Dynamic Panel Estimation and Homogenity Testing Under Cross Section Dependence

Peter Phillips* Donggyu Sul†

*University of Auckland, peter.phillips@yale.edu

†University of Auckland, d.sul@auckland.ac.nz

Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence*

Peter C.B. Phillips
Cowles Foundation, Yale University
University of Auckland & University of York

Donggyu Sul
Department of Economics
University of Auckland

December 20, 2001

Abstract

Least squares bias in autoregression and dynamic panel regression is shown to be exacerbated in case of cross section dependence. The bias is substantial and is shown to have serious effects in applications like HAC estimation and dynamic half-life response estimation. To address the bias problem, this paper develops a panel approach to median unbiased estimation that takes into account cross section dependence. The new estimators given here considerably reduce the effects of bias and gain precision from estimating cross section error correlation. The 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 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.

JEL Classification Numbers: C32 Time Series Models. C33 Panel Data

*Computational work was performed in GAUSS. Phillips thanks the NSF for support under Grant # SES 0092509.

1 Introduction

This paper studies pooled least squares and proposes new median unbiased estimators (MUE's) for dynamic (autoregressive) panel models under conditions that include cross sectional dependence. The well known bias of ordinary least squares (OLS) in a first order autoregression is shown by simulation and by large cross section (N)/fixed time series (T) asymptotics to be exacerbated in panels with cross section dependence, making the need for procedures that correct for bias more urgent in panel estimation. To address this problem, the paper introduces some new panel estimation procedures that are based on the idea of median unbiased estimation (Lehmann, 1959; Andrews, 1993). The methods developed here include a pooled median unbiased estimator (MUE), a pooled feasible generalized MUE, and a seemingly unrelated regression (SUR) MUE. A large T/fixed N asymptotic theory is given for the SUR-MUE procedure which is useful in testing homogeneity across section.

The problem of small sample bias in the 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 small sample bias (hereafter, SB) were worked out by Kendall (1954) and Marriot and Pope (1954). Orcutt (1948) showed that fitting an intercept in an autoregression produced an additional source of bias that can exacerbate the SB problem, and this was confirmed in a later simulation study by Orcutt and Winokur (1969). The point was echoed in Andrews' (1993) more recent study, which provided further simulations that included the case of a fitted linear trend and explored the alternate procedure of using median unbiased estimation along lines that were suggested originally by Lehmann (1959).

In dynamic panel models with fixed time dimension T and cross section dimension N, Nickell (1981) showed that the presence of heterogeneous intercepts (fixed effects) causes the OLS estimator of a common autoregressive coefficient to be inconsistent as $N \to \infty$ for fixed T. Nickell computed an expression for the bias under the assumption of cross section independence. The existence of this bias in panel regressions is now well known, but it is often neglected or assumed to be of minor importance in empirical studies. A recent discussion of some of the issues in the context of growth rate convergence applications is given by Nerlove (2000). The present paper shows how this bias is exacerbated when there is cross section dependence or when there is unit root nonstationarity, it derives new asymptotic formulae for these cases and it gives some numerical examples to illustrate how the bias can make a huge difference in estimation and testing.

First, a common feature of recent work on the estimation of long run and HAC variances (e.g., Andrews, 1991, Andrews and Monahan, 1992, den Hann and Levin, 1996, Lee and Phillips, 1994) is the use of simple prewhitening filters like AR(1) or VAR(1) regressions. Such filters will also suffer from SB bias problems when the methods are used in a panel context and a homogeneous (common) long run variance is assumed (e.g. Hadri, 2001). In such situations, the SB problem can cause serious bias in the estimation of the long run variance and statistical tests that rely on these estimates will, in turn, suffer size distortions. Some calculations illustrating how large these effects can be are given in Table 1.

Second, the availability of panel data has tempted empirical macroeconomists (eg, Frankel and Rose, 1996, Evans and Karras 1996, and Papell and Murray, 2001) to implement pooling methods to obtain more precise estimates of dynamic response times by imposing a homogeneity restriction on the autoregressive parameter in the estimation of the half-life of dynamic responses. Again, ignoring the SB problem in autoregressive estimation can lead to serious problems of bias in half-life response time estimation. For example, computations show that when the true half-life is 69 years (for an autoregressive coefficient of 0.99 and where the SB is only -0.066) the estimated median half-life is as low as 8 years when there are 20 cross sectional units and 50 time series observations. Moreover, the upper 95% of the estimated half-life distribution is still less than 17 years, revealing an extraordinary level of bias in panel estimation. Some further illustrations of the potential effects of bias in dynamic panel regressions are given below in Table 1 of Section 2.

To address the SB bias problem in dynamic panel estimation and the difficulties that can arise from it, this paper proposes some panel MUE's that follow the approach taken by Andrews (1993) in the time series case. 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 and homogeneous trends. Our framework extends Cermeno's study by developing a class of panel MUE's 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. We also provide an asymptotic analysis of the SB bias problem under cross section dependence and an asymptotic theory for the estimators and homogeneity tests that are proposed here, including the panel unit root tests.

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.

First, we consider the case where cross sectional dependence occurs but is ignored in the panel regression analysis. We find that in this case the pooled OLS estimator provides little gain in precision compared to single equation OLS. 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 SB 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. In view of the potential gains from pooling and the changes in the limit theory in the nonstationary 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, the present 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[1] is developed, which enables the development of a general class of unit root tests for panel models when there is cross section dependence.

The remainder of the paper is organized as follows. The next section shows how even a small time series SB 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 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 alternate 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 provides an empirical application of the estimators to the growth convergence problem. Section 7 concludes. Derivations and some additional technical material are given in the Appendices: A derives some invariance results; B provides extensions of the Nickell (1981) bias formula to cases where there is cross section dependence, unit root nonstationarity and heterogeneous errors; C develops limit theory for the stationary and unit root nonstationary cases; D provides an algorithm for estimating the cross section dependence coefficients.

2 Dynamic Panel Models and Bias Illustrations

2.1 Model Definition

Three basic models are considered. These are panel versions of the models in Andrews (1993). As in that work, Gaussianity is assumed in order to construct the median unbiased estimator. Each of the basic models involves a latent panel $\{y_{i,t}^* : t = 0, 1, ...T;\ i = 1, ..., n\}$ 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

$$y_{i,t}^* = \rho y_{i,t-1}^* + u_{i,t}, \quad \text{for } t = 1, \cdots, T, \quad \text{and } i = 1, \cdots, N, \quad \text{where } \rho \in (-1, 1]. \tag{1}$$

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

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

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

[1] When this paper was in the final stages of completion, the authors learnt that Moon and Perron (2001) have independently proposed the same procedure for unit root testing in the context of dynamic panels with multiple factors.

over i. The observed panel data $\{y_{i,t} : t = 0, 1, ...T;\ i = 1, ..., n\}$ are defined in terms of $y^*_{i,t}$ as follows:

M1: $y_{i,t} = y^*_{i,t}$ for $t = 0, \cdots, T$ and $i = 1, \cdots, N$. and $\rho \in (-1, 1)$

M2: $y_{i,t} = \mu_i + y^*_{i,t}$ for $t = 0, \cdots, T$, $i = 1, \cdots, N$, $\mu_i \in R$ and $\rho \in (-1, 1]$

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

M2 $y_{i,t} = \underline{\mu}_i + \rho y_{i,t-1} + u_{it}$ for $t = 1, \cdots, T$, $i = 1, \cdots, N$, 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 $t = 1, \cdots, T$, $i = 1, \cdots, N$, 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 $y_{i,0} \sim N(0, \sigma_i^2/(1 - \rho^2))$ when $\rho \in (-1, 1)$ and $y_{i,0} = O_p(1)$ when $\rho = 1$.

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

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

The exact quantiles of $\hat{\rho}_{pols}$ 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).

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 $\hat{\rho}_{pols}$ 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 $\hat{\rho}_{pols}$ becomes tighter as N increases. In contrast to model M1, $\hat{\rho}_{pols}$ suffers from substantial SB in model M2 even when N or T are moderately large. But, as in Model M1, the distribution of $\hat{\rho}_{pols}$ 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 $\hat{\rho}_{pols}$ are 0.94, 0.89, 0.88 and 0.85 for N= 1, 10, 20, and 30, respectively when $\rho = 0.9$. This problem becomes more severe for model M3, where the upper 95% points of $\hat{\rho}_{pols}$ are 0.904, 0.843, 0.831 and 0.825 for N= 1, 10, 20, and 30.

Table 1: Downward Bias in Dynamic Panel Estimation

Part A: Quantiles of $\hat{\rho}_{pols}$ for $\rho = 0.9$

Sample	Model M1			Model M2			Model M3		
	5%	50%	95%	5%	50%	95%	5%	50%	95%
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 $\widehat{h}$ 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
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

The bias and concentration of the pooled estimator $\hat{\rho}_{pols}$ 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 $\hat{\rho}_{pols}$ are $\widehat{h} = \ln 0.5/\ln \hat{\rho}_{pols}$ and $\widehat{lrv} = 1/(1-\hat{\rho}_{pols})^2$. As is apparent from Tables 1(B) and 1(C), even a small SB can have large effects on these derived functions in the panel case because of the concentration of the estimate $\hat{\rho}_{pols}$ and the nonlinearity of the functions. As discussed in the last paragraph, the upper 95% point of the distribution of $\hat{\rho}_{pols}$ is smaller than ρ when N is moderately large, and then 95% of the distribution of $\widehat{h}$ 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 $\widehat{h}$ 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 0.89. Table 1(C) shows how 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 45% 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 panel median unbiased estimators. The first estimator is a panel exactly median unbiased (PEMU) estimator, constructed under the assumptions of a homogenous AR(1) parameter and cross sectional independence. This estimator is a panel version of Andrews' exactly median unbiased estimator in the time series case. It is of interest to see how this procedure is affected by panel observations. As mentioned in the introduction, Cermeno (1999) has independently proposed the use of a PEMU estimator for dynamic panel models with a common time effect and homogeneous trends and 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 common time effect) 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 common time effect. We will calibrate the performance of the new median unbiased estimators 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 median unbiased estimators 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 median unbiased estimators 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 1,2 and 3 satisfy such distributional invariance properties. These invariance results extend to the pooled panel forms of the least squares estimators in models 1,2 and 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 $\hat{\rho}_{pols}$ for each of the three models M1, M2, and M3; but when there is possible ambiguity, we use an additional subscript and write $\hat{\rho}_{polsj}$ for the POLS estimator of ρ in model j.

Invariance Property IP1: *Under the assumption of cross section independence, the distribution of $\hat{\rho}_{polsj}$ depends only on ρ when model j is correct and the error variance $\sigma_{ii} = \sigma^2$ for all i. When y_{it} is stationary, it does not depend on the common variance σ_{ii} for model M1, or (σ_{ii}, μ_i) for model M2, or $(\sigma_{ii}, \mu_i, \beta_i)$ 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 *iid* draws from a distribution with common scale. For example, if σ_i^2/σ^2 are *iid* 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 $\hat{\rho}_{pols}$ may then be rescaled by $1/\sigma^2$ and it is apparent that IP1 continues to hold, as shown in the Appendix. 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 $\hat{\rho}_{pols}$. This extension and other generalizations of $\hat{\rho}_{pols}$ that are better suited to empirical applications are discussed later. For the time being, we confine our discussion to the estimator $\hat{\rho}_{pols}$ 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 $\hat{\rho}_{pols}$ has a median function $m(\rho) = m_{T,N}(\rho)$ which simulation shows to be strictly increasing[2] in ρ on the parameter space

[2] 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

$\rho \in (-1, 1]$. Using this function (which depends on T and N), the panel median-unbiased estimator $\hat{\rho}_{pemu}$ can be defined as follows;

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

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 $\hat{\rho}_{pols}$. Define

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

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

Then, $\widehat{c}^{U}_{PU}$ and $\widehat{c}^{L}_{PU}$ provide upper and lower confidence limits and the $100(1-p)\%$ confidence interval for ρ is $\{\rho : \widehat{c}^{L}_{PU} \leq \rho \leq \widehat{c}^{U}_{PU}\}$. 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 $\hat{\rho}_{pols}$.

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 common time effect 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 common time effect that can impact individual series differently. Specifically, the model for the regression errors has the form

$$u_{it} = \delta_i \theta_t + \varepsilon_{it}, \quad \theta_t \sim iid\ N(0, 1) \text{ over } t, \tag{6}$$

in which θ_t is a common time effect, whose variance is normalized to be unity for identification purposes and whose coefficients, δ_i, may be regarded as 'idiosyncratic share' parameters that

values $T \geq 20$ and $N \geq 5$. 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.

measure the impact of the common time effect on series i. The δ_i are assumed to be non-stochastic and we let $\delta = (\delta_1, ..., \delta_N)$. In (6) the general error component ε_{it} is assumed to satisfy

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

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 δ. Setting $u_t = (u_{1t}, ..., u_{Nt})'$ we have the conditional covariance matrix

$$V_u = E\left(u_t u_t' | \sigma_1^2, ..., \sigma_N^2\right) = \Sigma + \delta\delta', \quad \Sigma = diag\left(\sigma_1^2, ..., \sigma_N^2\right). \tag{8}$$

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. 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 (2001) 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 δ.

As in the earlier case with cross sectional independence, it will be convenient in what follows to assume that the individual error variances σ_i^2 are iid draws from a distribution with common scale. More particularly, we assume that $\tau_i = \sigma_i^2/\sigma^2$ are iid draws from an independent distribution with density $f(\tau)$ that does not involve further nuisance parameters and whose first moment is finite. Then, the standardized error component

$$\frac{u_{it}}{\sigma} = \underline{\delta}_i \theta_t + \frac{\varepsilon_{it}}{\sigma_i}\frac{\sigma_i}{\sigma},$$

where $\underline{\delta}_i = \delta_i/\sigma$, has unconditional variance matrix

$$E\left(\frac{u_t u_t'}{\sigma^2}\right) = \int_0^\infty \left[\tau I + \underline{\delta\delta'}\right] f(\tau)\, d\tau = E(\tau)\, I_N + \underline{\delta\delta'}.$$

with $\underline{\delta} = \delta/\sigma$.

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

$$\hat{\rho}_{pgls} = \frac{\sum_{t=1}^{T} \widehat{y}_{t-1}' V_u^{-1} \widehat{y}_t}{\sum_{t=1}^{T} \widehat{y}_{t-1}' V_u^{-1} \widehat{y}_{t-1}}, \tag{9}$$

where $\widehat{y}_t = (\widehat{y}_{1t}, ..., \widehat{y}_{Nt})'$ and where $\widehat{y}_{it}$ 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 $\hat{\rho}_{polsj}$ depends only on $(\rho, \underline{\delta} = \delta/\sigma)$ when model j is correct and the error variance ratios $\tau_i = \sigma_i^2/\sigma^2$ are iid draws from an independent distribution with density $f(\tau)$ that does not involve further nuisance parameters. Further, the distribution of the panel GLS estimator $\hat{\rho}_{pgls}$ depends only on ρ when model j is correct. When $\rho = 1$ and y_{it} is non-stationary, the distributions of $\hat{\rho}_{pols}$ and $\hat{\rho}_{pgls}$ for models 2 and 3 do not depend on the value of y_{i0}.*

According to this proposition, the bias of $\hat{\rho}_{pols}$ depends on the nuisance parameters $\underline{\delta}_i$ that bring cross sectional dependence to the data. On the other hand, the panel GLS estimator depends only on ρ. Accordingly, we propose an iterative procedure that involves the use of a feasible GLS estimator, $\widehat{\rho}_{pfgls}$, whose form is specified below in (10). Our objective is to reduce the SB bias problems of these least squares procedures by constructing a feasible generalized version of the PMU estimator of ρ.

The first stage in this iteration uses the residuals from a panel regression in which we use our median unbiased estimator $\hat{\rho}_{pemu}$ rather than OLS to reduce the SB bias problems in this primary stage. Simulations we have conducted that are reported below (see Fig.2) indicate that 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 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 $\widehat{\rho}_{pfgls}$, which simulations show to be strictly increasing in ρ on the parameter space $\rho \in (-1, 1]$. Using this median function (which depends on T and N), the panel feasible generalized median-unbiased estimator, $\widehat{\rho}_{pfgmu}$, 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 $\hat{\rho}_{pemu}$ and using the residuals from this regression construct the error covariance matrix estimate $\widehat{V}_{pemu}$.

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

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

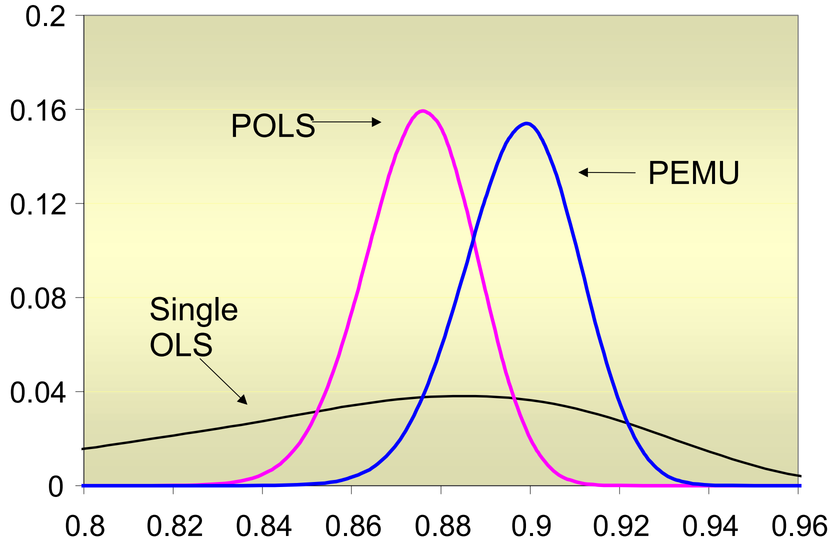

Figure 1: Empirical Distributions of Single OLS, POLS and PEMU under No Cross-sectional Dependence (N=20,T=100, and $\rho = 0.9$)

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

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

Fig. 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 $\hat{\rho}_{pols}$ is more concentrated than single equation OLS (which does not use the additional cross section data) but is badly biased biased downwards. The bias is sufficiently serious that almost the entire distribution of $\hat{\rho}_{pols}$ lies below the true value of ρ.

Fig. 2 shows the distributions of the POLS and PMU estimators for the same parameter configuration as Fig. 1 and based on the same number of replications, but with high cross sectional correlation [3]. As shown in Appendix A, the POLS bias in the case of cross section dependence always exceeds the bias in the cross section independent case. However, as is apparent from Fig. 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 0.9) 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.

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

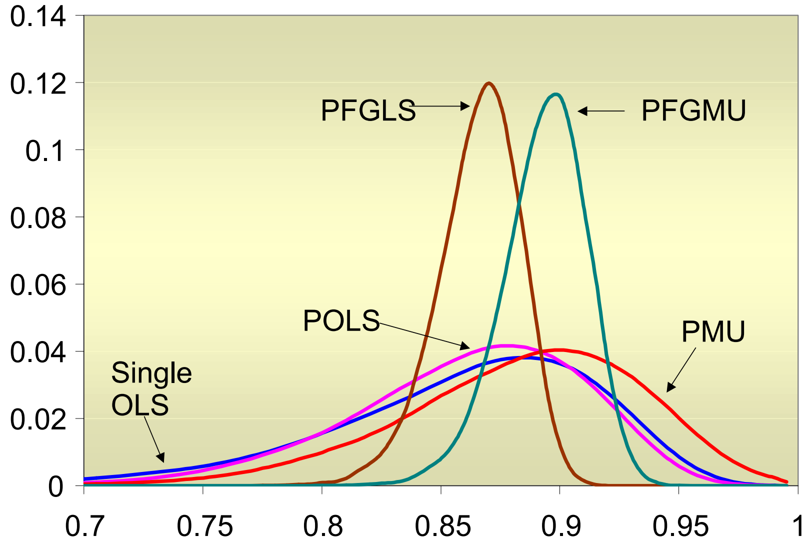

Figure 2: Empirical Distributions of POLS, PFGLS, and PFGMU under High Cross-sectional Dependence(N=10,T=100 and $\rho = 0.9$.)

Apparently, the PMU estimator still seems to be quite effective in removing the bias of POLS even under cross section dependence.

One might conjecture that, since the POLS and PMU estimators have little efficiency gain over OLS from pooling due to the presence of cross sectional dependence, there should be an advantage to the use of feasible GLS methods. In particular, one might expect the PFGLS estimator $\widehat{\rho}_{pfgls}$ to restore some of the advantages of pooling over single OLS. Fig. 2 also displays the distribution of $\widehat{\rho}_{pfgls}$. Evidently, PFGLS does restore much of the original gains from pooling in terms of variance reduction that were apparent in Fig. 1 for $\hat{\rho}_{pols}$. But, as is also apparent from Fig. 2, the distribution of $\widehat{\rho}_{pfgls}$ is seriously downward biased. Use of the PFGMU median unbiased procedure described above now corrects for this bias while retaining the concentration of the GLS estimator. In particular, Fig. 2 shows that the distribution of $\widehat{\rho}_{pfgmu}$ is well centered about the true value and has concentration close to that of the median unbiased estimator $\hat{\rho}_{pemu}$ under cross sectional independence (Fig. 1).

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 SB 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 SB 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}^* :

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

in which the regression errors

$$u_t \sim iid\ N(0, V_u), \qquad \text{for } t = 1, \cdots, T, \tag{12}$$

where $u_t = (u_{it,} ..., u_{Nt})'$. 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

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

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 $\hat{\rho}_{iemu}$ for each series $i = 1, ..., N$ (and the appropriate model) and use the regression residuals to construct the error covariance matrix estimate $\widehat{V}_{EMU}$.

Step 2: Using $\widehat{V}_{EMU}$ perform a conventional seemingly unrelated regression on the panel and obtain the SUR estimates of the ρ_i, $\widehat{\rho}_{isur}$.

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

Step 4: Repeat steps 1-3 until $\widehat{\rho}_{isurmu}$ converges.

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 nonstationary cases and shows how cross section dependencies affect the asymptotic theory under nonstationarity. We then propose an alternative Wald procedure for testing homogeneity that utilizes the structure of the cross section dependence in the construction of the Wald statistic.

4.1 The Wald Test and its Asymptotic Properties

The Stationary Case

Using the unrestricted estimates $\widehat{\rho}_{isurmu}$ 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 $\underline{\widehat{\rho}}_{surmu} = (\widehat{\rho}_{isurmu})$ be the SUR-MU estimate of the vector $\underline{\rho} = (\rho_1, ..., \rho_N)'$ and write the restrictions in H_0 as $D\underline{\rho} = 0$ where $D = [i_{N-1}, -I_{N-1}]$ and i_A has A unit elements. Under Gaussianity and in the stationary case where $|\rho_i| < 1$ for all i, the SUR-MU estimator $\underline{\widehat{\rho}}_{surmu}$ is asymptotically ($T \to \infty$, N fixed) equivalent to the unconstrained maximum likelihood estimate[4] of $\underline{\rho}$. In that case, standard stationary asymptotics and some algebraic manipulations (outlined in Appendix C) lead to the limit theory

$$\sqrt{T}\left(\underline{\widehat{\rho}}_{surmu} - \underline{\rho}\right) \to_d N\left(0, V_{SUR}\right), \tag{14}$$

where

$$V_{SUR}^{-1} = \left[\left(v_u^{ij} E\left(y_{it} y_{jt}\right)\right)_{ij}\right] = V_u^{-1} * E\left(y_t y_t'\right). \tag{15}$$

In (15) the operator $*$ is the Hadamard product, v_u^{ij} is the ij'th element of V_u^{-1}, where $V_u = E(u_t u_t') = \Sigma + \delta\delta'$ as in (8), and

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

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

so that

$$E\left(y_t y_t'\right) = \left(\Sigma + \delta\delta'\right) * R, \quad \text{where } R = (r_{ij}) \text{ and } r_{ij} = \frac{1}{1 - \rho_i \rho_j}. \tag{16}$$

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

$$V_{SUR}^{-1} = \frac{1}{1 - \rho^2} V_u^{-1} * V_u, \tag{17}$$

which depends on the common ρ and again on the cross section dependence parameter δ.

The Wald statistic for testing H_0 is

$$W_{surmu} = \widehat{\underline{\rho}}_{surmu}' D' \left[D \widehat{V}_{SURMU} D' \right]^{-1} D \widehat{\underline{\rho}}_{surmu},$$

where

$$\widehat{V}_{SURMU} = \left[\sum_{t=1}^{T} Z_t' \widehat{V}_u^{-1} Z_t \right]^{-1},$$

in which $Z_t = diag(y_{1t-1}, ..., y_{Nt-1})$ and $\widehat{V}_u$ 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, it is straightforward to show that traditional chi-squared limit theory for W_{surmu} holds, i.e. $W_{surmu} \rightarrow \chi_N^2$.

The Unit Root Case

In the nonstationary $\rho = 1$ 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 - c.f. (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 C) using unit root limit theory deliver the limit distribution of the SUR estimator $\widehat{\underline{\rho}}_{sur}$. This estimator is defined as

$$\widehat{\underline{\rho}}_{sur} = \left(\sum_{t=1}^{T} Z_t' \widehat{V}_u^{-1} Z_t \right)^{-1} \left(\sum_{t=1}^{T} Z_t' \widehat{V}_u^{-1} y_t \right),$$

where $\widehat{V}_u$ is an estimate of V_u based on residuals from a first stage regression. We find the following asymptotic distribution for $\widehat{\underline{\rho}}_{sur}$

$$T\left(\widehat{\underline{\rho}}_{sur} - i_N\right) \xrightarrow{d} \left[V_u^{-1} * \int_0^1 BB' \right]^{-1} \left[\int_0^1 B * \left(V_u^{-1} dB \right) \right] = \xi, \tag{18}$$

where B is vector Brownian motion with covariance matrix V_u. It is clear from (**??**) that the limit distribution of $T\left(\widehat{\underline{\rho}}_{SUR} - i_N\right)$ depends on the cross section dependence parameter δ even in the homogeneous case where $\rho_i = 1 \ \forall i$. Correspondingly, the asymptotic distribution of $\widehat{\rho}_{surmu}$ 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

$$\begin{aligned} W_{sur} &= \widehat{\underline{\rho}}_{SUR}' D' \left[D \widehat{V}_{SUR} D' \right]^{-1} D \underline{\widehat{\rho}}_{SUR} \\ &\xrightarrow{d} \xi' D' \left[D \left(V_u^{-1} * \int_0^1 BB' \right)^{-1} D' \right]^{-1} D\xi. \end{aligned} \tag{19}$$

where

$$\widehat{V}_{SUR} = \left(\sum_{t=1}^{T} Z_t' \widehat{V}_u^{-1} Z_t \right)^{-1}.$$

The limit distribution (19) also depends on nuisance parameters.

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

$$\widehat{\rho} = \left(\sum_{t=1}^{T} y_{t-1}' V_u^{-1} y_{t-1} \right)^{-1} \left(\sum_{t=1}^{T} y_{t-1}' V_u^{-1} y_t \right),$$

with a corresponding feasible SUR version. By straightforward derivation (see Appendix C), we find that

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

where $W = (W_i)$ is standard Brownian motion with covariance matrix 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 above, 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 alternate 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

$$y_t = Z_t \underline{\rho} + u_t, \quad Z_t = diag\left(y_{1t-1}, ..., y_{Nt-1}\right), \quad \underline{\rho} = \left(\rho_1, ..., \rho_N\right)'. \tag{21}$$

Let $\hat{\rho}_i$ (respectively $\underline{\widehat{\rho}}$) be the OLS estimate of ρ_i ($\underline{\rho}$) Then

$$\underline{\widehat{\rho}} = \left(\sum_{t=1}^{T} Z_t' Z_t \right)^{-1} \left(\sum_{t=1}^{T} Z_t' y_t \right).$$

Let $\widehat{\underline{\rho}}_{emu}$ be the corresponding vector of median unbiased estimates of ρ_i. Under the null hypothesis of homogenous autoregressive coefficients $\rho_i = \rho \, \forall i$, 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 median unbiased estimators $\widehat{\rho}_{iemu}$. Under cross section independence and 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

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

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

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

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

$$\begin{aligned}
\sqrt{T}\left(\underline{\hat{\rho}} - \rho i_N\right) &= \left(\frac{1}{T}\sum_{t=1}^{T} Z_t' Z_t\right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t' u_t\right) \\
&\to_d N\left(0, D_y^{-1}\left[V_u * E\left(y_t y_t'\right)\right] D_y^{-1}\right) \\
&= N(\left(0, \left(1-\rho^2\right) R_V * R_V\right), \qquad (22)
\end{aligned}$$

where $D_y = diag(E(y_{1t}^2), ..., E(y_{Nt}^2))$ and the matrix R_V has ij'th element $v_{ij}/\{v_{ii}v_{jj}\}^{1/2}$. It follows that

$$\sum_{i=1}^{N} \frac{\sqrt{T}(\hat{\rho}_i - \rho)}{\sqrt{1-\rho^2}} \to_d N(0, i_N' \left(R_V * R_V\right) i_N).$$

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

We propose to construct an estimate of the matrix R_V that appears in the asymptotic covariance matrix of (22) and use this estimate to develop an alternate test of H_0. The following moment based procedure may be used[5].

Moment Based Estimation of (δ, Σ)

[5] Apendix D gives an 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.

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

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

in both stationary and nonstationary cases.

Step 2: Construct the moment matrix of residuals $M_T = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$, which is a consistent (as $T \to \infty$, N fixed) estimate of V_u. Let m_{Tij} be the ij'th element of M_T.

Step 3: Estimate the cross section coefficients δ and the diagonal elements of Σ using the following moment procedure that finds the least squares best fit to the matrix M_T, that is

$$\left(\widehat{\delta}, \widehat{\Sigma}\right) = \arg \min_{\delta, \Sigma} tr \left[\left(M_T - \Sigma - \delta\delta' \right) \left(M_T - \Sigma - \delta\delta' \right)' \right]. \tag{23}$$

The solution of (23) satisfies the system of equations

$$\hat{\delta} = (M_T \hat{\delta} - \Sigma \hat{\delta}) / \hat{\delta}' \hat{\delta}, \quad \hat{\sigma}_i^2 = M_{Tii} - \hat{\delta}_i^2, \ i = 1, ..., N$$

and this can be solved using the iteration

$$\begin{aligned} \delta^{(r)} &= (M_T \delta^{(r-1)} - \Sigma \delta^{(r-1)}) / \delta^{(r-1)\prime} \delta^{(r-1)}, \\ \sigma_i^{(r)2} &= M_{Tii} - \delta_i^{(r)2}, \end{aligned} \tag{24}$$

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

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

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

$$\sqrt{T} \left(\widetilde{\rho} - \rho \right) = \left(\frac{1}{T} \sum_{t=1}^{T} y_{t-1}' V_u^{-1} y_{t-1} \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t-1}' V_u^{-1} u_t \right) \to_d N \left(0, \left\{ \text{trace} \left[V_u^{-1} E \left(y_t y_t' \right) \right] \right\}^{-1} \right).$$

Since

$$E \left(y_t y_t' \right) = \left(\Sigma + \sigma^2 \delta\delta' \right) * R = V_u * R = \frac{1}{1 - \rho^2} V_u,$$

under H_0, we end up with the simple result

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

Next consider the asymptotic covariance

$$\begin{aligned}
& \text{Acov}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t' u_t, \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}' V_u^{-1} u_t\right) \\
= \; & \frac{1}{T}\sum_{t=1}^{T} Z_t' E\left(u_t u_t'\right) V_u^{-1} y_{t-1} = \frac{1}{T}\sum_{t=1}^{T} Z_t' y_{t-1} \rightarrow \begin{bmatrix} E\left(y_{1t}^2\right) \\ E\left(y_{2t}^2\right) \\ \vdots \\ E\left(y_{Nt}^2\right) \end{bmatrix} = D_y i_N,
\end{aligned}$$

from which we deduce that

$$\begin{aligned}
& \text{Acov}\left(\sqrt{T}\left(\underline{\hat{\rho}} - \rho i_N\right), \sqrt{T}\left(\widetilde{\rho} - \rho\right)\right) \\
= \; & D_y^{-1}\left[D_y i_N\right]\left\{\text{trace}\left[V_u^{-1} E\left(y_t y_t'\right)\right]\right\}^{-1} \\
= \; & i_N\left(1 - \rho^2\right). \qquad (25)
\end{aligned}$$

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

$$\sqrt{T}\left(\underline{\hat{\rho}}_{emu} - \widetilde{\rho} i_N\right) = \sqrt{T}\left(\underline{\hat{\rho}}_{emu} - \rho i_N\right) - \sqrt{T}\left(\widetilde{\rho} - \rho\right) i_N,$$

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

$$\frac{\sqrt{T}\left(\underline{\hat{\rho}}_{emu} - \widetilde{\rho} i_N\right)}{\sqrt{1 - \widetilde{\rho}^2}} = \frac{\sqrt{T}\left(\underline{\hat{\rho}}_{emu} - \rho i_N\right)}{\sqrt{1 - \widetilde{\rho}^2}} - \frac{\sqrt{T}\left(\widetilde{\rho} - \rho\right)}{\sqrt{1 - \widetilde{\rho}^2}} i_N \rightarrow_d N(\left(0, R_V * R_V - \frac{1}{N} i_N i_N'\right). \qquad (26)$$

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

$$G = \frac{T}{1 - \widetilde{\rho}^2}\left(\underline{\hat{\rho}}_{emu} - \widetilde{\rho} i_N\right)'\left\{\left[\widehat{R}_V * \widehat{R}_V\right]^{-1} - \frac{1}{N} i_N i_N'\right\}\left(\underline{\hat{\rho}}_{emu} - \widetilde{\rho} i_N\right), \qquad (27)$$

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

$$G_{pfmgu} \rightarrow \chi_N^2, \text{ as } T \rightarrow \infty. \qquad (28)$$

One practical difficulty that can arise with (27) is that the variance matrix $\left[\widehat{R}_V * \widehat{R}_V\right]^{-1} - \frac{1}{N} i_N i_N'$ 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: $H_o : \rho_i = 1, \forall i$

As shown in Appendix C, 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 (62)

$$\frac{1}{\sqrt{T}}y_{[Tr]} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} u_t \rightarrow_d B(r) = BM(V_u).$$

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

$$B(r) = \delta B_\theta(r) + B_\varepsilon(r), \tag{29}$$

where

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \theta_t \rightarrow_d B_\theta(r) = BM\left(\sigma^2\right), \text{ and } \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \varepsilon_t \rightarrow_d B_\varepsilon(r) = BM(\Sigma).$$

Let $\delta_\perp$ be an $N \times (N-1)$ matrix that spans the orthogonal complement of the vector δ. Then

$$\left[(\delta_\perp' \Sigma \delta_\perp)^{-1/2} \delta_\perp'\right] \frac{1}{\sqrt{T}} y_{[Tr]} \rightarrow_d (\delta_\perp' \Sigma \delta_\perp)^{-1/2} \delta_\perp' B(r) = (\delta_\perp' \Sigma \delta_\perp)^{-1/2} \delta_\perp' B_\varepsilon(r) = W_\perp(r), \tag{30}$$

where $W_\perp(r) = BM(I_{N-1})$, or $(N-1)$ - vector standard Brownian motion. The transformation matrix that appears in (30) 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 $M_T = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$. As in the stationary case, M_T is a consistent (as $T \to \infty$, N fixed) estimate of V_u. Again, let m_{Tij} be the ij'th element of M_T.

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

Step 4: Using $\widehat{\Sigma}$ and $\widehat{\delta}$, construct[6] $\widehat{\delta}_\perp$ and $\widehat{F}_\delta = \left(\widehat{\delta}_\perp' \widehat{\Sigma} \widehat{\delta}_\perp\right)^{-1/2} \widehat{\delta}_\perp'$. Clearly,

$$\widehat{F}_\delta = \left(\widehat{\delta}_\perp' \widehat{\Sigma} \widehat{\delta}_\perp\right)^{-1/2} \widehat{\delta}_\perp' \rightarrow_p (\delta_\perp' \Sigma \delta_\perp)^{-1/2} \delta_\perp', \tag{31}$$

as $T \to \infty$.

Using $\widehat{F}_\delta$ we transform the data y_t (or demeaned/detrended data in the case of models M2 and M3) giving $y_t^+ = \widehat{F}_\delta y_t$. As is apparent from (30), the transformation $\widehat{F}_\delta$ asymptotically

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

removes cross section dependence in the panel and y_t^+ is asymptotically cross section independent as $T \to \infty$. Using y_t^+ we may now construct estimates of the autoregressive coefficients. Let $\hat{\rho}_i^+$ (respectively $\widehat{\underline{\rho}}^+$) be the OLS estimate of $\rho_i = 1$ ($\underline{\rho} = i_{N-1}$). Then, in an obvious notation,

$$\widehat{\underline{\rho}}^+ = \left(\sum_{t=1}^{T} Z_t^{+\prime} Z_t^+ \right)^{-1} \left(\sum_{t=1}^{T} Z_t^{+\prime} y_t^+ \right).$$

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

$$G_H^+ = T^2 \left(\hat{\underline{\rho}}_{emu}^+ - \widetilde{\rho}^+ i_{N-1} \right)' \left(\hat{\underline{\rho}}_{emu}^+ - \widetilde{\rho}^+ i_{N-1} \right), \tag{32}$$

As shown in Appendix C

$$G_H^+ \to_d \Xi_{N-1}' \Xi_{N-1} \tag{33}$$

where

$$\Xi_{N-1} = \begin{bmatrix} \left[\int_0^1 W_{\perp,1}^2\right]^{-1} \left[\int_0^1 W_{\perp,1} dW_{\perp,1}\right] - \left[\int_0^1 W_\perp' W_\perp\right]^{-1} \left[\int_0^1 W_\perp{}' dW_\perp\right] \\ \vdots \\ \left[\int_0^1 W_{\perp,N-1}^2\right]^{-1} \left[\int_0^1 W_{\perp,N-1} dW_{\perp,N-1}\right] - \left[\int_0^1 W_\perp' W_\perp\right]^{-1} \left[\int_0^1 W_\perp{}' dW_\perp\right] \end{bmatrix}, \tag{34}$$

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

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

$$G_{ols}^+ = \sum_{i=1}^{N-1} \frac{\hat{\rho}_i^+ - i_{N-1}}{\hat{\sigma}_{\hat{\rho}^+}} \tag{35}$$

$$G_{emu}^+ = \sum_{i=1}^{N-1} \frac{\hat{\rho}_{i,emu}^+ - i_{N-1}}{\hat{\sigma}_{\hat{\rho}_{i,emu}^+}} \tag{36}$$

In contrast to (32), the test statistics (35) and (36) do not involve a pooled estimate of the homogeneous unit root parameter. As shown in Appendix C, for fixed N we have the following limit theory for these statistics as $T \to \infty$

$$G_{ols}^+, \to_d \sum_{i=1}^{N-1} \xi_i, \quad , G_{emu}^+ \to_d \sum_{i=1}^{N-1} \xi_i^- \tag{37}$$

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 (37) 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 (34) - (36). 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 (35) and (36) have the advantage that they lend themselves to simple large N asymptotics. In particular, the means and variances

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

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

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

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

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

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 p_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

$$\Delta y_{it} = \underline{\mu}_i + \underline{\beta}_i t + (\rho - 1) y_{it-1} + \sum_{j=1}^{p_i} \phi_{ij} \Delta y_{it-j} + u_{it}. \tag{38}$$

The OP procedure leading to (31) 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 (38).

Since the transformed data y_{it}^{+} are asymptotically uncorrelated across i, regressions like (38) 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 (31) continues to eliminate the contribution of the common factor component θ_t to the limit Brownian motion in (29). 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.[7] 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

$$P = -2 \sum_{i=1}^{N-1} \ln(p_i), \tag{39}$$

$$P_m = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} [\ln(p_i) + 1] \tag{40}$$

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} \Phi^{-1}(p_i) \tag{41}$$

The P test is called the inverse chi-square test or Fisher test after Fisher (1932). The P_m test statistic is a centered 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 (2001), we have the following limit distributions for P and Z as $T \to \infty$

$$P \to_d \chi^2_{2(N-1)}, \quad Z \to_d N(0,1) \quad \text{for fixed } N, \tag{42}$$

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

$$P_m, Z \to_d N(0,1). \tag{43}$$

Each of these tests and the limit theory applies under the null hypothesis to dynamic panel autoregressions like (38) 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

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

(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 present paper. For instance, we could use an ADF-t statistic based not on OLS estimation but instead on EMU estimation as explained earlier (c.f. Andrews and Chen, 1994).

Im, Pesaran and Shin (1997, IPS) 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 SB 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 Model 1 and 2. For example, for an AR(2), the OLS estimator of ρ_i in (38) 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 heavily rely 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 6 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

$$y_{it} = \rho_i y_{it-1} + u_{it}, \tag{44}$$

$$u_{it} = \delta_i\theta_t + \varepsilon_{it}, \tag{45}$$

where $\varepsilon_{it} \sim iid\, N(0,1)$ over i and t, $\theta_t \sim iid\, N(0,1)$ over t, and for (ρ_i, δ_i) parameter selections that are detailed below. The primary distinction is between the homogeneous case where $\rho_i = \rho$

for all i and the heterogeneous 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:

$$\begin{array}{ll} y_{it} = a_i + \rho_i y_{it-1} + u_{it} & \text{for Model M2} \\ y_{it} = a_i + b_i t + \rho_i y_{it-1} + u_{it} & \text{for Model M3} \end{array}$$

Panel data are generated under four specifications which differ according to their 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

$$\begin{array}{rcll} y_{i,t} & = & \rho_i y_{i,t-1} + v_{i,t}, & \\ v_{it} & = & \phi_i v_{it-1} + u_{it} & \text{AR(1) errors} \quad (46) \\ v_{it} & = & \kappa_i u_{it-1} + u_{it} & \text{MA(1) errors} \quad (47) \\ u_{it} & = & \delta_i \theta_t + \varepsilon_{it}, & \end{array}$$

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 (46) and (47) are set as follows:

$\phi_i \sim U[0, 0.4]$	AR(1) errors
$\kappa_i \sim U[0, 0.4]$	MA(1) errors, $\kappa_i > 0$
$\kappa_i \sim U[-0.4, 0]$	MA(1) errors, $\kappa_i < 0$

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 MSE's 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 MSE's that are 5 to 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 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 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, 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 with negative coefficients but these are not reported here. An alternative approach to removing serial dependence, such as the nonparametric 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. These models seem particularly well suited to studying issues such as the potential for growth convergence across economic regions and nation states. By providing a mechanism for pooling time series information across section, they also offer the opportunity for improved estimation of such quantities as the half-life of dynamic response times and long run variance parameters. In all these cases, there is substantial interest in the estimation of autoregressive coefficients. The bias in the estimation of these parameters from least squares regression has long been recognised as a problem in time series analysis and in dynamic panels with fixed effects. The present paper shows that this bias problem is substantially exacerbated when there is cross section dependence. It is, in fact, so serious that the empirical distribution of pooled least squares estimates all but excudes the true autoregressive coefficient in many cases. Hence, there is an urgent need for corrective action that allows for models and data sets where there is cross sectional dependence.

The solutions offered in the present paper largely involve the use of median unbiased estimation procedures for estimation, testing and confidence interval construction. On the whole, the new estimation methods work very well in correcting for bias and accounting for cross section dependence in conditions (viz. correct specification and Gaussianity) that might be described as 'ideal' for these methods. On the other hand, Wald tests for homogeneity (just like those that are based on least squares procedures in conventional regression models) show evidence of unacceptable size distortion even under ideal conditions, including stationarity. We therefore propose a modfied Hausman test for homogeneity that utilizes a pooled panel MUE estimator that is asymptotically efficient under the null, in conjuction with MUE estimates that are robust to heterogeneity and moment based estimates of the cross section dependence parameters. Simulations indicate that this homogeneity test, whose limit distribution is chi-squared, works very well except in cases where N and T are both small.

In the important case of tests for homogeneous 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, Pesaran and Shin, 1997) suffering serious size distortion. The p-value based meta Z test of Choi (2001) is found to work particularly well with stable size and good power and is easy to compute and apply in practice. Moon and Perron (2001) 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 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

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 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

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).

	Size							
	Model 2				Model 3			
Sample	1%	2.5%	5%	10%	1%	2.5%	5%	10%
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							
	Model 2				Model 3			
Sample	1%	2.5%	5%	10%	1%	2.5%	5%	10%
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

Table 6: Tests for Homogeneous Panel Unit Roots under Cross-Section Dependence (Case VI): Cross-Sectional Correlation (min= 0.52, med= 0.82, max=0.94) -

Panel A: Model M2 - Fitted Intercept Case					
			Size: 5%		
Sample	IPS	G^{++}_{ols}	G^{++}_{emu}	P	Z
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			
Sample	IPS	G^{++}_{ols}	G^{++}_{emu}	P	Z
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%		
Sample	IPS	G_{ols}^{++}	G_{emu}^{++}	P	Z
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		
Sample	IPS	G_{ols}^{++}	G_{emu}^{++}	P	Z
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
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 Homogeneous Panel Unit Roots under Cross-Section Dependence & AR(1) Errors (Case VII). Cross-Sectional Correlation (min= 0.52, med= 0.82, max=0.94)-

Panel A: Fitted Intercept						
			Size			
Sample	IPS		*P*		*Z*	
	5%	10%	5%	10%	5%	10%
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			
Sample	IPS		*P*		*Z*	
	5%	10%	5%	10%	5%	10%
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

Panel B: Fitted Intercept and Trend						
	Size					
Sample	IPS		P		Z	
	5%	10%	5%	10%	5%	10%
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					
Sample	IPS		P		Z	
	5%	10%	5%	10%	5%	10%
N=10,T= 50	0.168	0.259	0.976	0.987	0.973	0.985
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

Table 8: Tests for Homogeneous Panel Unit Roots under Cross-Section Dependence & MA(1) Errors (Case VII). Cross-Sectional Correlation (min= 0.52, med= 0.82, max=0.94)-

Panel A: Fitted Intercept						
			Size			
Sample	IPS		P		Z	
	5%	10%	5%	10%	5%	10%
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			
Sample	IPS		P		Z	
	5%	10%	5%	10%	5%	10%
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					
Sample	IPS		P		Z	
	5%	10%	5%	10%	5%	10%
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					
Sample	IPS		P		Z	
	5%	10%	5%	10%	5%	10%
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
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
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

7 Appendix A

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

$$\hat{\rho}_{pols2} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - y_{i.-1})(y_{it} - y_{i.})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - y_{i.-1})^2}, \tag{48}$$

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 $t = 1, .., T$, the residuals are linear functions of the y_{it}^* and are invariant to (μ_i, β_i). Let Q_t be the orthogonal projection matrix onto the othogonal complement of the space spanned by the matrix $X = [x_1, ..., x_T]'$ and let $y_i = (y_{i1}, ..., y_{iT})'$, $y_{i,-1} = (y_{i0}, ..., y_{iT-1})'$, with a corresponding notation for y_i^* and $y_{i,-1}^*$. The residual vectors from these detrending regressions are

$$\widehat{y}_i = Q_t y_i = Q_t y_i^* = \widehat{y}_i^*,$$

and

$$\widehat{y}_{i,-1} = Q_t y_{i,-1} = Q_t y_{i,-1}^* = \widehat{y}_i^*.$$

The POLS estimator in Model M3 is

$$\begin{aligned} \hat{\rho}_{pols3} &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{y}_{it-1}\widehat{y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{y}_{it-1}^2} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{y}_{it-1}^*\widehat{y}_{it}^*}{\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{y}_{it-1}^{*2}} \\ &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\widehat{y}_{it-1}^*\widehat{y}_{it}^*}{\sigma_i^2}\frac{\sigma_i^2}{\sigma^2}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\widehat{y}_{it-1}^{*2}}{\sigma_i^2}\frac{\sigma_i^2}{\sigma^2}}, \end{aligned} \tag{49}$$

and invariance to $(\mu_i, \beta_i, \sigma^2)$ is clear. Proofs for the nonstationary 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 $(\mu_i, \beta_i, \sigma^2)$ follows precisely as in the proof of Property IP1. From (1) and (6) we have, in the stationary case,

$$\begin{aligned} \frac{y_{it}^*}{\sigma} &= \sum_{j=0}^{\infty}\rho^j\frac{u_{i,t-j}}{\sigma} = \sum_{j=0}^{\infty}\rho^j\left[\frac{\delta_i}{\sigma}\theta_{t-j} + \frac{\varepsilon_{it-j}}{\sigma_i}\frac{\sigma_i}{\sigma}\right] \\ &= \underline{\delta}_i\sum_{j=0}^{\infty}\rho^j\theta_{t-j} + \sum_{j=0}^{\infty}\rho^j\frac{\varepsilon_{it-j}}{\sigma_i}\frac{\sigma_i}{\sigma} \\ &\sim \int_0^{\infty} N\left(0, \frac{\tau + \underline{\delta}_i^2}{1-\rho^2}\right) f(\tau)\, d\tau, \end{aligned}$$

with a similar expression for y_{it}^*/σ, so that both depend only on (ρ, δ_i). It follows that the POLS estimator depends on $(\rho, \underline{\delta}_1, ..., \underline{\delta}_N)$ in each of the Models 1-3.

Next, let $y_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$, $\widehat{y}_t^* = (\widehat{y}_{1t}^*, ..., \widehat{y}_{Nt}^*)'$ where $\widehat{y}_{it}^*$ denotes y_{1t}^* or demeaned or detrended y_{it}^*, respectively for Models M1,M2 and M3, with corresponding notation for y_t and $\widehat{y}_t$. Let $D_\tau = diag\left(\tau_1, ..., \tau_N\right)$ and let Ω be the matrix whose ij'th element is $\rho^{|i-j|}/(1-\rho^2)$. Note that

$$\frac{y_t^*}{\sigma}, \frac{y_{t-1}^*}{\sigma} \sim \int_0^\infty ... \int_0^\infty N\left(0, \frac{1}{1-\rho^2}\left[D_\tau + \underline{\delta\delta'}\right]\right) f\left(\tau_1\right) ... f\left(\tau_N\right) d\tau_1 ... d\tau_N,$$

and, vectorizing $Y^*/\sigma = \left[y_1^*, ..., y_T^*\right]/\sigma$ by columns, we have

$$vec\left(Y^*/\sigma\right) \sim \int_0^\infty .. \int_0^\infty N\left(0, \Omega \otimes \left[D_\tau + \underline{\delta\delta'}\right]\right) f\left(\tau_1\right) ... f\left(\tau_N\right) d\tau_1 ... d\tau_N,$$

which depends on (ρ, δ). Now $\widehat{Y} = YQ_t = Y^*Q_t$ and $vec(\widehat{Y}) = (Q_t \otimes I)vec(Y)$ and so

$$vec\left(\widehat{Y}/\sigma\right) \sim \int_0^\infty .. \int_0^\infty N\left(0, Q_t\Omega Q_t \otimes \left[D_\tau + \underline{\delta\delta'}\right]\right) f\left(\tau_1\right) ... f\left(\tau_N\right) d\tau_1 ... d\tau_N.$$

Similarly, if $u_t = (u_{1t}, ..., u_{Nt})'$ and $U = \left[u_1, ..., u_T\right]$, we have

$$vec\left(U/\sigma\right) \sim \int_0^\infty N\left(0, I \otimes \left[\tau I + \underline{\delta\delta'}\right]\right) f\left(\tau\right) d\tau.$$

As in (8), set

$$V = E\left(u_t u_t' | \sigma_1^2, ..., \sigma_N^2\right) = \Sigma + \delta\delta' = \sigma^2\left[D_\tau + \underline{\delta\delta'}\right] := \sigma^2 V_\tau.$$

The GLS estimator of ρ then has the form

$$\hat{\rho}_{pgls} = \frac{\sum_{t=1}^T \widehat{y}_{t-1}' V^{-1} \widehat{y}_t}{\sum_{t=1}^T \widehat{y}_{t-1}' V^{-1} \widehat{y}_{t-1}} = \frac{\frac{1}{\sigma^2}\sum_{t=1}^T \widehat{y}_{t-1}^{*\prime} V_\tau^{-1} \widehat{y}_t^*}{\frac{1}{\sigma^2}\sum_{t=1}^T \widehat{y}_{t-1}^{*\prime} V_\tau^{-1} \widehat{y}_{t-1}^*},$$

which depends only on ρ.

Again, proofs in the nonstationary 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.

8 Appendix B

Extensions of the Nickell Bias Formula

This section provides some analytic extensions of the Nickell (1981) bias formula to cases where there is error heterogeneity, cross section dependence and unit root dynamics.

Stationary, Cross Section Independent Case First consider the homogeneous stationary case where $\rho_i = \rho$ and $|\rho| < 1$ and where there is no cross section dependence. It is not necessary to assume that the errors $u_{it} = \varepsilon_{it}$ are $iid(0, \sigma^2)$. Instead, we assume that the ε_{it} have zero mean, finite $2 + 2\nu$ moments for some $\nu > 0$ and are independent over i and t with $E(\varepsilon_{it}^2) = \sigma_{ii}$ for all t. We also assume that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \sigma_{ii} = \bar{\sigma}. \tag{50}$$

Then, the $\varepsilon_{it}^2 - \sigma_{ii}$ are independent with mean zero and finite $1+\nu$ moments, and we have by standard theory (e.g., the Markov law)

$$\frac{1}{N}\sum_{i=1}^{N}\left(\varepsilon_{it}^2 - \sigma_i^2\right) \to_p 0, \quad \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}^2 = \sigma^2.$$

We develop a formula for the bias in large cross section (N) asymptotics, as in Nickell (1981), for the pooled least squares estimate. To illustrate, we work with model M2 where the estimate has the form

$$\begin{aligned} \hat{\rho} &= \rho + \frac{\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it-1} - y_{i\cdot-1})(u_{it} - u_{i\cdot})}{\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it-1} - y_{i\cdot-1})^2} \\ &= \rho + \frac{A_{NT}}{B_{NT}} = \rho + \frac{\frac{1}{N}A_{NT}}{\frac{1}{N}B_{NT}}, \end{aligned} \tag{51}$$

where $y_{i\cdot-1} = \frac{1}{T}\sum_{t=1}^{T} y_{it-1}$, and $u_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} u_{it}$. Without loss of generality, set $\mu = 0$ in M2 so that $y_{it} = \sum_{j=0}^{\infty}\rho^j u_{it-j}$. Some calculations analogous to those in Nickell (1981) show that the probability limits of the numerator and denominator in (51) as $N \to \infty$ with T fixed are

$$\text{plim}_{N\to\infty}\frac{1}{N}A_{NT} = -\frac{\sigma^2}{T}\frac{1}{1-\rho}\left[T - \frac{1-\rho^T}{1-\rho}\right], \tag{52}$$

and

$$\text{plim}_{N\to\infty}\frac{1}{N}B_{NT} = \sigma^2\frac{T-1}{1-\rho^2}\left\{1 - \frac{1}{T-1}\frac{2\rho}{1-\rho}\left[1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right]\right\}. \tag{53}$$

Combining (52) and (53) we get

$$\begin{aligned} \text{plim}_{N\to\infty}\frac{\frac{1}{N}A_{NT}}{\frac{1}{N}B_{NT}} &= -\frac{1+\rho}{T-1}\left[1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right]\left\{1 - \frac{1}{T-1}\frac{2\rho}{1-\rho}\left[1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right]\right\}^{-1} \\ &= -\frac{1+\rho}{T-1} + O\left(\frac{1}{T^2}\right), \end{aligned}$$

the formula given in Nickell (1981) for the homogeneous case.

Unit root, Cross Section Independent Case The case $\rho = 1$ can be handled in the same way, although this case was not considered by Nickell (1981). Using E_i to signify expectations with respect to the probability measure for individual i, and setting $y_{i0} = u_{i0} = 0$, we get:

$$\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}A_{NT} &= -\frac{1}{T}E_i\left[\sum_{t=1}^{T}(\sum_{j=0}^{t-1}u_{it-j-1})\sum_{t=1}^{T}u_{it}\right] \\
&= -\frac{\sigma^2}{T}\left[1+2+3+\cdots+T-1\right] \\
&= -\frac{\sigma^2}{T}\sum_{i=1}^{T-1}t = -\frac{\sigma^2}{2}(T-1),
\end{aligned}$$

and

$$\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}B_{NT} &= E_i\sum_{t=1}^{T}(y_{it-1}-y_{i\cdot-1})(y_{it-1}-y_{i\cdot-1}) \\
&= E_i\sum_{t=1}^{T}(y_{it-1}-\frac{1}{T}\sum_{t=1}^{T}y_{it-1})(y_{it-1}-\frac{1}{T}\sum_{t=1}^{T}y_{it-1}) \\
&= E_i\left[\sum_{t=1}^{T}(\sum_{j=0}^{t-1}u_{it-j-1})^2-\frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{t-1}u_{it-j-1})\right)^2\right] \\
&= B_{aT}-B_{bT},
\end{aligned}$$

where

$$\begin{aligned}
B_{aT} &= E_i\left[\sum_{t=1}^{T}(\sum_{j=0}^{t-1}u_{it-j-1})^2\right] \\
&= \sum_{t=1}^{T}E_i(\sum_{j=0}^{t-1}u_{it-j-1})^2 \\
&= \sum_{t=1}^{T}E_i\left[(u_{it-1}+u_{it-2}+\cdots+u_{i0})^2\right] \\
&= \sigma^2\sum_{t=1}^{T}(t-1) = \sigma^2\frac{T(T-1)}{2},
\end{aligned}$$

and

$$\begin{aligned}
B_{bT} &= \frac{1}{T}E_i\left(\sum_{t=1}^{T}y_{it-1}\right)^2 = E_i\left(y_{i0}+\cdots+y_{iT-1}\right)^2 \\
&= \frac{1}{T}E_i\left(\sum_{t=1}^{T}y_{it-1}^2+2\sum_{t=1}^{T-1}y_{it-1}y_{it}+2\sum_{t=1}^{T-2}y_{it-1}y_{it+1}+\cdots+2\sum_{t=1}^{1}y_{it-1}y_{iT-1}\right) \\
&= \frac{\sigma^2}{T}\left[\sum_{t=1}^{T}(t-1)+2\sum_{t=1}^{T-1}(t-1)+2\sum_{t=1}^{T-2}(t-1)+\cdots+2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sigma^2}{T}\left[\frac{T(T-1)}{2} + 2\sum_{s=1}^{T-2}\sum_{s=0}^{s} r\right] = \frac{\sigma^2}{T}\left[\frac{T(T-1)}{2} + 2\sum_{s=1}^{T-2}\frac{s(s+1)}{2}\right] \\
&= \frac{\sigma^2 (T-1)(2T+3)}{6}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}B_{NT} &= \sigma^2\frac{T(T-1)}{2} - \frac{\sigma^2(T-1)(2T+3)}{6} \\
&= \sigma^2\frac{(T-1)(T-3)}{6},
\end{aligned}
$$

and so the bias in the unit root case is

$$
\text{plim}_{N\to\infty}\frac{\frac{1}{N}A_{NT}}{\frac{1}{N}B_{NT}} = -\frac{3}{T-3},
$$

which exceeds the bias in the stationary, cross section independent case.

Stationary, Cross Section Dependent Case Let $|\rho| < 1$ and suppose (50) holds and the δ_i satisfy $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2 = \mu_{\delta 2}$. Proceeding as above, we find

$$
\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^{C} &= \text{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1} - y_{i\cdot-1})(u_{it} - u_{i\cdot}) \\
&= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j u_{it-j-1})u_{it} - \frac{1}{T}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j u_{it-j-1})\sum_{t=1}^{T}u_{it}\right] \\
&= -\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j u_{it-j-1})\sum_{t=1}^{T}u_{it} \\
&= -\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\left[\frac{\delta_i^2}{T}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})\sum_{t=1}^{T}\theta_t\right] + \left[\frac{1}{T}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})\sum_{t=1}^{T}\varepsilon_{it}\right]\right)
\end{aligned}
$$

Using (50) and

$$
\sum_{t=1}^{T-1} t\rho^t = \frac{\rho\left(1-\rho^{T-1}\right)}{(1-\rho)^2} - \frac{(T-1)\rho^T}{1-\rho}, \tag{54}
$$

we obtain

$$
\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\delta_i^2}{T}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})\sum_{s=1}^{T}\theta_s\right] &= \mu_{\delta 2}\frac{1}{T}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})\sum_{s=1}^{T}\theta_s \\
&= \mu_{\delta 2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\theta_s \\
&\to_d \mu_{\delta 2}\xi_\theta\eta_\theta
\end{aligned} \tag{55}
$$

as $T\to\infty$, where by standard central limit theory

$$
\left[\begin{array}{c} \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\theta_s \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1}) \end{array}\right] \to_d \left[\begin{array}{c}\xi_\theta \\ \eta_\theta\end{array}\right] = N\left(0, \left[\begin{array}{cc} \sigma_\theta^2 & \frac{\sigma_\theta^2}{1-\rho} \\ \frac{\sigma_\theta^2}{1-\rho} & \frac{\sigma_\theta^2}{1-\rho^2}\end{array}\right]\right). \tag{56}
$$

By suitable augmentation of the probability space and embedding arguments, we may write the convergence in (56) as an almost sure convergence and then we may write (55) as

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\delta_i^2}{T}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})\sum_{s=1}^{T}\theta_s\right]=\mu_{\delta 2}\xi_\theta\eta_\theta+o_{a.s}(1).$$

where the final term is $o_{a..s.}(1)$ as $T\to\infty$. By similar augmentation of the space and embedding we get

$$\left[\begin{array}{c}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{is}\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})\end{array}\right]\to_{a.s.}\left[\begin{array}{c}\xi_i\\ \eta_i\end{array}\right]=N\left(0,\left[\begin{array}{cc}\sigma_{ii} & \frac{\sigma_{ii}}{1-\rho}\\ \frac{\sigma_{ii}}{1-\rho} & \frac{\sigma_{ii}}{1-\rho^2}\end{array}\right]\right),$$

and then

$$\begin{aligned}
& \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})\sum_{s=1}^{T}\varepsilon_{is}\right]\\
= & \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}[\xi_i\eta_i+o_{a..s.}(1)]\\
= & \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_{ii}}{1-\rho}+\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\xi_i\eta_i-\frac{\sigma_{ii}}{1-\rho}+o_{a..s.}(1)\right]\\
= & \frac{\bar{\sigma}}{1-\rho}+o_{a..s.}(1).
\end{aligned}$$

Hence

$$\text{plim}_{N\to\infty}\frac{1}{N}A_{NT}^C=-\mu_{\delta 2}\xi_\theta\eta_\theta-\frac{\bar{\sigma}}{1-\rho}+o_{a..s.}(1)$$

Note that the leading term in this limit is a random quantity.

Similary

$$\begin{aligned}
\text{plim}_{N\to\infty}\frac{1}{N}B_{NT}^C &= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1}-y_{i\cdot-1})(y_{it-1}-y_{i\cdot-1})\\
&= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j u_{it-j-1})^2-\frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j u_{it-j-1})\right)^2\right]\\
&= \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})^2-\frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})\right)^2\right]+\\
&\quad \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})^2-\frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})\right)^2\right].
\end{aligned}$$

Setting $Z_{\theta t}=\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1}$ and letting $T\to\infty$ we have

$$\frac{1}{T}\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})^2-\frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})\right)^2\right]=\frac{1}{T}\sum_{t=1}^{T}\left(Z_{\theta t}-\bar{Z}_\theta\right)^2\to_{a.s.}E\left(Z_{\theta t}^2\right)=\frac{\sigma_\theta^2}{1-\rho^2},$$

and then

$$\mathrm{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})^2-\frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\theta_{t-j-1})\right)^2\right]=T\left[\mu_{\delta 2}\frac{\sigma_\theta^2}{1-\rho^2}+o_{a.s.}(1)\right].$$

Next, let $Z_{it}=\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1}$ and letting $T\to\infty$ we have

$$\frac{1}{T}\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})^2-\frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})\right)^2\right]=\frac{1}{T}\sum_{t=1}^{T}(Z_{it}-\bar{Z}_{i.})^2\to_{a.s.}E\left(Z_{it}^2\right)=\frac{\sigma_{ii}}{1-\rho^2},$$

so that

$$\mathrm{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})^2-\frac{1}{T}\left(\sum_{t=1}^{T}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-1})\right)^2\right]=T\left[\frac{\bar{\sigma}}{1-\rho^2}+o_{a.s.}(1)\right].$$

Hence

$$\begin{aligned}\mathrm{plim}_{N\to\infty}(\hat{\rho}_{pols}-\rho) &= \mathrm{plim}_{N\to\infty}\frac{\frac{1}{N}A_{NT}^C}{\frac{1}{N}B_{NT}^C}=-\frac{\mu_{\delta 2}\xi_\theta\eta_\theta+\frac{\bar{\sigma}}{1-\rho}+o_{a..s.}(1)}{T\left[\mu_{\delta 2}\frac{\sigma_\theta^2}{1-\rho^2}+\frac{\bar{\sigma}}{1-\rho^2}+o_{a.s.}(1)\right]}\\ &= -\frac{1}{T}\frac{(1+\rho)+\frac{\mu_{\delta 2}}{\bar{\sigma}}\xi_\theta\eta_\theta}{1+\frac{\mu_{\delta 2}}{\bar{\sigma}}\sigma_\theta^2}+o_{a.s.}\left(\frac{1}{T}\right)\\ &= -\frac{1}{T}\frac{(1+\rho)+\frac{\mu_{\delta 2}}{\bar{\sigma}}\frac{\sigma_\theta^2}{1-\rho}}{1+\frac{\mu_{\delta 2}}{\bar{\sigma}}\sigma_\theta^2}\\ &\quad -\frac{1}{T}\frac{\frac{\mu_{\delta 2}}{\bar{\sigma}}}{1+\frac{\mu_{\delta 2}}{\bar{\sigma}}\sigma_\theta^2}\left[\xi_\theta\eta_\theta-E\left(\xi_\theta\eta_\theta\right)\right]+o_{a.s.}\left(\frac{1}{T}\right) \qquad (57)\end{aligned}$$

where

$$E\left(\xi_\theta\eta_\theta\right)=\frac{\sigma_\theta^2}{1-\rho}.$$

Note that the probability limit (57) is a random variable whose second term has expectation zero, so that the bias to $O\left(T^{-1}\right)$ is given by the first term. It is easily seen that for all $\rho\in(-1,1)$

$$\frac{(1+\rho)+\frac{\mu_{\delta 2}}{\bar{\sigma}}\frac{\sigma_\theta^2}{1-\rho}}{1+\frac{\mu_{\delta 2}}{\bar{\sigma}}\sigma_\theta^2}>1+\rho,$$

so that the POLS bias in the case of cross section dependence always exceeds the bias in the cross section independent case.

9 Appendix C

Derivation of SUR Limit Theory

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

$$y_{it}^* = \rho_i y_{it-1}^* + u_{it}, \quad \text{for } t = 1, \cdots, T, \quad \text{and } i = 1, \cdots, N, \tag{58}$$

in which the regression errors are from (6)

$$u_{it} = \delta_i \theta_t + \varepsilon_{it}, \quad \theta_t \sim iid\ N(0, \sigma^2) \text{ over } t, \tag{59}$$

and

$$\varepsilon_{i,t} \sim iid\ N(0, \sigma_i^2) \text{ over } t, \text{ and } \varepsilon_{i,t} \text{ is independent of } \varepsilon_{j,s} \text{ and } \theta_s \text{ for all } i \neq j \text{ and for all } s, t. \tag{60}$$

The proof in the case of models M2 and M3 is a straightforward extension. From (59) and (60)

$$u_t \sim iid\ N(0, V_u), \qquad \text{for } t = 1, \cdots, T,$$

where, as in (8), we have

$$V_u = E\left(u_t u_t' | \sigma^2, \sigma_1^2, ..., \sigma_N^2\right) = \Sigma + \sigma^2 \delta\delta', \quad \Sigma = diag\left(\sigma_1^2, ..., \sigma_N^2\right).$$

Now write (58) in vector form as

$$y_t = Z_t \underline{\rho} + u_t, \quad Z_t = diag\left(y_{1t-1}, ..., y_{Nt-1}\right), \quad \underline{\rho} = (\rho_1, ..., \rho_N)' \tag{61}$$

Then the GLS estimate is

$$\widehat{\underline{\rho}} = \left(\sum_{t=1}^{T} Z_t' V_u^{-1} Z_t\right)^{-1} \left(\sum_{t=1}^{T} Z_t' V_u^{-1} y_t\right)$$

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

$$\sqrt{T}\left(\widehat{\underline{\rho}} - \underline{\rho}\right) \rightarrow_d N\left(0, V_{SUR}\right)$$

with

$$V_{SUR} = p \lim_{T \to \infty} \left(\frac{1}{T} \sum_{t=1}^{T} Z_t' V_u^{-1} Z_t\right)^{-1}.$$

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

$$p \lim_{T \to \infty} \left(\frac{1}{T} \sum_{t=1}^{T} Z_t' V_u^{-1} Z_t\right) = \left[\left(v_u^{ij} E\left(y_{it} y_{jt}\right)\right)_{ij}\right],$$

where v_u^{ij} is the ij'th element of V_u, so that

$$V_{SUR}^{-1} = \left[\left(v_u^{ij} E\left(y_{it} y_{jt}\right) \right)_{ij} \right] = V_u^{-1} * E\left(y_t y_t'\right).$$

Next note that

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

so that

$$E\left(y_t y_t'\right) = \left(\Sigma + \sigma^2 \delta \delta'\right) * R,$$

with $R = [(r_{ij})]$ and $r_{ij} = \frac{1}{1-\rho_i \rho_j}$. Note that

$$V_u^{-1} = \Sigma^{-1} - \frac{\sigma^2 \Sigma^{-1} \delta \delta' \Sigma^{-1}}{1 + \sigma^2 \delta' \Sigma^{-1} \delta}.$$

The same result 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

$$\frac{1}{\sqrt{T}} y_{[Tr]} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \rightarrow_d B(r) = BM\left(V_u\right). \tag{62}$$

Setting i_N to be vector with N unit components, the centred GLS and feasible SUR estimates have the form

$$\underline{\widehat{\rho}}_{sur} - i_N = \left(\sum_{t=1}^{T} Z_t' V_u^{-1} Z_t \right)^{-1} \left(\sum_{t=1}^{T} Z_t' V_u^{-1} u_t \right).$$

Now

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^{T} Z_t' V_u^{-1} Z_t &= \left[\left(v_u^{ij} \frac{1}{T^2} \sum_{t=1}^{T} y_{it-1} y_{jt-1} \right)_{ij} \right] \rightarrow_d \left[\left(v_u^{ij} \int_0^1 B_i B_j \right)_{ij} \right] \\ &= V_u^{-1} * \int_0^1 BB', \end{aligned}$$

and

$$\begin{aligned} \frac{1}{T} \left(\sum_{t=1}^{T} Z_t' V_u^{-1} u_t \right) &= \left[\left(\sum_{j=1}^{N} v_u^{ij} \frac{1}{T} \sum_{t=1}^{T} y_{it-1} u_{jt} \right)_i \right] \rightarrow_d \left[\left(\sum_{j=1}^{N} v_u^{ij} \int_0^1 B_i dB_j \right)_i \right] \\ &= \int_0^1 B * V_u^{-1} dB. \end{aligned}$$

This gives the stated limit result

$$T\left(\underline{\widehat{\rho}}_{sur} - i_N\right) \to_d \left[V_u^{-1} * \int_0^1 BB'\right]^{-1} \left[\int_0^1 B * V_u^{-1} dB\right] = \xi.$$

Note that the quadratic variation process of the stochastic integral $\int_0^r B * V_u^{-1} dB$ is

$$\left[\int_0^r B * V_u^{-1} dB\right]_r = V_u^{-1} * \int_0^r BB',$$

so the matrix $V_u^{-1} * \int_0^1 BB'$ is a suitable metric for $\int_0^1 B * V_u^{-1} dB$. The joint Wald test for unit roots is

$$\begin{aligned} W_{SUR} &= \left(\underline{\widehat{\rho}}_{sur} - i\right)' \left[\sum_{t=1}^{T} Z_t' V_u^{-1} Z_t\right] \left(\underline{\widehat{\rho}}_{sur} - i\right) \\ &\to_d \left[\int_0^1 B * V_u^{-1} dB\right]' \left[V_u^{-1} * \int_0^1 BB'\right]^{-1} \left[\int_0^1 B * V_u^{-1} dB\right], \end{aligned}$$

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

$$D\left(\underline{\widehat{\rho}}_{sur} - i_N\right) = D\underline{\widehat{\rho}}_{sur}$$

we would have the statistic

$$\begin{aligned} WD_{sur} &= \underline{\widehat{\rho}}_{sur}{}' D' \left[D\widehat{V}_{SUR} D'\right]^{-1} D\underline{\widehat{\rho}}_{sur} \\ &\to_d \xi' D' \left[D\left(V_u^{-1} * \int_0^1 BB'\right)^{-1} D'\right]^{-1} D\xi. \end{aligned}$$

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

$$\widehat{\rho} = \left(\sum_{t=1}^{T} y_{t-1}' V_u^{-1} y_{t-1}\right)^{-1} \left(\sum_{t=1}^{T} y_{t-1}' V_u^{-1} y_t\right)$$

and by straightforward derivation

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}' V_u^{-1} y_{t-1} \to_d \int_0^1 B' V_u^{-1} B =_d \int_0^1 W'W = \sum_{i=1}^{N} \int_0^1 W_i^2,$$

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}' V_u^{-1} u_t \to_d \int_0^1 B' V_u^{-1} dB =_d \int_0^1 W' dW = \sum_{i=1}^{N} \int_0^1 W_i dW_i$$

and so

$$T\left(\widehat{\rho} - 1\right) \to_d \frac{\int_0^1 W' dW}{\int_0^1 W'W} = \frac{\sum_{i=1}^{N} \int_0^1 W_i dW_i}{\sum_{i=1}^{N} \int_0^1 W_i^2}, \tag{63}$$

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

Hausman Test Limit Theory (Unit Root Case) The Hausman statistic relies on the difference

$$\sqrt{T}\left(\underline{\hat{\rho}}_{emu}-i_N\right)=\sqrt{T}\left(\underline{\hat{\rho}}_{emu}-i_N\right)$$

From (63) we have

$$T\left(\widetilde{\rho}-1\right)\rightarrow_d \frac{\int_0^1 W'dW}{\int_0^1 W'W},$$

and

$$\begin{aligned}
T\left(\underline{\hat{\rho}}_{emu}-i_N\right) &= T\left(\underline{\hat{\rho}}-i_N\right)+o_p(1)\\
&= \left(\frac{1}{T^2}\sum_{t=1}^{T}Z_t^{'}Z_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}Z_t^{'}u_t\right)\\
&\rightarrow_d \left[\int_0^1 D_{B^2}\right]^{-1}\left[\int_0^1 D_B dB\right]\\
&= \begin{bmatrix}\left[\int_0^1 B_1^2\right]^{-1}\left[\int_0^1 B_1 dB_1\right]\\ \vdots \\ \left[\int_0^1 B_N^2\right]^{-1}\left[\int_0^1 B_N dB_N\right]\end{bmatrix},
\end{aligned}\tag{64}$$

where $D_{B^2}=diag\left(B_1(r)^2,...,B_N(r)^2\right)$, $D_B=diag\left(B_1(r),...,B_N(r)\right)$. In view of the correlation between the Brownian motions $\{B_i : i=1,...,N\}$ the limit distribution (64) is dependent on nuisance parameters arising from the cross section dependence.

We also have the joint convergence

$$\begin{bmatrix}T\left(\underline{\hat{\rho}}_{emu}-i_N\right)\\ T\left(\widetilde{\rho}-1\right)\end{bmatrix}\rightarrow_d \begin{bmatrix}\left[\int_0^1 D_{B^2}\right]^{-1}\left[\int_0^1 D_B dB\right]\\ \left[\int_0^1 B'B\right]^{-1}\left[\int_0^1 B'dB\right]\end{bmatrix}$$

and then

$$T\left(\underline{\hat{\rho}}_{emu}-i_N\right)\overset{d}{\rightarrow}\begin{bmatrix}\left[\int_0^1 B'B\right]^{-1}\left[\int_0^1 B'dB\right]\\ \vdots \\ \left[\int_0^1 B'B\right]^{-1}\left[\int_0^1 B'dB\right]\end{bmatrix},$$

Again, this limit distribution is dependent on nuisance parameters arising from cross section dependence. Thus, the Hausman statistic has a limit theory that depends on nuisance parameters in the unit root case.

Modified Hausman and Panel Unit Root Tests Limit Theory First note that we have the joint convergence

$$\begin{bmatrix}T\left(\underline{\hat{\rho}}_{emu}^{+}-i_{N-1}\right)\\ T\left(\widetilde{\rho}^{+}-1\right)\end{bmatrix}\rightarrow_d \begin{bmatrix}\left[\int_0^1 D_{W_\perp^2}\right]^{-1}\left[\int_0^1 D_{W_\perp}dW_\perp\right]\\ \left[\int_0^1 W_\perp' W_\perp\right]^{-1}\left[\int_0^1 W_\perp{}'dW_\perp\right]\end{bmatrix}$$

which is free of nuisance parameters. Then

$$\begin{aligned}
T\left(\underline{\hat{\rho}}^{+}_{emu}-\widetilde{\rho}^{+}i_{N-1}\right) &\to_{d} \left[\int_0^1 D_{W_\perp^2}\right]^{-1}\left[\int_0^1 D_{W_\perp} dW_\perp\right]-\left[\int_0^1 W_\perp' W_\perp\right]^{-1}\left[\int_0^1 W_\perp{}' dW_\perp\right] i_{N-1} \\
&= \begin{bmatrix} \left[\int_0^1 W_{\perp,1}^2\right]^{-1}\left[\int_0^1 W_{\perp,1} dW_{\perp,1}\right]-\left[\int_0^1 W_\perp' W_\perp\right]^{-1}\left[\int_0^1 W_\perp{}' dW_\perp\right] \\ \vdots \\ \left[\int_0^1 W_{\perp,N-1}^2\right]^{-1}\left[\int_0^1 W_{\perp,N-1} dW_{\perp,N-1}\right]-\left[\int_0^1 W_\perp' W_\perp\right]^{-1}\left[\int_0^1 W_\perp{}' dW_\perp\right] \end{bmatrix} \\
&: = \Xi_{N-1}, \qquad (65)
\end{aligned}$$

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

$$G_H^{+}=T^2\left(\underline{\hat{\rho}}^{+}_{emu}-\widetilde{\rho}^{+}i_{N-1}\right)'\left(\underline{\hat{\rho}}^{+}_{emu}-\widetilde{\rho}^{+}i_{N-1}\right)\to_d \Xi_{N-1}'\Xi_{N-1}.$$

Similarly, the modified unit root tests have the limit

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

10 Appendix D

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.

$$y_t = Z_t \underline{\rho} + u_t, \quad Z_t = diag\left(y_{1t-1}, ..., y_{Nt-1}\right), \quad \underline{\rho} = \left(\rho_1, ..., \rho_N\right)',$$

with errors u_t that are *iid* $N(0, V_u)$ where $V_u = \Sigma + \delta\delta'$ and $\Sigma = diag(\sigma_1^2, ..., \sigma_N^2)$. The log likelihood function has the form

$$\begin{aligned}
\ell_{NT}(\underline{\rho}, \Sigma, \delta) &= -\frac{NT}{2} \log 2\pi - \frac{T}{2} \log V_u - \frac{1}{2} \sum_{t=1}^{T} \left(y_t - Z_t\underline{\rho}\right)' V_u^{-1} \left(y_t - Z_t\underline{\rho}\right) \\
&= -\frac{NT}{2} \log 2\pi - \frac{T}{2} \log V_u - \frac{T}{2} tr\left[V_u^{-1} M_T\right],
\end{aligned}$$

where $M_T\left(\underline{\rho}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - Z_t\underline{\rho}\right)\left(y_t - Z_t\underline{\rho}\right)'$. First order conditions for maximization of $\ell_{NT}(\underline{\rho}, \Sigma, \delta)$ lead to

$$\widehat{\underline{\rho}} = \left(\sum_{t=1}^{T} Z_t' \hat{V}_u^{-1} Z_t\right)^{-1} \left(\sum_{t=1}^{T} Z_t' \hat{V}_u^{-1} y_t\right), \tag{66}$$

and

$$tr\left[\left(\hat{V}_u^{-1} - \hat{V}_u^{-1} M_T \hat{V}_u^{-1}\right) dV_u\right] = 0, \tag{67}$$

where $\hat{V}_u = \hat{\Sigma} + \hat{\delta}\hat{\delta}'$, $\hat{\Sigma} = diag(\hat{\sigma}_1^2, ..., \hat{\sigma}_N^2)$ and $dV_u = d\Sigma + d\delta\delta' + \delta d\delta'$. Expanding (67) leads to the following system of equations

$$\hat{\sigma}_i^2 \left[1 - \frac{\hat{\delta}_i^2 / \hat{\sigma}_i^2}{1 + \hat{\delta}'\hat{\Sigma}^{-1}\hat{\delta}}\right] = \left[e_i' - \frac{\hat{\delta}_i \hat{\delta}' \hat{\Sigma}^{-1}}{1 + \hat{\delta}'\hat{\Sigma}^{-1}\hat{\delta}}\right] M_T\left(\widehat{\underline{\rho}}\right) \left[e_i - \frac{\hat{\Sigma}^{-1}\hat{\delta}\hat{\delta}_i}{1 + \hat{\delta}'\hat{\Sigma}^{-1}\hat{\delta}}\right], \quad i = 1, ..., N \tag{68}$$

$$\hat{\delta} = \frac{M_T\left(\widehat{\underline{\rho}}\right) \hat{\Sigma}^{-1}\hat{\delta}}{1 + \hat{\delta}'\hat{\Sigma}^{-1}\hat{\delta}}, \tag{69}$$

which we may solve by the following iteration

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

$$\hat{\delta}^{(j)} = \frac{M_T\left(\widehat{\underline{\rho}}\right) \left(\hat{\Sigma}^{(j-1)}\right)^{-1} \hat{\delta}^{(j-1)}}{1 + \hat{\delta}^{(j-1)\prime} \left(\hat{\Sigma}^{(j-1)}\right)^{-1} \hat{\delta}^{(j-1)}},$$

which we continue until convergence. For starting values we may choose $\hat{\Sigma}^{(0)} = \hat{\sigma}^2 I_N$ where $\hat{\sigma}^2 = \frac{1}{N} tr\left[M_T\right]$ and $\hat{\delta}^{(0)}$ is the largest eigenvector of M_T. In place of the the residual moment matrix, $M_T\left(\widehat{\underline{\rho}}\right)$, from maximum likelihood estimation that appears in (68) and (69), we propose that the matrix $M_T\left(\hat{\underline{\rho}}_{emu}\right)$ corresponding to the median unbiased estimates $\hat{\underline{\rho}}_{emu}$ be used.

Note that in the special case where $\Sigma = \sigma^2 I_N$, the first order equations lead to the following system simplifying (68) and (69)

$$\hat{\sigma}^2 \left[N - \frac{\hat{\delta}'\hat{\delta}}{\hat{\sigma}^2 + \hat{\delta}'\hat{\delta}}\right] = tr\left[\left(I_N - \frac{\hat{\delta}\hat{\delta}'}{\hat{\sigma}^2 + \hat{\delta}'\hat{\delta}}\right) M_T\left(\widehat{\underline{\rho}}\right)\left(I_N - \frac{\hat{\delta}\hat{\delta}'}{\hat{\sigma}^2 + \hat{\delta}'\hat{\delta}}\right)\right],$$

and

$$\hat{\delta} = \frac{M_T\hat{\delta}}{\hat{\sigma}^2 + \hat{\delta}'\hat{\delta}}.$$

References

[1] Andrews, D. W. K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica* 59, 817–858.

[2] Andrews, D. W. K. (1993), "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models." *Econometrica*, 61, 139-165.

[3] Andrews, D. W. K. and H. Chen (1994), "Approximately Median-Unbiased Estimation of Autoregssive Models." *Journal of Business & Economic Statistics 12, 187-204.*

[4] Andrews, D. W. K., and J.C. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica* 60, 953-966.

[5] Bai, J. and S. Ng (2001). "Determining the Number of Factors in Approximate Factor Models," forthcoming in *Econometrica.*

[6] Baltagi (2001) "Econometric Analysis of the Panel Data" Baffins Lane Chichester, John Wiley & Sons Ltd.

[7] Cermeno, R. (1999). "Median Unbiased Estimation in Fixed Effects Dynamic Panels," *Annales D'Economie Et De Statistique*, pp. 351–368

[8] Choi, I (2001a). "Unit Root Tests for Panel Data," *Journal of International Money and Finance, 20, 249-272.*

[9] Choi, I (2001b). "Unit Root Tests for Cross-Sectionally Correlated Panels," *Mimeo.* Kookmin University.

[10] den Haan, W. J. and A. Levin (1996). "Inferences from parametric and nonparametric covevariance matrix estimation procedures". Federal Reserve Board, mimeographed.

[11] Dickey, D. A. and W. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, 427-31

[12] Evans, P. and G. Karras (1996), "Convergence Revisited," *Journal of Monetary Economics*, 37, 249-265.

[13] Frankel, J. and A. Rose (1996), "Panel Projection on Purchasing Power Parity: Mean Reversion within and between Countries," *Journal of International Economics*, 40, 209-224.

[14] Hadri, K. (2001). "Testing for stationarity in heterogeneious panel data". *Econometrics Journal*, 1, 1-14.

[15] Hurvicz, L. (1950). "Least squares bias in time series," in T. Koopmans (ed.) *Statistical Inference in Dynamic Economic Models*, New York: Wiley, 365-383.

[16] Im, K.S., M.H. Pesaran, and Y. Shin (1997). "Testing for Unit Roots in Heterogeneous Panels," *Mimeo.*

[17] Kendall, M. G. (1954). "Note on the bias in the estimation of autocorrelation," *Biometrika*, 41, 403-404.

[18] Lee, C. C. and P. C. B. Phillips (1994), "An ARMA Prewhitened Long-run Variance Estimator," Manuscript, Yale University.

[19] Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley and Sons

[20] 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, 631-652.

[21] Marriott, F. H. C. and J. A. Pope (1954). "Bias in the estimation of autocorrelations," *Biometrika*, 41, 393-403

[22] Moon, H. R and B. Perron (2001). "Testing for a Unit Root in Panels with Dynamic Factors," USC, *mimeo.*

[23] Nerlove, M. (2000) "Growth rate convergence, fact or artifact? An essay on panel data econometrics", in J. Krishnakumar and E. Ronchetti (eds.) *Panel Data Econometrics. Papers in Honour of Professor Pietro Balestra.* Amsterdam: North Holland.

[24] Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1417–1426.

[25] 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.

[26] Orcutt, G. H. and H. S. Winokur (1969). "First order autoregression: inference, estimation and prediction". *Econometrica*, 37, 1-14.

[27] Papell, D and C. Murray (2001), "The Purchasing Power Parity Persistence Paradigm," forthcoming in *Journal of International Economics.*

[28] Shaman, P. and R.A. Stine (1988), "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association, 83, 842-848.*

[29] 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.*