Econometrics Journal (2003), volume **6**, pp. 217–259.

Dynamic panel estimation and homogeneity testing under cross section dependence*

PETER C. B. PHILLIPS[†] AND DONGGYU SUL[‡]

[†]*Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-8281, USA;*
University of Auckland & University of York
E-mail: peter.phillips@yale.edu

[‡]*Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand*
E-mail: d.sul@auckland.ac.nz

Received: October 2002

Summary This paper deals with cross section dependence, homogeneity restrictions and small sample bias issues in dynamic panel regressions. To address the bias problem we develop a panel approach to median unbiased estimation that takes account of cross section dependence. The estimators given here considerably reduce the effects of bias and gain precision from estimating cross section error correlation. This paper also develops an asymptotic theory for tests of coefficient homogeneity under cross section dependence, and proposes a modified Hausman test to test for the presence of homogeneous unit roots. An orthogonalization procedure, based on iterated method of moments estimation, is developed to remove cross section dependence and permit the use of conventional and meta unit root tests with panel data. Some simulations investigating the finite sample performance of the estimation and test procedures are reported.

Keywords: *Autoregression, Bias, Cross section dependence, Dynamic factors, Dynamic panel estimation, GLS estimation, Homogeneity tests, Median unbiased estimation, Modified Hausman tests, Median unbiased SUR estimation, Orthogonalization procedure, Panel unit root test.*

1. INTRODUCTION

This paper suggests some simple and practical methods for treating three important and thorny issues that arise in estimation and testing with dynamic panel models: cross section dependence, homogeneity testing and small sample bias (hereafter SSB) problems. Each of these issues is individually important in dynamic panel regression and has received attention; particularly the SSB problem on which there is a large literature. But the problems are not independent and, when they are taken together, they substantially complicate estimation and inference in dynamic panel models. The rapidly growing number of applied panel studies in growth economics, international finance and empirical labor economics in recent years accentuates the need for these issues to

*The first draft of this paper was written in August, 2000. It was presented at the Midwest Econometrics Conference, October, 2001, and the York Econometrics Conference, June 2002.

be addressed in a systematic fashion. As yet, however, there have been few attempt to address these issues at the same time and this paper is a small step in that direction, offering some new possibilities in estimation and inference. We start by noting the following implications.

First, when there is cross section dependence in panel data, commonly used econometric estimators and tests about parameters of interest generally rely on the nuisance parameters of cross section dependence. As we will show, one of the most striking effects of cross section dependence is that the pooled ordinary least squares (OLS) estimator provides little gain in precision compared with single equation OLS when cross sectional dependence occurs but is ignored in the panel regression. Another effect is that commonly used panel unit root tests are no longer asymptotically similar. These effects are easily demonstrated using a simple but intuitive parametric structure for the cross section dependence.

Second, the well known SSB problem in least squares estimation of the coefficients in dynamic models is much more serious in panel models than it is in univariate autoregressions (Nickell, 1981). In some cases the bias is so marked that the true autoregressive coefficient lies completely outside the empirical distribution of the pooled OLS estimator of the coefficient. To address this problem, this paper introduces some new panel estimation procedures that are based on the idea of median unbiased estimation (Lehmann, 1959; Andrews, 1993). This approach works well in the context of panel models with a simple dynamic structure and no additional regressors, and provides a benchmark for other procedures which have greater flexibility for application in more general models but which also need extension to allow for cross section dependence, like the bias corrected IV/GMM estimators considered in recent work by Hahn and Kuersteiner (2002).

Third, homogeneity assumptions in dynamic panel models are convenient and commonly employed to take advantage of pooling in panel regression. But these restrictions are sometimes not well supported by the data and they can produce misleading results and invalidate inference, as argued for example, by Durlauf and Quah (1999) in connection with homogeneity restrictions used in the economic growth and convergence literatures. Of particular importance in applied work is the need to take account of cross section dependence in testing homogeneity restrictions in non-stationary panels, especially in connection with panel unit root testing. This paper shows how to test for panel unit roots in the presence of cross section dependence and proposes two types of test statistic. The first type is based on median unbiased correction after eliminating cross section dependence. The second type involves the use of meta statistics which seek to avoid small sample biases rather than correct for them.

This paper gives precedence initially to the treatment of the SSB problem. This is not because this issue is more important than that of cross section dependence or homogeneity, but because the SSB problem arises irrespective of homogeneity testing or the presence of cross section dependence. Further, as is already well recognized, bias can make a huge difference in applied work, as the examples of HAC and dynamic response time estimation given in the next section illustrate.[1]

To handle the SSB problem in dynamic panel estimation and the difficulties that can arise from it, this paper proposes some panel median unbiased estimators (MUEs) that follow the

[1]The SSB problem in least squares estimation of the coefficients in an autoregression has a long history, two important early contributions being Hurvicz (1950) and Orcutt (1948). In simple autoregressions, asymptotic formulae for the SSB were worked out by Kendall (1954) and Marriott and Pope (1954). Orcutt (1948) was the first to show that fitting an intercept in an autoregression produced an additional source of bias that can exacerbate SSB, and this was confirmed in a later simulation study by Orcutt and Winokur (1969). The point was echoed in Andrews (1993), which provided further simulations that included the case of a fitted linear trend.

approach taken by Andrews (1993) in the time series case.[2] Our starting point is a panel version of the MUE of Andrews in which the innovations in the panel are assumed to be free of cross sectional dependence and the autoregressive coefficient is assumed to be homogenous across cross sectional units. Since both these assumptions are strong and are unlikely to be satisfied in empirical work, we explore the consequences of relaxing these assumptions and develop some alternate MUE procedures that are more suitable in that event.

For this purpose, we use a generalized CTE model to parameterize the structure of cross section dependence (see equation (6) below). This structure has been used in practical work (for example, Barro and Sala-i-Martin (1992)) because of its simplicity and economic interpretability. Also, other authors (e.g., Im *et al.* (1997)) have suggested this parametric structure as a possible model for cross section dependence and have indicated, without providing analysis, that such formulations can be expected to complicate asymptotics in both stationary and non-stationary cases. Under this structure, we find that pooling GLS (which takes account of the dependence) reduces variance, but the pooled GLS estimator suffers from downward bias. To deal with these effects of cross section dependence, we develop a panel generalized MUE and find that this procedure restores the precision gains from pooling in the panel and largely removes the bias in GLS. Next, we consider the more realistic case in empirical research where there is cross sectional dependence among the innovations and heterogeneity in the autoregressive coefficients. In this case, we provide a seemingly unrelated MUE that deals with heterogeneity and cross section dependence in much the same way as the conventional SUR estimator, while also addressing the SSB bias problem.

In panel applications it is often of interest to test whether the data support homogeneity restrictions on the coefficients, an important example being that of panel unit roots, as mentioned above. In view of the potential gains from pooling and the changes in the limit theory in the non-stationary case, homogeneity of the autoregressive coefficients in a panel is an important restriction in dynamic panel models. In developing tests of such restrictions in dynamic panels it is particularly important in empirical applications to allow for cross section dependence. To this end, this paper investigates the properties of Wald and Hausman-type tests of homogeneity under cross section dependence and proposes a modified Hausman test procedure that helps to deal with the effects of such dependence in testing for the presence of homogeneous unit roots. An orthogonalization procedure is developed which validates the implementation unit root tests for panel models when there is cross section dependence. The procedure involves an iterative method of moments approach to estimate the cross section dependence parameters and removes cross section dependence by means of a suitable projection. Moon and Perron (2002) have independently suggested the same approach but use principal components methods rather than iterative method of moments estimation in their implementation of the procedure.

The remainder of this paper is organized as follows. The next section shows how even a small time series SSB can make a large difference in estimation and testing in the context of panel pooling. Section 3 studies the invariance properties of the panel MUE under the assumption of cross sectional independence. Since invariance breaks down under cross sectional dependence, this section also investigates alternative invariance properties that hold in the presence of cross section dependence and proposes two new estimators for this case—a pooled feasible generalized

[2]Our work is also related to some recent independent work by Cermeno (1999). Using simulation methods, Cermeno investigates the use of MUE estimation in a dynamic panel regression with fixed effects, a common time effect (CTE) and homogeneous trends. Our framework extends Cermeno's study by developing a class of PMUEs that address a more general case of cross section dependence and that enable tests of homogeneity restrictions on the dynamics, including the important case of unit root homogeneity.

MUE and a seemingly unrelated MUE. Section 4 considers the asymptotic properties of Wald and Hausman tests for homogeneity under cross section dependence and develops some alternative procedures that offer advantages, especially in the case of unit roots. In Section 5, we report the results of a simulation experiment examining the bias and efficiency of the various panel estimators and the performance of the tests of cross section homogeneity. Section 6 concludes. Derivations and some additional technical results are given in the Appendices: A derives some invariance results; B develops limit theory for the stationary and unit root non-stationary cases; C provides an iterative algorithm for estimating the cross section dependence coefficients.

2. DYNAMIC PANEL MODELS AND BIAS ILLUSTRATIONS

2.1. Model definitions

Three basic models are considered. These are panel versions of the models given in Andrews (1993). As in that work, Gaussianity is assumed in order to construct the MUE. Each of the basic models involves a latent panel $\{y^*_{i,t} : t = 0, 1, \ldots, T; i = 1, \ldots, N\}$ that is generated over time as an AR(1) with errors that are independent across section. The more complex case of cross section error dependence is taken up in Section 3.2 and allowance for more general time series effects is considered in Section 4.3.

The model for $y^*_{i,t}$ is

$$y^*_{i,t} = \rho y^*_{i,t-1} + u_{i,t}, \qquad \text{for } t = 1, \ldots, T, \qquad \text{and } i = 1, \ldots, N, \qquad \text{where } \rho \in (-1, 1], \tag{1}$$

$u_{i,t} \sim i.i.d.\ N(0, \sigma_i^2)$ over t and $u_{i,t}$ is independent of $u_{j,s}$ for all $i \neq j$ and for all s, t and initialization is as follows:

$$y^*_{i,0} \sim \begin{cases} N\left(0, \frac{\sigma_i^2}{1-\rho^2}\right) & \rho \in (-1, 1) \\ O_p(1) & \rho = 1. \end{cases}$$

When $\rho \in (-1, 1)$, $y^*_{i,t}$ is a zero mean, Gaussian panel that follows an AR(1) structure over time and that is independent over i. When $\rho = 1$, $y^*_{i,t}$ is a Gaussian panel random walk starting from a (possibly random) initialization $y^*_{i,0}$ (not necessarily Gaussian) and that is independent over i. The observed panel data $\{y_{i,t} : t = 0, 1, \ldots T;\ i = 1, \ldots, N\}$ are defined in terms of $y^*_{i,t}$ as follows:

M1: $y_{i,t} = y^*_{i,t}$ for $\{t = 0, \ldots, T;\ i = 1, \ldots, N\}$ and $\rho \in (-1, 1)$
M2: $y_{i,t} = \mu_i + y^*_{i,t}$ for $t = 0, \ldots, T$, $i = 1, \ldots, N$, $\mu_i \in R$ and $\rho \in (-1, 1]$
M3: $y_{i,t} = \mu_i + \beta_i t + y^*_{i,t}$ for $t = 0, \ldots, T$, $i = 1, \ldots, N$, $\mu_i, \beta_i \in R$ and $\rho \in (-1, 1]$.

In each case, there is an equivalent dynamic panel representation in terms of $y_{i,t}$:

M1 $y_{i,t} = \rho y_{i,t-1} + u_{it}$ for $t = 1, \ldots, T$, $i = 1, \ldots, N$ and $\rho \in (-1, 1)$
M2 $y_{i,t} = \underline{\mu}_i + \rho y_{i,t-1} + u_{it}$ for $t = 1, \ldots, T$, $i = 1, \ldots, N$, with $\underline{\mu}_i = \mu_i(1 - \rho)$ and $\rho \in (-1, 1]$
M3 $y_{i,t} = \underline{\mu}_i + \underline{\beta}_i t + \rho y_{i,t-1} + u_{it}$ for $t = 1, \ldots, T$, $i = 1, \ldots, N$, with $\underline{\mu}_i = \mu_i(1 - \rho) + \rho\beta_i$, $\underline{\beta}_i = \beta_i(1 - \rho)$, and $\rho \in (-1, 1]$.

In M1–M3, the initialization is $y_{i,0} \sim N(0, \sigma_i^2/(1-\rho^2))$ when $\rho \in (-1, 1)$ and $y_{i,0} = O_p(1)$ when $\rho = 1$.

2.2. Pooled estimation and bias illustrations

Denote the pooled panel least squares (POLS) estimator of ρ by $\hat{\rho}_{\text{pols}}$ in each of the three models M1, M2 and M3. In M2, for instance, $\hat{\rho}_{\text{pols}}$ has the form

$$\hat{\rho}_{\text{pols}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - y_{i.-1})(y_{it} - y_{i.})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - y_{i.-1})^2},$$

$$\text{where } y_{i.} = T^{-1}\sum_{t=1}^{T} y_{it}, \text{ and } y_{i.-1} = T^{-1}\sum_{t=1}^{T} y_{it-1}. \quad (2)$$

The exact quantiles of $\hat{\rho}_{\text{pols}}$ were computed by simulation using 100,000 replications for a selection of N, T and ρ values and for $\sigma_i^2 = 1$. We report some summary statistics here (detailed results are available upon request) and make the following general observations: (i) the median values of the pooled OLS estimators are less than the true values for all models and all cases; (ii) the difference between the median value and the true value (which we call the median bias) is increasing as the true value of ρ increases for all configurations of (N, T). These observations reflect what is known about the bias from the asymptotic formulae. Nickell (1981) first obtained the asymptotic bias of $\hat{\rho}_{\text{pols}}$ under time series stationarity and cross section independence. Nickell's formula has recently been extended by Phillips and Sul (2002, 2003) to cases of cross section dependence, where the bias formula (as $N \to \infty$) is the same up to the first order in T^{-1}, and to the non-stationary $\rho = 1$ case.

Table 1 shows the bias of the POLS estimator for each model when $\rho = 0.9$. For model M1, the bias of the OLS estimator vanishes for moderate sizes of N and T. For example, the median values of $\hat{\rho}_{\text{pols}}$ are 0.88 for $N = 1$, $T = 50$, 0.89 for $N = 1$, $T = 100$ and 0.90 for $N = 10$, $T = 50$. Also, the empirical distribution of $\hat{\rho}_{\text{pols}}$ becomes tighter as N increases. In contrast to model M1, $\hat{\rho}_{\text{pols}}$ suffers from substantial SSB in model M2 even when N or T are moderately large. But, as in model M1, the distribution of $\hat{\rho}_{\text{pols}}$ concentrates quickly as N increases. In several cases, the bias and concentration of the POLS estimator are such that the true value of ρ lies almost completely outside the empirical distribution for moderate N. For example, for $T = 50$, the upper 95% points of $\hat{\rho}_{\text{pols}}$ are 0.94, 0.89, 0.88 and 0.88 for N= 1, 10, 20 and 30, respectively, when $\rho = 0.9$. Even for $T = 200$ and $N = 30$, 95% of the distribution of $\hat{\rho}_{\text{pols}}$ is below the true value. This problem becomes more severe for model M3, where the upper 95% points of $\hat{\rho}_{\text{pols}}$ are 0.904, 0.843, 0.831 and 0.825 for N = 1, 10, 20 and 30 and $T = 50$.

The bias and concentration of the pooled estimator $\hat{\rho}_{\text{pols}}$ are pertinent in applications where they influence the distribution of derived statistics such as impulse responses, cumulative impulse response functions, the half-life of a unit shock (h) and the long run variance (lrv). We provide some brief illustrations of these effects in the case of h and lrv. In the panel AR models above, the h and lrv estimates based on $\hat{\rho}_{\text{pols}}$ are $\hat{h} = \ln 0.5/\ln\hat{\rho}_{\text{pols}}$ and $\widehat{lrv} = 1/(1 - \hat{\rho}_{\text{pols}})^2$. As is apparent from Table 1(B) and 1(C), even a small SSB can have large effects on these derived functions in the panel case because of the concentration of the estimate $\hat{\rho}_{\text{pols}}$ and the non-linearity of the functions. As discussed in the last paragraph, the upper 95% point of the distribution of $\hat{\rho}_{\text{pols}}$ is smaller than ρ when N is moderately large, and then 95% of the distribution of $\hat{h}$ is less than the true half-life h. In model M3, for example, when $\rho = 0.9$, $N = 10$ and $T = 100$, 95% of the distribution of $\hat{h}$ is less than 5.518, whereas the actual half-life is $h = 6.597$. Similarly, for the same model and parameter values, 95% of the distribution of $\widehat{lrv}/lrv$ lies below 0.696. Even for $N = 30$, $T = 200$, 95% of the distribution of $\widehat{lrv}/lrv$ lies below 0.81. Table 1(C) shows how

Table 1. Downward bias in dynamic panel estimation.

Sample	Model M1			Model M2			Model M3		
	5%	50%	95%	5%	50%	95%	5%	50%	95%
Part A: quantiles of $\hat{\rho}_{\text{pols}}$ for $\rho = 0.9$									
$N = 1, T = 50$	0.710	0.883	0.962	0.628	0.830	0.937	0.548	0.772	0.904
$N = 1, T = 100$	0.787	0.891	0.948	0.749	0.868	0.935	0.713	0.842	0.920
$N = 1, T = 200$	0.829	0.896	0.938	0.814	0.885	0.931	0.798	0.874	0.924
$N = 10, T = 50$	0.858	0.898	0.928	0.799	0.850	0.889	0.735	0.795	0.843
$N = 10, T = 100$	0.874	0.899	0.920	0.847	0.877	0.902	0.820	0.853	0.882
$N = 10, T = 200$	0.882	0.900	0.915	0.870	0.890	0.906	0.858	0.879	0.897
$N = 20, T = 50$	0.872	0.899	0.921	0.816	0.850	0.880	0.755	0.796	0.831
$N = 20, T = 100$	0.882	0.900	0.915	0.857	0.878	0.896	0.830	0.854	0.874
$N = 20, T = 200$	0.888	0.900	0.911	0.876	0.890	0.902	0.864	0.878	0.892
$N = 30, T = 50$	0.878	0.900	0.917	0.824	0.851	0.875	0.763	0.796	0.825
$N = 30, T = 100$	0.885	0.900	0.913	0.861	0.878	0.893	0.835	0.853	0.870
$N = 30, T = 200$	0.890	0.900	0.909	0.879	0.890	0.900	0.868	0.879	0.890
Part B: quantiles of $\hat{h}$ when $\rho = 0.9$ and $h = 6.579$									
$N = 1, T = 50$	2.027	5.569	18.036	1.487	3.709	10.730	1.153	2.685	6.905
$N = 1, T = 100$	2.890	6.029	13.034	2.403	4.895	10.393	2.051	4.033	8.342
$N = 1, T = 200$	3.704	6.303	10.783	3.366	5.670	9.698	3.071	5.130	8.734
$N = 10, T = 50$	4.532	6.465	9.244	3.086	4.250	5.897	2.248	3.024	4.071
$N = 10, T = 100$	5.130	6.502	8.332	4.184	5.293	6.753	3.487	4.362	5.518
$N = 10, T = 200$	5.524	6.549	7.764	4.995	5.921	7.041	4.520	5.352	6.364
$N = 20, T = 50$	5.073	6.479	8.454	3.407	4.257	5.422	2.462	3.033	3.745
$N = 20, T = 100$	5.530	6.550	7.799	4.477	5.310	6.305	3.717	4.377	5.164
$N = 20, T = 200$	5.831	6.557	7.410	5.254	5.922	6.689	4.745	5.348	6.042
$N = 30, T = 50$	5.313	6.556	8.019	3.573	4.306	5.171	2.561	3.046	3.614
$N = 30, T = 100$	5.698	6.554	7.617	4.645	5.321	6.095	3.847	4.372	4.973
$N = 30, T = 200$	5.957	6.573	7.242	5.391	5.934	6.555	4.882	5.360	5.920
Part C: quantiles of $\frac{\widehat{lrv}}{lrv}$ when $\rho = 0.9$ and $lrv = 100$									
$N = 1, T = 50$	0.113	0.763	7.047	0.064	0.339	2.580	0.040	0.182	1.091
$N = 1, T = 100$	0.206	0.863	3.880	0.147	0.575	2.501	0.109	0.403	1.643
$N = 1, T = 200$	0.337	0.918	2.608	0.282	0.753	2.127	0.235	0.616	1.726
$N = 10, T = 50$	0.501	0.965	1.933	0.235	0.425	0.810	0.129	0.220	0.390
$N = 10, T = 100$	0.620	0.986	1.565	0.420	0.656	1.035	0.292	0.449	0.696
$N = 10, T = 200$	0.717	0.994	1.385	0.587	0.810	1.137	0.489	0.669	0.928
$N = 20, T = 50$	0.615	0.981	1.596	0.281	0.432	0.670	0.152	0.223	0.331
$N = 20, T = 100$	0.711	0.988	1.382	0.478	0.658	0.908	0.333	0.449	0.616
$N = 20, T = 200$	0.791	0.996	1.264	0.649	0.815	1.029	0.537	0.670	0.845
$N = 30, T = 50$	0.671	0.990	1.479	0.307	0.435	0.626	0.164	0.225	0.309

Table 1. Continued.

Sample	Model M1			Model M2			Model M3		
	5%	50%	95%	5%	50%	95%	5%	50%	95%
	Part C: quantiles of $\frac{\widehat{lrv}}{lrv}$ when $\rho = 0.9$ and $lrv = 100$								
$N = 30, T = 100$	0.759	0.993	1.302	0.510	0.663	0.866	0.352	0.453	0.587
$N = 30, T = 200$	0.824	0.993	1.201	0.678	0.814	0.986	0.557	0.670	0.810

serious the bias in $\widehat{lrv}$ can be. When $T = 50$ and $N = 1$, the median value of $\widehat{lrv}$ for model M2 is about 76% of the true lrv. For model M3, it is less than 20% of the true value when $T = 50$ and $N = 1$, and still less than 46% when $T = 100$ and $N = 30$. Thus, when estimation of the lrv is based on panel data with fitted fixed effects or individual trends, the estimated lrv suffers from serious downward bias. We can expect test statistics that rely on these lrv estimates to be correspondingly affected.

3. PANEL MEDIAN UNBIASED ESTIMATION

This section proposes three PMU estimators. As in Andrews (1993), the basic idea is that the median function $m(\rho)$ in the relation $P[\hat{\rho}_{\text{pols}} < m(\rho) \mid \rho] = \frac{1}{2}$ can be inverted to give an estimator $\hat{\rho}_{\text{pemu}} = m^{-1}(\hat{\rho}_{\text{pols}})$ for which the relation $P[\hat{\rho}_{\text{pemu}} < \rho \mid \rho] = \frac{1}{2}$, or median unbiasedness, holds. The first estimator considered is the panel exactly median unbiased (PEMU) estimator, $\hat{\rho}_{\text{pemu}}$, constructed under the assumptions of a homogenous AR(1) parameter and cross sectional independence. This estimator is simply a panel version of Andrews' exactly MUE for the time series case. Our interest is in how well this procedure works in a panel data set up and what can be done to allow for cross section dependence. As mentioned in the introduction, Cermeno (1999) has independently proposed the use of a PEMU estimator for dynamic panel models with a CTE, homogeneous trends and no cross section dependence. He shows in simulations that the approach can work well in models of this type.

The PEMU estimator is based on the assumption of cross section independence (or the presence of a CTE) which will often be too strong in practical work, particularly with macroeconomic panels. In such applications, PEMU is likely to be less relevant than our second and third estimators, which are designed to take account of cross section dependence that is more general than a CTE. We will calibrate the performance of the new MUEs against that of the conventional POLS estimator in cases where there is cross sectional dependence amongst the regression errors. This comparison will highlight the gains of working with MUEs in the panel context, especially when there is cross section dependence.

3.1. Panel exactly median unbiased estimation

As discussed in Andrews (1993), it is useful in the construction of MUEs for the distribution of the least squares estimator to be invariant to scale and other nuisance parameters. It is well known (e.g. Dickey and Fuller (1979)) that least squares estimates of the autoregressive coefficient in pure time series versions of models 1–3 satisfy such distributional invariance properties. These invariance results extend to the pooled panel forms of the least squares estimators in models 1–3 under certain conditions, which we now provide. The following property is a panel version of the property given in Andrews (1993) for the time series case. As before, the POLS estimator of ρ is generally denoted by $\hat{\rho}_{\text{pols}}$ for each of the three models M1, M2 and M3; but when there is possible ambiguity, we use an additional subscript and write $\hat{\rho}_{\text{polsj}}$ for the POLS estimator of ρ in model j.

Invariance Property IP1. *Under the assumption of cross section independence, the distribution of $\hat{\rho}_{\text{polsj}}$ depends only on ρ when model j is correct and the error variance $\sigma_i^2 = \sigma^2$ for all i. When y_{it} is stationary, it does not depend on the common variance σ_i^2 for model M1, or (σ_i^2, μ_i) for model M2, or $(\sigma_i^2, \mu_i, \beta_i)$ for model M3, nor on the value of y_{i0} when $\rho = 1$ and y_{it} is non-stationary.*

The common variance condition in IP1 is a strong one and will be inappropriate in many applications. It may be relaxed by allowing the individual error variances σ_i^2 to be $i.i.d.$ draws from a known distribution f with common scale. For example, if σ_i^2/σ^2 are $i.i.d.$ χ_1^2, then $u_{it}/\sigma = (u_{it}/\sigma_i)(\sigma_i/\sigma)$, which is independent of nuisance parameters. The numerator and denominator of $\hat{\rho}_{\text{pols}}$ may then be rescaled by $1/\sigma^2$ and it is apparent that IP1 continues to hold, as shown in the Appendix. In this case, the distribution f is assumed known, like the normal distribution of the errors, so that the median function of the POLS estimator can be constructed. For more general cases of variation in σ_i^2 over i, we may use weighted least squares in the construction of the panel estimator. This extension and other generalizations of $\hat{\rho}_{\text{pols}}$ that are better suited to empirical applications are discussed in the consideration of the GLS approach later. For the time being, we confine our discussion to the estimator $\hat{\rho}_{\text{pols}}$ and those cases where property IP1 holds.

Property IP1 enables the construction of a panel version of the exactly median unbiased estimator (PEMU) in Andrews (1993). We start by noting that $\hat{\rho}_{\text{pols}}$ has a median function $m(\rho) = m_{T,N}(\rho)$ which simulation shows to be strictly increasing in ρ on the parameter space $\rho \in (-1, 1]$.[3] Using this function (which depends on T and N), the panel median-unbiased estimator $\hat{\rho}_{\text{pemu}}$ can be defined as follows:

$$\hat{\rho}_{\text{pemu}} = \begin{cases} 1 & \text{if} \quad \hat{\rho}_{\text{pols}} > m(1), \\ m^{-1}(\hat{\rho}_{\text{pols}}) & \text{if} \quad m(-1) < \hat{\rho}_{\text{pols}} \le m(1), \\ -1 & \text{if} \quad \hat{\rho}_{\text{pols}} \le m(-1), \end{cases} \tag{3}$$

where $m(-1) = \lim_{\rho \to -1} m(\rho)$ and m^{-1} is the inverse function of $m(\cdot) = m_{T,N}(\cdot)$ so that $m^{-1}(m(\rho)) = \rho$. Furthermore, a $100(1-p)\%$ confidence interval for ρ in model j can be constructed as follows. Let $q_L(\cdot)$ and $q_U(\cdot)$ be the lower and upper quantile functions for $\hat{\rho}_{\text{pols}}$. Define

$$\hat{c}_{PU}^{L} = \begin{cases} 1 & \text{if} \quad \hat{\rho}_{\text{pols}} > q_U(1), \\ q_U^{-1}(\hat{\rho}_{\text{pols}}) & \text{if} \quad q_U(-1) < \hat{\rho}_{\text{pols}} \le q_U(1), \\ -1 & \text{if} \quad \hat{\rho}_{\text{pols}} \le q_U(-1), \end{cases} \tag{4}$$

[3] An analytic demonstration of this property would be useful but is not presently available either in the panel or the pure time series case (Andrews, 1993). The simulation evidence is strongly confirmatory at least for values $T \ge 20$ and $N \ge 5$. There seems to be some evidence from simulations that the property fails for small T when $N = 1$. Andrews (1993, fn. 4) reports that the 0.95 quantile function appears to dip slightly for values of ρ close to unity for small values of T.

$$\hat{c}_{PU}^{U} = \begin{cases} 1 & \text{if} \quad \hat{\rho}_{\text{pols}} > q_L(1), \\ q_L^{-1}(\hat{\rho}_{\text{pols}}) & \text{if} \quad q_L(-1) < \hat{\rho}_{\text{pols}} \leq q_L(1), \\ -1 & \text{if} \quad \hat{\rho}_{\text{pols}} \leq q_L(-1). \end{cases} \tag{5}$$

Then, $\hat{c}_{PU}^{U}$ and $\hat{c}_{PU}^{L}$ provide upper and lower confidence limits and the $100(1-p)\%$ confidence interval for ρ is $\{\rho : \hat{c}_{PU}^{L} \leq \rho \leq \hat{c}_{PU}^{U}\}$. This construction follows Andrews (1993). The intervals are obtained in precisely the same way as in that paper, but use tables of the quantiles of the panel estimator $\hat{\rho}_{\text{pols}}$.

3.2. Panel feasible generalized median unbiased estimator

The assumption of no cross sectional correlation among the regression residuals is a strong one and is unlikely to hold in many applications. When the structure of cross sectional dependence among the regression errors is completely unknown, it is generally infeasible to deal with the correlations because of degrees of freedom constraints. Hence, it is common to assume some simplifying form of dependence structure. The most conventional way to handle cross section dependence has been to include a common time dummy in the panel regression. The justification for the CTE is that certain co-movements of multivariate time series may be due to a common factor. For example, in cross country panels it might be argued that the time dummy represents a common international effect (e.g. a global shock or a common business cycle factor), or in a panel study of purchasing power parity it may represent the numeraire currency.

The model we use here allows for a CTE that can impact individual series differently. Specifically, the model for the regression errors has the form

$$u_{it} = \delta_i \theta_t + \varepsilon_{it}, \qquad \theta_t \sim i.i.d.\ N(0, 1) \text{ over } t, \tag{6}$$

in which θ_t is a CTE, whose variance is normalized to be unity for identification purposes and whose coefficients, δ_i, may be regarded as 'idiosyncratic share' parameters that measure the impact of the common time effect on series i. The δ_i are assumed to be non-stochastic and we let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)$. In (6) the general error component ε_{it} is assumed to satisfy

$$\varepsilon_{i,t} \sim i.i.d. N(0, \sigma_i^2) \text{ over } t, \text{ and } \varepsilon_{i,t} \text{ is independent of } \varepsilon_{j,s} \text{ and } \theta_s \text{ for all } i \neq j \text{ and for all } s, t.$$

In this formulation, the source of the cross sectional dependence is generated from the common stochastic series θ_t and the extent of the dependence is measured by the coefficients δ_i. In particular, the covariance between u_{it} and u_{jt} $(i \neq j)$ is given by

$$E(u_{it} u_{jt}) = \delta_i \delta_j. \tag{7}$$

There is no cross sectional correlation when $\delta_i = 0$ for all i, and there is identical cross sectional correlation when $\delta_i = \delta_j = \delta_0$ for all i and j. Thus, the degree of cross sectional correlation is controlled by the components of $\boldsymbol{\delta}$. Setting $\mathbf{u}_t = (u_{1t}, \ldots, u_{Nt})'$ we have the conditional covariance matrix

$$\mathbf{V}_u = E(\mathbf{u}_t \mathbf{u}_t' \mid \sigma_1^2, \ldots, \sigma_N^2) = \boldsymbol{\Sigma} + \boldsymbol{\delta\delta}', \qquad \boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2). \tag{8}$$

The model (6) can be regarded as a single factor model in which θ_t is the common factor and δ_i is the factor loading for series i. It has been used in empirical research in studying growth convergence by Barro and Sala-i-Martin (1992). More general versions of this model that allow for weakly dependent time series effects and multiple factors have been considered in recent work by Bai and Ng (2001) and Moon and Perron (2002) that concentrates on model determination issues relating to the number of factors and panel unit root testing. The models used by these authors are more complex than (6), especially with regard to time series properties. Nonetheless, (6) is general enough to allow for interesting cases of high and low cross sectional dependence and yet simple enough to enable us to develop good procedures for bias removal in dynamic panel regressions where cross section dependence arises. In the panel unit root case, we show later in the paper that time series effects in ε_{it} can be treated by a simple augmented dynamic panel regression and that time series effects in θ_t can be treated simply by projecting on the space orthogonal to $\boldsymbol{\delta}$.

With this formulation for the error variances, the numerator and denominator of $\hat{\rho}_{\text{pols}}$ may be rescaled by $1/\sigma^2$, giving some invariance characteristics to the panel estimator $\hat{\rho}_{\text{pols}}$. Stronger invariance properties apply to the panel generalized least squares estimator $\hat{\rho}_{\text{pgls}}$ defined by

$$\hat{\rho}_{\text{pgls}} = \frac{\sum_{t=1}^{T} \hat{\mathbf{y}}_{t-1}' \mathbf{V}_u^{-1} \hat{\mathbf{y}}_t}{\sum_{t=1}^{T} \hat{\mathbf{y}}_{t-1}' \mathbf{V}_u^{-1} \hat{\mathbf{y}}_{t-1}}, \tag{9}$$

where $\hat{\mathbf{y}}_t = (\hat{y}_{1t}, \ldots, \hat{y}_{Nt})'$ and where $\hat{y}_{it}$ denotes y_{it} or demeaned or detrended y_{it}, respectively for models M1, M2 and M3. In particular, we have the following property.

Invariance Property IP2. *Under cross sectional dependence of the form* (6), *the distribution of the panel GLS estimator* $\hat{\rho}_{\text{pgls}}$ *depends only on* ρ. *When* $\rho = 1$ *and* y_{it} *is non-stationary, the distribution of* $\hat{\rho}_{\text{pgls}}$ *for models* 2 *and* 3 *does not depend on the value of* y_{i0}.

Since the distribution of the panel GLS estimator depends only on ρ, we now propose an iterative procedure that involves the use of a feasible GLS estimator, $\hat{\rho}_{\text{pfgls}}$, whose form is specified in what follows in (10). Our objective is to reduce the SSB problem of the least squares procedure by constructing a feasible generalized version of the PMU estimator of ρ. It should be pointed out that, while the distribution of the panel GLS estimator $\hat{\rho}_{\text{pgls}}$ depends only on ρ, this is not necessarily true of a feasible GLS procedure. However, provided a consistent estimator of the covariance structure is employed, this property will hold asymptotically. Moreover, it is known that covariance matrix estimation generally only has a second-order effect on the distribution of feasible GLS estimates (see Phillips (1977, 1993), Rothenberg (1984)), although such results have not yet been shown for the dynamic panel model considered here. For these reasons, use of feasible GLS in the construction of a corresponding panel median unbiased procedure seems promising.

The first stage in the iteration we propose uses the residuals from a panel regression in which we use our MUE $\hat{\rho}_{\text{pemu}}$ rather than OLS to reduce the SSB problem in the primary stage. Simulations we have conducted that are reported below (see Figure 2) indicate that the use of the PMU estimator in the first stage helps to remove bias and improve estimates of the error variance matrix even in the presence of cross section dependence. The error variance matrix is estimated by an iterated method of moments procedure which is explained in Section 4.2 below. The next stage of the iteration involves the construction of a panel feasible generalized median unbiased (PFGMU) estimator that utilizes this estimated error covariance matrix. In this construction, we use the median function $m(\rho) = m_{T,N}(\rho)$ of the estimator $\hat{\rho}_{\text{pfgls}}$, which simulations show to be strictly increasing in ρ on the parameter space $\rho \in (-1, 1]$. Using this median function (which

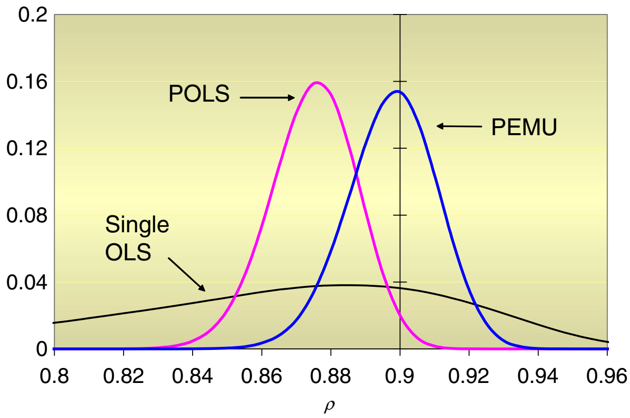

Figure 1. Empirical distributions of single equation OLS, POLS and PEMU under no cross sectional dependence ($T = 100$, $N = 20$, $\rho = 0.9$).

depends on T and N), the PFGMU estimator, $\hat{\rho}_{\text{pfgmu}}$, can be defined as in (3). The process can be continued, revising the estimate of the error covariance matrix in each iteration.

To fix ideas, the steps in the iteration are laid out as follows:

Step 1: Obtain the estimator $\hat{\rho}_{\text{pemu}}$ and using the residuals from this regression construct the error covariance matrix estimate $\hat{\mathbf{V}}_{\text{pemu}}$ by the method explained in Section 4.2.

Step 2: Using $\hat{\mathbf{V}}_{\text{pemu}}$, perform panel generalized least squares as in (9) and obtain the PFGLS estimate of ρ defined by

$$\hat{\rho}_{\text{pfgls}} = \frac{\sum_{t=1}^{T} \hat{\mathbf{y}}_{t-1}' \hat{\mathbf{V}}_{\text{pemu}}^{-1} \hat{\mathbf{y}}_t}{\sum_{t=1}^{T} \hat{\mathbf{y}}_{t-1}' \hat{\mathbf{V}}_{\text{pemu}}^{-1} \hat{\mathbf{y}}_{t-1}}. \tag{10}$$

Step 3: The PFGMU estimator is now calculated as $\hat{\rho}_{\text{pfgmu}} = m^{-1}(\hat{\rho}_{\text{pfgls}})$ just as in (3) but using the median function $m(\rho) = m_{T,N}(\rho)$ of the estimator $\hat{\rho}_{\text{pfgls}}$.

Step 4: Repeat steps 1–3 (using updated estimates of ρ in the first stage rather than $\hat{\rho}_{\text{pemu}}$) until $\hat{\rho}_{\text{pfgmu}}$ converges.

Figure 1 displays a kernel estimate of the distribution of POLS based on 100,000 replications with $N = 20$, $T = 100$, $\rho = 0.9$ when there is no cross sectional dependence. Apparently, the POLS estimator $\hat{\rho}_{\text{pols}}$ is more concentrated than single equation OLS (which does not use the additional cross section data) but is badly biased downwards. The bias is sufficiently serious that almost the entire distribution of $\hat{\rho}_{\text{pols}}$ lies below the true value of ρ.

Figure 2 shows the distributions of the POLS and PMU estimators for the same parameter configuration as Figure 1 and based on the same number of replications, but with high cross sectional correlation.[4] As shown in Phillips and Sul (2002), the POLS bias in the case of cross section dependence is the same to first order as the bias in the cross section independent case, and this bias equivalence between the two cases is born out by the simulation results. As is apparent

[4]When $\delta_i \in (1, 4)$ in (6), the average cross sectional correlation is around 0.82.

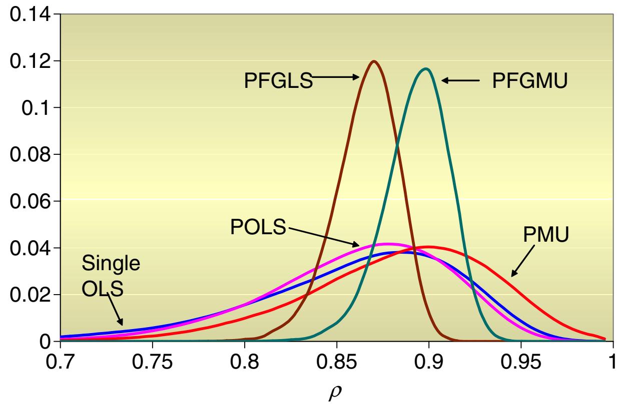

Figure 2. Empirical distributions of POLS, PFGLS and PFGMU under high cross section dependence ($T = 100$, $N = 20$, $\rho = 0.9$).

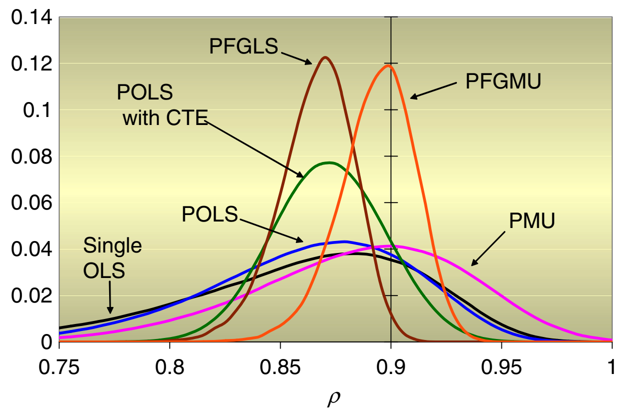

Figure 3. Same as in Figure 2 with the addition of POLS with a CTE under high cross section dependence ($T = 100$, $N = 20$, $\rho = 0.9$).

from Figure 2, the main effect of the cross sectional dependence is to increase the variation of both the POLS and PMU estimators. In fact, in the displayed case (where the average cross section correlation is around 0.82) the POLS and PMU estimators show only a slight gain in concentration over single equation OLS. In other words, if there is high cross sectional correlation, there is not much efficiency gain from pooling in the POLS estimator. Figure 3 shows the distribution of the POLS estimator in which a CTE has been estimated. While this estimator is obviously inappropriate under the general form of cross section dependence considered in (6), it is a commonly used procedure in practice and is applicable when the elements of δ all take on a common value. As is apparent from Figure 3, this estimator successfully reduces variance even though the presence of a CTE in estimation provides only a crude approximation to the error structure (6).

Figures 2 and 3 show that the PMU estimator is still quite effective in removing the bias of POLS even under cross section dependence. However, its high variance makes it a less appealing

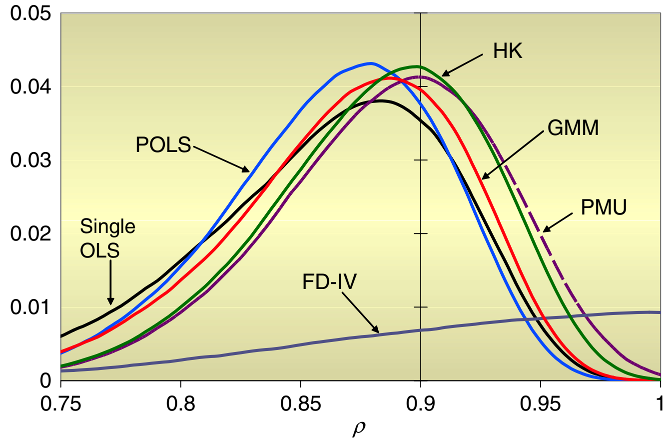

Figure 4. Extended comparison of PMU with common panel IV estimators under high cross section dependence ($T = 100$, $N = 20$, $\rho = 0.9$).

estimator for applications than our PFGMU estimator, which reduces variance and removes bias, as we now discuss. Figures 2 and 3 show the distributions of both the feasible GLS procedures, PFGLS and PFGMU. Evidently, the PFGLS estimator $\hat{\rho}_{\text{pfgls}}$ does restore much of the original gains from pooling in terms of variance reduction that were apparent in Figure 1 for $\hat{\rho}_{\text{pols}}$. But, as is also apparent from Figure 2, the distribution of $\hat{\rho}_{\text{pfgls}}$ is seriously downward biased. Use of the PFGMU median unbiased procedure corrects for this bias while retaining the concentration gains of the GLS estimator. In particular, the distribution of $\hat{\rho}_{\text{pfgmu}}$ is well centered about the true value and has concentration close to that of the MUE $\hat{\rho}_{\text{pemu}}$ under cross sectional independence (Figure 1).

Figure 4 shows some comparisons of POLS and PMU in the cross section dependent case against some alternative procedures that have been suggested for dynamic panel regression. The first of these is the crude first difference instrumental variable (FD-IV) estimator which uses y_{it-2} as an instrument in a first differenced form of the model. Apparently, FD-IV has variation substantially in excess of all the other estimators. The commonly used GMM estimator which uses the full set of instruments $\{y_{is} : s = 0, 1, \ldots, t-2\}$ shows downward bias, although not as severely as POLS, and it seems to have comparable variance. HK is the bias corrected GMM estimator suggested in Hahn and Kuersteiner (2002) and Hahn *et al.* (2001) and this estimator apparently has performance closest to that of the PMU estimator. All these procedures clearly show inferior performance to the $\hat{\rho}_{\text{pfgmu}}$ estimator under high cross section dependence.

3.3. Seemingly unrelated median unbiased estimation

The results above indicate that, if we are to gain from panel estimation by pooling cross section and time series information when there is cross section dependence, we need to take account of the dependence in estimation. In contrast, most empirical studies that utilize dynamic panels in the international finance and the macroeconomic growth literatures tend to ignore issues of cross sectional dependence when pooling. Our results indicate that there is information in cross sectional correlation that is valuable in pooled estimation and that it can be accounted for, at least in situations where the cross section sample size N is not too large. Moreover, one can utilize this information and at the same time deal with SSB bias problems in dynamic panel estimation.

Notwithstanding these potential advantages of pooling dependent data and adjusting for bias in dynamic panels, perhaps the most important issue in pooled regressions relates to the justification of the homogeneity restriction on the autoregressive coefficient ρ. In the absence of this restriction, it might be thought that there would be little gain from pooling time series and cross section data. However, because of cross section dependence, there are advantages to pooling panel data even in the estimation of heterogeneous coefficients. The reasoning is the same as that of a conventional seemingly unrelated regression (SUR) system. But in a dynamic panel context there are still SSB bias problems that need attention. This section shows that these can be addressed using a SUR version of the panel median unbiased procedure.

An additional advantage to performing heterogenous coefficient estimation is that it facilitates testing of the homogeneity restriction. Therefore, this section also proposes a test for homogeneity that is based on the seemingly unrelated panel median-unbiased (SUR-MU) estimator.

We start the discussion by combining models M1, M2 and M3 with the following heterogenous autoregressive panel model for the latent panel variable y_{it}^* :

$$y_{it}^* = \rho_i y_{it-1}^* + u_{it}, \qquad \text{for } t = 1, \ldots, T, \qquad \text{and } i = 1, \ldots, N, \tag{11}$$

in which the regression errors

$$\mathbf{u}_t \sim i.i.d.\ N(\mathbf{0}, \mathbf{V}_u), \qquad \text{for } t = 1, \ldots, T, \tag{12}$$

where $\mathbf{u}_t = (u_{it}, \ldots, u_{Nt})'$. This formulation allows for a general form of cross section error correlation as well as the more specific set up (6). The same range of ρ values as before is permitted for each of the models.

When $|\rho_i| < 1$ for all i, the cross section error correlations are higher than the cross section correlations among the regressors y_{it-1}. To see this, note that the correlation between y_{it} and y_{jt} is given by

$$\gamma_{i,j}^y = \frac{E(y_{it}y_{jt})}{\{E(y_{it}^2)E(y_{jt}^2)\}^{\frac{1}{2}}} = \gamma_{ij}\frac{\sqrt{1-\rho_i{}^2}\sqrt{1-\rho_j{}^2}}{1-\rho_i\rho_j} < \gamma_{ij}, \tag{13}$$

where $\gamma_{ij} = E(u_{it}u_{jt})/\{E(u_{it}^2)E(u_{jt}^2)\}^{\frac{1}{2}}$. We might therefore anticipate the potential gains from SUR estimation to be substantial—the regressors are different and less correlated across individual equations in the panel for which the errors are more correlated. In consequence, we propose a SUR-MU estimator based on the following iteration.

Step 1: Obtain the time series panel median unbiased estimates $\hat{\rho}_{\text{iemu}}$ for each series $i = 1, \ldots, N$ (and the appropriate model) and use the regression residuals to construct the error covariance matrix estimate $\hat{\mathbf{V}}_{EMU}$ as explained in Section 4.2.

Step 2: Using $\hat{\mathbf{V}}_{EMU}$ perform a conventional SUR on the panel and obtain the SUR estimates of the ρ_i, $\hat{\rho}_{\text{isur}}$.

Step 3: The panel SUR-MU estimator is now calculated as $\hat{\rho}_{\text{isurmu}} = m^{-1}(\hat{\rho}_{\text{isur}})$ just as in (3) but using the median function $m(\rho) = m_{T,N}(\rho)$ of the estimator $\hat{\rho}_{\text{isur}}$ for each i.

Step 4: Repeat steps 1–3 until $\hat{\rho}_{\text{isurmu}}$ converges.

The limit theory for this estimator and some associated tests of homogeneity are derived in Appendix B and are discussed in the following section. Finite sample performance is considered in Section 5.

4. TESTING HOMOGENEITY RESTRICTIONS

Using unrestricted estimates of the coefficients ρ_i in the heterogeneous dynamic panel model (11), Wald tests can be constructed to test the homogeneity restriction $H_0 : \rho_i = \rho$ for all i. It is well known that in finite samples, Wald tests suffer from size distortion that is sometimes serious even in simple univariate regressions. For the panel regression case here we have found that the size distortion of Wald tests becomes even more serious as the cross section sample size N increases. This section first investigates the asymptotic properties of Wald tests based on the SUR approach in both the stationary and non-stationary cases and shows how cross section dependencies affect the asymptotic theory under non-stationarity. We then propose an alternative Hausman-type procedure for testing homogeneity that utilizes the structure of the cross section dependence. Again this approach is affected by nuisance parameters in the non-stationary case. To address these difficulties with conventional approaches, we propose an orthogonalization process that enables panel unit root testing under cross section dependence. These issues are considered sequentially in the following sections. Derivations are given in Appendix B.

4.1. The Wald test and its asymptotic properties

The stationary case. Using the unrestricted estimates $\hat{\rho}_{\text{isurmu}}$ of the coefficients ρ_i in the heterogeneous dynamic panel model (11), Wald tests can be constructed to test the homogeneity restriction $H_0 : \rho_i = \rho$ for all i. More specifically, let $\underline{\hat{\boldsymbol{\rho}}}_{\text{surmu}} = (\hat{\rho}_{\text{isurmu}})$ be the SUR-MU estimate of the vector $\underline{\boldsymbol{\rho}} = (\rho_1, \ldots, \rho_N)'$ and write the restrictions in H_0 as $\mathbf{D}\underline{\boldsymbol{\rho}} = 0$ where $\mathbf{D} = [\boldsymbol{\iota}_{N-1}, -\mathbf{I}_{N-1}]$ and $\boldsymbol{\iota}_A$ has A unit elements. Under Gaussianity and in the stationary case where $|\rho_i| < 1$ for all i, the SUR-MU estimator $\underline{\hat{\boldsymbol{\rho}}}_{\text{surmu}}$ is asymptotically ($T \to \infty$, N fixed) equivalent to the unconstrained maximum likelihood estimate of $\underline{\rho}$.[5] In that case, standard stationary asymptotics and some algebraic manipulations (outlined in Appendix B) lead to the limit theory

$$\sqrt{T}\left(\underline{\hat{\boldsymbol{\rho}}}_{\text{surmu}} - \underline{\boldsymbol{\rho}}\right) \to_d N(\mathbf{0}, \mathbf{V}_{\text{SUR}}), \tag{14}$$

where

$$\mathbf{V}_{\text{SUR}}^{-1} = [(v_u^{ij} E(y_{it} y_{jt}))_{ij}] = \mathbf{V}_u^{-1} \odot E(\mathbf{y}_t \mathbf{y}_t'). \tag{15}$$

In (15) the operator $\odot$ is the Hadamard product, v_u^{ij} is the ijth element of $\mathbf{V}_u^{-1}$, where $\mathbf{V}_u = E(\mathbf{u}_t \mathbf{u}_t') = \boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}'$ as in (8), and

$$E(y_{it} y_{jt}) = \begin{cases} \frac{\delta_i \delta_j}{1-\rho_i \rho_j} & i \neq j \\ \frac{\sigma_i^2 + \delta_i^2}{1-\rho_i^2} & i = j, \end{cases}$$

so that

$$E(\mathbf{y}_t \mathbf{y}_t') = (\boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}') \odot \mathbf{R}, \qquad \text{where } \mathbf{R} = (r_{ij}) \qquad \text{and } r_{ij} = \frac{1}{1 - \rho_i \rho_j}. \tag{16}$$

[5]Note that the median function $m(\cdot)$ is asymptotically ($T \to \infty$, N fixed) the identity function and the SUR estimator of $\underline{\boldsymbol{\rho}}$ is the vector of Gaussian maximum likelihood estimators of the autoregressive coefficients in the unconstrained models.

From (15) and (16) it is apparent that the covariance matrix V_{SUR} depends on both $\underline{\rho}$ and δ as well as $\mathbf{\Sigma}$. When H_0 holds, $E(\mathbf{y}_t\mathbf{y}_t') = (\mathbf{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}')/(1-\rho^2)$ and $\mathbf{V}_{\mathrm{SUR}}$ has a simpler form in which

$$\mathbf{V}_{\mathrm{SUR}}^{-1} = \frac{1}{1-\rho^2}\mathbf{V}_u^{-1} \odot \mathbf{V}_u, \tag{17}$$

which depends on the common ρ and again on the cross section dependence parameter $\boldsymbol{\delta}$.

The Wald statistic for testing H_0 is

$$W_{\mathrm{surmu}} = \underline{\hat{\boldsymbol{\rho}}}_{\mathrm{surmu}}'\mathbf{D}'[\mathbf{D}\hat{\mathbf{V}}_{SURMU}\mathbf{D}']^{-1}\mathbf{D}\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{surmu}},$$

where

$$\hat{\mathbf{V}}_{SURMU} = \left[\sum_{t=1}^{T}\mathbf{Z}_t'\hat{\mathbf{V}}_u^{-1}\mathbf{Z}_t\right]^{-1},$$

in which $\mathbf{Z}_t = \mathrm{diag}(y_{1t-1}, \ldots, y_{Nt-1})$ and $\hat{\mathbf{V}}_u$ is an estimate of the error covariance matrix V_u computed from the SUR-MU regression residuals. Under H_0 and in the stationary case, the traditional chi-squared limit theory holds for W_{surmu}, i.e. $W_{\mathrm{surmu}} \rightarrow_d \chi_N^2$.

The unit root case. In the non-stationary $\rho = 1$ case, the asymptotic results depend, as might be expected, on whether M1, M2 or M3 is employed in estimation and also on the boundary condition that arises in the transition from the SUR estimator to SUR-MU—cf. (3). In addition, the asymptotic theory for the SUR estimator is more complex than that of a traditional unit root model when there is cross section dependence. For instance, when model M1 is used and the null hypothesis $H_0 : \rho_i = 1\ \forall i$ holds, derivations (outlined in Appendix B) using standard unit root limit theory deliver the limit distribution of the SUR estimator $\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}}$. This estimator is defined as

$$\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}} = \left(\sum_{t=1}^{T}\mathbf{Z}_t'\hat{\mathbf{V}}_u^{-1}\mathbf{Z}_t\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{Z}_t'\hat{\mathbf{V}}_u^{-1}\mathbf{y}_t\right),$$

where $\hat{\mathbf{V}}_u$ is an estimate of $\mathbf{V}_u$ based on residuals from a first stage regression. Appendix B gives the following asymptotic distribution for $\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}}$

$$T\left(\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}} - \boldsymbol{\iota}_N\right) \xrightarrow{d} \left[\mathbf{V}_u^{-1} \odot \int_0^1 \mathbf{B}\mathbf{B}'\right]^{-1}\left[\int_0^1 \mathbf{B} \odot (\mathbf{V}_u^{-1}d\mathbf{B})\right] = \boldsymbol{\xi}, \tag{18}$$

where $\mathbf{B}$ is vector Brownian motion with covariance matrix $\mathbf{V}_u$. It is clear from (18) that the limit distribution of $T(\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{SUR}} - \boldsymbol{\iota}_N)$ depends on the cross section dependence parameter $\boldsymbol{\delta}$ even in the homogeneous case where $\rho_i = 1\ \forall i$. Correspondingly, the asymptotic distribution of $\hat{\rho}_{\mathrm{surmu}}$ in the unit root case also depends on cross section dependence and error variance nuisance parameters. The Wald statistic, W_{sur}, for testing H_0 is given by

$$\begin{aligned} W_{\mathrm{sur}} &= \underline{\hat{\boldsymbol{\rho}}}_{\mathrm{SUR}}'\mathbf{D}'[\mathbf{D}\hat{\mathbf{V}}_{\mathrm{SUR}}\mathbf{D}']^{-1}\mathbf{D}\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{SUR}} \\ &\xrightarrow{d} \boldsymbol{\xi}'\mathbf{D}'\left[\mathbf{D}\left(\mathbf{V}_u^{-1} \odot \int_0^1 \mathbf{B}\mathbf{B}'\right)^{-1}\mathbf{D}'\right]^{-1}\boldsymbol{D}\boldsymbol{\xi}, \end{aligned} \tag{19}$$

where $\hat{\mathbf{V}}_{\mathrm{SUR}} = (\sum_{t=1}^{T} \mathbf{Z}_t' \hat{\mathbf{V}}_u^{-1} \mathbf{Z}_t)^{-1}$, and again the limit distribution (19) depends on nuisance parameters.

In contrast, in the unit root case where homogeneity of ρ across i is imposed, the pooled GLS estimator of ρ is

$$\hat{\rho} = \left(\sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{V}_u^{-1} \mathbf{y}_{t-1} \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{V}_u^{-1} \mathbf{y}_t \right),$$

with a corresponding feasible SUR version. By straightforward derivation detailed in Appendix B, we find that

$$T(\hat{\rho} - 1) \xrightarrow{d} \frac{\int_0^1 \mathbf{W}' d\mathbf{W}}{\int_0^1 \mathbf{W}'\mathbf{W}} = \frac{\sum_{i=1}^{N} \int_0^1 W_i \, dW_i}{\sum_{i=1}^{N} \int_0^1 W_i^2}, \tag{20}$$

where $\mathbf{W} = (W_i)$ is standard Brownian motion with covariance matrix $\mathbf{I}_N$. The limit (20) here depends only on the cross section sample size N.

4.2. Hausman and modified Hausman tests under cross section dependence

The stationary panel case: $H_0 : \rho_i = \rho$. The main problem with the conventional Wald test, as mentioned earlier, is that size distortion can be serious and it typically increases with the number of restrictions. Also, the Wald test based on SUR or SUR-MU estimation requires $N < T$, and is heavily influenced by the nuisance parameters of cross section correlation. This section proposes an alternative procedure for dealing with cross section dependence that takes into account the structure of the dependence.

Start by writing the model M1 (with suitable adjustments for models M2 and M3) in vector form as

$$\mathbf{y}_t = \mathbf{Z}_t \underline{\boldsymbol{\rho}} + \mathbf{u}_t, \qquad \mathbf{Z}_t = \mathrm{diag}(y_{1t-1}, \ldots, y_{Nt-1}), \qquad \underline{\boldsymbol{\rho}} = (\rho_1, \ldots, \rho_N)'. \tag{21}$$

Let $\hat{\rho}_i$ (resp. $\hat{\underline{\boldsymbol{\rho}}}$) be the OLS estimate of ρ_i ($\underline{\boldsymbol{\rho}}$) Then

$$\hat{\underline{\boldsymbol{\rho}}} = \left(\sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{Z}_t \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{y}_t \right).$$

Let $\hat{\underline{\boldsymbol{\rho}}}_{\mathrm{emu}}$ be the corresponding vector of median unbiased estimates of ρ_i. Under the null hypothesis of homogenous autoregressive coefficients $\rho_i = \rho \ \forall i$, and as $T \to \infty$, we have $\sqrt{T}(\hat{\rho}_i - \rho) \to_d N(0, 1 - \rho^2)$ for models M1, M2 and M3, with the same result for the MUEs $\hat{\rho}_{\mathrm{iemu}}$. Under cross section independence and as $T \to \infty$ for finite N, we have

$$\sum_{i=1}^{N} \frac{\sqrt{T}(\hat{\rho}_i - \rho)}{\sqrt{1 - \rho^2}} \to_d N(0, N).$$

On the other hand, if there is cross section dependence of the form implied by (6), then in the stationary case for model M1 we have

$$y_{it} = \sum_{j=0}^{\infty} \rho^j \left(\delta_i \theta_{t-j} + \varepsilon_{it-j} \right) = \delta_i \sum_{j=0}^{\infty} \rho^j \theta_{t-j} + \sum_{j=0}^{\infty} \rho^j \varepsilon_{it-j} = \delta_i \mu_t + \eta_{it}, \qquad \text{say.}$$

It follows that the asymptotic covariance between $\hat{\rho}_i$ and $\hat{\rho}_j$ is given by

$$\text{acov}\,(\hat{\rho}_i, \hat{\rho}_j) = \frac{1}{T}\frac{(\delta_i\delta_j)^2(1-\rho^2)}{(\delta_i^2+\sigma_i^2)(\delta_j^2+\sigma_j^2)} = \frac{1}{T}\frac{v_{ij}^2}{v_{ii}v_{jj}}(1-\rho^2),$$

where v_{ij} is the ijth element of $\mathbf{V}_u = \boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}'$. Setting $\underline{\hat{\boldsymbol{\rho}}} = (\hat{\rho}_1, \ldots, \hat{\rho}_N)'$ and letting i_N be an N-vector with unit elements, we find that standard derivations lead to the following limit theory:

$$\begin{aligned}\sqrt{T}(\underline{\hat{\boldsymbol{\rho}}} - \rho\boldsymbol{\iota}_N) &= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{Z}_t\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{u}_t\right)\\ &\rightarrow_d N\left(\mathbf{0}, \mathbf{D}_y^{-1}[\mathbf{V}_u \odot E(\mathbf{y}_t\mathbf{y}_t')]\mathbf{D}_y^{-1}\right)\\ &= N(\mathbf{0}, (1-\rho^2)\mathbf{R}_V \odot \mathbf{R}_V), \end{aligned} \tag{22}$$

where $\mathbf{D}_y = \text{diag}(E(y_{1t}^2), \ldots, E(y_{Nt}^2))$ and the matrix $\mathbf{R}_V$ has ijth element $v_{ij}/\{v_{ii}v_{jj}\}^{1/2}$. It follows that

$$\sum_{i=1}^{N}\frac{\sqrt{T}(\hat{\rho}_i - \rho)}{\sqrt{1-\rho^2}} \rightarrow_d N(0, \boldsymbol{\iota}_N'(\mathbf{R}_V \odot \mathbf{R}_V)\boldsymbol{\iota}_N).$$

The same result applies when the median unbiased estimates $\hat{\rho}_{\text{iemu}}$ are used in place of $\hat{\rho}_i$.

We propose to construct an estimate of the matrix $\mathbf{R}_V$ that appears in the asymptotic covariance matrix of (22) and use this estimate to develop an alternative test of H_0. The following moment based procedure may be used.[6]

Moment based estimation of $(\boldsymbol{\delta}, \boldsymbol{\Sigma})$.

Step 1: Estimate the ρ_i by using OLS or EMU and obtain the regression residuals $\hat{u}_{it} = y_{it} - \hat{\rho}_i y_{it-1}$, which are asymptotically equivalent to OLS residuals and consistent (as $T \to \infty$, N fixed) for u_{it}. In particular,

$$\hat{u}_{it} = u_{it} + (\rho_i - \hat{\rho}_i)y_{it-1} = u_{it} + o_p(1)$$

in both stationary and non-stationary cases.

Step 2: Construct the moment matrix of residuals $\mathbf{M}_T = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t'$, which is a consistent (as $T \to \infty$, N fixed) estimate of $\mathbf{V}_u$. Let m_{Tij} be the ijth element of $\mathbf{M}_T$.

Step 3: Estimate the cross section coefficients $\boldsymbol{\delta}$ and the diagonal elements of $\boldsymbol{\Sigma}$ using the following moment procedure that finds the least squares best fit to the matrix $\mathbf{M}_T$, that is

$$(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Sigma}}) = \arg\min_{\boldsymbol{\delta},\boldsymbol{\Sigma}} \text{tr}[(\mathbf{M}_T - \boldsymbol{\Sigma} - \boldsymbol{\delta}\boldsymbol{\delta}')(\mathbf{M}_T - \boldsymbol{\Sigma} - \boldsymbol{\delta}\boldsymbol{\delta}')']. \tag{23}$$

The solution of (23) satisfies the system of equations

$$\hat{\boldsymbol{\delta}} = (\mathbf{M}_T\hat{\boldsymbol{\delta}} - \boldsymbol{\Sigma}\hat{\boldsymbol{\delta}})/\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\delta}}, \qquad \hat{\sigma}_i^2 = M_{Tii} - \hat{\delta}_i^2, i = 1, \ldots, N$$

[6] Appendix C gives an explicit algorithm for Gaussian maximum likelihood estimation of the cross section coefficients. Simulation results indicate that the moment based method described here gave superior results, especially for large N.

and this can be solved using the iteration

$$\boldsymbol{\delta}^{(r)} = (\mathbf{M}_T \boldsymbol{\delta}^{(r-1)} - \boldsymbol{\Sigma}\boldsymbol{\delta}^{(r-1)})/\boldsymbol{\delta}^{(r-1)\prime}\boldsymbol{\delta}^{(r-1)},$$
$$\sigma_i^{(r)2} = M_{Tii} - \delta_i^{(r)2}, \tag{24}$$

starting from some initialization $\delta^{(0)}$ (such as the largest eigenvector of M_T) until convergence. Since $\mathbf{M}_T \to_p \mathbf{V}_u = \boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}'$ as $T \to \infty$, it follows that $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Sigma}}) \to_p (\boldsymbol{\delta}, \boldsymbol{\Sigma})$ as $T \to \infty$, with N fixed. Since $\hat{\boldsymbol{\Sigma}} \to_p \boldsymbol{\Sigma} > 0$ as $T \to \infty$, $\hat{\boldsymbol{\Sigma}}$ will be positive definite for large enough T.

Step 4: Construct the variance matrix estimate $\hat{\mathbf{V}}_u = \hat{\boldsymbol{\Sigma}} + \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}'$. Let $\hat{v}_{ij}$ be the ijth element of $\hat{\mathbf{V}}_u$ and construct the estimate $\hat{\mathbf{R}}_V$ whose ijth element is $\hat{v}_{ij}/\{\hat{v}_{ii}\hat{v}_{jj}\}^{1/2}$.

Since $\hat{\mathbf{V}}_u \to_p \mathbf{V}_u$, we have $\hat{\mathbf{R}}_V \to_p \mathbf{R}_V$ as $T \to \infty$. Now let $\tilde{\rho}$ be the PFMGU estimate of ρ under the assumption of homogeneity. Under H_0, the pooled estimate $\tilde{\rho}$ is asymptotically equivalent to GLS and then by standard limit theory

$$\sqrt{T}(\tilde{\rho} - \rho) = \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{y}'_{t-1}\mathbf{V}_u^{-1}\mathbf{y}_{t-1}\right)^{-1} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{y}'_{t-1}\mathbf{V}_u^{-1}\mathbf{u}_t\right)$$
$$\to_d N(\mathbf{0}, \{\text{trace}[\mathbf{V}_u^{-1}E(\mathbf{y}_t\mathbf{y}'_t)]\}^{-1}).$$

Since

$$E(\mathbf{y}_t\mathbf{y}'_t) = (\boldsymbol{\Sigma} + \boldsymbol{\sigma}^2\boldsymbol{\delta}\boldsymbol{\delta}') \odot \mathbf{R} = \mathbf{V}_u \odot \mathbf{R} = \frac{1}{1-\rho^2}\mathbf{V}_u,$$

under H_0, we end up with the simple result

$$\sqrt{T}(\tilde{\rho} - \rho) \to_d N\left(0, \frac{1-\rho^2}{N}\right).$$

Next consider the asymptotic covariance

$$\text{acov}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}'_t\mathbf{u}_t, \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{y}'_{t-1}\mathbf{V}_u^{-1}\mathbf{u}_t\right)$$
$$= \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}'_t E(\mathbf{u}_t\mathbf{u}'_t)\mathbf{V}_u^{-1}\mathbf{y}_{t-1} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}'_t\mathbf{y}_{t-1} \to \begin{bmatrix} E(y_{1t}^2) \\ E(y_{2t}^2) \\ \vdots \\ E(y_{Nt}^2) \end{bmatrix} = \mathbf{D}_y\boldsymbol{\iota}_N,$$

as $T \to \infty$, from which we deduce that

$$\text{acov}\ (\sqrt{T}(\underline{\hat{\boldsymbol{\rho}}} - \rho\boldsymbol{\iota}_N), \sqrt{T}(\tilde{\rho} - \rho))$$
$$= \mathbf{D}_y^{-1}[\mathbf{D}_y\boldsymbol{\iota}_N]\{\text{trace}\ [\mathbf{V}_u^{-1}E(\mathbf{y}_t\mathbf{y}'_t)]\}^{-1}$$
$$= \boldsymbol{\iota}_N(1-\rho^2). \tag{25}$$

Our test statistic for H_0 is based on the difference between the estimates

$$\sqrt{T}(\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}} - \tilde{\rho}\boldsymbol{\iota}_N) = \sqrt{T}(\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}} - \rho\boldsymbol{\iota}_N) - \sqrt{T}(\tilde{\rho} - \rho)\boldsymbol{\iota}_N,$$

and from (22), (25) and joint convergence we find that

$$\frac{\sqrt{T}(\hat{\underline{\boldsymbol{\rho}}}_{\text{emu}} - \tilde{\rho}\boldsymbol{\iota}_N)}{\sqrt{1-\tilde{\rho}^2}} = \frac{\sqrt{T}(\hat{\underline{\boldsymbol{\rho}}}_{\text{emu}} - \rho\boldsymbol{\iota}_N)}{\sqrt{1-\tilde{\rho}^2}} - \frac{\sqrt{T}(\tilde{\rho} - \rho)}{\sqrt{1-\tilde{\rho}^2}}\boldsymbol{\iota}_N \to_d N\left(\mathbf{0}, \mathbf{R}_V \odot \mathbf{R}_V - \frac{1}{N}\boldsymbol{\iota}_N\boldsymbol{\iota}_N'\right). \tag{26}$$

It follows that we may construct the Hausman-type test statistic

$$G = \frac{T}{1-\tilde{\rho}^2}(\hat{\underline{\boldsymbol{\rho}}}_{\text{emu}} - \tilde{\rho}\boldsymbol{\iota}_N)'\left\{[\hat{\mathbf{R}}_V \odot \hat{\mathbf{R}}_V]^{-1} - \frac{1}{N}\boldsymbol{\iota}_N\boldsymbol{\iota}_N'\right\}(\hat{\underline{\boldsymbol{\rho}}}_{\text{emu}} - \tilde{\rho}\boldsymbol{\iota}_N), \tag{27}$$

which is based on the difference between the robust-to-heterogeneity estimate $\hat{\underline{\boldsymbol{\rho}}}_{\text{emu}}$ of $\underline{\boldsymbol{\rho}}$ and the efficient estimate $\tilde{\rho}$ of ρ under the null, and which uses the moment based procedure outlined above to construct estimates of $\mathbf{V}_u$ and $\mathbf{R}_V$. We use the notation G_{pfmgu} to indicate that the pooled estimate $\tilde{\rho}$ in (27) is the PFMGU estimate of the (common) ρ. Then, in view of (26) and the consistency of $\hat{\mathbf{R}}_V$, we have

$$G_{\text{pfmgu}} \to \chi_N^2, \qquad \text{as } T \to \infty. \tag{28}$$

One practical difficulty that can arise with (27) is that the variance matrix $[\hat{\mathbf{R}}_V \odot \hat{\mathbf{R}}_V]^{-1} - \frac{1}{N}\boldsymbol{\iota}_N\boldsymbol{\iota}_N'$ is not necessarily positive definite and, in our simulations negative values of G have occasionally occurred when N and T are small ($N = 10$, $T = 50$).

The panel unit root case ($H_o : \rho_i = 1, \forall i$) and orthogonalization. As shown in Appendix B, the Hausman test has a limit distribution in the unit root ($\rho_i = 1, \forall i$) case that is dependent on the cross section nuisance parameters. It is therefore unsuitable for testing homogeneity. However, there is a simple way of constructing a modified test that is free of nuisance parameters, which we now describe.

Under the null hypothesis, we have as in (B.6)

$$\frac{1}{\sqrt{T}}\mathbf{y}_{[\text{Tr}]} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[\text{Tr}]}\mathbf{u}_t \to_d \mathbf{B}(r) = \mathbf{BM}(\mathbf{V}_u). \tag{29}$$

Note that we can decompose B into component Brownian motions as follows:

$$\mathbf{B}(r) = \delta\mathbf{B}_\theta(r) + \mathbf{B}_\varepsilon(r), \tag{30}$$

where

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[\text{Tr}]}\theta_t \to_d B_\theta(r) = BM(\sigma^2), \qquad \text{and } \frac{1}{\sqrt{T}}\sum_{t=1}^{[\text{Tr}]}\varepsilon_t \to_d \mathbf{B}_\varepsilon(r) = \mathbf{BM}(\boldsymbol{\Sigma}).$$

Let $\boldsymbol{\delta}_\perp$ be an $N \times (N-1)$ matrix that spans the orthogonal complement of the vector $\boldsymbol{\delta}$. Then

$$[(\boldsymbol{\delta}_\perp'\boldsymbol{\Sigma}\boldsymbol{\delta}_\perp)^{-1/2}\boldsymbol{\delta}_\perp']\frac{1}{\sqrt{T}}\mathbf{y}_{[\text{Tr}]} \to_d (\boldsymbol{\delta}_\perp'\boldsymbol{\Sigma}\boldsymbol{\delta}_\perp)^{-1/2}\boldsymbol{\delta}_\perp'\mathbf{B}(r) = (\boldsymbol{\delta}_\perp'\boldsymbol{\Sigma}\boldsymbol{\delta}_\perp)^{-1/2}\boldsymbol{\delta}_\perp'\mathbf{B}_\varepsilon(r) = \mathbf{W}_\perp(r), \tag{31}$$

where $\mathbf{W}_\perp(r) = \mathbf{BM}(\mathbf{I}_{N-1})$, or $(N-1)$ – vector standard Brownian motion. The transformation matrix that appears in (31) can be estimated by implementing the following modification of our earlier procedure.

Orthogonalization procedure (OP).

Step 1: Construct the moment matrix of differences (for models M1 and M2) or demeaned differences (for model M3) which we write as $\mathbf{M}_T = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t'$. As in the stationary case, $\mathbf{M}_T$ is a consistent (as $T \to \infty$, N fixed) estimate of $\mathbf{V}_u$. Again, let m_{Tij} be the ijth element of $\mathbf{M}_T$.

Step 2: Estimate the cross section coefficients $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}$ by moment based optimization as in (23) leading to $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Sigma}})$. As before, $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Sigma}}) \to_p (\boldsymbol{\delta}, \boldsymbol{\Sigma})$ as $T \to \infty$, with N fixed, and $\hat{\boldsymbol{\Sigma}}$ is positive definite for large enough T.

Step 3: Using $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\delta}}$, construct[7] $\hat{\boldsymbol{\delta}}_\perp$ and $\hat{\mathbf{F}}_\delta = (\hat{\boldsymbol{\delta}}_\perp'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\delta}}_\perp)^{-1/2}\hat{\boldsymbol{\delta}}_\perp'$. Clearly,

$$\hat{\mathbf{F}}_\delta = (\hat{\boldsymbol{\delta}}_\perp'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\delta}}_\perp)^{-1/2}\hat{\boldsymbol{\delta}}_\perp' \to_p (\boldsymbol{\delta}_\perp'\boldsymbol{\Sigma}\boldsymbol{\delta}_\perp)^{-1/2}\boldsymbol{\delta}_\perp', \tag{32}$$

as $T \to \infty$.

Using $\hat{\mathbf{F}}_\delta$ we transform the data $\mathbf{y}_t$ (or demeaned/detrended data in the case of models M2 and M3) giving $\mathbf{y}_t^+ = \hat{\mathbf{F}}_\delta\mathbf{y}_t$. As is apparent from (31), the transformation $\hat{\mathbf{F}}_\delta$ asymptotically removes cross section dependence in the panel and $\mathbf{y}_t^+$ is asymptotically cross section independent as $T \to \infty$. Using $\mathbf{y}_t^+$ we may now construct estimates of the autoregressive coefficients. Let $\hat{\rho}_i^+$ (resp. $\underline{\hat{\boldsymbol{\rho}}}^+$) be the OLS estimate of $\rho_i = 1$ ($\underline{\boldsymbol{\rho}} = \boldsymbol{\iota}_{N-1}$). Then, in an obvious notation,

$$\underline{\hat{\boldsymbol{\rho}}}^+ = \left(\sum_{t=1}^{T}\mathbf{Z}_t^{+\prime}\mathbf{Z}_t^+\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{Z}_t^{+\prime}\mathbf{y}_t^+\right).$$

Let $\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}}^+$ be the corresponding vector of median unbiased estimates of ρ_i. Similarly, let $\tilde{\rho}^+$ be the PFMGU estimate of ρ obtained from the transformed data $\mathbf{y}_t^+$ under the assumption of homogeneous unit roots. The modified Hausman statistic is defined as

$$G_H^+ = T^2(\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}}^+ - \tilde{\rho}^+\boldsymbol{\iota}_{N-1})'(\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}}^+ - \tilde{\rho}^+\boldsymbol{\iota}_{N-1}). \tag{33}$$

As shown in Appendix B

$$G_H^+ \to_d \boldsymbol{\Xi}_{N-1}'\boldsymbol{\Xi}_{N-1}, \tag{34}$$

where

$$\boldsymbol{\Xi}_{N-1} = \begin{bmatrix} \left[\int_0^1 W_{\perp,1}^2\right]^{-1}\left[\int_0^1 W_{\perp,1}dW_{\perp,1}\right] - \left[\int_0^1 \mathbf{W}_\perp'\mathbf{W}_\perp\right]^{-1}\left[\int_0^1 \mathbf{W}_\perp' d\mathbf{W}_\perp\right] \\ \vdots \\ \left[\int_0^1 W_{\perp,N-1}^2\right]^{-1}\left[\int_0^1 W_{\perp,N-1}dW_{\perp,N-1}\right] - \left[\int_0^1 \mathbf{W}_\perp'\mathbf{W}_\perp\right]^{-1}\left[\int_0^1 \mathbf{W}_\perp' d\mathbf{W}_\perp\right] \end{bmatrix}, \tag{35}$$

and where $\{W_{\perp,i} : i = 1, \ldots, N-1\}$ are the components of the $N-1$ vector standard Brownian motion $\mathbf{W}_\perp$ Clearly, G_H^* is free of nuisance parameters in the limit and is suitable for testing the null $H_0 : \rho_i = 1\ \forall i$.

[7] The orthogonal complement matrix $\hat{\boldsymbol{\delta}}_\perp$ can be constructed by taking the eigenvectors of the projection matrix $\mathbf{P}_{\hat{\delta}} = \mathbf{I} - \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\delta}})^{-1}\hat{\boldsymbol{\delta}}'$ corresponding to unit eigenvalues.

An alternative approach is to construct panel unit root test statistics directly by taking the sum of the differences between the estimates $\hat{\rho}_i^+$, $\hat{\rho}_{i,\mathrm{emu}}^+$ and their limits under the null, viz.

$$G_{\mathrm{ols}}^+ = \sum_{i=1}^{N-1} \frac{\hat{\rho}_i^+ - 1}{\hat{\sigma}_{\hat{\rho}^+}} \tag{36}$$

$$G_{\mathrm{emu}}^+ = \sum_{i=1}^{N-1} \frac{\hat{\rho}_{i,\mathrm{emu}}^+ - 1}{\hat{\sigma}_{\hat{\rho}_{i,\mathrm{emu}}^+}}. \tag{37}$$

In contrast to (33), the test statistics (36) and (37) do not involve a pooled estimate of the homogenous unit root parameter. As shown in Appendix B, for fixed N we have the following limit theory for these statistics as $T \to \infty$

$$G_{\mathrm{ols}}^+, \to_d \sum_{i=1}^{N-1} \xi_i, \qquad G_{\mathrm{emu}}^+ \to_d \sum_{i=1}^{N-1} \xi_i^-, \tag{38}$$

where $\xi_i = (\int_0^1 W_i^2)^{-1}(\int_0^1 W_i dW_i)$ and

$$\xi_i^- = \begin{cases} \xi_i & \xi_i < 0 \\ 0 & \xi_i \geq 0. \end{cases}$$

The limits in (38) depend only on N. Both G_{ols}^+, G_{emu}^+ are therefore suitable for testing the null H_0.

Note that there are only $N - 1$ elements in (36)–(38). This is because the panel system has been transformed to dimension $N - 1$ in Step 4 above in order to remove the effects of cross section dependence in the limit.

The tests (36) and (37) have the advantage that they lend themselves to simple large N asymptotics. In particular, the means and variances

$$E(\xi_i), E(\xi_i^-) = \mu_\xi, \mu_{\xi^-} \qquad \mathrm{Var}(\xi_i), \mathrm{Var}(\xi_i^-) = \sigma_\xi^2, \sigma_{\xi^-}^2$$

can be computed and, noting that ξ_i, ξ_i^- are *i.i.d.* over i, we have the large N limit theory

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} (\xi_i - \mu_\xi) \to_d N(0, \sigma_\xi^2), \qquad \frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} (\xi_i^- - \mu_{\xi^-}) \to_d N(0, \sigma_{\xi^-}^2).$$

It follows that in sequential asymptotics (see Phillips and Moon (1999)) as $(T, N \to \infty)_{\mathrm{seq}}$

$$\left.\begin{aligned} G_{\mathrm{ols}}^{++} &= \tfrac{1}{\sqrt{N}\sigma_\xi} \textstyle\sum_{i=1}^{N-1} \left[\frac{\hat{\rho}_i^+ - 1}{\hat{\sigma}_{\hat{\rho}^+}} - \mu_\xi \right] \\ G_{\mathrm{emu}}^{++} &= \tfrac{1}{\sqrt{N}\sigma_{\xi^-}} \textstyle\sum_{i=1}^{N-1} \left[\frac{\hat{\rho}_{i,\mathrm{emu}}^+ - 1}{\hat{\sigma}_{\hat{\rho}_{i,\mathrm{emu}}^+}} - \mu_{\xi^-} \right] \end{aligned}\right\} \to_d N(0, 1).$$

All of these procedures are easy to implement. Their finite sample performance is assessed in Section 6 below. As shown in the next section, once the OP procedure has been applied to the data, a wide class of panel unit root and stationarity tests become applicable.

4.3. Dynamic AR(p) panels with cross section dependence

The procedures outlined above for panel unit root testing under cross section dependence may be applied to cases of higher order panel dynamics and cases where the common factor component θ_t is weakly dependent. Specifically, consider a panel of dynamic panel autoregressions with (possibly) heterogenous lag orders ℓ_i for each i and allow for cross section dependence of the same form as (6) above. The model is written in augmented format as

$$\Delta y_{it} = \underline{\mu}_i + \underline{\beta}_i t + (\rho - 1) y_{it-1} + \sum_{j=1}^{\ell_i} \phi_{ij} \Delta y_{it-j} + u_{it}. \tag{39}$$

The OP procedure leading to (32) above is the same as that laid out above except for the first step. Here, instead of using the moment matrix of differences or demeaned differences, one simply uses the moment matrix of the regression residuals $\hat{u}_{it}$ obtained under the (null hypothesis) restriction $\rho = 1$ in (39).

Since the transformed data y_{it}^{+} are asymptotically uncorrelated across i, regressions like (39) of y_{it}^{+} on y_{it-1}^{+} and the lagged differences Δy_{it-j}^{+} do not suffer (asymptotically) from cross section dependence. Importantly, this will be so even when the common time series factor θ_t is weakly dependent rather than uncorrelated over time. This is because the transformation procedure leading to (32) continues to eliminate the contribution of the common factor component θ_t to the limit Brownian motion in (30). It follows that several existing panel unit root tests that were designed to work with data that are independent across section can now be applied to test for panel unit roots when there is cross section dependence. Accordingly, we consider here two broad types of panel unit root tests.

Meta-analysis tests for panel unit roots and stationarity under cross section dependence. The first type of test is based on meta-analysis, wherein the P-values of tests for each cross section individual i are combined to construct a new test. Tests of this type were suggested in Choi (2001a) and Maddala and Wu (1999) for use in testing unit roots with panel data under cross section independence.[8] These tests apply here under cross section dependence after our OP orthogonalization procedure has been implemented. Choi (2001a) provides a full discussion of tests of this type and his simulation results suggest use of the three tests that we concentrate on here.

Let p_i be the P-value of a unit root test associated with cross section element i. Define

$$P = -2 \sum_{i=1}^{N-1} \ln(p_i), \tag{40}$$

$$P_m = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} [\ln(p_i) + 1], \tag{41}$$

and

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} \Phi^{-1}(p_i). \tag{42}$$

[8] Choi (2001b) considers several statistics based on meta-analysis with random individual and time effects in (1).

The P test is called the inverse chi-square test or Fisher test after Fisher (1932). The P_m test statistic is a centred and normalized version of P that is useful for large N. The Z test is called the inverse normal test, following Stouffer *et al.* (1949). As discussed in Choi (2001a), we have the following limit distributions for P and Z as $T \to \infty$

$$P \to_d \chi^2_{2(N-1)}, \qquad Z \to_d N(0,1) \qquad \text{for fixed } N, \tag{43}$$

leading to the following sequential limit theory as $(T, N \to \infty)_{\text{seq}}$

$$P_m, Z \to_d N(0,1). \tag{44}$$

Each of these tests and the limit theory applies under the null hypothesis to dynamic panel autoregressions like (39) with cross section dependence after the OP procedure has been implemented.

Other tests for panel unit roots. In fact, after transforming the data using the OP procedure, we can apply most other methods for testing panel unit roots that are valid under cross section independence. Baltagi (2001) provides a recent discussion and overview of these tests, which generally take the form of cross section averages of time series test statistics and have the generic form

$$G_\tau = \frac{1}{N-1}\sum_{i=1}^{N-1} \tau_i,$$

where τ_i stands for an individual unit root test statistic. This class of tests can also be extended by using the bias reduction techniques discussed earlier in this paper. For instance, we could use an ADF-*t* statistic based not on OLS estimation but instead on EMU estimation as explained earlier (cf. Andrews and Chen (1994)).

Im *et al.* (1997) use two cross sectional average tests constructed like G_τ and study their small sample properties using simulations. Without modification, this type of test typically suffers from serious size distortion in small samples due to SSB bias. IPS use simulation to calculate the mean and variance of the G_τ statistics and they employ bias correction in the implementation of these procedures. However, in the dynamic panel AR(p) case, the means and variances of the G_τ statistics heavily depend on the nuisance parameters that arise in the augmented dynamic terms. Tanaka (1984) and Shaman and Stine (1988) provide formulae for the mean bias for cases up to an AR(6) for models 1 and 2. For example, for an AR(2), the OLS estimator of ρ_i in (39) will be biased downward when the true coefficient on y^+_{it-2} is negative, while it will be biased upward when the true coefficient on y^+_{it-2} is large and positive. IPS also found that the size distortion problem of their G_τ tests relies heavily on the sign of the true coefficient on y^+_{it-2}. Since their Monte Carlo studies are based on AR(2) process, their size distortion corrections are based on the sign and magnitude of the coefficient on y^+_{it-2}. For general dynamic panel AR(p) processes, the size of the G_τ test will depend on all the nuisance parameters arising in the augmented terms and, in the absence of analytic formulae, extensive simulations are needed to make the appropriate corrections in such cases.

The finite sample performance of these panel unit root tests and, more generally, tests of homogeneity are considered in the simulation experiments reported in Section 5 below.

5. SIMULATION EXPERIMENTS

This section consists of three parts. First, we report the finite sample performance of the three panel median unbiased estimators. Second, we show the finite sample performance of the Wald statistic W_{surmu} and the G_{pfmgu} statistic. Finally, we examine the small sample performance of the panel unit root tests G^{++}_{emu}, G^{++}_{ols}, P_m and Z and show how well the orthogonalization procedure for handling cross sectional dependence works.

5.1. Design of data generating process

The data generating process for the first two parts is given by

$$y_{it} = \rho_i y_{it-1} + u_{it}, \tag{45}$$

$$u_{it} = \delta_i \theta_t + \varepsilon_{it}, \tag{46}$$

where $\varepsilon_{it} \sim i.i.d.\ N(0, 1)$ over i and t, $\theta_t \sim i.i.d.\ N(0, 1)$ over t, and for (ρ_i, δ_i) parameter selections are used that are detailed below. The primary distinction is between the homogenous case where $\rho_i = \rho$ for all i and the heterogenous case where the ρ_i differ across individuals i. We also distinguish cases of high and low cross section dependence according to the value of δ_i. Estimation is based on the following two regression models that involve a fitted mean and trend:

$$y_{it} = a_i + \rho_i y_{it-1} + u_{it} \qquad \text{for model M2}$$

$$y_{it} = a_i + b_i t + \rho_i y_{it-1} + u_{it} \qquad \text{for model M3.}$$

Panel data are generated under four specifications which differ according to the degree of the cross sectional dependence and whether or not the homogeneity restriction is imposed on ρ. These specifications are as follows:

Case I (Homogeneity and Low Cross sectional Dependence). The homogeneity restriction is imposed and we set $\rho_1 = \rho_2 = \cdots = \rho_N = 0.9$, and allow low cross sectional dependence by setting $\delta_i \sim U[0, 0.2]$, where $U[a, b]$ represents the uniform distribution over the interval $[a, b]$. In this experiment, the average error (u_{it}) cross sectional dependence has correlation coefficient around 0.03.

Case II (Homogeneity and High Cross sectional Dependence). Again, we set $\rho_i = 0.9$ for all i and $\delta_i \sim U[1, 4]$. Here, the lowest error (u_{it}) cross sectional correlation is around 0.52, the median is around 0.82 and the highest is around 0.94.

Case III (Heterogeneity and Low Cross sectional Dependence). Here, $\rho_i \sim U[0.7, 0.9]$, and $\delta_i \sim U[0, 0.2]$.

Case IV (Heterogeneity and High Cross sectional Dependence). Here $\rho_i \sim U[0.7, 0.9]$ and $\delta_i \sim U[1, 4]$.

Case V (Testing Homogeneity under Stationarity). Under the null hypothesis of homogeneity of ρ, we set $\rho_i = 0.8$ for all i to investigate test size. Under the alternative, we set $\rho_i \sim U[0.7, 0.9]$ and consider test power.

Each experiment involves 5,000 replications of panel samples of (N, T) observations. We use $N = 10, 20, 30$ and $T = 50, 100, 200$.

The third part of the simulation has two sections. In the first section the fitted models have intercepts and trends (as in M2 and M3) and the DGP is based on (45) and (46) with the following parameter settings:

Case VI (Testing Panel Unit Roots under Cross sectional Dependence). Here, $\rho_i = 1.0$ for all i under the null, and we set $\delta_i \sim U[1, 4]$ for high cross sectional dependence. We use $\rho_i \sim U[0.8, 1.0]$ as the alternative hypothesis to calculate the power of the tests.

In the second section, the fitted models again have intercepts and trends (as in M2 and M3) and the DGP is based on

$$\begin{aligned} y_{i,t} &= \rho_i y_{i,t-1} + v_{i,t}, \\ v_{it} &= \phi_i v_{it-1} + u_{it} \qquad \text{AR(1) errors,} \qquad (47) \\ v_{it} &= \kappa_i u_{it-1} + u_{it} \qquad \text{MA(1) errors,} \qquad (48) \\ u_{it} &= \delta_i \theta_t + \varepsilon_{it}, \end{aligned}$$

with the following parameter settings:

Case VII (Testing Panel Unit Roots under Cross sectional Dependence and Weak Dependence). As in Case VI, $\rho_i = 1.0$ for all i under the null, $\delta_i \sim U[1, 4]$ for high cross sectional dependence and $\rho_i \sim U[0.8, 1.0]$ is used as the alternative hypothesis. In addition the parameters of the time series models in (47) and (48) are set as follows:

$$\begin{aligned} &\phi_i \sim U[0, 0.4] && \text{AR(1) errors,} \\ &\kappa_i \sim U[0, 0.4] && \text{MA(1) errors, } \kappa_i > 0, \\ &\kappa_i \sim U[-0.4, 0] && \text{MA(1) errors, } \kappa_i < 0. \end{aligned}$$

5.2. Finite sample properties

Table 2 reports mean square errors (MSE's) of the POLS, PFGLS and PFGMU estimators. The first column shows the MSE $\times\ 10^2$ of the POLS estimator, and the second and third columns show the ratios of the MSE of the other estimators to that of the POLS estimator. When the degree of cross sectional dependence is low, the PFGLS estimator becomes less efficient than the POLS since the MSE ratio is greater than one in all these cases. Surprisingly, two panel median unbiased estimators have much better MSEs than POLS even for low degrees of cross sectional dependence. The ordering among the estimators in terms of MSE performance (higher is better) is PFGLS < POLS < PFGMU for both models M2 and M3. When there are high degrees of cross sectional dependence, the performance ordering changes to POLS < PFGLS < PFGMU. The performance of the PFGMU estimator is substantially better than POLS in all cases, yielding MSEs that are 5–20 times better than POLS.

Table 3 shows the average MSE of the OLS, EMU, SUR and SUR-MU estimators over N. When the degree of cross sectional dependence is low (Case III), the order among the estimators in terms of MSE performance (again, higher is better in what follows) is SUR < OLS < SUR-MU < EMU. When there are high degrees of cross sectional dependence, this ordering changes

Table 2. Monte Carlo performance of POLS, PFGLS and panel FGMU estimators under homogenous ρ (cases I & II): MSE and MSE ratios.

Sample size	Only constant			Constant and trend		
	MSE	MSE ratio		MSE	MSE ratio	
	POLS	PFGLS	PFGMU	POLS	PFGLS	PFGMU
			Low cross sectional dependence: Case I			
$N = 10, T = 50$	0.372	1.294	0.331	1.282	1.336	0.183
$N = 20, T = 50$	0.306	1.725	0.208	1.174	1.719	0.137
$N = 30, T = 50$	0.279	2.136	0.177	1.140	2.017	0.168
$N = 10, T = 100$	0.082	1.161	0.401	0.269	1.189	0.189
$N = 20, T = 100$	0.067	1.360	0.261	0.247	1.414	0.106
$N = 30, T = 100$	0.060	1.581	0.208	0.233	1.636	0.081
$N = 10, T = 200$	0.025	1.070	0.544	0.063	1.086	0.252
$N = 20, T = 200$	0.016	1.182	0.393	0.052	1.208	0.151
$N = 30, T = 200$	0.016	1.261	0.302	0.052	1.309	0.110
			High cross sectional dependence: Case II			
$N = 10, T = 50$	1.210	0.515	0.139	2.585	0.779	0.113
$N = 20, T = 50$	1.224	0.730	0.188	2.654	1.033	0.143
$N = 30, T = 50$	1.172	1.013	0.318	2.583	1.299	0.238
$N = 10, T = 100$	0.368	0.324	0.108	0.668	0.544	0.085
$N = 20, T = 100$	0.327	0.379	0.092	0.626	0.648	0.070
$N = 30, T = 100$	0.340	0.465	0.121	0.623	0.790	0.090
$N = 10, T = 200$	0.124	0.216	0.103	0.192	0.370	0.081
$N = 20, T = 200$	0.120	0.202	0.066	0.191	0.381	0.050
$N = 30, T = 200$	0.118	0.214	0.059	0.180	0.437	0.048

to OLS < EMU < SUR < SUR-MU. Overall, the SUR-MU estimator has MSE performance that is 5 times better than that of the OLS estimator and twice as good as that of the SUR estimator.

Table 4 displays finite sample properties of the Wald test for dynamic homogeneity, i.e. $H_0 : \rho_i = \rho$ for all i with $\rho = 0.7$ (Case V). As mentioned earlier, the size distortion of the Wald test is substantial and the distortion gets larger and becomes very serious as the number of cross sectional units increases. Even for large values of T the size distortion is considerable. It is also worse for the fitted trend case. Interestingly, the size distortion is worse under low cross sectional dependence than it is under high dependence. We deduce that the Wald test for homogeneity in dynamic panels is very unreliable and not to be recommended.

In contrast, Table 5 shows much more reasonable finite sample performance of the G statistic in the stationary case. As N becomes large for small T, the size of the G test increases, due to reduced degrees of freedom. But for moderate T, the G test suffers only mild size distortion and the size is conservative for larger T. Moreover, the size adjusted power of the G test is nearly unity in all the cases considered.

Table 3. Monte Carlo performance of OLS, MU, SUR and SUR-MU estimators under heterogeneous ρ_i (cases III & IV): MSE and MSE ratios.

Sample size	Constant				Constant and trend			
	MSE	MSE ratio			MSE	MSE ratio		
	OLS	MU	SUR	SUR-MU	OLS	MU	SUR	SUR-MU
	Low cross sectional dependence: Case III							
$N = 10, T = 50$	1.691	0.812	1.134	1.028	2.827	0.660	1.108	0.846
$N = 20, T = 50$	1.740	0.807	1.212	1.351	2.923	0.654	1.153	1.114
$N = 30, T = 50$	1.727	0.806	1.222	1.827	2.876	0.650	1.130	1.453
$N = 10, T = 100$	0.610	0.856	1.066	0.936	0.858	0.717	1.057	0.796
$N = 20, T = 100$	0.603	0.856	1.144	1.079	0.870	0.715	1.121	0.930
$N = 30, T = 100$	0.601	0.859	1.195	1.217	0.863	0.717	1.168	1.062
$N = 10, T = 200$	0.242	0.921	1.044	0.966	0.302	0.803	1.039	0.845
$N = 20, T = 200$	0.241	0.919	1.079	1.002	0.302	0.800	1.070	0.878
$N = 30, T = 200$	0.239	0.922	1.117	1.048	0.299	0.806	1.106	0.925
	High cross sectional dependence: Case IV							
$N = 10, T = 50$	1.734	0.815	0.484	0.308	2.856	0.658	0.584	0.355
$N = 20, T = 50$	1.736	0.801	0.530	0.353	2.916	0.642	0.599	0.506
$N = 30, T = 50$	1.732	0.813	0.617	0.616	2.913	0.656	0.632	0.793
$N = 10, T = 100$	0.633	0.863	0.383	0.265	0.900	0.726	0.458	0.229
$N = 20, T = 100$	0.613	0.866	0.381	0.248	0.861	0.730	0.462	0.221
$N = 30, T = 100$	0.606	0.873	0.413	0.259	0.853	0.729	0.488	0.242
$N = 10, T = 200$	0.241	0.925	0.349	0.284	0.302	0.813	0.400	0.246
$N = 20, T = 200$	0.242	0.915	0.317	0.244	0.303	0.798	0.373	0.213
$N = 30, T = 200$	0.249	0.922	0.305	0.228	0.311	0.805	0.361	0.202

Table 6 deals with the panel unit root case and shows the size and size adjusted power of the IPS, G^{++}_{ols}, G^{++}_{emu}, P and Z tests in respective columns. Overall, G^{++}_{emu} shows better performance than G^{++}_{ols} in terms of both size and power comparisons. The P and Z tests are in turn superior to the G tests and have considerably greater power. All of these tests outrank the IPS test, which shows considerable size distortion as well as lower power. Generally, the power of the tests for model M2 (the fitted intercept case) is higher than that for model M3 (fitted constant and linear trend). The results for the P and Z tests are particularly good and indicate that these panel unit root tests work well in the presence of cross section dependence.

Tables 7 and 8 report further results for the P, Z and IPS tests in the case where the model has AR(1) and MA(1) errors, respectively. Apparently, both P and Z tests work very well in terms of size and power for AR(1) errors. This is not unexpected given that the ADF procedure is used to obtain estimates of the errors in the first stage of the procedure leading to these tests. On the other hand, neither the P nor Z tests work well for MA(1) errors, both tests showing size distortion in this case. Similar results were obtained for the case of MA(1) errors

Table 4. Wald test for homogeneity (case V). $H_0 : \rho_i = \rho = 0.7$. Cross sectional correlation (min = 0.52, med = 0.82, max = 0.94).

Sample size	Constant		Constant and trend	
	Size (5%)	Size (2.5%)	Size (5%)	Size (2.5%)
	Low cross sectional dependence			
$N = 10, T = 50$	0.466	0.369	0.571	0.474
$N = 10, T = 100$	0.185	0.123	0.225	0.153
$N = 10, T = 200$	0.103	0.051	0.115	0.059
$N = 20, T = 50$	0.983	0.973	0.982	0.971
$N = 20, T = 100$	0.584	0.488	0.653	0.555
$N = 20, T = 200$	0.253	0.174	0.285	0.198
$N = 30, T = 50$	1.000	1.000	0.998	0.996
$N = 30, T = 100$	0.906	0.781	0.937	0.855
$N = 30, T = 200$	0.433	0.207	0.478	0.251
	High cross sectional dependence			
$N = 10, T = 50$	0.351	0.263	0.522	0.440
$N = 10, T = 100$	0.155	0.107	0.176	0.120
$N = 10, T = 200$	0.096	0.059	0.101	0.063
$N = 20, T = 50$	0.873	0.820	0.959	0.938
$N = 20, T = 100$	0.421	0.341	0.464	0.377
$N = 20, T = 200$	0.226	0.153	0.236	0.163
$N = 30, T = 50$	1.000	0.995	0.979	0.968
$N = 30, T = 100$	0.703	0.503	0.742	0.558
$N = 30, T = 200$	0.337	0.159	0.341	0.162

with negative coefficients but these are not reported here. An alternative approach to removing serial dependence, such as the non-parametric adjustments used in Phillips (1987), may be more successful in this case, although we have not implemented that procedure in the present work. The IPS test shows substantially greater size distortion in all cases and generally seems to be inferior to the other tests.

6. CONCLUDING REMARKS

Panel models with dynamic autoregressive components are now extensively used in empirical research in growth economics and international finance, both areas where cross section dependence is likely to be important. In the absence of alternative approaches, it is often convenient in such studies to deal with cross section dependence by means of a CTE, to ignore issues of bias and to presume the validity of homogeneity restrictions. The bias problem in dynamic panel regressions with fixed effects is shown here to persist and be compounded by high variance when

Table 5. G-test for homogeneity (case V) $H_0 : \rho_i = \rho = 0.8$ with cross sectional correlation (min $= 0.52$, med $= 0.82$, max $= 0.94$).

	Model 2				Model 3			
Sample	1%	2.5%	5%	10%	1%	2.5%	5%	10%
				Size				
$N = 10, T = 50$	0.028	0.043	0.065	0.092	0.027	0.046	0.068	0.089
$N = 20, T = 50$	0.051	0.075	0.100	0.136	0.047	0.069	0.094	0.126
$N = 30, T = 50$	0.082	0.110	0.136	0.172	0.071	0.092	0.114	0.140
$N = 10, T = 100$	0.017	0.032	0.050	0.080	0.019	0.030	0.052	0.077
$N = 20, T = 100$	0.015	0.027	0.045	0.069	0.017	0.035	0.048	0.076
$N = 30, T = 100$	0.025	0.039	0.056	0.085	0.028	0.043	0.055	0.086
$N = 10, T = 200$	0.003	0.014	0.024	0.043	0.004	0.014	0.024	0.043
$N = 20, T = 200$	0.008	0.015	0.028	0.044	0.008	0.016	0.028	0.051
$N = 30, T = 200$	0.008	0.016	0.024	0.046	0.008	0.016	0.027	0.046
				Size adjusted power				
$N = 10, T = 50$	0.981	0.972	0.959	0.920	0.972	0.959	0.944	0.906
$N = 20, T = 50$	0.990	0.984	0.978	0.968	0.991	0.985	0.980	0.964
$N = 30, T = 50$	0.999	0.998	0.997	0.995	0.999	0.997	0.996	0.996
$N = 10, T = 100$	0.988	0.979	0.961	0.941	0.984	0.972	0.952	0.924
$N = 20, T = 100$	0.994	0.987	0.979	0.968	0.995	0.989	0.978	0.969
$N = 30, T = 100$	0.999	0.999	0.998	0.998	0.999	0.999	0.999	0.998
$N = 10, T = 200$	0.997	0.993	0.987	0.978	0.978	0.966	0.957	0.932
$N = 20, T = 200$	0.995	0.992	0.989	0.981	0.995	0.992	0.988	0.982
$N = 30, T = 200$	0.999	0.998	0.997	0.996	0.999	0.999	0.997	0.996

there is cross section dependence. Tests for homogeneity are also affected by cross section dependence, including the case of homogenous unit roots. These are issues that need corrective action in applied work.

The solutions offered in this paper to address these issues start with the use of median unbiased estimation procedures for estimation, testing and confidence interval construction. On the whole, the new estimation methods work well to correct for bias and reduce variance, accounting for cross section dependence in conditions (viz. correct specification, no additional regressors, and Gaussianity) that might be described as 'ideal' for these methods. When conditions are not 'ideal', for example when the distributional assumptions underlying the median function are incorrect, other bias correction methods such as those based on higher order expansions (e.g Hahn and Kuersteiner (2002)) may be useful. The present analysis is useful in calibrating how well such methods can work in relation to the median unbiased approach.

On the other hand, the present paper shows that Wald tests for homogeneity suffer from unacceptable size distortions even under ideal conditions, including stationarity. We have therefore proposed a modified Hausman test for homogeneity that utilizes a pooled panel MUE estimator that is asymptotically efficient under the null, in conjunction with MUE estimates that are robust to heterogeneity and moment based estimates of the cross section dependence

Table 6. Tests for homogenous panel unit roots under cross section dependence (case VI): cross sectional correlation (min = 0.52, med = 0.82, max = 0.94).

Sample	IPS	G^{++}_{ols}	G^{++}_{emu}	P	Z
Panel A: model M2—fitted intercept case					
			Size: 5%		
$N = 10, T = 50$	0.257	0.052	0.052	0.044	0.046
$N = 20, T = 50$	0.353	0.061	0.046	0.044	0.050
$N = 30, T = 50$	0.367	0.061	0.041	0.044	0.049
$N = 10, T = 100$	0.263	0.047	0.063	0.045	0.047
$N = 20, T = 100$	0.333	0.051	0.055	0.044	0.049
$N = 30, T = 100$	0.376	0.054	0.057	0.039	0.048
$N = 10, T = 200$	0.242	0.046	0.054	0.041	0.047
$N = 20, T = 200$	0.337	0.043	0.049	0.044	0.044
$N = 30, T = 200$	0.391	0.049	0.047	0.046	0.049
			Size adjusted power		
$N = 10, T = 50$	0.247	0.252	0.270	0.997	0.996
$N = 20, T = 50$	0.223	0.329	0.330	0.988	0.974
$N = 30, T = 50$	0.256	0.519	0.532	0.978	0.969
$N = 10, T = 100$	0.646	0.687	0.739	1.000	1.000
$N = 20, T = 100$	0.627	0.692	0.779	0.997	0.993
$N = 30, T = 100$	0.587	0.811	0.866	0.991	0.987
$N = 10, T = 200$	0.991	0.970	0.983	1.000	1.000
$N = 20, T = 200$	0.989	0.934	0.968	0.999	0.998
$N = 30, T = 200$	0.986	0.975	0.988	1.000	0.999
Panel B: model M3—fitted intercept and trend					
			Size: 5%		
$N = 10, T = 50$	0.278	0.077	0.072	0.043	0.048
$N = 20, T = 50$	0.366	0.086	0.073	0.044	0.049
$N = 30, T = 50$	0.390	0.098	0.067	0.046	0.052
$N = 10, T = 100$	0.280	0.062	0.073	0.049	0.052
$N = 20, T = 100$	0.357	0.064	0.063	0.044	0.047
$N = 30, T = 100$	0.379	0.078	0.068	0.049	0.053
$N = 10, T = 200$	0.260	0.049	0.062	0.046	0.049
$N = 20, T = 200$	0.313	0.044	0.056	0.042	0.045
$N = 30, T = 200$	0.363	0.047	0.055	0.042	0.046
			Size adjusted power		
$N = 10, T = 50$	0.122	0.086	0.088	0.985	0.983
$N = 20, T = 50$	0.142	0.097	0.095	0.969	0.947

Table 6. Continued.

Sample	IPS	G^{++}_{ols}	G^{++}_{emu}	P	Z
			Size adjusted power		
$N = 30, T = 50$	0.133	0.158	0.160	0.960	0.943
$N = 10, T = 100$	0.349	0.342	0.380	0.998	0.996
$N = 20, T = 100$	0.350	0.413	0.435	0.990	0.975
$N = 30, T = 100$	0.344	0.558	0.609	0.981	0.971
$N = 10, T = 200$	0.885	0.853	0.890	1.000	1.000
$N = 20, T = 200$	0.881	0.815	0.878	0.999	0.994
$N = 30, T = 200$	0.886	0.892	0.938	0.998	0.993

Table 7. Tests for homogenous panel unit roots under cross section dependence & AR(1) errors (case VII). Cross sectional correlation (min = 0.52, med = 0.82, max = 0.94).

Sample	IPS		P		Z	
	5%	10%	5%	10%	5%	10%
Panel A: fitted intercept						
			Size			
$N = 10, T = 50$	0.202	0.272	0.056	0.112	0.057	0.111
$N = 20, T = 50$	0.329	0.381	0.057	0.110	0.055	0.113
$N = 30, T = 50$	0.374	0.412	0.066	0.117	0.064	0.115
$N = 10, T = 100$	0.188	0.256	0.047	0.094	0.046	0.099
$N = 20, T = 100$	0.315	0.364	0.047	0.099	0.049	0.100
$N = 30, T = 100$	0.363	0.402	0.047	0.094	0.048	0.095
$N = 10, T = 200$	0.198	0.261	0.042	0.093	0.047	0.095
$N = 20, T = 200$	0.330	0.382	0.040	0.091	0.049	0.100
$N = 30, T = 200$	0.373	0.412	0.043	0.088	0.046	0.092
			Power			
$N = 10, T = 50$	0.294	0.415	0.993	0.997	0.992	0.998
$N = 20, T = 50$	0.225	0.343	0.984	0.991	0.979	0.986
$N = 30, T = 50$	0.199	0.325	0.981	0.989	0.981	0.988
$N = 10, T = 100$	0.632	0.763	1.000	1.000	0.999	1.000
$N = 20, T = 100$	0.592	0.706	0.998	0.999	0.995	0.997
$N = 30, T = 100$	0.539	0.689	0.997	0.999	0.995	0.997
$N = 10, T = 200$	0.984	0.994	1.000	1.000	1.000	1.000
$N = 20, T = 200$	0.967	0.987	1.000	1.000	1.000	1.000
$N = 30, T = 200$	0.967	0.987	1.000	1.000	1.000	1.000

Table 7. Continued.

Sample	IPS		*P*		*Z*	
	5%	10%	5%	10%	5%	10%
Panel B: fitted intercept and trend						
			Size			
$N = 10, T = 50$	0.218	0.279	0.051	0.100	0.050	0.096
$N = 20, T = 50$	0.327	0.372	0.049	0.096	0.049	0.098
$N = 30, T = 50$	0.382	0.414	0.054	0.107	0.056	0.104
$N = 10, T = 100$	0.205	0.259	0.047	0.091	0.050	0.098
$N = 20, T = 100$	0.319	0.366	0.049	0.092	0.051	0.100
$N = 30, T = 100$	0.360	0.393	0.048	0.094	0.053	0.100
$N = 10, T = 200$	0.193	0.254	0.039	0.084	0.042	0.085
$N = 20, T = 200$	0.312	0.355	0.037	0.083	0.044	0.093
$N = 30, T = 200$	0.365	0.402	0.040	0.086	0.045	0.091
			Power			
$N = 10, T = 50$	0.168	0.259	0.976	0.987	0.973	0.985
			Power			
$N = 20, T = 50$	0.143	0.229	0.953	0.978	0.938	0.960
$N = 30, T = 50$	0.116	0.206	0.955	0.973	0.938	0.961
$N = 10, T = 100$	0.400	0.535	0.993	0.997	0.988	0.995
$N = 20, T = 100$	0.353	0.477	0.986	0.991	0.970	0.984
$N = 30, T = 100$	0.334	0.467	0.988	0.993	0.974	0.983
$N = 10, T = 200$	0.890	0.940	1.000	1.000	1.000	1.000
$N = 20, T = 200$	0.831	0.903	1.000	1.000	0.997	0.998
$N = 30, T = 200$	0.813	0.895	1.000	1.000	0.998	0.999

parameters. Simulations indicate that this homogeneity test, whose limit distribution is chi-squared, works well except in cases where N and T are both small.

In the important case of tests for homogenous panel unit roots, we utilize the same moment based approach to estimation of the cross section dependence parameters δ and use these estimates to project on the space orthogonal to the common time effect in the panel. After this data transformation, it becomes possible to employ conventional panel unit root tests that have been developed under the assumption of independence. Simulations reveal that there are major differences between test procedures in practice, with some procedures (like the IPS test of Im *et al.* (1997)) suffering serious size distortion. The P-value based meta Z test of Choi (2001a) is found to work particularly well with stable size and good power and is easy to compute and apply in practice. Moon and Perron (2002) have independently suggested a related procedure for panel unit root testing that involves principal components estimation. They show that the approach may be used in dynamic panels with multiple factors in which the rank of the factor space itself has to be estimated.

Table 8. Tests for homogenous panel unit roots under cross section dependence & MA(1) errors (case VII). Cross sectional correlation (min = 0.52, med = 0.82, max = 0.94).

Sample	IPS		*P*		*Z*	
	5%	10%	5%	10%	5%	10%
Panel A: fitted intercept						
			Size			
$N = 10, T = 50$	0.247	0.323	0.083	0.150	0.084	0.151
$N = 20, T = 50$	0.371	0.421	0.090	0.173	0.089	0.163
$N = 30, T = 50$	0.421	0.466	0.108	0.192	0.110	0.193
$N = 10, T = 100$	0.235	0.315	0.072	0.131	0.071	0.137
$N = 20, T = 100$	0.344	0.404	0.083	0.159	0.086	0.161
$N = 30, T = 100$	0.430	0.467	0.100	0.173	0.101	0.169
$N = 10, T = 200$	0.242	0.305	0.066	0.131	0.073	0.134
$N = 20, T = 200$	0.366	0.414	0.081	0.153	0.090	0.161
$N = 30, T = 200$	0.409	0.450	0.092	0.170	0.103	0.177
			Power			
$N = 10, T = 50$	0.284	0.433	0.998	1.000	0.998	0.999
$N = 20, T = 50$	0.233	0.367	0.988	0.993	0.982	0.987
$N = 30, T = 50$	0.246	0.359	0.993	0.997	0.987	0.992
$N = 10, T = 100$	0.695	0.821	1.000	1.000	1.000	1.000
$N = 20, T = 100$	0.639	0.773	0.999	1.000	0.997	0.998
$N = 30, T = 100$	0.590	0.723	1.000	1.000	0.997	0.999
$N = 10, T = 200$	0.998	1.000	1.000	1.000	1.000	1.000
$N = 20, T = 200$	0.987	0.996	1.000	1.000	1.000	1.000
$N = 30, T = 200$	0.986	0.996	1.000	1.000	1.000	1.000
Panel B: fitted intercept and trend						
			Size			
$N = 10, T = 50$	0.290	0.358	0.087	0.164	0.088	0.158
$N = 20, T = 50$	0.387	0.431	0.111	0.206	0.107	0.198
$N = 30, T = 50$	0.458	0.492	0.155	0.253	0.152	0.250
$N = 10, T = 100$	0.280	0.336	0.087	0.164	0.090	0.165
$N = 20, T = 100$	0.390	0.434	0.111	0.201	0.121	0.207
$N = 30, T = 100$	0.460	0.495	0.143	0.234	0.155	0.248
$N = 10, T = 200$	0.257	0.325	0.082	0.150	0.086	0.163
$N = 20, T = 200$	0.384	0.430	0.099	0.189	0.111	0.197
$N = 30, T = 200$	0.438	0.474	0.131	0.225	0.142	0.239
			Power			
$N = 10, T = 50$	0.131	0.225	0.990	0.996	0.990	0.994
$N = 20, T = 50$	0.123	0.217	0.958	0.976	0.939	0.959

Table 8. Continued.

Sample	IPS		P		Z	
	5%	10%	5%	10%	5%	10%
$N = 30, T = 50$	0.130	0.215	0.969	0.984	0.952	0.971
$N = 10, T = 100$	0.406	0.528	1.000	1.000	1.000	1.000
			Power			
$N = 20, T = 100$	0.361	0.506	0.985	0.990	0.970	0.978
$N = 30, T = 100$	0.349	0.481	0.992	0.996	0.980	0.986
$N = 10, T = 200$	0.934	0.974	1.000	1.000	1.000	1.000
$N = 20, T = 200$	0.853	0.928	0.999	1.000	0.995	0.997
$N = 30, T = 200$	0.849	0.931	1.000	1.000	0.999	0.999

ACKNOWLEDGEMENTS

Our thanks go to two referees for helpful comments and to Feng Zhu for pointing out some errors in the original version of this paper. Computational work was performed in GAUSS. Phillips thanks the NSF for support under Grant # SES 0092509.

REFERENCES

Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica 61*, 139–65.

Andrews, D. W. K. and H. Chen (1994). Approximately median-unbiased estimation of autoregssive models. *Journal of Business & Economic Statistics 12*, 187–204.

Bai, J. and S. Ng (2001). Determining the number of factors in approximate factor models. *Econometrica*, forthcoming.

Baltagi (2001). *Econometric Analysis of Panel Data*. Baffins Lane Chichester: John Wiley & Sons Ltd.

Barro, R. J. and X. Sala-i-Martin (1992). Convergence. *Journal of Political Economy 100*, 223–51.

Cermeno, R. (1999). Median unbiased estimation in fixed effects dynamic panels. *Annales D'Economie Et De Statistique*. pp. 351–68.

Choi, I. (2001a). Unit root tests for panel data. *Journal of International Money and Finance 20*, 249–72.

Choi, I (2001b). Unit root tests for cross-sectionally correlated panels. Mimeo, Kookmin University.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 427–31.

Durlauf, S. N. and D. T. Quah (1999). The new emprics of economic growth. *Handbook of Macroeconomics 1A*, Chapter 4, Amsterdam: North-Holland Elsevier.

Fisher, R. A. (1932). *Statistical Methods for Research Workers*. London: Oliver and Boyd.

Hahn, J., J. Hausman and G. Kuersteiner (2001). Bias corrected instrumental variables estimation for dynamic panel models with fixed effects. MIT Working Paper No 01-24.

Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects. *Econometrica 70*, 1639–58.

Hurvicz, L. (1950). Least squares bias in time series. In T. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*. pp. 365–83. New York: Wiley.

Im, K. S., M. H. Pesaran and Y. Shin (1997). Testing for unit roots in heterogeneous panels. Mimeo.
Kendall, M. G. (1954). Note on the bias in the estimation of autocorrelation. *Biometrika 41*, 403–4.
Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. New York: John Wiley and Sons.
Maddala, G. S. and S. Wu (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics 61*, 631–52.
Marriott, F. H. C. and J. A. Pope (1954). Bias in the estimation of autocorrelations. *Biometrika 41*, 393–403.
Moon, H. R and B. Perron (2002). Testing for a unit root in panels with dynamic factors, USC. Mimeo.
Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417–26.
Orcutt, G. H. (1948). A study of the autoregressive nature of the times series used for Tinbergen's model of the economic system of the United States. *Journal of the Royal Statistical Society, Series B* 1–45.
Orcutt, G. H. and H. S. Winokur (1969). First order autoregression: inference, estimation and prediction. *Econometrica 37*, 1–14.
Phillips, P. C. B. (1977). An approximation to the finite sample distribution of Zellner's seemingly unrelated regression estimator. *Journal of Econometrics 6*, 147–64.
Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica 55*, 277–301.
Phillips, P. C. B. (ed.) (1993). Operational algebra and regression t tests. *Models, Methods and Applications of Econometrics*, Chapter 10, Oxford: Blackwell.
Phillips, P. C. B. and H. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*, 1057–111.
Phillips, P. C. B. and D. Sul (2002). Dynamic panel estimation and homogeneity testing under cross section dependence. Cowles Foundation Discussion Paper #1362, Yale University.
Phillips, P. C. B. and D. Sul (2003). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence, Mimeo, Yale University.
Rothenberg, T. J. (1984). Approximate normality of generalized least squares estimates. *Econometrica 52*, 811–26.
Shaman, P. and R. A. Stine (1988). The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association 83*, 842–8.
Stouffer, S. A., E. A. Suchman, L. C. DeVinney, S. A. Star and R. M. Williams (1949). *The American Soldier, I*. Princeton: Princeton University Press.
Tanaka, K. (1984). An asymptotic expansion associated with the maximum likelihood estimators in ARMA models. *Journal of the Royal Statistical Society, Series B 46*, 58–67.

APPENDIX A

Proof of Property IP1. For model M1, the result follows directly by scaling. For model M2, we have

$$\hat{\rho}_{\text{pols2}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - y_{i.-1})(y_{it} - y_{i.})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - y_{i.-1})^2}. \tag{A.1}$$

Now, $y_{it} = \mu_i + y_{it}^* = \mu_i + \sum_{j=0}^{\infty}\rho^j u_{i,t-j}$ and so $y_{it} - y_{i.} = y_{it}^* - y_{i.}^*$ and $y_{it-1} - y_{i.-1} = y_{it-1}^* - y_{i.-1}^*$ are both invariant to μ_i. Also

$$\frac{y_{it-1} - y_{i.-1}}{\sigma} = \frac{y_{it-1} - y_{i.-1}}{\sigma_i}\frac{\sigma_i}{\sigma},$$

whose factors are invariant to μ_i, σ_i and σ. For model M3, we have in the stationary case

$$y_{it} = \mu_i + \beta_i t + y_{it}^* = \mu_i + \beta_i t + \sum_{j=0}^{\infty}\rho^j u_{i,t-j}.$$

When we regress y_{it} and y_{it-1} on $x_t' = (1, t)$ for $t = 1, \ldots, T$, the residuals are linear functions of the y_{it}^* and are invariant to (μ_i, β_i). Let $\mathbf{Q}_t$ be the orthogonal projection matrix onto the orthogonal complement of the space spanned by the matrix $\mathbf{X} = [x_1, \ldots, x_T]'$ and let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, \ldots, y_{iT-1})'$, with a corresponding notation for $\mathbf{y}_i^*$ and $\mathbf{y}_{i,-1}^*$. The residual vectors from these detrending regressions are

$$\hat{\mathbf{y}}_i = \mathbf{Q}_t \mathbf{y}_i = \mathbf{Q}_t \mathbf{y}_i^* = \hat{\mathbf{y}}_i^*,$$

and

$$\hat{\mathbf{y}}_{i,-1} = \mathbf{Q}_t \mathbf{y}_{i,-1} = \mathbf{Q}_t \mathbf{y}_{i,-1}^* = \hat{\mathbf{y}}_{i,-1}^*.$$

The POLS estimator in model M3 is

$$\hat{\rho}_{\text{pols3}} = \frac{\sum_{i=1}^N \sum_{t=1}^T \hat{y}_{it-1}\hat{y}_{it}}{\sum_{i=1}^N \sum_{t=1}^T \hat{y}_{it-1}^2} = \frac{\sum_{i=1}^N \sum_{t=1}^T \hat{y}_{it-1}^*\hat{y}_{it}^*}{\sum_{i=1}^N \sum_{t=1}^T \hat{y}_{it-1}^{*2}} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{\hat{y}_{it-1}^*\hat{y}_{it}^*}{\sigma_i^2}\frac{\sigma_i^2}{\sigma^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{\hat{y}_{it-1}^{*2}}{\sigma_i^2}\frac{\sigma_i^2}{\sigma^2}}, \tag{A.2}$$

and invariance to $(\mu_i, \beta_i, \sigma^2)$ is clear. Proofs for the non-stationary case ($\rho = 1$) for models 2 and 3 carry over in a similar fashion using $y_{it}^* - y_{i0}^* = \sum_{j=0}^{t-1} u_{i,t-j}$ and the fact that y_{i0}^* is removed by the demeaning and detrending filters.

Proof of Property IP2. Invariance to (μ_i, β_i) follows precisely as in the proof of Property IP1. Let $\mathbf{y}_t^* = (y_{1t}^*, \ldots, y_{Nt}^*)'$, and $\hat{\mathbf{y}}_t^* = (\hat{y}_{1t}^*, \ldots, \hat{y}_{Nt}^*)'$ where $\hat{y}_{it}^*$ denotes y_{1t}^* or demeaned or detrended y_{it}^*, respectively for models M1, M2 and M3, with corresponding notation for $\mathbf{y}_t$ and $\hat{\mathbf{y}}_t$. Let $\mathbf{\Omega}$ be the matrix whose ijth element is $\rho^{|i-j|}/(1-\rho^2)$. In the stationary case, $\mathbf{y}_t^* \sim N\left(\mathbf{0}, \frac{1}{1-\rho^2}\mathbf{V}_u\right)$, and, vectorizing $[y_1^*, \ldots, y_T^*]$ into the column $\mathbf{y}^* \sim N(\mathbf{0}, \mathbf{\Omega} \otimes \mathbf{V}_u)$, we have $(\mathbf{I}_N \otimes \mathbf{V}_u^{-\frac{1}{2}})\mathbf{y}^* \sim N(\mathbf{0}, \mathbf{\Omega} \otimes \mathbf{I}_N)$, which depends only on ρ. A similar result holds for the vectorized column $\mathbf{y}_{-1}^*$ of lagged variables $[y_0^*, \ldots, y_{T-1}^*]$. Next, using the notation $\hat{\mathbf{y}}^* = (\mathbf{Q}_t \otimes \mathbf{I}_N)\mathbf{y}^*$ and $\hat{\mathbf{y}}_{-1}^* = (\mathbf{Q}_t \otimes \mathbf{I}_N)\mathbf{y}_{-1}^*$, we have

$$(\mathbf{I}_N \otimes \mathbf{V}_u^{-\frac{1}{2}})\hat{\mathbf{y}}^*, \qquad (\mathbf{I}_N \otimes \mathbf{V}_u^{-\frac{1}{2}})\hat{\mathbf{y}}_{-1}^* \sim N(\mathbf{0}, \mathbf{Q}_t \mathbf{\Omega} \mathbf{Q}_t \otimes \mathbf{I}_n),$$

and the GLS estimator

$$\hat{\rho}_{\text{pgls}} = \frac{\sum_{t=1}^T \hat{\mathbf{y}}_{t-1}' \mathbf{V}_u^{-1} \hat{\mathbf{y}}_t}{\sum_{t=1}^T \hat{\mathbf{y}}_{t-1}' \mathbf{V}_u^{-1} \hat{\mathbf{y}}_{t-1}} = \frac{\sum_{t=1}^T \hat{\mathbf{y}}_{t-1}^{*\prime} \mathbf{V}_u^{-1} \hat{\mathbf{y}}_t^*}{\sum_{t=1}^T \hat{\mathbf{y}}_{t-1}^{*\prime} \mathbf{V}_u^{-1} \hat{\mathbf{y}}_{t-1}^*} = \frac{\hat{\mathbf{y}}_{-1}^{*\prime}(\mathbf{I}_T \otimes \mathbf{V}_u^{-1})\hat{\mathbf{y}}^*}{\hat{\mathbf{y}}_{-1}^{*\prime}(\mathbf{I}_T \otimes \mathbf{V}_u^{-1})\hat{\mathbf{y}}_{-1}^*},$$

is seen to depend only on ρ.

Again, proofs in the non-stationary case ($\rho = 1$) for models 2 and 3 carry over in a similar fashion using $y_{it}^* - y_{i0}^* = \sum_{j=0}^{t-1} u_{i,t-j}$ and the fact that y_{i0}^* is removed by the demeaning and detrending filters.

APPENDIX B

Derivation of SUR limit theory

Stationary case. We use the heterogenous model for SUR estimation with $y_{it} = y_{it}^*$ (i.e. model M1)

$$y_{it}^* = \rho_i y_{it-1}^* + u_{it}, \qquad \text{for } t = 1, \ldots, T, \qquad \text{and } i = 1, \ldots, N, \tag{B.1}$$

in which the regression errors are from (6)

$$u_{it} = \delta_i \theta_t + \varepsilon_{it}, \qquad \theta_t \sim i.i.d.\ N(0,1) \text{ over } t, \tag{B.2}$$

and

$$\varepsilon_{i,t} \sim i.i.d.\ N(0, \sigma_i^2) \text{ over } t, \text{ and } \varepsilon_{i,t} \text{ is independent of } \varepsilon_{j,s} \text{ and } \theta_s \text{ for all } i \neq j \text{ and for all } s, t. \tag{B.3}$$

The proof in the case of models M2 and M3 is a straightforward extension. From (B.2) and (B.3)

$$\mathbf{u}_t \sim i.i.d.\ N(\mathbf{0}, \mathbf{V}_u), \qquad \text{for } t = 1, \ldots, T,$$

where, as in (8), we have

$$\mathbf{V}_u = E(\mathbf{u}_t \mathbf{u}_t' \mid \sigma_1^2, \ldots, \sigma_N^2) = \boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}', \qquad \boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_N^2).$$

Now write (B.1) in vector form as

$$\mathbf{y}_t = \mathbf{Z}_t \underline{\boldsymbol{\rho}} + \mathbf{u}_t, \qquad \mathbf{Z}_t = \operatorname{diag}(y_{1t-1}, \ldots, y_{Nt-1}), \qquad \underline{\boldsymbol{\rho}} = (\rho_1, \ldots, \rho_N)'. \tag{B.4}$$

Then the GLS estimate is

$$\hat{\underline{\boldsymbol{\rho}}} = \left(\sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{V}_u^{-1} \mathbf{Z}_t \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{V}_u^{-1} \mathbf{y}_t \right),$$

and the SUR estimate is simply a feasible version of this estimate with V_u estimated by a consistent estimate. GLS and SUR are obviously asymptotically equivalent.

Under stationarity $|\rho_i| < 1$ for all i we have by standard theory that

$$\sqrt{T}(\hat{\underline{\boldsymbol{\rho}}} - \underline{\boldsymbol{\rho}}) \to_d N(\mathbf{0}, \mathbf{V}_{\mathrm{SUR}}), \tag{B.5}$$

with

$$\mathbf{V}_{\mathrm{SUR}} = \operatorname{plim}_{T\to\infty} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{V}_u^{-1} \mathbf{Z}_t \right)^{-1}.$$

We can calculate the inverse of this matrix as follows. Note that

$$\operatorname{plim}_{T\to\infty} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{V}_u^{-1} \mathbf{Z}_t \right) = [(v_u^{ij} E(y_{it} y_{jt}))_{ij}],$$

where v_u^{ij} is the ijth element of $\mathbf{V}_u$, so that

$$\mathbf{V}_{\mathrm{SUR}}^{-1} = [(v_u^{ij} E(y_{it} y_{jt}))_{ij}] = \mathbf{V}_u^{-1} \odot E(\mathbf{y}_t \mathbf{y}_t').$$

Next note that

$$\begin{aligned} E(y_{it} y_{jt}) &= E\left(\sum_{s=0}^{\infty} \rho_i^s u_{it-s} \sum_{p=0}^{\infty} \rho_j^p u_{jt-p} \right) \\ &= \frac{E(u_{it} u_{jt})}{1 - \rho_i \rho_j} = \begin{cases} \frac{\sigma_i^2 + \delta_i^2}{1 - \rho_i^2} & i = j \\ \frac{\delta_i \delta_j}{1 - \rho_i \rho_j} & i \neq j, \end{cases} \end{aligned}$$

so that

$$E(\mathbf{y}_t \mathbf{y}_t') = (\boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}') \odot \mathbf{R},$$

with $\mathbf{R} = [(r_{ij})]$ and $r_{ij} = \frac{1}{1 - \rho_i \rho_j}$. Note that $\mathbf{V}_u^{-1} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}}{1 + \boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}}$. The same result (B.5) holds for models M2 and M3 in the stationary case as trend elimination does not affect the limit theory.

Unit root case. When $\rho_i = 1$ for all i, we have the functional law

$$\frac{1}{\sqrt{T}}\mathbf{y}_{[\mathrm{Tr}]} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[\mathrm{Tr}]}\mathbf{u}_t \rightarrow_d \mathbf{B}(r) = \mathbf{BM}(\mathbf{V}_u). \tag{B.6}$$

Setting $\boldsymbol{\iota}_N$ to be vector with N unit components, the centred GLS and feasible SUR estimates have the form

$$\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}} - \boldsymbol{\iota}_N = \left(\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{V}_u^{-1}\mathbf{Z}_t\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{V}_u^{-1}\mathbf{u}_t\right).$$

Now

$$\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{V}_u^{-1}\mathbf{Z}_t = \left[\left(v_u^{ij}\frac{1}{T^2}\sum_{t=1}^{T}y_{it-1}y_{jt-1}\right)_{ij}\right] \rightarrow_d \left[\left(v_u^{ij}\int_0^1 B_i B_j\right)_{ij}\right]$$
$$= \mathbf{V}_u^{-1}\odot\int_0^1 \mathbf{BB}',$$

and

$$\frac{1}{T}\left(\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{V}_u^{-1}\mathbf{u}_t\right) = \left[\left(\sum_{j=1}^{N}v_u^{ij}\frac{1}{T}\sum_{t=1}^{T}y_{it-1}u_{jt}\right)_i\right] \rightarrow_d \left[\left(\sum_{j=1}^{N}v_u^{ij}\int_0^1 B_i\,dB_j\right)_i\right]$$
$$= \int_0^1 \mathbf{B}\odot\mathbf{V}_u^{-1}d\mathbf{B}.$$

This gives the stated limit result

$$T(\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}} - \boldsymbol{\iota}_N) \rightarrow_d \left[\mathbf{V}_u^{-1}\odot\int_0^1\mathbf{BB}'\right]^{-1}\left[\int_0^1\mathbf{B}\odot\mathbf{V}_u^{-1}d\mathbf{B}\right] = \boldsymbol{\xi}.$$

Note that the quadratic variation process of the stochastic integral $\int_0^r \mathbf{B}\odot\mathbf{V}_u^{-1}d\mathbf{B}$ is

$$\left[\int_0^r \mathbf{B}\odot\mathbf{V}_u^{-1}d\mathbf{B}\right]_r = \mathbf{V}_u^{-1}\odot\int_0^r\mathbf{BB}',$$

so the matrix $\mathbf{V}_u^{-1}\odot\int_0^1\mathbf{BB}'$ is a suitable metric for $\int_0^1\mathbf{B}\odot\mathbf{V}_u^{-1}d\mathbf{B}$. The joint Wald test for unit roots is

$$W_{\mathrm{SUR}} = (\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}} - \boldsymbol{\iota}_N)'\left[\sum_{t=1}^{T}\mathbf{Z}_t'\mathbf{V}_u^{-1}\mathbf{Z}_t\right](\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}} - \boldsymbol{\iota}_N)$$
$$\rightarrow_d \left[\int_0^1\mathbf{B}\odot\mathbf{V}_u^{-1}d\mathbf{B}\right]'\left[\mathbf{V}_u^{-1}\odot\int_0^1\mathbf{BB}'\right]^{-1}\left[\int_0^1\mathbf{B}\odot\mathbf{V}_u^{-1}d\mathbf{B}\right],$$

which is dependent on nuisance parameters. Also, if we were to test homogeneity using the SUR estimate $\underline{\hat{\boldsymbol{\rho}}}$, then noting that

$$\mathbf{D}(\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}} - \boldsymbol{\iota}_N) = \mathbf{D}\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{sur}},$$

we would have the statistic

$$WD_{\text{sur}} = \underline{\hat{\boldsymbol{\rho}}}_{\text{sur}}' \mathbf{D}' [\mathbf{D}\hat{\mathbf{V}}_{\text{SUR}}\mathbf{D}']^{-1} \mathbf{D} \underline{\hat{\boldsymbol{\rho}}}_{\text{sur}}$$

$$\to_d \boldsymbol{\xi}'\mathbf{D}' \left[\mathbf{D} \left(\mathbf{V}_u^{-1} \odot \int_0^1 \mathbf{B}\mathbf{B}' \right)^{-1} \mathbf{D}' \right]^{-1} \mathbf{D}\xi.$$

Next consider the pooled estimate of ρ when H_0 holds. In this case, we have

$$\hat{\boldsymbol{\rho}} = \left(\sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{V}_u^{-1} \mathbf{y}_{t-1} \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{V}_u^{-1} \mathbf{y}_t \right)$$

and by straightforward derivation

$$\frac{1}{T^2} \sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{V}_u^{-1} \mathbf{y}_{t-1} \to_d \int_0^1 \mathbf{B}' \mathbf{V}_u^{-1} \mathbf{B} =_d \int_0^1 \mathbf{W}'\mathbf{W} = \sum_{i=1}^{N} \int_0^1 W_i^2,$$

$$\frac{1}{T} \sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{V}_u^{-1} \mathbf{u}_t \to_d \int_0^1 \mathbf{B}' \mathbf{V}_u^{-1} d\mathbf{B} =_d \int_0^1 \mathbf{W}' d\mathbf{W} = \sum_{i=1}^{N} \int_0^1 W_i \, dW_i,$$

and so

$$T\left(\hat{\rho} - 1\right) \to_d \frac{\int_0^1 \mathbf{W}' d\mathbf{W}}{\int_0^1 \mathbf{W}'\mathbf{W}} = \frac{\sum_{i=1}^{N} \int_0^1 W_i \, dW_i}{\sum_{i=1}^{N} \int_0^1 W_i^2}, \tag{B.7}$$

where $\mathbf{W} = (W_i)$ is standard Brownian motion with covariance matrix $\mathbf{I}_N$. Hence, the limit distribution of $T\left(\hat{\rho} - 1\right)$ is free of nuisance parameters.

Hausman Test Limit Theory (Unit Root Case)

The Hausman statistic relies on the difference $\sqrt{T}(\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}} - \boldsymbol{\iota}_N) = \sqrt{T}(\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}} - \boldsymbol{\iota}_N)$. From (B.7) we have

$$T(\tilde{\rho} - 1) \to_d \frac{\int_0^1 \mathbf{W}' d\mathbf{W}}{\int_0^1 \mathbf{W}'\mathbf{W}},$$

and

$$T\left(\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}} - \boldsymbol{\iota}_N\right) = T(\underline{\hat{\boldsymbol{\rho}}} - \boldsymbol{\iota}_N) + o_p(1)$$

$$= \left(\frac{1}{T^2} \sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{Z}_t \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{u}_t \right)$$

$$\to_d \left[\int_0^1 \mathbf{D}_{B^2} \right]^{-1} \left[\int_0^1 \mathbf{D}_B \, d\mathbf{B} \right]$$

$$= \begin{bmatrix} \left[\int_0^1 B_1^2 \right]^{-1} \left[\int_0^1 B_1 \, dB_1 \right] \\ \vdots \\ \left[\int_0^1 B_N^2 \right]^{-1} \left[\int_0^1 B_N \, dB_N \right] \end{bmatrix}, \tag{B.8}$$

where we use the notation $\mathbf{D}_A = \mathrm{diag}(A_1(r), \ldots, A_N(r))$. In view of the correlation between the Brownian motions $\{B_i : i = 1, \ldots, N\}$ the limit distribution (B.8) is dependent on nuisance parameters arising from the cross section dependence.

We also have the joint convergence

$$\begin{bmatrix} T(\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{emu}} - \boldsymbol{\iota}_N) \\ T(\tilde{\rho} - 1) \end{bmatrix} \to_d \begin{bmatrix} \left[\int_0^1 \mathbf{D}_{B^2}\right]^{-1} \left[\int_0^1 \mathbf{D}_B d\mathbf{B}\right] \\ \left[\int_0^1 \mathbf{B}'\mathbf{B}\right]^{-1} \left[\int_0^1 \mathbf{B}' d\mathbf{B}\right] \end{bmatrix},$$

and then

$$T(\underline{\hat{\boldsymbol{\rho}}}_{\mathrm{emu}} - \boldsymbol{\iota}_N) \xrightarrow{d} \begin{bmatrix} \left[\int_0^1 \mathbf{B}'\mathbf{B}\right]^{-1} \left[\int_0^1 \mathbf{B}' d\mathbf{B}\right] \\ \vdots \\ \left[\int_0^1 \mathbf{B}'\mathbf{B}\right]^{-1} \left[\int_0^1 \mathbf{B}' d\mathbf{B}\right] \end{bmatrix}.$$

Again, this limit distribution is dependent on nuisance parameters arising from cross section dependence. Thus, the Hausman statistic does not produce an asymptotically similar test in the unit root case.

Modified Hausman and Panel Unit Root Tests Limit Theory

First note that we have the joint convergence

$$\begin{bmatrix} T(\underline{\hat{\boldsymbol{\rho}}}^+_{\mathrm{emu}} - \boldsymbol{\iota}_{N-1}) \\ T(\tilde{\rho}^+ - 1) \end{bmatrix} \to_d \begin{bmatrix} \left[\int_0^1 \mathbf{D}_{W_\perp^2}\right]^{-1} \left[\int_0^1 \mathbf{D}_{W_\perp} d\mathbf{W}_\perp\right] \\ \left[\int_0^1 \mathbf{W}_\perp'\mathbf{W}_\perp\right]^{-1} \left[\int_0^1 \mathbf{W}_\perp{}' d\mathbf{W}_\perp\right] \end{bmatrix},$$

which is free of nuisance parameters. Then

$$\begin{aligned} T(\underline{\hat{\boldsymbol{\rho}}}^+_{\mathrm{emu}} - \tilde{\rho}^+\boldsymbol{\iota}_{N-1}) &\to_d \left[\int_0^1 \mathbf{D}_{W_\perp^2}\right]^{-1} \left[\int_0^1 \mathbf{D}_{W_\perp} d\mathbf{W}_\perp\right] - \left[\int_0^1 \mathbf{W}_\perp'\mathbf{W}_\perp\right]^{-1} \left[\int_0^1 \mathbf{W}_\perp{}' d\mathbf{W}_\perp\right] \boldsymbol{\iota}_{N-1} \\ &= \begin{bmatrix} \left[\int_0^1 W_{\perp,1}^2\right]^{-1} \left[\int_0^1 W_{\perp,1} dW_{\perp,1}\right] - \left[\int_0^1 \mathbf{W}_\perp'\mathbf{W}_\perp\right]^{-1} \left[\int_0^1 \mathbf{W}_\perp{}' d\mathbf{W}_\perp\right] \\ \vdots \\ \left[\int_0^1 W_{\perp,N-1}^2\right]^{-1} \left[\int_0^1 W_{\perp,N-1} dW_{\perp,N-1}\right] - \left[\int_0^1 \mathbf{W}_\perp'\mathbf{W}_\perp\right]^{-1} \left[\int_0^1 \mathbf{W}_\perp{}' d\mathbf{W}_\perp\right] \end{bmatrix} \\ &:= \Xi_{N-1}, \end{aligned} \tag{B.9}$$

and it follows that the modified Hausman test has the following limit:

$$G_H^+ = T^2(\underline{\hat{\boldsymbol{\rho}}}^+_{\mathrm{emu}} - \tilde{\rho}^+\boldsymbol{\iota}_{N-1})'(\underline{\hat{\boldsymbol{\rho}}}^+_{\mathrm{emu}} - \tilde{\rho}^+\boldsymbol{\iota}_{N-1}) \to_d \boldsymbol{\Xi}_{N-1}'\boldsymbol{\Xi}_{N-1}.$$

Similarly, the modified unit root tests have the limit

$$G^+_{\mathrm{ols}}, G^+_{\mathrm{emu}} \to_d \sum_{i=1}^{N-1} \left[\int_0^1 W_i^2\right]^{-1/2} \left[\int_0^1 W_i dW_i\right], \qquad \text{for fixed } N.$$

APPENDIX C

Algorithm for MLE Estimation of Cross Section Dependence Coefficients

We develop here an iterative procedure for estimating the cross section dependence coefficient vector δ using maximum likelihood. As above, we work with model M1 and make suitable adjustments in the case of models M2 and M3. Write the model in vector form as in (21) above, viz.

$$\mathbf{y}_t = \mathbf{Z}_t \underline{\boldsymbol{\rho}} + \mathbf{u}_t, \qquad \mathbf{Z}_t = \operatorname{diag}(y_{1t-1}, \ldots, y_{Nt-1}), \qquad \underline{\boldsymbol{\rho}} = (\rho_1, \ldots, \rho_N)',$$

with errors $\mathbf{u}_t$ that are $i.i.d.\ N(\mathbf{0}, \mathbf{V}_u)$ where $\mathbf{V}_u = \boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}'$ and $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. The log likelihood function has the form

$$\begin{aligned}\ell_{NT}(\underline{\boldsymbol{\rho}}, \boldsymbol{\Sigma}, \boldsymbol{\delta}) &= -\frac{NT}{2}\log 2\pi - \frac{T}{2}\log \mathbf{V}_u - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{Z}_t\underline{\boldsymbol{\rho}})'\mathbf{V}_u^{-1}(\mathbf{y}_t - \mathbf{Z}_t\underline{\boldsymbol{\rho}}) \\ &= -\frac{NT}{2}\log 2\pi - \frac{T}{2}\log \mathbf{V}_u - \frac{T}{2}\operatorname{tr}[\mathbf{V}_u^{-1}\mathbf{M}_T],\end{aligned}$$

where $\mathbf{M}_T(\underline{\boldsymbol{\rho}}) = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{Z}_t\underline{\boldsymbol{\rho}})(\mathbf{y}_t - \mathbf{Z}_t\underline{\boldsymbol{\rho}})'$. First-order conditions for maximization of $\ell_{NT}(\underline{\boldsymbol{\rho}}, \boldsymbol{\Sigma}, \boldsymbol{\delta})$ lead to

$$\hat{\underline{\boldsymbol{\rho}}} = \left(\sum_{t=1}^{T}\mathbf{Z}_t'\hat{\mathbf{V}}_u^{-1}\mathbf{Z}_t\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{Z}_t'\hat{\mathbf{V}}_u^{-1}\mathbf{y}_t\right), \tag{C.1}$$

and

$$\operatorname{tr}\left[\left(\hat{\mathbf{V}}_u^{-1} - \hat{\mathbf{V}}_u^{-1}\mathbf{M}_T\hat{\mathbf{V}}_u^{-1}\right)d\mathbf{V}_u\right] = 0, \tag{C.2}$$

where $\hat{\mathbf{V}}_u = \hat{\boldsymbol{\Sigma}} + \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}'$, $\hat{\boldsymbol{\Sigma}} = \operatorname{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2)$ and $d\mathbf{V}_u = d\boldsymbol{\Sigma} + d\boldsymbol{\delta}\boldsymbol{\delta}' + \boldsymbol{\delta}d\boldsymbol{\delta}'$. Expanding (C.2) leads to the following system of equations:

$$\hat{\sigma}_i^2\left[1 - \frac{\hat{\delta}_i^2/\hat{\sigma}_i^2}{1 + \hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\delta}}}\right] = \left[e_i' - \frac{\hat{\delta}_i\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}^{-1}}{1 + \hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\delta}}}\right]M_T(\hat{\underline{\rho}})\left[e_i - \frac{\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\delta}}\hat{\delta}_i}{1 + \hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\delta}}}\right], \quad i = 1, \ldots, N \tag{C.3}$$

$$\hat{\boldsymbol{\delta}} = \frac{\mathbf{M}_T(\hat{\underline{\boldsymbol{\rho}}})\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\delta}}}{1 + \hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\delta}}}, \tag{C.4}$$

which we may solve by the following iteration:

$$\begin{aligned}&\hat{\sigma}_2^{2(j)}\left[1 - \frac{(\hat{\delta}_i^{(j-1)})^2/\hat{\sigma}_i^{2(j-1)}}{1 + \hat{\boldsymbol{\delta}}^{(j-1)\prime}\left(\hat{\boldsymbol{\Sigma}}^{(j-1)}\right)^{-1}\hat{\boldsymbol{\delta}}^{(j-1)}}\right] \\ &\quad = \left[e_i' - \frac{\hat{\delta}_j^{(j-1)}\hat{\boldsymbol{\delta}}^{(j-1)\prime}\left(\hat{\boldsymbol{\Sigma}}^{(j-1)}\right)^{-1}}{1 + \hat{\boldsymbol{\delta}}^{(j-1)\prime}\left(\hat{\boldsymbol{\Sigma}}^{(j-1)}\right)^{-1}\hat{\boldsymbol{\delta}}^{(j-1)}}\right]\mathbf{M}_T(\hat{\underline{\boldsymbol{\rho}}})\left[e_i - \frac{\left(\hat{\boldsymbol{\Sigma}}^{(j-1)}\right)^{-1}\hat{\boldsymbol{\delta}}^{(j-1)}\hat{\delta}_i^{(j-1)}}{1 + \hat{\boldsymbol{\delta}}^{(j-1)\prime}\left(\hat{\boldsymbol{\Sigma}}^{(j-1)}\right)^{-1}\hat{\boldsymbol{\delta}}^{(j-1)}}\right],\end{aligned}$$

$$\hat{\boldsymbol{\delta}}^{(j)} = \frac{\mathbf{M}_T(\hat{\underline{\boldsymbol{\rho}}})\left(\hat{\boldsymbol{\Sigma}}^{(j-1)}\right)^{-1}\hat{\boldsymbol{\delta}}^{(j-1)}}{1 + \hat{\boldsymbol{\delta}}^{(j-1)\prime}\left(\hat{\boldsymbol{\Sigma}}^{(j-1)}\right)^{-1}\hat{\boldsymbol{\delta}}^{(j-1)}},$$

which is continued until convergence. For starting values we may choose $\hat{\mathbf{\Sigma}}^{(0)} = \hat{\sigma}^2 \mathbf{I}_N$ where $\hat{\sigma}^2 = \frac{1}{N} tr[\mathbf{M}_T]$ and $\hat{\boldsymbol{\delta}}^{(0)}$ is the largest eigenvector of $\mathbf{M}_T$. In place of the the residual moment matrix, $\mathbf{M}_T(\underline{\hat{\boldsymbol{\rho}}})$, from maximum likelihood estimation that appears in (C.3) and (C.4), we propose that the matrix $\mathbf{M}_T(\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}})$ corresponding to the median unbiased estimates $\underline{\hat{\boldsymbol{\rho}}}_{\text{emu}}$ be used.

Note that in the special case where $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_N$, the first-order equations lead to the following system simplifying (C.3) and (C.4):

$$\hat{\sigma}^2 \left[N - \frac{\hat{\boldsymbol{\delta}}' \hat{\boldsymbol{\delta}}}{\hat{\sigma}^2 + \hat{\boldsymbol{\delta}}' \hat{\boldsymbol{\delta}}} \right] = \text{tr} \left[\left(\mathbf{I}_N - \frac{\hat{\boldsymbol{\delta}} \hat{\boldsymbol{\delta}}'}{\hat{\sigma}^2 + \hat{\boldsymbol{\delta}}' \hat{\boldsymbol{\delta}}} \right) \mathbf{M}_T \left(\underline{\hat{\boldsymbol{\rho}}} \right) \left(\mathbf{I}_N - \frac{\hat{\boldsymbol{\delta}} \hat{\boldsymbol{\delta}}'}{\hat{\sigma}^2 + \hat{\boldsymbol{\delta}}' \hat{\boldsymbol{\delta}}} \right) \right],$$

and

$$\hat{\boldsymbol{\delta}} = \frac{\mathbf{M}_T \hat{\boldsymbol{\delta}}}{\hat{\sigma}^2 + \hat{\boldsymbol{\delta}}' \hat{\boldsymbol{\delta}}}.$$