, in the above scheme, $\theta = 1 - 10^{-8}$, say, without technical difficulty.

Table 1 in the Appendix reports estimates, based on 10000 Monte Carlo runs, of the bias and the root-MSE (RMSE) of $\hat{\theta}$, $\hat{\theta}^h$ and $\hat{\theta}^{cc}$. Here and later, figures in boldface indicate which one of $\hat{\theta}$, $\hat{\theta}^h$ and $\hat{\theta}^{cc}$ has the smallest bias (in absolute value) and RMSE. A very clear and striking picture emerges from the simulations. For *all* points in the design, $\hat{\theta}$ has a larger bias than $\hat{\theta}^h$ and $\hat{\theta}^{cc}$, and for *almost all* points, $\hat{\theta}^{cc}$ has a smaller bias than $\hat{\theta}^h$. In fact, the bias of $\hat{\theta}^{cc}$ is uniformly very small, whereas the bias of $\hat{\theta}^h$, while much smaller than that of $\hat{\theta}$, is not negligibly small. Thus, the bias-corrections effectively reduce the bias even for values of T as small as 3, 4 or 5, and N as small as 5. It is surprising that the second-order correction, entirely motivated on asymptotic grounds, yields a systematically smaller bias than the first-order correction. For $\hat{\theta}$, the bias term always dominates in the RMSE, as expected. For $\hat{\theta}^h$, the bias term starts to become important in the RMSE from $N = 50$ onward, and for $\hat{\theta}^{cc}$ from $N = 500$ onward. The results for the RMSE also reveal that the bias-correction indeed inflates the variance, and more so for $\hat{\theta}^{cc}$ than for $\hat{\theta}^h$. As a result, for $N = 5$, $\hat{\theta}^h$ has the smallest RMSE, although only marginally smaller than that of $\hat{\theta}$. For larger values, $N = 50$ or 500, $\hat{\theta}^{cc}$ starts to outperform $\hat{\theta}^h$ in terms of RMSE. For $N = 5000$, $\hat{\theta}^{cc}$ has by far the smallest RMSE, except in some cases where the bias of $\hat{\theta}^h$ is, incidentally, close to zero. From $N = 50$ and $T = 5$ onward, the RMSE of $\hat{\theta}$ is almost always at least twice as large as that of $\hat{\theta}^h$ and $\hat{\theta}^{cc}$. Finally, we note that the presence of a unit root poses no particular difficulty for the panel jackknife. If $\theta = 1$, then, with $N = 5$ and $T = 12$, for example, $E(\hat{\theta}) = 0.741$, $E(\hat{\theta}^h) = 0.951$ and $E(\hat{\theta}^{cc}) = 0.991$. Thus, $\hat{\theta}^{cc}$ is nearly unbiased in the unit root case.

5.2 Binary AR(1) panel models

The binary AR(1) panel model with fixed effects is

$$y_{it} = 1(\alpha_i + \theta y_{it-1} + \epsilon_{it}), \qquad t = 1, ..., T, \qquad i = 1, ..., N,$$

where ϵ_{it} is iid with cdf $F(\cdot)$ (here taken to be the standard normal or the logistic cdf, leading to the probit and logit AR(1) models) and y_{i0} are independent draws from the stationary Bernoulli distribution $P(y_{it} = 1) = F(\alpha_i)/(1 + F(\alpha_i) - F(\alpha_i + \theta))$. We let $N \in \{25, 100, 250, 500, 1000\}$, $T \in \{4, 8, 16, 32\}$ and $\theta \in \{0.5, 1\}$. The MLE is severely downward biased and not invariant with respect to the α_i's. To avoid arbitrary choices, we set $\alpha_i = 0$. We also experimented with fixed effects equidistantly distributed on $[-1, 1]$ (not reported here), which led to similar conclusions.

Tables 2 and 3 report the results for the logit and probit models, respectively, all based on 10000 Monte Carlo runs. As in the Gaussian AR(1) model the bias corrections effectively reduce the bias even for relatively short panels. When $T \geq 8$, $\hat{\theta}$ always has a larger bias than $\hat{\theta}^h$; in turn, when $T \geq 16$, $\hat{\theta}^h$ has a larger bias than $\hat{\theta}^{cc}$. The bias of $\hat{\theta}^h$ and, in particular, of $\hat{\theta}^{cc}$, rapidly becomes vanishingly small as T increases. Turning to the RMSE we see that the RMSE of $\hat{\theta}$ is always dominated by the bias. For $\hat{\theta}^h$, the bias term starts to become important in the RMSE from $N = 100$ onward, and for $\hat{\theta}^{cc}$ only from $N = 1000$. The results again show an inflation of the variance due to the bias correction. In contrast to the Gaussian AR(1) model, $\hat{\theta}^{cc}$ does not outperform $\hat{\theta}^h$ in terms of RMSE for any sample size considered here. Yet, whenever the bias of $\hat{\theta}^{cc}$ is smaller than that of $\hat{\theta}^h$, which frequently happens, for large enough N, $\hat{\theta}^{cc}$ *will* outperform $\hat{\theta}^h$ in terms of RMSE. For the sample sizes considered here, except when $T = 4$, $\hat{\theta}^h$ has the smallest RMSE. Although the improvement in RMSE of $\hat{\theta}^h$ compared to $\hat{\theta}$ is only marginal for $N = 25$, the RMSE of $\hat{\theta}^h$ is already more than twice as small as the RMSE of $\hat{\theta}$ when $N = 100$ and the difference increases further as N grows.

We conclude from the experiments conducted for the logit and probit AR(1) models that the bias corrections remove the bias of the MLE to a very large extent, but at the cost of variance inflation, especially as concerns the second-order bias correction implemented through $\hat{\theta}^{cc}$. Perhaps, as the asymptotic theory suggests, the nested half-panel estimator, $\hat{\theta}^{hh}$, is less prone to variance inflation. This will require further analysis and experiments.

6 Concluding remarks

We proposed dynamic panel jackknife estimators that reduce the bias of the MLE in panel models with fixed effects. The estimators are applicable to parametric models in general, whether linear or non-linear, dynamic or static, with iid or dependent data. The half-panel jackknife, which we emphasised, can be applied in static models (but with possibly dependent covariates) as soon as we have 3 time-series observations; if the model has first-order dynamic features, at least 4 time-series observations are required, and so on. Implementing the estimators poses no analytical nor computational difficulty beyond the requirement to compute the MLE. Thus, the dynamic panel jackknife is very easy to implement.

Although the dynamic panel jackknife was motivated by a sequentially large N, large T argument, we have in mind microeconometric applications where N is relatively large and T relatively small. Further simulations in a range of different models need to shed light on the question whether, with very small T, the dynamic panel jackknife can effectively reduce the bias of the MLE without inflating the variance too much. With small N, the variance is likely to be much larger. The power of the jackknife with panel data, as opposed to single time-series or single cross-section data, lies in the fact that, even if the variance of the jackknife estimator is *relatively* much larger than that of the MLE, both variances tend to zero as $N \to \infty$. Hence any variance inflation is eventually offset by the bias reduction—if there is any.

Much remains to be done. Primitive conditions under which the expansion of the asymptotic bias of the MLE holds up to the required order need to be established. Further, a rigorous asymptotic distribution theory for the dynamic panel jackknife estimators, as $N \to \infty$ and $T \to \infty$, possibly under different scenarios for the relative rate, has to be developed. This is important from the point of view of hypothesis testing and the construction of confidence sets. It would also be of interest to compare the half-panel jackknife (with half-panels of equal length) with the analytical bias reduction approach of Hahn and Kuersteiner (2004). Of course, the two methods should be equivalent up to and including the order $O(T^{-1})$, but they may be different at the order $O(T^{-2})$. Finally, with large N and very small T, any theoretical guidance to choose between first-order and higher-order bias correction methods would be of great interest.

References

[1] Arellano, M. (2003), Discrete choices with panel data, *Investigaciones Económicas*, 27, 423–458.

[2] Arellano, M. and J. Hahn (2005), Understanding bias in nonlinear panel models: Some recent developments, Invited Lecture at the Econometric Society World Congress, London, 2005.

[3] Chamberlain, G. (1993), Feedback in panel data models, unpublished manuscript.

[4] Hahn, J. and G. Kuersteiner (2002), Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, *Econometrica*, 70, 1639–1658.

[5] Hahn, J. (2001), The information bound of a dynamic panel logit model with fixed effects, *Econometric Theory*, 17, 913-932.

[6] Hahn, J. and G. Kuersteiner (2004), Bias reduction for dynamic nonlinear panel models with fixed effects, unpublished manuscript.

[7] Hahn, J. and W.K. Newey (2004), Jackknife and analytical bias reduction for nonlinear panel models, *Econometrica*, 72, 1295–1319.

[8] Honoré, B. E. and E. Kyriazidou (2000), Panel data discrete choice models with lagged dependent variables, *Econometrica*, 68, 839–874.

[9] Kiviet, J. (1995), On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics*, 68, 53–78.

[10] Miller, R.G. (1971), The jackknife – a review, *Biometrika*, 61,1-15.

[11] Neyman, J. and E. L. Scott (1948), Consistent estimates based on partially consistent observations, *Econometrica*, 16, 1–32.

[12] Quenouille, M.H. (1956), Notes on bias in estimation, *Biometrika*, 43, 353–360.

[13] Schucany, W.R., H.L. Gray and D.B. Owen (1971), On bias reduction in estimation, *Journal of the American Statistical Association*, 66, 524-533.

[14] Tukey, J.W. (1958), Bias and confidence in not-quite large samples (Abstract), *Annals of Mathematical Statistics*, 29, 614.

[15] Woutersen, T. (2002), Robustness against incidental parameters, unpublished manuscript.

Table 1: Bias and RMSE of MLE and jackknife estimators for the Gaussian AR(1) fixed effects model

N	T	θ	Bias $\hat{\theta}$	Bias $\hat{\theta}^h$	Bias $\hat{\theta}^{cc}$	RMSE $\hat{\theta}$	RMSE $\hat{\theta}^h$	RMSE $\hat{\theta}^{cc}$
5	3	0.5	−0.555	**−0.303**		**0.639**	0.740	
5	3	0.9	−0.740	**−0.468**		**0.817**	0.838	
5	3	1	−0.779	**−0.506**		0.852	**0.841**	
5	4	0.5	−0.435	**−0.119**		**0.501**	0.521	
5	4	0.9	−0.596	**−0.242**		0.654	**0.576**	
5	4	1	−0.642	**−0.280**		0.697	**0.584**	
5	5	0.5	−0.353	−0.137	**0.011**	0.411	**0.375**	0.788
5	5	0.9	−0.504	−0.253	**−0.024**	0.551	**0.451**	0.789
5	5	1	−0.546	−0.290	**−0.063**	0.592	**0.471**	0.770
5	6	0.5	−0.294	−0.041	**0.010**	0.349	**0.323**	0.489
5	6	0.9	−0.431	−0.121	**−0.023**	0.473	**0.359**	0.482
5	6	1	−0.474	−0.161	**−0.052**	0.515	**0.369**	0.484
5	7	0.5	−0.256	−0.075	**0.016**	0.309	**0.275**	0.651
5	7	0.9	−0.377	−0.146	**−0.005**	0.415	**0.313**	0.609
5	7	1	−0.415	−0.177	**−0.038**	0.451	**0.325**	0.592
5	8	0.5	−0.223	**−0.011**	0.028	0.274	**0.250**	0.428
5	8	0.9	−0.335	−0.073	**−0.004**	0.371	**0.275**	0.416
5	8	1	−0.369	−0.098	**−0.020**	0.401	**0.273**	0.402
5	10	0.5	−0.179	**−0.003**	0.013	0.226	**0.204**	0.401
5	10	0.9	−0.271	−0.041	**0.008**	0.301	**0.221**	0.374
5	10	1	−0.306	−0.069	**−0.016**	0.333	**0.220**	0.355
5	12	0.5	−0.148	**0.002**	0.014	0.194	**0.177**	0.381
5	12	0.9	−0.231	−0.029	**0.006**	0.258	**0.191**	0.351
5	12	1	−0.259	−0.049	**−0.009**	0.283	**0.186**	0.327
50	3	0.5	−0.538	**−0.274**		0.546	**0.324**	
50	3	0.9	−0.710	**−0.407**		0.717	**0.443**	
50	3	1	−0.756	**−0.447**		0.763	**0.480**	
50	4	0.5	−0.415	**−0.078**		0.422	**0.164**	
50	4	0.9	−0.565	**−0.178**		0.571	**0.230**	
50	4	1	−0.605	**−0.209**		0.610	**0.252**	
50	5	0.5	−0.333	−0.111	**0.027**	0.340	**0.155**	0.222
50	5	0.9	−0.467	−0.204	**−0.018**	0.472	0.233	**0.203**
50	5	1	−0.504	−0.235	**−0.045**	0.509	0.259	**0.193**
50	6	0.5	−0.279	−0.019	**0.030**	0.285	**0.102**	0.144
50	6	0.9	−0.399	−0.087	**−0.007**	0.403	0.135	**0.132**
50	6	1	−0.433	−0.111	**−0.032**	0.437	0.147	**0.130**
50	7	0.5	−0.238	−0.055	**0.030**	0.244	**0.101**	0.199
50	7	0.9	−0.346	−0.118	**0.000**	0.351	**0.146**	0.174
50	7	1	−0.379	−0.144	**−0.018**	0.383	0.166	**0.161**
50	8	0.5	−0.207	**−0.001**	0.024	0.213	**0.079**	0.134
50	8	0.9	−0.305	−0.046	**0.005**	0.309	**0.094**	0.118
50	8	1	−0.337	−0.071	**−0.016**	0.340	**0.104**	0.109
50	10	0.5	−0.164	**0.006**	0.017	0.170	**0.065**	0.124
50	10	0.9	−0.246	−0.025	**0.010**	0.250	**0.072**	0.110
50	10	1	−0.276	−0.047	**−0.009**	0.279	**0.078**	0.099
50	12	0.5	−0.135	**0.008**	0.010	0.141	**0.057**	0.118
50	12	0.9	−0.206	−0.014	**0.013**	0.209	**0.060**	0.104
50	12	1	−0.233	−0.034	**−0.004**	0.236	**0.063**	0.092

N	T	θ	Bias $\hat{\theta}$	Bias $\hat{\theta}^h$	Bias $\hat{\theta}^{cc}$	RMSE $\hat{\theta}$	RMSE $\hat{\theta}^h$	RMSE $\hat{\theta}^{cc}$
500	3	0.5	−0.536	**−0.268**		0.537	**0.274**	
500	3	0.9	−0.707	**−0.403**		0.708	**0.407**	
500	3	1	−0.751	**−0.439**		0.751	**0.442**	
500	4	0.5	−0.412	**−0.073**		0.412	**0.086**	
500	4	0.9	−0.560	**−0.171**		0.561	**0.177**	
500	4	1	−0.601	**−0.200**		0.601	**0.205**	
500	5	0.5	−0.332	−0.110	**0.031**	0.332	0.115	**0.076**
500	5	0.9	−0.463	−0.200	**−0.018**	0.464	0.203	**0.064**
500	5	1	−0.500	−0.229	**−0.043**	0.501	0.232	**0.073**
500	6	0.5	−0.276	**−0.016**	0.030	0.277	0.036	**0.053**
500	6	0.9	−0.394	−0.082	**−0.006**	0.395	0.088	**0.041**
500	6	1	−0.429	−0.107	**−0.028**	0.429	0.112	**0.048**
500	7	0.5	−0.235	−0.052	**0.028**	0.236	**0.058**	0.068
500	7	0.9	−0.342	−0.114	**0.003**	0.343	0.117	**0.054**
500	7	1	−0.375	−0.141	**−0.019**	0.376	0.143	**0.054**
500	8	0.5	−0.206	**0.001**	0.023	0.206	**0.025**	0.047
500	8	0.9	−0.302	−0.043	**0.008**	0.303	0.050	**0.038**
500	8	1	−0.334	−0.067	**−0.014**	0.334	0.072	**0.036**
500	10	0.5	−0.162	**0.007**	0.016	0.163	**0.022**	0.042
500	10	0.9	−0.243	−0.023	**0.011**	0.244	**0.031**	0.037
500	10	1	−0.273	−0.046	**−0.008**	0.273	0.050	**0.032**
500	12	0.5	−0.134	**0.008**	0.011	0.135	**0.019**	0.039
500	12	0.9	−0.203	**−0.012**	0.013	0.204	**0.022**	0.035
500	12	1	−0.231	−0.033	**−0.004**	0.231	0.037	**0.029**
5000	3	0.5	−0.536	**−0.268**		0.536	**0.269**	
5000	3	0.9	−0.707	**−0.402**		0.707	**0.403**	
5000	3	1	−0.750	**−0.438**		0.750	**0.438**	
5000	4	0.5	−0.411	**−0.073**		0.411	**0.074**	
5000	4	0.9	−0.560	**−0.171**		0.561	**0.171**	
5000	4	1	−0.600	**−0.200**		0.600	**0.201**	
5000	5	0.5	−0.331	−0.109	**0.031**	0.331	0.110	**0.038**
5000	5	0.9	−0.463	−0.200	**−0.018**	0.463	0.200	**0.026**
5000	5	1	−0.500	−0.229	**−0.042**	0.500	0.229	**0.046**
5000	6	0.5	−0.276	**−0.015**	0.030	0.276	**0.018**	0.033
5000	6	0.9	−0.394	−0.081	**−0.006**	0.394	0.082	**0.014**
5000	6	1	−0.429	−0.107	**−0.029**	0.429	0.108	**0.031**
5000	7	0.5	−0.235	−0.052	**0.027**	0.235	0.053	**0.033**
5000	7	0.9	−0.342	−0.114	**0.003**	0.342	0.115	**0.017**
5000	7	1	−0.375	−0.141	**−0.019**	0.375	0.141	**0.024**
5000	8	0.5	−0.205	**0.001**	0.023	0.205	**0.008**	0.027
5000	8	0.9	−0.302	−0.043	**0.007**	0.302	0.044	**0.014**
5000	8	1	−0.333	−0.067	**−0.014**	0.333	0.067	**0.018**
5000	10	0.5	−0.162	**0.007**	0.016	0.162	**0.009**	0.020
5000	10	0.9	−0.243	−0.023	**0.012**	0.243	0.024	**0.016**
5000	10	1	−0.273	−0.045	**−0.008**	0.273	0.046	**0.012**
5000	12	0.5	−0.134	**0.008**	0.011	0.134	**0.010**	0.016
5000	12	0.9	−0.203	**−0.012**	0.013	0.203	**0.013**	0.017
5000	12	1	−0.231	−0.033	**−0.005**	0.231	0.033	**0.010**

Table 2: Bias and RMSE of MLE and jackknife estimators for the panel AR(1) logit model with fixed effects

N	T	θ	Bias $\hat{\theta}$	Bias $\hat{\theta}^h$	Bias $\hat{\theta}^{cc}$	RMSE $\hat{\theta}$	RMSE $\hat{\theta}^h$	RMSE $\hat{\theta}^{cc}$
25	4	0.5	**−1.449**	40.438		**1.754**	40.761	
25	4	1	**−1.558**	38.439		**1.827**	39.529	
25	8	0.5	−0.601	**0.251**	−1.081	**0.689**	0.807	4.269
25	8	1	−0.657	**0.219**	−1.080	0.745	**0.685**	4.291
25	16	0.5	−0.275	0.048	**−0.001**	0.352	**0.248**	0.710
25	16	1	−0.307	0.046	**−0.005**	0.385	**0.270**	0.758
25	32	0.5	−0.132	0.012	**0.004**	0.198	**0.155**	0.630
25	32	1	−0.151	0.009	**−0.001**	0.219	**0.170**	0.679
100	4	0.5	**−1.395**	43.697		**1.426**	43.874	
100	4	1	**−1.496**	43.541		**1.526**	43.579	
100	8	0.5	−0.595	0.190	**−0.152**	0.617	**0.304**	0.503
100	8	1	−0.657	0.177	**−0.129**	0.680	**0.298**	0.502
100	16	0.5	−0.276	0.041	**−0.008**	0.296	**0.127**	0.351
100	16	1	−0.310	0.039	**−0.007**	0.330	**0.138**	0.373
100	32	0.5	−0.132	0.011	**−0.003**	0.152	**0.078**	0.311
100	32	1	−0.151	0.008	**0.004**	0.170	**0.084**	0.334
250	4	0.5	**−1.384**	45.819		**1.397**	45.968	
250	4	1	**−1.489**	45.267		**1.501**	45.278	
250	8	0.5	−0.593	0.189	**−0.146**	0.602	**0.249**	0.338
250	8	1	−0.657	0.173	**−0.116**	0.666	**0.229**	0.333
250	16	0.5	−0.276	0.041	**−0.003**	0.284	**0.087**	0.225
250	16	1	−0.310	0.038	**−0.006**	0.319	**0.092**	0.240
250	32	0.5	−0.134	0.009	**−0.001**	0.142	**0.050**	0.201
250	32	1	−0.150	0.009	**0.001**	0.158	**0.054**	0.213
500	4	0.5	**−1.379**	46.903		**1.385**	47.014	
500	4	1	**−1.488**	46.429		**1.494**	46.430	
500	8	0.5	−0.591	0.193	**−0.149**	0.596	**0.227**	0.277
500	8	1	−0.658	0.171	**−0.109**	0.662	**0.199**	0.249
500	16	0.5	−0.276	0.040	**−0.007**	0.280	**0.068**	0.159
500	16	1	−0.309	0.039	**−0.004**	0.313	**0.071**	0.171
500	32	0.5	−0.134	0.009	**0.001**	0.138	**0.036**	0.139
500	32	1	−0.150	0.009	**0.001**	0.154	**0.040**	0.151
1000	4	0.5	**−1.380**	48.018		**1.383**	48.084	
1000	4	1	**−1.486**	48.433		**1.489**	48.434	
1000	8	0.5	−0.592	0.192	**−0.140**	0.594	**0.213**	0.223
1000	8	1	−0.658	0.169	**−0.108**	0.661	**0.186**	0.192
1000	16	0.5	−0.276	0.039	**−0.007**	0.278	**0.056**	0.114
1000	16	1	−0.310	0.038	**−0.002**	0.312	**0.058**	0.121
1000	32	0.5	−0.134	0.009	**0.000**	0.136	**0.026**	0.100
1000	32	1	−0.150	0.009	**−0.001**	0.152	**0.029**	0.107

Table 3: Bias and RMSE of MLE and jackknife estimators for the panel AR(1) probit model with fixed effects

N	T	θ	Bias $\hat{\theta}$	Bias $\hat{\theta}^h$	Bias $\hat{\theta}^{cc}$	RMSE $\hat{\theta}$	RMSE $\hat{\theta}^h$	RMSE $\hat{\theta}^{cc}$
25	4	0.5	**−0.921**	10.516		**1.009**	10.726	
25	4	1	**−1.110**	7.973		**1.269**	9.399	
25	8	0.5	−0.394	**0.127**	−0.252	0.448	**0.322**	0.977
25	8	1	−0.512	**0.067**	−0.404	0.568	**0.399**	1.327
25	16	0.5	−0.187	0.022	**−0.010**	0.232	**0.158**	0.444
25	16	1	−0.251	0.011	**0.009**	0.298	**0.193**	0.542
25	32	0.5	−0.089	0.007	**0.002**	0.130	**0.101**	0.405
25	32	1	−0.121	0.007	**0.001**	0.163	**0.120**	0.484
100	4	0.5	**−0.902**	11.404		**0.920**	11.419	
100	4	1	**−1.067**	11.278		**1.088**	11.310	
100	8	0.5	−0.393	0.113	**−0.089**	0.407	**0.180**	0.306
100	8	1	−0.509	**0.049**	−0.104	0.523	**0.168**	0.363
100	16	0.5	−0.186	0.022	**0.000**	0.198	**0.082**	0.223
100	16	1	−0.248	0.010	**−0.003**	0.261	**0.099**	0.266
100	32	0.5	−0.090	0.006	**0.001**	0.101	**0.051**	0.202
100	32	1	−0.121	0.007	**0.006**	0.133	**0.061**	0.242
250	4	0.5	**−0.898**	11.735		**0.905**	11.738	
250	4	1	**−1.066**	11.487		**1.074**	11.579	
250	8	0.5	−0.394	0.110	**−0.082**	0.400	**0.140**	0.202
250	8	1	−0.509	**0.028**	−0.102	0.514	**0.141**	0.241
250	16	0.5	−0.185	0.023	**−0.003**	0.190	**0.055**	0.141
250	16	1	−0.250	0.007	**−0.002**	0.255	**0.066**	0.170
250	32	0.5	−0.090	0.005	**0.000**	0.095	**0.032**	0.127
250	32	1	−0.121	0.006	**0.003**	0.126	**0.039**	0.151
500	4	0.5	**−0.897**	11.814		**0.901**	11.815	
500	4	1	**−1.065**	11.642		**1.069**	11.791	
500	8	0.5	−0.393	0.110	**−0.080**	0.396	**0.127**	0.155
500	8	1	−0.509	**0.018**	−0.097	0.512	**0.140**	0.183
500	16	0.5	−0.185	0.023	**−0.001**	0.188	**0.042**	0.100
500	16	1	−0.249	0.009	**0.000**	0.251	**0.051**	0.125
500	32	0.5	−0.090	0.005	**−0.001**	0.093	**0.023**	0.091
500	32	1	−0.121	0.007	**0.003**	0.123	**0.028**	0.107
1000	4	0.5	**−0.896**	12.120		**0.898**	12.121	
1000	4	1	**−1.064**	11.593		**1.066**	11.804	
1000	8	0.5	−0.393	0.109	**−0.079**	0.395	**0.118**	0.121
1000	8	1	−0.509	**0.010**	−0.097	0.510	**0.146**	0.149
1000	16	0.5	−0.185	0.023	**−0.002**	0.186	**0.034**	0.071
1000	16	1	−0.248	0.010	**0.002**	0.250	**0.038**	0.091
1000	32	0.5	−0.090	0.005	**−0.001**	0.091	**0.017**	0.064
1000	32	1	−0.121	0.006	**0.003**	0.122	**0.021**	0.075