Pooled Mean Group Estimation of Dynamic Heterogeneous Panels
Author(s): M. Hashem Pesaran, Yongcheol Shin and Ron P. Smith
Source: *Journal of the American Statistical Association*, Vol. 94, No. 446 (Jun., 1999), pp. 621-634
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2670182
Accessed: 14-05-2020 10:48 UTC

Pooled Mean Group Estimation of Dynamic Heterogeneous Panels

M. Hashem PESARAN, Yongcheol SHIN, and Ron P. SMITH

It is now quite common to have panels in which both T, the number of time series observations, and N, the number of groups, are quite large and of the same order of magnitude. The usual practice is either to estimate N separate regressions and calculate the coefficient means, which we call the mean group (MG) estimator, or to pool the data and assume that the slope coefficients and error variances are identical. In this article we propose an intermediate procedure, the pooled mean group (PMG) estimator, which constrains long-run coefficients to be identical but allows short-run coefficients and error variances to differ across groups. We consider both the case where the regressors are stationary and the case where they follow unit root processes, and for both cases derive the asymptotic distribution of the PMG estimators as T tends to infinity. We also provide two empirical applications: aggregate consumption functions for 24 Organization for Economic Cooperation and Development economies over the period 1962–1993, and energy demand functions for 10 Asian developing economies over the period 1974–1990.

KEY WORDS: Consumption functions; Energy demand; Heterogeneous dynamic panels; $I(0)$ and $I(1)$ regressors; Pooled mean group estimator.

1. INTRODUCTION

Recent years have brought increasing interest in dynamic panel data models, where the number of time series observations, T, is relatively large and of the same order of magnitude as N, the number of groups. Such panels arise particularly in cross-country analyses. In most applications of this type, the parameters of interest are the long-run effects and the speed of adjustment to the long run. An example is the large literature on testing purchasing power parity in panels, where according to economic theory the long-run coefficients of the logarithms of domestic prices, foreign prices, and exchange rates should be $(1, -1, -1)$, with the speed of adjustment being of central policy concern. (A recent review of this literature was provided in Rogoff 1996; some of the related econometric issues were discussed in Boyd and Smith 1998.) Another prominent example is the Fisher equation, which postulates a unit coefficient for the long-run effect of (expected) inflation on the nominal rate of interest, but is silent as to the magnitude of the short-run effects of changes in inflation on interest rates.

There are two procedures commonly used for such panels. At one extreme, one can estimate separate equations for each group and examine the distribution of the estimated coefficients across groups. Of particular interest will be the mean of the estimates, which we call the mean group (MG) estimator. In earlier work, Pesaran and Smith (1995) we showed that the MG estimator will produce consistent estimates of the average of the parameters. This estimator, however, does not take account of the fact that certain parameters may be the same across groups. At the other extreme are the traditional pooled estimators, such as the fixed and random effects estimators, where the intercepts are allowed to differ across groups while all other coefficients and error variances are constrained to be the same. In this paper, we consider an intermediate estimator, which we call the pooled mean group (PMG) estimator because it involves both pooling and averaging. This estimator allows the intercepts, short-run coefficients, and error variances to differ freely across groups, but constrains the long-run coefficients to be the same. There are often good reasons to expect the long-run equilibrium relationships between variables to be similar across groups, due to budget or solvency constraints, arbitrage conditions, or common technologies influencing all groups in a similar way. The reasons for assuming that short-run dynamics and error variances should be the same tend to be less compelling. Not imposing equality of short-run slope coefficients also allows the dynamic specification (e.g., the number of lags included) to differ across groups.

We use two important empirical examples to compare the MG, PMG, and dynamic fixed effect (DFE) estimators. The first is the consumption function in Organization for Economic Cooperation and Development (OECD) countries. Determination of saving rates is an important policy issue and, as with the example of purchasing power parity mentioned earlier, the theory makes strong predictions: The long-run income elasticity of consumption should be unity in all countries, despite their important institutional and cultural differences. Otherwise, national (private) saving rates will be rising or falling indefinitely. The PMG estimator allows us to estimate this common long-run coefficient without making the less plausible assumption of identical dynamics in each country. In this application, both $N = 24$ and $T = 32$ are quite large. The second example is

M. Hashem Pesaran is Professor, Faculty of Economics and Politics, Trinity College, University of Cambridge CB3 9DD, U.K., and University of Southern California (E-mail: *hashem.pesaran@econ.cam.ac.uk*). Yongcheol Shin is Lecturer, Department of Economics, University of Edinburgh, EH8 9JY, U.K. (E-mail: *Yongcheol.Shin@ed.ac.uk*). Ron P. Smith is Professor, Department of Economics, Birkbeck College, University of London, W1P 1PA, U.K. (E-mail: *R.Smith@bbk.ac.uk*). Partial financial support from the ESRC (grant H519255003) and the Isaac Newton Trust of Trinity College, Cambridge is gratefully acknowledged. The first author also acknowledges the hospitality of the IMF Research Department while some of the ideas behind this article were formalized. The authors are also grateful for the efforts of Sunil Sharma and Mansour Gill in compiling the OECD dataset; for helpful discussions with Michael Binder, Nadeem U. Haque, Robert MacKay, Sunil Sharma, and Richard Smith; and for valuable comments on an earlier version from the editors and a referee.

Journal of the American Statistical Association
June 1999, Vol. 94, No. 446, Theory and Methods

energy demand in developing Asian economies. Prior to the financial crises of the summer of 1997, this was the world's fastest growing region, and the growth in its energy demand will have serious implications both for world energy markets and greenhouse gas emissions. As with the consumption function example, long-run responses of energy demand to income and relative energy prices are likely to be similar across countries, although short-run adjustments in energy demand, depending on patterns of investment in energy using equipment and supply constraints, is unlikely to be homogeneous across countries. Again, the PMG estimator allows us to investigate long-run homogeneity without imposing parameter homogeneity in the short run. In this example, both $N = 10$ and $T = 17$ are quite small. Thus there is a contrast in the panel dimensions in the two examples.

The article is organized as follows: Section 2 provides a brief review of alternative panel data estimators and discusses how the PMG estimator is related to them. Section 3 sets out the model and its underlying assumptions, and derives the log-likelihood function. Section 4 develops the general theory of PMG estimation, in both stationary regressors and unit root processes. Section 5 discusses a number of modeling issues and presents the empirical applications, and Section 6 contains some concluding remarks. Mathematical proofs are provided in an Appendix.

The following notation is used throughout: The symbol $\xrightarrow{p}$ signifies convergence in probability; $\Rightarrow$, weak convergence in probability measure; $\overset{a}{\sim}$, asymptotic equality in distributions; MN, mixture normal distribution; $\mathbf{I}_m$, an identity matrix of order m; diag[$\cdot$], a diagonal matrix; and $I(d)$, an integrated variable of order d.

2. ALTERNATIVE ESTIMATORS OF DYNAMIC PANELS

Numerous dynamic panel data estimators exist. To see the relationship of the PMG estimator advanced in this article to the ones proposed in the literature, it is convenient to divide the latter into three categories, differing in their assumptions about the relative magnitudes of N and T. First, there is the "small N, large T" time-series literature devoted to estimating long-run effects. In the case of individual time-series where $N = 1$, the traditional approach was to estimate an autoregressive distributed lag (ARDL) model. In recent years, the emphasis has shifted to estimating cointegrating relationships from a single time series. The relation between these two approaches has been discussed in earlier work (Pesaran and Shin 1999). For $N > 1$, following Zellner (1962), the seemingly unrelated regression equation (SURE) procedure is often used. The main attraction of the SURE procedure is that it allows the contemporaneous error covariances to be freely estimated. But this is possible only when N is reasonably small relative to T. When N is of the same order of magnitude as T—the case we are interested in—SURE is not feasible. (It is also worth noting that most of the SURE literature is concerned with linear cross-equation restrictions, whereas our concern with common long-run coefficients implies nonlinear restrictions across the different equations.) Other approaches must be sought. It is often likely that the cause of nonzero error covariances may be due to omitted common effects that impact all groups; an indication of misspecification rather than error correlation. In such cases it is more appropriate to model the effects that account for between-group nonzero error covariances explicitly. This can be done in a number of ways. One possibility is to include cross-sectional means of the included regressors as additional regressors in the model. For instance, in an earlier work (Pesaran and Smith 1995), aggregate output was included in addition to industry output in industry employment equations. A second possibility is to express all of the variables as deviations from their respective cross-sectional means in each period. In the special case where the slope coefficients are identical across groups, the common time-period effects will be completely eliminated by this (cross-sectional) demeaning procedure. In cases where the slope coefficients differ, such demeaning will reduce (but not eliminate) the common time-specific effects. A third possibility is using explicit spatial models of the interactions between neighbors. In what follows, we assume that the use of such procedures has made it reasonable to assume that the errors are independent between groups.

Second, there is the "small T, large N" dynamic panel literature. In the case where T is small, we (Pesaran and Smith 1995) have shown that under certain assumptions, the cross-section regression based on time-averages of the variables will provide consistent estimates of the long-run coefficients. However, these assumptions are quite strong. In particular, they require that the group-specific parameters are distributed independently of the regressors, and that the regressors are strictly exogenous. For larger T, we (Pesaran and Smith 1995) showed that the traditional procedures for estimation of pooled models, such as the fixed effects, instrumental variables, and generalized method-of-moments (GMM) estimators proposed, by among others, Ahn and Schmidt (1995), Anderson and Hsiao (1981, 1982), Arellano (1989), Arellano and Bover (1995), and Keane and Runkle (1992) can produce inconsistent, and potentially very misleading estimates of the average values of the parameters in dynamic panel data models unless the slope coefficients are in fact identical. (This result holds in the case of dynamic random coefficient models even if it is assumed that the random coefficients and the regressors are independently distributed.) But tests on most panels of this sort, indicate that these parameters differ significantly across groups. Thus an estimator that imposes weaker homogeneity assumptions would be useful.

The third literature, and the one closest to our concerns, is the Bayes and empirical Bayes estimators proposed by Hsiao and Tahmiscioglu (1997) and Hsiao, Pesaran and Tahmiscioglu (1999). These estimators build on the early work of Lindley and Smith (1972) and Swamy (1970). Hsiao et al. (1999) considered Bayes estimation of short-run coefficients in dynamic heterogenous panels, and established the asymptotic equivalence of the Bayes estimator and the MG estimator. In particular, they showed that the MG estimator is asymptotically normal for large N and large T as long as $\sqrt{N}/T \rightarrow 0$ as both N and $T \rightarrow \infty$. Using Monte Carlo experiments, they also showed that although the MG estimator is consistent, it is unlikely to be a good estimator when either N or T is small. The main difference between the approach adopted in this article and the literature on Bayes estimation is that whereas we regard the parameters as fixed, this literature regards them as random: drawn from some distribution with a finite number of parameters. Given that the possibility of random parameters is admitted, the issue is not primarily one of classical versus Bayesian approaches, as classical estimators can be given a Bayesian interpretation. For example, the Swamy (1970) estimator for random coefficient models, motivated by classical generalized least squares arguments, can also be viewed as an empirical Bayes estimator.

The choice between fixed- and random-effects formulations, which has been extensively discussed in the literature, depends on a number of related considerations. First, the purpose of the exercise matters, as the two approaches try to answer distinctly different questions. Hsiao (1996, pp. 93–94) gave a number of examples in which the purpose of analysis will determine the choice between the two formulations. Stoker (1993, p. 1848) pointed out that if one wishes to make inference about macro relationships from micro estimates based on a subset of the population—a common problem—then the effects must be treated as random.

Second, the perceived degree of commonality in the parameters matters. There is a continuum, with at one extreme fixed and common parameters and the other extreme fixed and heterogenous parameters, with the random coefficient model fitting somewhere in between (see, e.g., Maddala and Hu 1996). In practice, it is often difficult to determine where on this continuum a particular case lies. The examples that we use are not samples, but rather almost the whole population of countries in a particular category: members of the OECD or Asian developing countries. This suggests treating their parameters as fixed. The difficulty of making the distinction precise was illustrated by Hsiao et al. (1995), who proposed a procedure for choosing between fixed- and random-effects models, which they evaluated by Monte Carlo methods. But in the simulations, the fixed effects are actually generated randomly either as a mixture of normal distributions or as a function of the regressors. Although this design effectively captures the situations where traditional random-effects estimators do badly, the fact that there is no obvious way to generate fixed-effects data for heterogenous panels indicates the conceptual difficulty.

Third are considerations of estimation. The fixed-effects approach uses the conditional likelihood (conditional on the particular effects); the random approach uses the unconditional or marginal likelihood. Given the correct specification, the latter will be more efficient. When N is large relative to T, the fixed-effects approach can be very inefficient.

Fourth, and an important practical issue, is whether the effects are independent of the regressors and this is the basis of a number of tests between the two specifications.

Dynamic models add further complications. The initial conditions can also be treated as fixed or random, and the long-run parameters are nonlinear functions of the short-run parameters. Thus, when using the random coefficient approach, one must ensure that the joint distribution for the short-run parameters implies a meaningful joint distribution for the long-run parameters, or vice versa. In the case that we consider—homogenous long-run parameters—the long-run parameters across the groups have a degenerate distribution, which raises some technical difficulties in treating the heterogenous short-run parameters as random. Although a number of issues remain to be resolved, and there is unlikely to be a single estimator that is appropriate for all dynamic heterogenous panel problems, we believe that the PMG estimator that we propose may be useful in a number of cases that are important in practice.

3. THE MODEL

Suppose that given data on time periods, $t = 1, 2, \ldots, T$, and groups, $i = 1, 2, \ldots, N$, we wish to estimate an ARDL$(p, q, q, \ldots, q)$ model,

$$y_{it} = \sum_{j=1}^{p} \lambda_{ij} y_{i,t-j} + \sum_{j=0}^{q} \boldsymbol{\delta}_{ij}' \mathbf{x}_{i,t-j} + \mu_i + \varepsilon_{it}, \qquad (1)$$

where $\mathbf{x}_{it}$ $(k \times 1)$ is the vector of explanatory variables (regressors) for group i; μ_i represent the fixed effects; the coefficients of the lagged dependent variables, λ_{ij}, are scalars; and $\boldsymbol{\delta}_{ij}$ are $k \times 1$ coefficient vectors. T must be large enough such that we can estimate the model for each group separately. For notational convenience, we use a common T and p across groups and a common q across groups and regressors, but this is not necessary. Similarly, time trends or other types of fixed regressors, such as seasonal dummies, can be included in (1). But to keep the notations simple, we do not allow for such effects.

It is convenient to work with the following reparameterization of (1):

$$\Delta y_{it} = \phi_i y_{i,t-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \sum_{j=1}^{p-1} \lambda_{ij}^{*} \Delta y_{i,t-j} + \sum_{j=0}^{q-1} \boldsymbol{\delta}_{ij}^{*\prime} \Delta \mathbf{x}_{i,t-j} + \mu_i + \varepsilon_{it}, \qquad (2)$$

$i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$, where $\phi_i = -(1 - \sum_{j=1}^{p} \lambda_{ij}), \boldsymbol{\beta}_i = \sum_{j=0}^{q} \boldsymbol{\delta}_{ij}$,

$$\lambda_{ij}^{*} = -\sum_{m=j+1}^{p} \lambda_{im}, \qquad j = 1, 2, \ldots, p-1,$$

and

$$\boldsymbol{\delta}_{ij}^{*} = -\sum_{m=j+1}^{q} \boldsymbol{\delta}_{im}, \qquad j = 1, 2, \ldots, q-1. \qquad (3)$$

If we stack the time-series observations for each group, then (2) can be written as

$$\Delta \mathbf{y}_i = \phi_i \mathbf{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta}_i + \sum_{j=1}^{p-1} \lambda_{ij}^* \Delta \mathbf{y}_{i,-j} + \sum_{j=0}^{q-1} \Delta \mathbf{X}_{i,-j} \boldsymbol{\delta}_{ij}^* + \mu_i \boldsymbol{\iota} + \boldsymbol{\varepsilon}_i, \quad (4)$$

$i = 1, 2, \ldots, N$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ is a $T \times 1$ vector of the observations on the dependent variable of the ith group, $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})'$ is a $T \times k$ matrix of observations on the regressors that vary both across groups and time periods, $\boldsymbol{\iota} = (1, \ldots, 1)'$ is a $T \times 1$ vector of 1s, $\mathbf{y}_{i,-j}$ and $\mathbf{X}_{i,-j}$ are j period lagged values of $\mathbf{y}_i$ and $\mathbf{X}_i$, and $\Delta \mathbf{y}_i = \mathbf{y}_i - \mathbf{y}_{i,-1}$, $\Delta \mathbf{X}_i = \mathbf{X}_i - \mathbf{X}_{i,-1}, \Delta \mathbf{y}_{i,-j}$ and $\Delta \mathbf{X}_{i,-j}$ are j period lagged values of $\Delta \mathbf{y}_i$ and $\Delta \mathbf{X}_i$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$.

We make the following assumptions.

Assumption 1. The disturbances $\varepsilon_{it}, i = 1, 2, \ldots, N, t = 1, 2, \ldots, T$, in (1) are independently distributed across i and t, with means 0, variances $\sigma_i^2 > 0$, and finite fourth-order moments. They are also distributed independently of the regressors, $\mathbf{x}_{it}$.

The assumption that the disturbances ε_{it} are distributed independently across groups was discussed in Section 2. The assumption that they are independent across time is also not very restrictive, and can be satisfied in most applications by increasing the distributed lag orders on y_{it} and $\mathbf{x}_{it}$. The independence of the disturbances and the regressors is needed for the consistent estimation of the short-run coefficients, but, as shown by Pesaran (1997), it is relatively straightforward to allow for the possible dependence of $\mathbf{x}_{it}$ on ε_{it} when estimating the long-run coefficients, as long as $\mathbf{x}_{it}$ have finite-order autoregressive representations.

Assumption 2. The ARDL$(p, q, q, \ldots, q)$ model (1) is stable in that the roots of

$$\sum_{j=1}^{p} \lambda_{ij} z^j = 1, \qquad i = 1, 2, \ldots, N,$$

lie outside the unit circle.

This assumption ensures that $\phi_i < 0$, and hence there exists a long-run relationship between y_{it} and $\mathbf{x}_{it}$ defined by

$$y_{it} = -(\boldsymbol{\beta}_i'/\phi_i)\mathbf{x}_{it} + \eta_{it},$$

for each $i = 1, 2, \ldots, N$, where η_{it} is a stationary process. Pesaran, Shin, and R. J. Smith (1999) provide a general framework for testing assumption 2, irrespective of whether the regressors, $\mathbf{x}_{it}$, are $I(1)$ or $I(0)$. Assumption 2 also ensures that the order of integration of y_t is at most equal to that of $\mathbf{x}_{it}$.

Assumption 3 (Long-Run Homogeneity). The long-run coefficients on $\mathbf{X}_i$, defined by $\boldsymbol{\theta}_i = -\boldsymbol{\beta}_i/\phi_i$, are the same across the groups, namely

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}, \qquad i = 1, 2, \ldots, N. \quad (5)$$

Under Assumptions 2 and 3, relation (4) can be written more compactly as

$$\Delta \mathbf{y}_i = \phi_i \boldsymbol{\xi}_i(\boldsymbol{\theta}) + \mathbf{W}_i \boldsymbol{\kappa}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, 2, \ldots, N, \quad (6)$$

where

$$\boldsymbol{\xi}_i(\boldsymbol{\theta}) = \mathbf{y}_{i,-1} - \mathbf{X}_i \boldsymbol{\theta}, \qquad i = 1, 2, \ldots, N, \quad (7)$$

is the error correction component, $\mathbf{W}_i = (\Delta \mathbf{y}_{i,-1}, \ldots, \Delta \mathbf{y}_{i,-p+1}, \Delta \mathbf{X}_i, \Delta \mathbf{X}_{i,-1}, \ldots, \Delta \mathbf{X}_{i,-q+1}, \boldsymbol{\iota})$, and $\boldsymbol{\kappa}_i = (\lambda_{i1}^*, \ldots, \lambda_{i,p-1}^*, \boldsymbol{\delta}_{i0}^{*\prime}, \boldsymbol{\delta}_{i1}^{*\prime}, \ldots, \boldsymbol{\delta}_{i,q-1}^{*\prime}, \mu_i)'$. The group-specific equations in the panel (6) are nonlinear in ϕ_i and $\boldsymbol{\theta}$, and since $\boldsymbol{\theta}$ is common across groups the panel is subject to cross-equation parameter restrictions. In what follows we also allow the error variances, $\text{var}(\varepsilon_{it}) = \sigma_i^2$, to differ across groups.

To estimate the model, we adopt a likelihood approach and initially assume that the disturbances ε_{it} are normally distributed, though this assumption is not required for the asymptotic results. Under Assumption 1, the likelihood of the panel data model can be written as the product of the likelihoods for each group. Because the parameters of interest are the long-run effects and adjustment coefficients, we work directly with the concentrated log-likelihood function. Given normality, we have

$$l_T(\boldsymbol{\varphi}) = -\frac{T}{2} \sum_{i=1}^{N} \ln 2\pi\sigma_i^2 - \frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (\Delta \mathbf{y}_i - \phi_i \boldsymbol{\xi}_i(\boldsymbol{\theta}))' \mathbf{H}_i (\Delta \mathbf{y}_i - \phi_i \boldsymbol{\xi}_i(\boldsymbol{\theta})), \quad (8)$$

where $\mathbf{H}_i = \mathbf{I}_T - \mathbf{W}_i(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'$, $\boldsymbol{\varphi} = (\boldsymbol{\theta}', \boldsymbol{\phi}', \boldsymbol{\sigma}')'$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_N)'$, and $\boldsymbol{\sigma} = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)'$.

For the proof of consistency and asymptotic normality of the PMG estimators, we also require the following additional assumptions.

Assumption 4. $\boldsymbol{\varphi} \in \Theta_{\varphi} = \Theta_{\theta} \times \Theta_{\phi} \times \Theta_{\sigma}$, where Θ_{φ} is a compact subset of $\mathbb{R}^{n_{\varphi}}$ with $n_{\varphi} = k + 2N$. The true value of $\boldsymbol{\varphi}$, denoted by $\boldsymbol{\varphi}_{\mathbf{0}} = (\boldsymbol{\theta}_{\mathbf{0}}', \boldsymbol{\phi}_{\mathbf{0}}', \boldsymbol{\sigma}_{\mathbf{0}}')'$, is an interior point of Θ_{φ}.

Assumption 5.

a. For a given sample size, T, and all values of $\boldsymbol{\varphi} \in \Theta_{\varphi}$, the redefined observation matrix $\mathbf{Z}(\boldsymbol{\varphi}) = [\bar{\mathbf{X}}(\boldsymbol{\phi}, \boldsymbol{\sigma}), \check{\boldsymbol{\xi}}(\boldsymbol{\theta}, \boldsymbol{\sigma}), \check{\mathbf{W}}(\boldsymbol{\sigma})]$ has a full column rank, where

$$\bar{\mathbf{X}}(\boldsymbol{\phi}, \boldsymbol{\sigma}) = -[(\phi_1/\sigma_1)\mathbf{X}_1', (\phi_2/\sigma_2)\mathbf{X}_2', \ldots, (\phi_N/\sigma_N)\mathbf{X}_N']'$$

and $\check{\mathbf{W}}(\boldsymbol{\sigma})$ and $\check{\boldsymbol{\xi}}(\boldsymbol{\theta}, \boldsymbol{\sigma})$ are block diagonal matrices with their ith blocks given by $\mathbf{W}_i/\sigma_i$ and $\boldsymbol{\xi}_i(\boldsymbol{\theta})/\sigma_i$.

b. When $\mathbf{x}_{it}$ are stationary, $T^{-1}\mathbf{Z}'(\boldsymbol{\varphi})\mathbf{Z}(\boldsymbol{\varphi})$ converges in probability to a positive definite matrix as $T \to \infty$.

c. When $\mathbf{x}_{it}$ are $I(1)$, as $T \to \infty, \mathbf{K}_z\mathbf{Z}'(\boldsymbol{\varphi})\mathbf{Z}(\boldsymbol{\varphi})\mathbf{K}_z$ weakly converges in probability to a positive definite random matrix with probability 1, where $\mathbf{K}_z = \text{diag}\,(T^{-1}\mathbf{I}_k, T^{-1/2}\mathbf{I}_N, T^{-1/2}\mathbf{I}_{N(p+kq)})$.

Assumption 5 sets out standard identification conditions and rules out the possibility of exact multicollinearity. In the case where $\mathbf{x}_{it}$ are $I(1)$, part c of this assumption ensures that $\mathbf{x}_{it}$ are not themselves cointegrated. Notice also that this assumption is weaker than is needed for identification of the group-specific long-run coefficients, $\boldsymbol{\theta}_i$, if the long-run homogeneity assumption 3 is not imposed.

Necessary conditions for parts b and c of assumption 5 to hold are given in the Appendix, Section A.1. In the case where $\mathbf{x}_{it}$ are $I(0)$, the pooled cross-product observation matrix on the regressors, $N^{-1}\sum_{i=1}^{N}(\phi_i^2/\sigma_i^2)T^{-1}\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i$, must converge in probability to a fixed positive definite matrix as $T \to \infty$. In the case where $\mathbf{x}_{it}$ are $I(1)$, $N^{-1}\sum_{i=1}^{N}(\phi_i^2/\sigma_i^2)T^{-2}\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i$ must weakly converge to a random positive definite matrix with probability 1 as $T \to \infty$. These conditions also cover the case where both T and N are large. The conditions should hold for all feasible values of ϕ_i and σ_i^2.

4. THE POOLED MEAN GROUP ESTIMATOR

Maximum likelihood (ML) estimation of the long-run coefficients, $\boldsymbol{\theta}$, and the group-specific error-correction coefficients, ϕ_i, can be computed by maximizing (8) with respect to $\boldsymbol{\varphi}$. These ML estimators (MLEs) are termed the *pooled mean group* (PMG) estimators to highlight both the pooling implied by the homogeneity restrictions on the long-run coefficients and the averaging across groups used to obtain means of the estimated error-correction coefficients and the other short-run parameters of the model.

The PMG estimators can be computed by the familiar Newton–Raphson algorithm, which makes use of both the first and second derivatives. Alternatively, they can be computed by a "back-substitution" algorithm that makes use of only the first derivatives of (8). In this case, setting the first derivatives of the concentrated log-likelihood function with respect to $\boldsymbol{\varphi}$ to $\mathbf{0}$ yields the following relations in $\hat{\boldsymbol{\theta}}$, $\hat{\phi}_i$, and $\hat{\sigma}_i^2$, which need to be solved iteratively:

$$\hat{\boldsymbol{\theta}} = -\left\{\sum_{i=1}^{N}\frac{\hat{\phi}_i^2}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i\right\}^{-1} \times \left\{\sum_{i=1}^{N}\frac{\hat{\phi}_i}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i(\Delta\mathbf{y}_i - \hat{\phi}_i\mathbf{y}_{i,-1})\right\}, \quad (9)$$

$$\hat{\phi}_i = (\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\hat{\boldsymbol{\xi}}_i)^{-1}\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\Delta\mathbf{y}_i, \qquad i = 1,2,\ldots,N, \quad (10)$$

and

$$\hat{\sigma}_i^2 = T^{-1}(\Delta\mathbf{y}_i - \hat{\phi}_i\hat{\boldsymbol{\xi}}_i)'\mathbf{H}_i(\Delta\mathbf{y}_i - \hat{\phi}_i\hat{\boldsymbol{\xi}}_i), \qquad i = 1,2,\ldots,N, \quad (11)$$

where $\hat{\boldsymbol{\xi}}_i = \mathbf{y}_{i,-1} - \mathbf{X}_i\hat{\boldsymbol{\theta}}$. To simplify the notations, we denote $\boldsymbol{\xi}_i(\hat{\boldsymbol{\theta}})$ by $\hat{\boldsymbol{\xi}}_i$. Starting with an initial estimate of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}^{(0)}$, estimates of ϕ_i and σ_i^2 can be computed using (10) and (11), which can then be substituted in (9) to obtain a new estimate of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}^{(1)}$, and so on until convergence is achieved.

In order to derive the asymptotic distribution of the PMG estimators we distinguish between the cases of stationary and nonstationary regressors, $\mathbf{x}_{it}$. Although in principle the same algorithm can be used to compute the PMG estimators irrespective of whether the regressors are $I(0)$ or $I(1)$, the underlying asymptotic theories for these two cases are fundamentally different and their derivations require separate treatments.

4.1 The Case of Stationary Regressors

In this case, under fairly standard conditions, the consistency and the asymptotic normality of the ML estimators of the parameters in (8) can be easily established. We set out the consistency and the asymptotic distribution of the MLEs of $\boldsymbol{\varphi}$, denoted by $\hat{\boldsymbol{\varphi}} = (\hat{\boldsymbol{\theta}}', \hat{\boldsymbol{\phi}}', \hat{\boldsymbol{\sigma}}')'$, in the following theorem.

Theorem 1. Under assumptions 1–4, and 5(a) and (b), and assuming that the regressors $\mathbf{x}_{it}$ are stationary, the MLE of φ in the dynamic heterogeneous panel data model (6) is consistent. Furthermore, as $T \to \infty$ for a fixed N, the PMG estimator of $\psi = (\boldsymbol{\theta}', \boldsymbol{\phi}')'$ has the following asymptotic distribution:

$$\sqrt{T}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) \overset{a}{\sim} \mathrm{N}\{\mathbf{0}, \mathbf{J}^{-1}(\boldsymbol{\psi}_0)\}, \quad (12)$$

where $\mathbf{J}(\boldsymbol{\psi}_0)$ is the $(k+N)\times(k+N)$ information matrix given by

$$\mathbf{J}(\boldsymbol{\psi}_0) \equiv \begin{bmatrix} \sum_{i=1}^{N}\frac{\phi_{i0}^2}{\sigma_{i0}^2}\,\mathbf{Q}_{X_iX_i} & -\frac{\phi_{10}}{\sigma_{10}^2}\,\mathbf{Q}_{X_1\xi_{10}} & \cdots & -\frac{\phi_{N0}}{\sigma_{N0}^2}\,\mathbf{Q}_{X_N\xi_{N0}} \\ -\frac{\phi_{10}}{\sigma_{10}^2}\,\mathbf{Q}_{X_1\xi_{10}}' & \frac{1}{\sigma_{10}^2}\,q_{\xi_{10}\xi_{10}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{\phi_{N0}}{\sigma_{N0}^2}\,\mathbf{Q}_{X_N\xi_{N0}}' & 0 & \cdots & \frac{1}{\sigma_{N0}^2}\,q_{\xi_{N0}\xi_{N0}} \end{bmatrix}, \quad (13)$$

with $\mathbf{Q}_{X_iX_i}$, $\mathbf{Q}_{X_i\xi_{i0}}$, and $q_{\xi_{i0}\xi_{i0}}$ being the probability limits of $T^{-1}\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i$, $T^{-1}\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\xi}_{i0}$, and $T^{-1}\boldsymbol{\xi}_{i0}'\mathbf{H}_i\boldsymbol{\xi}_{i0}$.

Proof. In this case the proof can be established using familiar results. (Details can be found in Pesaran, Shin, and Smith 1997, appendix.)

Remark 1. Focusing on the long-run coefficients, for a fixed N as $T \to \infty$ the pooled MLE $\hat{\boldsymbol{\theta}}$, defined by (9), is asymptotically distributed as

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{a}{\sim} \mathrm{N}\left\{\mathbf{0}, \left[\sum_{i=1}^{N}\frac{\phi_{i0}^2}{\sigma_{i0}^2}\,(\mathbf{Q}_{X_iX_i} - \mathbf{Q}_{X_i\xi_{i0}}q_{\xi_{i0}\xi_{i0}}^{-1}\mathbf{Q}_{X_i\xi_{i0}}')\right]^{-1}\right\}. \quad (14)$$

4.2 The Case of Nonstationary Regressors

Here the underlying regressors are assumed to follow integrated processes of order 1, namely $I(1)$. The asymptotic analysis in this case is more complicated, due to the fact that the MLEs of the long-run and the short-run parameters, $\boldsymbol{\theta}$ and $(\boldsymbol{\phi}', \boldsymbol{\sigma}')'$, converge to their true values at different rates, and the mere consistency of the MLEs is not sufficient to guarantee the weak convergence of the sample information matrix to its true value. In particular, to ensure such a convergence, the sample information matrix should satisfy the stochastic equicontinuity condition. (For a general treatment of these issues, see Pesaran and Shin 1999; Saikkonen 1995.)

The following theorem establishes the consistency, the relative rates of convergence, and the asymptotic distribution of the MLE of $\boldsymbol{\varphi}$.

Theorem 2. Under Assumptions 1–4 and 5(a) and (c), and assuming that $\mathbf{x}_{it}$ are $I(1)$, the MLE of the short-run coefficients $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$ in the dynamic heterogeneous panel data model (6) are $\sqrt{T}$ consistent and the MLE of $\boldsymbol{\theta}$ is T consistent, namely

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(T^{-1/2}), \qquad \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 = o_p(1)$$

and

$$\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0 = o_p(1). \tag{15}$$

Furthermore, for a fixed N and as $T \to \infty$, the MLE of $\boldsymbol{\psi} = (\boldsymbol{\theta}', \boldsymbol{\phi}')'$ asymptotically has the mixture-normal distribution

$$\mathbf{D}_{\psi}^{-1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) \overset{a}{\sim} \mathrm{MN}\{\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\psi}_0)\}, \tag{16}$$

where $\mathbf{D}_{\psi} = \mathrm{diag}(T^{-1}\mathbf{I}_k, T^{-1/2}\mathbf{I}_N)$ and $\mathcal{I}(\boldsymbol{\psi}_0)$ is the random information matrix defined by (A.20).

Proof. See the Appendix, Section A.2.

Remark 2. More specifically, the pooled MLE $\hat{\boldsymbol{\theta}}$, defined by (9), has the following large T asymptotic distribution:

$$T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{a}{\sim} \mathrm{MN}\left\{\mathbf{0}, \left(\sum_{i=1}^{N} \frac{\phi_{i0}^2}{\sigma_{i0}^2}\, \mathbf{R}_{X_i X_i}\right)^{-1}\right\}, \tag{17}$$

where $\mathbf{R}_{X_i X_i}, i = 1, 2, \ldots, N$, are the random probability limits defined by (A.1). Notice that it is not required that all the matrices $\mathbf{R}_{X_i X_i}$ be positive definite. It is, however, easily seen that by part (c) of Assumption 5, $\sum_{i=1}^{N} (\phi_{i0}^2/\sigma_{i0}^2)\mathbf{R}_{X_i X_i}$ is a positive definite matrix with probability 1. Also note that in the case of $I(1)$ regressors, the MLEs of the long-run and short-run parameters are asymptotically distributed independently of each other. [For a proof of (17), see Section A.3 in the Appendix.]

Once the pooled MLE of the long-run parameters, $\hat{\boldsymbol{\theta}}$, is successfully computed, the short-run coefficients, including the group-specific error-correction coefficients, ϕ_i can be consistently estimated by running the individual ordinary least squares (OLS) regressions of $\Delta \mathbf{y}_i$ on $(\hat{\boldsymbol{\xi}}_i, \mathbf{W}_i), i = 1, \ldots, N$, where $\hat{\boldsymbol{\xi}}_i = \mathbf{y}_{i,-1} - \mathbf{X}_i \hat{\boldsymbol{\theta}}$. The covariance matrix of the MLEs, $(\hat{\boldsymbol{\theta}}', \hat{\phi}_1, \ldots, \hat{\phi}_N, \hat{\boldsymbol{\kappa}}_1', \ldots, \hat{\boldsymbol{\kappa}}_N')'$, is then consistently estimated by the inverse of

$$\begin{bmatrix}
\sum_{i=1}^{N} \frac{\hat{\phi}_i^2 \mathbf{X}_i' \mathbf{X}_i'}{\hat{\sigma}_i^2} & \frac{-\hat{\phi}_1 \mathbf{X}_1' \hat{\xi}_1}{\hat{\sigma}_1^2} & \cdots & \frac{-\hat{\phi}_N \mathbf{X}_N' \hat{\xi}_N}{\hat{\sigma}_N^2} & \frac{-\hat{\phi}_1 \mathbf{X}_1' \mathbf{W}_1}{\hat{\sigma}_1^2} & \cdots & \frac{-\hat{\phi}_1 \mathbf{X}_N' \mathbf{W}_N}{\hat{\sigma}_N^2} \\
 & \frac{\hat{\xi}_1' \hat{\xi}_1}{\hat{\sigma}_1^2} & \cdots & 0 & \frac{\hat{\xi}_1' \mathbf{W}_1}{\hat{\sigma}_1^2} & \cdots & 0 \\
 & & \vdots & \vdots & & \ddots & \vdots \\
 & & & \frac{\hat{\xi}_N' \hat{\xi}_N}{\hat{\sigma}_N^2} & 0 & \cdots & \frac{\hat{\xi}_N' \mathbf{W}_N}{\hat{\sigma}_N^2} \\
 & & & & \frac{\mathbf{W}_1' \mathbf{W}_1}{\hat{\sigma}_1^2} & \cdots & 0 \\
 & & & & & \vdots & \vdots \\
 & & & & & & \frac{\mathbf{W}_N' \mathbf{W}_N}{\hat{\sigma}_N^2}
\end{bmatrix}.$$

4.3 The Case of Large *T* and Large *N*

In this case the mean of the error correction coefficients and the other short-run parameters can be estimated consistently by the unweighted average of the individual coefficients (or the MG estimators):

$$\hat{\phi}_{\mathrm{MG}} = N^{-1} \sum_{i=1}^{N} \hat{\phi}_i, \qquad \hat{\boldsymbol{\kappa}}_{\mathrm{MG}} = N^{-1} \sum_{i=1}^{N} \hat{\boldsymbol{\kappa}}_i.$$

The variance of these estimators can be consistently estimated along the lines suggested by Pesaran, Smith, and Im (1996). For example, in the case of $\hat{\phi}_{\mathrm{MG}}$, a consistent estimator of the variance of $\hat{\phi}_{\mathrm{MG}}$ is given by

$$\hat{\Delta}_{\hat{\phi}} = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\phi}_i - \hat{\phi}_{\mathrm{MG}})^2.$$

Alternatively, one can use the weighted estimator proposed by Swamy (1970) in the context of the static random coefficient models. But as was shown by Hsiao, Pesaran, and Tahmiscioglu (1999) and noted earlier in this article, the Swamy estimator (also known as the empirical Bayes estimator) and the MG estimator are asymptotically equivalent as $T \to \infty$ and $N \to \infty$ such that $\sqrt{N}/T \to 0$. Under these conditions, the asymptotic distribution of $\hat{\phi}_{\mathrm{MG}}$, for example, is given by

$$\sqrt{\mathrm{N}}(\hat{\phi}_{\mathrm{MG}} - \phi) \overset{a}{\sim} \mathrm{N}(0, \Delta_\phi),$$

where $\phi = E(\phi_i)$ and $\Delta_\phi = \mathrm{var}(\phi_i)$.

For the common long-run coefficients, $\boldsymbol{\theta}$, the pooled MLE is consistent as long as $T \to \infty$, irrespective of whether N is large or not. To obtain the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ when N is large, we need to assume that N is a monotonic function of T, say $\mathrm{N}(T)$, such that $N \to \infty$ *only* as $T \to \infty$. In this setting the rate of convergence of $\hat{\boldsymbol{\theta}}$ toward its true value is given by $\sqrt{NT}$ in the case where the regressors are $I(0)$. For this result to hold, the limit of

$$N^{-1} \sum_{i=1}^{N} \frac{\phi_{i0}^2}{\sigma_{i0}^2} \left(\mathbf{Q}_{X_i X_i} - \mathbf{Q}_{X_i \xi_{i0}} q_{\xi_{i0}\xi_{i0}}^{-1} \mathbf{Q}'_{X_i \xi_{i0}}\right)$$

as $N \to \infty$ must be a positive definite matrix [see (14)]. When the regressors are $I(1)$, the rate of convergence of $\hat{\boldsymbol{\theta}}$ is given by $T\sqrt{N}$ as long as

$$N^{-1} \sum_{i=1}^{N} \frac{\phi_{i0}^2}{\sigma_{i0}^2} \mathbf{R}_{X_i X_i}$$

tends to a positive definite matrix with probability 1, as $N \to \infty$ [see (17)]. Notice that $\hat{\boldsymbol{\theta}}$ will not be consistent for finite T, even if $N \to \infty$.

5. EMPIRICAL APPLICATIONS

Before considering the two examples, we briefly discuss three modeling issues. The first issue is that although the aforementioned theory assumes that all long-run coefficients are the same across groups, it is straightforward to adapt the PMG estimator to allow only a subset of the long-run coefficients to be the same while the others differ. Details have been given in the earlier version of this article (Pesaran, Shin, and Smith 1997). The second issue is that we need to choose between the alternative specifications. Tests of homogeneity of error variances and/or short- or long-run slope coefficients can be easily carried out using likelihood ratio and other classical statistical tests, because the PMG and fixed-effects estimators are restricted versions of the set of individual group equations. Although it is common to use pooled estimators without testing the implied restrictions, in the case of cross-country studies the likelihood ratio tests usually reject equality of error variances and/or slopes (short-run or long-run) at conventional significance levels. This is the case in both of our examples. The interpretation of this feature has been discussed extensively by Pesaran, Smith, and Akiyama (1998), and we return to this work in our conclusion. An alternative to likelihood ratio tests would be to use Hausman (1978)-type tests. The MG estimator provides consistent estimates of the mean of the long-run coefficients, though these will be inefficient if slope homogeneity holds. Under long-run slope homogeneity, the pooled estimators are consistent and efficient. Therefore, the effect of heterogeneity on the means of the coefficients can be determined by a Hausman-type test applied to the difference between the MG and the PMG or the fixed-effects estimators. (The details of such tests were discussed in Pesaran, Smith and Im 1996.) The third issue is that although the MG estimator is consistent for large N and T, for small T the familiar lagged dependent variable bias causes the estimates of λ_i and $\phi_i = (\lambda_i - 1)$ to underestimate their true values. Large N does not help with this problem, as all of the estimates are biased in the same direction. Pesaran and Zhao (1999) proposed a bias-corrected MG estimator that directly adjusts the long-run coefficient by an estimate of its bias. In the case of the pooled estimators, the downward-lagged dependent variable bias, which they also suffer, may offset the upward heterogeneity bias discussed by Pesaran and Smith (1995). In empirical applications it may be difficult to judge the relative effect of the two biases, making inference about the speed of adjustment difficult. This is a feature of both of our examples.

5.1 The Consumption Function in the OECD

The first example that we examine is a standard consumption function of the Davidson et al. (1978) type for a sample of OECD countries. Similar specifications have also been estimated for a number of developing countries by Haque and Montiel (1989). We assume that the long-run consumption function is given by

$$c_{it} = \theta_{0i} + \theta_{1i} y_{it}^d + \theta_{2i}\pi_{it} + u_{it},$$
$$i = 1, 2, \ldots, N, \qquad t = 1, 2, \ldots, T,$$

where c_{it} is the logarithm of real consumption per capita, y_{it}^d is the logarithm of real per capita disposable income, and π_{it} is the rate of inflation. Most theories of aggregate consumption would suggest that $\theta_{1i} = 1$. The PMG estimation procedure allows us to estimate a common long-run coefficient and test whether it is unity. The inflation variable, π_{it}, is a proxy for various wealth effects, and we would expect $\theta_{2i} < 0$. We assume that all of these variables are $I(1)$ and cointegrated, making u_{it} an $I(0)$ process for all i. In this application we take the maximum lag as being 1; thus the autoregressive distributed lag (ARDL) $(1, 1, 1)$ equation is

$$c_{it} = \mu_i + \delta_{10i} y_{it}^d + \delta_{11i} y_{i,t-1}^d + \delta_{20i}\pi_{it} + \delta_{21i}\pi_{i,t-1} + \lambda_i c_{i,t-1} + \varepsilon_{it},$$

and the error correction equation is

$$\Delta c_{it} = \phi_i(c_{i,t-1} - \theta_{0i} - \theta_{1i} y_{it} - \theta_{2i}\pi_{it}) - \delta_{11i}\Delta y_{it}^d - \delta_{21i}\Delta\pi_{it} + \varepsilon_{it},$$

where

$$\theta_{0i} = \frac{\mu_i}{1 - \lambda_i}, \qquad \theta_{1i} = \frac{\delta_{10i} + \delta_{11i}}{1 - \lambda_i},$$

$$\theta_{2i} = \frac{\delta_{20i} + \delta_{21i}}{1 - \lambda_i}, \qquad \phi_i = -(1 - \lambda_i).$$

The foregoing error correction equations are written in terms of current, rather than lagged levels of the exogenous regressors. This allows an ARDL$(1, 0, 0)$ as a special case.

We measure consumption by the logarithm of total real private consumption per capita and inflation by the change in the logarithm of the consumption deflator. The initial measure of income that we use is national disposable per capita income deflated by the consumption deflator, NDI. (The data are taken from the OECD National Accounts Statistics and were collected as part of an IMF project on international saving rates.) Using 1960 and 1961 to create lags, we have data for 32 years (1962–1993) for 23 countries and 31 years (1962–1992) for Belgium.

First, a common ARDL$(1,1,1)$ was run for each country separately. The results are given in the earlier version of this article (Pesaran, Shin, and Smith 1997). For a long-run relationship to exist, we require that $\phi_i \neq 0$. It is clear that in a number of countries, the hypothesis of no long-run relation would not be rejected. (Pesaran, Shin, and R. J. Smith [1999] develop a test for the existence of a long-run relationship in ARDLs of this form.) The long-run income elasticities range from .62 in Switzerland to 1.19 in the United States and are significantly less than unity in nine countries and significantly greater than unity in three (Italy, the United Kingdom, and the United States). The long-run inflation coefficient is more dispersed, ranging between 1.46 in Germany and -1.25 in Canada. However, it is negative in all but four countries. More than 60% of the change in the logarithm of per capita consumption is explained in all but three countries: Luxembourg, Norway, and Austria. The standard error of regression varies from .5% in France to 3.3% in Turkey, and the equality of error variances does not seem to be an appropriate assumption—a result born out by formal statistical tests. At the 5% level, there is evidence of serial correlation in the equations for two of the 24 countries, and of functional form misspecification in three, nonnormal errors in two, and heteroscedasticity in one. The tests have been described by Pesaran and Pesaran (1997). The RESET tests for functional form and the test of heteroscedasticity are based on the fitted values for the changes in the logarithm of per capita consumption. The fact that 17 of the 24 country-specific equations show no evidence of misspecification is reassuring.

Table 1 presents the alternative pooled estimates: MG, which imposes no restrictions; PMG, which imposes common long-run effects; and DFE, which constrains all of the slope coefficients and error variances to be the same. The PMG computations were carried out using the Newton-Raphson algorithm in a program written in Gauss. The program and data are available on *http://www.econ.cam.ac.uk/faculty/pesaran.* The program allows the user to start the iterations with different sets of initial values, depending on the nature of the ARDL specification across the groups. When the same ARDL is estimated across the countries, the MG and DFE estimates of the long-run coefficients can be used as initial values. In the examples considered here, these resulted in the same maximum for the pooled log-likelihood function. This has not been the case in some other examples.

Table 1. Alternative Pooled Estimates for ARDL(1, 1, 1) Consumption Functions for OECD Countries Over the Period 1962–1993 Using National Disposable Income

	MG	PMG	DFE
Income elasticity, θ_1	.918 (.027)	.904 (.010)	.912 (.045)
Inflation effect, θ_2	$-.353$ (.117)	$-.466$ (.063)	$-.266$ (.099)
Speed of adjustment, ϕ	$-.306$ (.030)	$-.200$ (.032)	$-.179$ (.042)
Maximized log likelihood	2,390	2,247	1,999
Number of parameters estimated	168	122	30

NOTE: Figures in parentheses are asymptotic standard errors. In the case of the dynamic fixed effects estimates the standard errors are corrected for possible heteroscedasticity in the cross-sectional error variances.

The DFE standard errors are corrected for the heteroscedasticity of error variances across countries; the uncorrected ones are substantially smaller. The heteroscedasticity robust standard errors are computed allowing for a general covariance matrix of the disturbances ε_{it} across i (see, e.g., Baltagi 1995, pp. 12–13). The three estimates of the long-run income elasticities are very similar, close to .91 and significantly less than unity. The long-run inflation coefficients are all significantly negative, though the estimates differ somewhat. As the econometric theory suggests, imposing homogeneity causes an upward bias in the coefficient of the lagged dependent variable, and the MG estimate suggests much faster adjustment than the PMG or DFE. The MG estimate of the adjustment coefficient also has the smallest standard error.

Imposing long-run homogeneity reduces the standard errors of the long-run coefficients, but does not change the estimates very much. This is confirmed by the Hausman test statistic of 1.19, which is $\chi^2(2)$ under the null hypothesis of no difference between the MG and PMG estimators. As is clear from Table 1, likelihood ratio tests at conventional significance levels, on the other hand, would strongly reject all of the restrictions, including homogeneity of long-run coefficients. This is true for both income and inflation. If homogeneity of just the long-run income elasticities is imposed (23 restrictions), then the maximized log-likelihood (MLL) falls from 2,390 to 2,270. If homogeneity of both income and inflation long-run effects is imposed (46 restrictions), then the MLL falls to 2,247.

One advantage of the PMG over the traditional DFE model is that it can allow the short-run dynamic specification to differ from country to country. The lag order was first chosen in each country on the unrestricted model by the Schwarz Bayesian criterion (SBC), subject to a maximum lag of 1. Then, using these SBC-determined lag orders, homogeneity was imposed. The most common choice, in half of the countries, was an ARDL$(1,1,0)$, one lag of consumption and income but only current inflation. Performance on the misspecification tests and estimates of the long-run effects were very similar to those obtained using an ARDL(1, 1, 1) model. Again, the Hausman test statistic does not reject equality between the MG and PMG estimates, whereas the likelihood ratio test statistic of 281 rejects equality of all of the long-run coefficients. Notice that under the null hypothesis, this statistic is asymptotically distributed as $\chi^2(46)$. In this example, estimates of the long-run coefficients are robust to the order of the ARDL. This is partly because T is large; as shown in the next example, when T is small, the choice of lag order is more important.

Although data for national disposable income is widely available, it is not the most appropriate theoretical measure of income. Thus we also estimated an equation using a better measure, private disposable labor income (interest and dividend income are excluded). Data availability reduces the sample to 15 countries, with some countries having as few as 8 observations (the Netherlands). The characteristics of these data allow us to investigate the sensitivity of the estimators to a number of interesting features. The panel is very unbalanced, with T very small in some cases; in these cases, choosing lag order by the SBC is more reliable than using a common ARDL(1, 1, 1), which tends to show misspecification. There is also an extreme outlier. In Sweden, where the SBC chose ARDL(1, 1, 0), the speed of adjustment was $-.03$ (.154), the long-run income elasticity was -3.74 (25.22), and the long-run inflation coefficient was -7.768 (40.36). Although this equation passes the four misspecification tests, the fit is very poor, and it distorts the MG estimates.

Table 2 presents the alternative pooled estimates of the long-run income elasticity for three groups of countries ($N = 15, 14$, and 9), for three ARDL specifications. The $N = 15$ group includes all the countries for which there are data on private disposable labor income. The $N = 14$ group excludes Sweden. The $N = 9$ group also excludes countries for which T is 13 or less (Italy, Netherlands, New Zealand, Portugal, Spain). The effect of Sweden on the MG estimator in the $N = 15$ group is very obvious, completely distorting the ARDL$(1, 1, 1)$ estimate and severely biasing the SBC estimate. The PMG estimates are much less sensitive, as one would expect, given the large standard errors on the Swedish estimates. Dropping Sweden from the sample hardly changes the PMG estimate, but changes the MG substantially. Again, the Hausman test statistic does not reject equality between the MG and PMG, whereas the likelihood ratio test statistic rejects equality of all of the long-run coefficients.

These results indicate that the PMG seems quite robust to outliers and to choice of lag order. Overall the estimates suggest that the average long-run income elasticity is rather larger than .9, but significantly less than unity on PMG standard errors. However, these standard errors may be underestimates. The average long-run inflation effect is significantly less than 0. These general conclusions seem robust to lag order, sample, and estimation method. The average adjustment coefficient seems to be significantly less than 0 and greater than $-.5$, but the estimates of the adjustment coefficients tend to show more sensitivity to the choice of the estimation method. The example of the consumption function was chosen to demonstrate the PMG estimator because of its familiarity and importance. However, it should be recognized that modeling consumption raises many difficult issues not captured by this simple model. Clearly, some important variables are omitted, including initial wealth, real interest rate effects, credit regulations, government budget surpluses (deficits), and demographics. This type of equation has proved structurally unstable in a number of the countries in our sample. In many of the countries one would not be able to reject the null hypothesis of no long-run relationship, though these tests have low power. The MG estimates are also subject to small T bias because of the familiar lagged dependent variable problem.

Table 2. Alternative Pooled Estimates of the Long-Run Consumption Income Elasticity Using per Capita Real Private Disposable Labor Income

N	ARDL	MG	PMG	DFE
15	1, 0, 0	.948 (.070)	.933 (.010)	.936 (.022)
15	1, 1, 1	−12.8 (13.8)	.918 (.016)	.935 (.023)
15	SBC	.669 (.319)	.938 (.013)	
14	1, 0, 0	.999 (.052)	.934 (.010)	.94 (.012)
14	1, 1, 1	.918 (.066)	.919 (.016)	.943 (.021)
14	SBC	.985 (.054)	.939 (.013)	
9	1, 0, 0	.993 (.060)	.926 (.011)	.931 (.014)
9	1, 1, 1	.916 (.100)	.927 (.015)	.925 (.025)
9	SBC	.995 (.059)	.941 (.013)	

NOTE: Figures in parentheses are asymptotic standard errors. In the case of the dynamic fixed effects estimates the standard errors are corrected for possible heteroscedasticity in the cross-sectional error variances.

5.2 Energy Demand in Asian Developing Countries

Our second example involves explaining the logarithm of energy consumption (measured in tons of oil equivalents [TOE] per capita), e_{it}, by the logarithm of real per capita output in international prices taken from the Penn World Tables, y_{it}, and by the logarithm of real energy prices (average energy prices per TOE in local currency, deflated by the consumer price index), p_{it}. The data are for 10 Asian developing countries for the years 1974–1990; thus both N and T are small. This example is drawn from Pesaran, Smith, and Akiyama (1998), who discussed the data in more detail and investigated functional form, dynamics, disaggregation, and various other issues.

The initial unrestricted model is a partial adjustment (ARDL(1, 0, 0)) model:

$$\Delta e_{it} = \phi_i(e_{it-1} - \theta_{0i} - \theta_{1i} y_{it} - \theta_{2i} p_{it}) + \varepsilon_{it}, \quad i = 1, 2, \ldots, 10.$$

Partial adjustment and homogeneity of degree zero in prices were not rejected in all but one of the 10 countries (Taiwan). In the country-specific regressions, the long-run income and price elasticity estimates for Malaysia, Sri Lanka, and Bangladesh are implausible, which could be due to data inadequacies and the high level of aggregation. Apart from these, the estimates are reasonable. The long-run income elasticity ranges from .835 to 1.564, with an average of 1.12; the long-run price elasticities range from .05 to $-.488$, with an average estimate of $-.22$. All of the estimates of the adjustment coefficients, ϕ_i, are negative and fall in the range $-.132$ to $-.825$. The standard error of the regression varies from 1.7% in the case of Thailand to 6.3% in Sri Lanka; so again, the assumption of constant error variances across countries seems inappropriate. Diagnostic tests reveal few problems: one failure for serial correlation and one for normality at the 5% level. When the order of each lag is chosen by the SBC for each country, starting with a general ARDL(1, 1, 1) specification, the ARDL(1, 0, 0) is chosen for seven countries, a static model is chosen for India and Sri Lanka, and an unstable ARDL(1, 0, 1) is chosen for Taiwan.

The alternative pooled estimators for the ARDL(1, 0, 0) specification are presented in Table 3. The table presents es-

Table 3. Alternative Pooled Estimators for ARDL(1, 0, 0) Energy Demand Function for 10 Asian Economies Over the Period 1974–1990

	MG estimators	PMG estimators	DFE estimators	Static fixed-effects estimators
Output	1.228	1.184	1.301	1.009
Elasticity ($\hat{\theta}_1$)	(.183)	(.039)	(.109)	(.037)
Price	−.261	−.339	−.365	−.067
Elasticity ($\hat{\theta}_2$)	(.118)	(.033)	(.097)	(.030)
Adjustment	−.524	−.298	−.235	−1
Coefficient ($\hat{\phi}$)	(.070)	(.063)	(.040)	(N/A)
Log-likelihood	347.12	322.79	288.36	186.95
$N \times T$	170	170	170	170
No. of estimated parameters	50	32	15	13

NOTE: Figures in parentheses are asymptotic standard errors. In the case of the dynamic fixed effects estimates the standard errors are corrected for possible heteroscedasticity in the cross-sectional error variances.

Table 4. Alternative Pooled Estimators for ARDL-SBC Energy Demand Function

	MG estimators	PMG estimators	DFE estimators ARDL(1, 1, 1)
Output	1.277	1.171	1.181
Elasticity ($\hat{\theta}_1$)	(.182)	(.036)	(.147)
Price	−.160	−.301	−.324
Elasticity ($\hat{\theta}_2$)	(.164)	(.028)	(.0137)
Adjustment	−.502	−.417	−.153
Coefficient ($\hat{\phi}$)	(.113)	(.121)	(.042)
Log-likelihood	356.81	315.41	300.27
$N \times T$	170	170	170
No. of estimated parameters	51	33	17

NOTE: Figures in parentheses are asymptotic standard errors. In the case of the dynamic fixed effects estimates the standard errors are corrected for possible heteroscedasticity in the cross-sectional error variances.

timates of the long-run output and price elasticities and the adjustment coefficients, together with the MLL, the number of observations, and the number of parameters estimated (including the error variances). As in the previous example, likelihood ratio tests reject the hypothesis of equality of long-run coefficients, whereas Hausman tests do not reject this hypothesis when the PMG and MG estimators are compared. Because the MG estimates have large standard errors, the Hausman test is likely to have low power. If the focus of the analysis is on average (across countries) income and price elasticities, then the PMG estimates are probably preferable to the MG estimates on the grounds of their better precision and the fact that they are less sensitive to outlier estimates. In this context, the Hausman test can be seen as providing formal statistical evidence that we are not in gross violation of the data by relying on PMG estimates rather than the MG estimates. The estimates suggest a long-run income elasticity of 1.18 (.039) and a price elasticity of −.34 (.033). The PMG estimates are close to the MG estimates, and the DFE estimates are rather larger. The dynamics clearly matter, and the static fixed-effects estimator has an insignificant price effect. The standard errors of both the PMG and the DFE are very much smaller than those of the MG; pooling sharpens the estimates considerably. As before, pooling also leads to a much smaller estimated speed of adjustment; the MG estimates suggest speeds of convergence to equilibrium of around 50% per year; the PMG, 30%; and the DFE, about 20%.

Table 4 gives MG and PMG estimates when the order of the specification in the individual countries is chosen by the SBC. This procedure cannot be used with the fixed-effects estimator; the ARDL(1, 1, 1) estimates are given instead. Again, the PMG estimates of the long-run price and income elasticities hardly change, though the estimated speed of adjustment is rather higher, partly because the SBC chooses the static model, with instantaneous adjustment, in the case of some countries. Using the SBC gives a substantial improvement in the MLL for the MG estimator (at 357 it is close to the ARDL(1, 1, 1) MLL of 362, with far fewer parameters) but less for the PMG, where the SBC MLL is worse than the ARDL(1, 0, 0) MLL. It is clear that homogeneity restrictions and dynamic specification interact in a complex way. What might be the optimal order for the country-specific estimates may not be the optimal order when cross-country homogeneity restrictions are imposed. The rather serious policy implications of an income elasticity around 1.2 and a price elasticity around −.3 have been discussed by Pesaran, Smith, and Akiyama (1998).

6. CONCLUDING REMARKS

The PMG estimator, which assumes homogeneous long-run coefficients, provides a useful intermediate alternative between estimating separate regressions, which allows all coefficients and error variances to differ across the groups, and conventional fixed-effects estimators, which assume that all slope coefficients and error variances are the same. It has the practical advantage in allowing the short-run dynamics to be data determined for each country, taking into account the number of time series observations available in each case. A number of unresolved issues remain, however. For small T, all of the estimators (group-specific, MG, PMG, and fixed-effects) will be subject to the familiar downward bias on the coefficient of the lagged dependent variable. Because the bias is in the same direction for each group, averaging or pooling does not reduce this bias. Bias corrections are available in the literature (e.g., Kiviet and Phillips 1993), but these apply to the short-run coefficients. Because the long-run coefficient is a nonlinear function of the short-run coefficients, procedures that remove the bias in the short-run coefficients can leave the long-run coefficient biased. Pesaran and Zhao (1999) discussed how the bias in the long-run coefficient can be reduced. The MG estimator used in this article is a simple unweighted mean of the coefficients. Weighted means are an obvious alternative; Hsiao et al. (1999) have considered Bayes estimation of the means of short-run coefficients in dynamic panel data models. Estimation in this article was conducted under the assumption that a long-run relationship existed. It would be useful to have a panel test for the existence of a long-run relationship when the variables are $I(1)$, similar to the panel unit root test suggested by Im, Pesaran, and Shin (1997).

Perhaps the most important issue is interpretation of the heterogeneity. Most studies that estimate separate relationships for a number of groups find differences in coefficients that are not only statistically significant, but also economically implausible. In our empirical examples such cases were the income elasticity of consumption out of labor income in Sweden and the income elasticity of demand for energy in Sri Lanka. Despite this, the MG and PMG estimates of the long-run coefficients tend to be sensible. One explanation is that the group-specific estimates are biased because of sample-specific omitted variables or measurement errors correlated with the regressors. If one is estimating an equation for a single group, then one might experiment with different specifications or alternative data until plausible estimates are obtained. When estimating equations for large numbers of groups, where no other data is available in many cases, this is not an option. If the coefficients really are the same and the bias-inducing correlations are not systematic (i.e., they average to 0 over groups), then pooled estimation will be appropriate despite the homogeneity restrictions being rejected. However, there is no obvious way of determining from the data that this is the case. This raises obvious problems for inference that require further analysis.

APPENDIX: MATHEMATICAL PROOFS

To save space, we give the proofs only for cases where the regressors are $I(1)$. Proof for the $I(0)$ case is given in the appendix to the earlier version of this article (Pesaran et al. 1997).

A.1 Preliminary Results

A.1.1 Some Useful Probability Limits

To simplify the notations in what follows, we denote $\boldsymbol{\xi}_i(\boldsymbol{\theta})$ by $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_i(\boldsymbol{\theta}_0)$ by $\boldsymbol{\xi}_{i0}$. When $\mathbf{x}_{it}$ are all $I(1)$, the following probability limits exist as $T \to \infty$, for each i:

$$T^{-2}\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i \Rightarrow \mathbf{R}_{X_iX_i}, \qquad T^{-1}\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\xi}_{i0} \Rightarrow \mathbf{R}_{X_i\xi_{i0}},$$
$$T^{-1}\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i = O_p(1), \tag{A.1}$$

where $\mathbf{R}_{X_iX_i}$ and $\mathbf{R}_{X_i\xi_{i0}}$ are $O_p(1)$ random matrices. We also have

$$T^{-1/2}\boldsymbol{\xi}_{i0}'\mathbf{H}_i\boldsymbol{\varepsilon}_i = O_p(1), \qquad T^{-1}\boldsymbol{\xi}_{i0}'\mathbf{H}_i\boldsymbol{\xi}_{i0} \xrightarrow{p} q_{\xi_{i0}\xi_{i0}},$$

and

$$T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i = \sigma_{i0}^2 + o_p(1), \qquad i = 1, \ldots, N. \tag{A.2}$$

A.1.2 Decomposition of the Log-Likelihood Ratio

Using (8), it is easily seen that

$$T^{-1}[l_T(\boldsymbol{\varphi}_0) - l_T(\boldsymbol{\varphi})] = \frac{1}{2}\,(\mathcal{A}_T + \mathcal{B}_T), \tag{A.3}$$

where

$$\mathcal{A}_T = \mathcal{A}_\sigma + \mathcal{C}_T, \tag{A.4}$$

$$\mathcal{A}_\sigma = \sum_{i=1}^N \left(\frac{\sigma_{i0}^2}{\sigma_i^2} - \ln\frac{\sigma_{i0}^2}{\sigma_i^2} - 1\right),$$

$$\mathcal{C}_T = \sum_{i=1}^N \left(\frac{1}{\sigma_i^2} - \frac{1}{\sigma_{i0}^2}\right)\left(\frac{\boldsymbol{\varepsilon}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i}{T} - \sigma_{i0}^2\right), \tag{A.5}$$

and

$$\mathcal{B}_T = T^{-1}\sum_{i=1}^N \frac{1}{\sigma_i^2}\,[(\Delta\mathbf{y}_i - \phi_i\boldsymbol{\xi}_i)'\mathbf{H}_i(\Delta\mathbf{y}_i - \phi_i\boldsymbol{\xi}_i) - \boldsymbol{\varepsilon}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i]. \tag{A.6}$$

Also, using the result

$$\begin{aligned}\phi_i\boldsymbol{\xi}_i &= (\phi_i - \phi_{i0} + \phi_{i0})(\boldsymbol{\xi}_i - \boldsymbol{\xi}_{i0} + \boldsymbol{\xi}_{i0}) \\ &= -\phi_i\mathbf{X}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + (\phi_i - \phi_{i0})\boldsymbol{\xi}_{i0} + \phi_{i0}\boldsymbol{\xi}_{i0},\end{aligned}$$

and noting that $\Delta\mathbf{y}_i = \phi_{i0}\boldsymbol{\xi}_{i0} + \mathbf{W}_i\boldsymbol{\kappa}_{i0} + \boldsymbol{\varepsilon}_i$, we have

$$\Delta\mathbf{y}_i - \phi_i\boldsymbol{\xi}_i = \boldsymbol{\varepsilon}_i + \mathbf{W}_i\boldsymbol{\kappa}_{i0} + \phi_i\mathbf{X}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - (\phi_i - \phi_{i0})\boldsymbol{\xi}_{i0}. \tag{A.7}$$

[See (6), and note that under the data-generating process, $\boldsymbol{\varphi} = \boldsymbol{\varphi}_0$.] Substituting (A.7) in (A.6), and after some algebra, we have

$$\begin{aligned}\mathcal{B}_T &= (\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\left(\sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2}\,\frac{\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i}{T}\right)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &\quad + \sum_{i=1}^N \left[\frac{(\phi_i - \phi_{i0})^2}{\sigma_i^2}\right]\left(\frac{\boldsymbol{\xi}_{i0}'\mathbf{H}_i\boldsymbol{\xi}_{i0}}{T}\right) \\ &\quad - 2\sum_{i=1}^N \left[\frac{\phi_i(\phi_i - \phi_{i0})}{\sigma_i^2}\right]\left(\frac{\boldsymbol{\xi}_{i0}'\mathbf{H}_i\mathbf{X}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{T}\right) \\ &\quad + 2(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\left(\sum_{i=1}^N \frac{\phi_i}{\sigma_i^2}\,\frac{\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i}{T}\right) \\ &\quad - 2\sum_{i=1}^N \left[\frac{(\phi_i - \phi_{i0})}{\sigma_i^2}\right]\left(\frac{\boldsymbol{\xi}_{i0}'\mathbf{H}_i\boldsymbol{\varepsilon}_i}{T}\right),\end{aligned} \tag{A.8}$$

which can also be written more compactly as [recall that $\boldsymbol{\psi} = (\boldsymbol{\theta}', \boldsymbol{\phi}')'$]

$$\mathcal{B}_T = (\boldsymbol{\psi} - \boldsymbol{\psi}_0)'\mathbf{G}_T(\boldsymbol{\psi} - \boldsymbol{\psi}_0) + 2(\boldsymbol{\psi} - \boldsymbol{\psi}_0)'\mathbf{f}_T, \tag{A.9}$$

where

$$\mathbf{G}_T = \begin{bmatrix} \sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2}\frac{\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i}{T} & -\frac{\phi_1}{\sigma_1^2}\frac{\mathbf{X}_1'\mathbf{H}_1\boldsymbol{\xi}_{10}}{T} & \cdots & -\frac{\phi_N}{\sigma_N^2}\frac{\mathbf{X}_N'\mathbf{H}_N\boldsymbol{\xi}_{N0}}{T} \\ -\frac{\phi_1}{\sigma_1^2}\frac{\boldsymbol{\xi}_{10}'\mathbf{H}_1\mathbf{X}_1}{T} & \frac{1}{\sigma_1^2}\frac{\boldsymbol{\xi}_{10}'\mathbf{H}_1\boldsymbol{\xi}_{10}}{T} & \cdots & 0 \\ -\frac{\phi_2}{\sigma_2^2}\frac{\boldsymbol{\xi}_{20}'\mathbf{H}_2\mathbf{X}_2}{T} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{\phi_N}{\sigma_N^2}\frac{\boldsymbol{\xi}_{N0}'\mathbf{H}_N\mathbf{X}_N}{T} & 0 & \cdots & \frac{1}{\sigma_N^2}\frac{\boldsymbol{\xi}_{N0}'\mathbf{H}_N\boldsymbol{\xi}_{N0}}{T} \end{bmatrix},$$

$$\mathbf{f}_T = \begin{bmatrix} \sum_{i=1}^N \frac{\phi_i}{\sigma_i^2}\frac{\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i}{T} \\ -\frac{1}{\sigma_1^2}\frac{\boldsymbol{\xi}_{10}'\mathbf{H}_1\boldsymbol{\varepsilon}_1}{T} \\ \vdots \\ -\frac{1}{\sigma_N^2}\frac{\boldsymbol{\xi}_{N0}'\mathbf{H}_N\boldsymbol{\varepsilon}_N}{T} \end{bmatrix}. \tag{A.10}$$

In terms of the matrix notations used in Section 3, $\mathbf{G}_T$ can also be written as

$$\mathbf{G}_T = \mathbf{Z}_1'(\varphi)[\mathbf{I} - \check{\mathbf{W}}(\sigma)(\check{\mathbf{W}}'(\sigma)\check{\mathbf{W}}(\sigma))^{-1}\check{\mathbf{W}}'(\sigma)]\mathbf{Z}_1(\varphi), \quad (A.11)$$

where $\mathbf{Z}_1(\varphi) = [\bar{\mathbf{X}}(\phi,\sigma)\vdots\check{\xi}(\theta,\sigma)]$, $\bar{\mathbf{X}}(\phi,\sigma) = -[(\phi_1/\sigma_1)\mathbf{X}_1', (\phi_2/\sigma_2)\mathbf{X}_2', \ldots, (\phi_N/\sigma_N)\mathbf{X}_N')]'$, and $\check{\mathbf{W}}(\sigma)$ and $\check{\xi}(\theta,\sigma)$ are block diagonal matrices with their ith blocks given by $\mathbf{W}_i/\sigma_i$ and $\xi_i(\theta)/\sigma_i$, with $\theta = \theta_0$. Therefore, $\mathbf{G}_T$ is the partitioned inverse of the upper left-hand corner of

$$\mathbf{Z}'(\varphi)\mathbf{Z}(\varphi) = \begin{pmatrix} \mathbf{Z}_1'(\varphi)\mathbf{Z}_1(\varphi) & \mathbf{Z}_1'(\varphi)\check{\mathbf{W}}(\sigma) \\ \mathbf{W}'(\sigma)\mathbf{Z}_1(\varphi) & \check{\mathbf{W}}'(\sigma)\check{\mathbf{W}}(\sigma) \end{pmatrix},$$

with $\varphi = \varphi_0$. Hence, under assumption 5, $\mathbf{G}_T$ is also a positive definite matrix.

A.2 Proof of Theorem 2: The Case of I(1) Regressors

Partition $\varphi = (\theta', \rho')'$ into the long-run parameters, θ, and the short-run parameters, $\rho = (\phi', \sigma')'$. Define the open ball $B(\theta_0, \delta_\theta) = \{\theta \in \Theta_\theta\colon \|\theta - \theta_0\| < \delta_\theta\}$ and its complement $\bar{B}(\theta_0, \delta_\theta) = \{\theta \in \Theta_\theta\colon \|\theta - \theta_0\| \geq \delta_\theta\}$.

To prove the consistency of the MLE of θ (denoted by $\hat{\theta}$), it is sufficient to show that for all values of $\rho \in \Theta_\rho = \Theta_\phi \times \Theta_\sigma$ and for every $\delta_\theta > 0$, (see, e.g., Saikkonen 1995, p. 903),

$$\lim_{T\to\infty} \Pr\{\inf_{\varphi\in\bar{B}(\theta_0,\delta_\theta)\times\Theta_\rho} T^{-1}[l_T(\varphi_0) - l_T(\varphi)] > 0\} = 1. \quad (A.12)$$

Using the results in (A.1) and (A.2) and under assumption 4, it is easily seen that except for the first term in (A.8), the inf over $\varphi \in \bar{B}(\theta_0, \delta_\theta) \times \Theta_\rho$ of all of the other terms are at most $O_p(1)$. Therefore,

$$\begin{aligned} &\inf_{\varphi\in\bar{B}(\theta_0,\delta_\theta)\times\Theta_\rho} T^{-1}[l_T(\varphi_0) - l_T(\varphi)] \\ &\geq T \inf_{\varphi\in\bar{B}(\theta_0,\delta_\theta)\times\Theta_\rho} \left\{(\theta - \theta_0)'\left(\sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2}\frac{\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i}{T^2}\right)(\theta - \theta_0)\right\} \\ &\quad + O_p(1), \\ &\geq T\delta_\theta^2\lambda_{\min}(\mathbf{R}_T) + O_p(1), \end{aligned} \quad (A.13)$$

where $\lambda_{\min}(\mathbf{R}_T)$ is the smallest eigenvalue of $\mathbf{R}_T = \sum_{i=1}^N (\phi_i^2/\sigma_i^2)(\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i/T^2)$ defined over Θ_ρ. As $T \to \infty$, $\mathbf{R}_T$ converges to $\mathbf{R} = \sum_{i=1}^N (\phi_i^2/\sigma_i^2)\mathbf{R}_{X_iX_i}$ with probability 1, where $\mathbf{R}_{X_iX_i}$ is given by (A.1). But under assumptions 4 and 5, $\sum_{i=1}^N (\phi_i^2/\sigma_i^2)\mathbf{R}_{X_iX_i}$ is a positive definite matrix with probability 1 for all values of $\rho \in \Theta_\rho$. Hence, as $T \to \infty$, $\lambda_{\min}(\mathbf{R}_T)$ also weakly converges to $\lambda_{\min}(\mathbf{R}) > 0$, and the right side of (A.13) will increase without bounds with probability 1, which establishes the consistency of $\hat{\theta}$.

To prove the superconsistency of $\hat{\theta}$, and the consistency of the MLEs of the other parameters, we define $B(\rho_0, \delta_\rho) = \{\rho \in \Theta_\rho\colon \|\rho - \rho_0\| < \delta_\rho\}$, $B(\phi_0, \delta_\phi) = \{\phi \in \Theta_\phi\colon \|\phi - \phi_0\| < \delta_\phi\}$, $B(\sigma_0, \delta_\sigma) = \{\sigma \in \Theta_\sigma\colon \|\sigma - \sigma_0\| < \delta_\sigma\}$ and their complements, $\bar{B}(\rho_0, \delta_\rho)$, $\bar{B}(\phi_0, \delta_\phi)$, and $\bar{B}(\sigma_0, \delta_\sigma)$, and follow Saikkonen (1995) and define the open shrinking ball $\mathrm{N}_T(\theta_0, \delta_d) = \{\theta \in \Theta_\theta\colon T^{1/2}\|\theta - \theta_0\| < \delta_d\}$ for θ and its complement $\bar{\mathrm{N}}_T(\theta_0, \delta_d) = \{\theta \in \Theta_\theta\colon T^{1/2}\|\theta - \theta_0\| \geq \delta_d\}$. Because the consistency of $\hat{\theta}$ is already established, we focus our analysis on values of θ close to θ_0, defined by $\theta = \theta_0 + T^{-1/2}\mathbf{d}$, where we take $\mathbf{d}$ to be a $k \times 1$ vector of fixed constants defined on a compact set. The case where elements of $\mathbf{d}$ are allowed to increase without bound has already been covered. Notice that on $\bar{\mathrm{N}}_T(\theta_0, \delta_d)$, we also have $\|\mathbf{d}\| \geq \delta_d$.

Let $C(\varphi_0, \delta_\rho, \delta_d) = \cup_{\delta_d,\delta_\rho} (\bar{N}_T(\theta_0, \delta_d) \times \bar{B}(\rho_0, \delta_\rho))$, where the union is taken over all values of δ_d and δ_ρ such that $\delta_\varphi = (\delta_d^2 + \delta_\rho^2)^{1/2}$ and $\delta_\rho = (\delta_\phi^2 + \delta_\sigma^2)^{1/2}$. We now prove that for every $\delta_\varphi > 0$,

$$\lim_{T\to\infty} \Pr\{\inf_{\varphi\in C(\varphi_0,\delta_\rho,\delta_d)} T^{-1}[l_T(\varphi_0) - l_T(\varphi)] > 0\} = 1. \quad (A.14)$$

Using (A.4) and (A.9) in (A.3), we first note that

$$\begin{aligned} &2\inf T^{-1}[l_T(\varphi_0) - l_T(\varphi)] \\ &\geq \inf(\mathcal{A}_\sigma) + \inf(\mathcal{C}_T) + \inf[(\psi - \psi_0)'\mathbf{G}_T(\psi - \psi_0)] \\ &\quad + 2\inf[(\psi - \psi_0)'\mathbf{f}_T], \end{aligned} \quad (A.15)$$

where it is now assumed that all of the inf operations are taken over the set $\varphi \in C(\varphi_0, \delta_\rho, \delta_d)$. But

$$(\psi - \psi_0)'\mathbf{f}_T = \mathbf{d}'\sum_{i=1}^N \frac{\phi_i}{\sigma_i^2}\frac{\mathbf{X}_i'\mathbf{H}_i\varepsilon_i}{T^{3/2}} - \sum_{i=1}^N \frac{(\phi_i - \phi_{i0})}{\sigma_i^2}\frac{\xi_{i0}'\mathbf{H}_i\varepsilon_i}{T}. \quad (A.16)$$

Therefore, using the results in (A.1) and (A.2), and recalling that $\mathbf{d}$ and ρ are defined on compact sets, it then follows that $\inf[(\psi - \psi_0)'\mathbf{f}_T] = o_p(1)$. Also from (A.2), it is easily seen that $\mathcal{C}_T = o_p(1)$. Using these results in (A.15), we now have

$$\begin{aligned} 2\inf T^{-1}[l_T(\varphi_0) - l_T(\varphi)] &\geq \inf(\mathcal{A}_\sigma) \\ &\quad + \inf[(\psi - \psi_0)'\mathbf{G}_T(\psi - \psi_0)] + o_p(1), \end{aligned} \quad (A.17)$$

where it is again assumed that all of the inf operations are taken over the set $\varphi \in C(\varphi_0, \delta_\rho, \delta_d)$. But the ith term in $\mathcal{A}_\sigma$—namely, $(\sigma_{i0}^2/\sigma_i^2) - \ln(\sigma_{i0}^2/\sigma_i^2) - 1$—attains its unique minimum at $(\sigma_{i0}^2/\sigma_i^2) = 1$ and is strictly positive for all feasible values of σ_i^2 not equal to σ_{i0}^2, for all $i = 1, 2, \ldots, N$. Hence $\inf(\mathcal{A}_\sigma) > 0 \Leftrightarrow \delta_\sigma > 0$. To establish (A.14), it thus is sufficient to prove that even if $\delta_\sigma = 0$, there exists $\delta_\varphi > 0$ such that

$$\lim_{T\to\infty} \Pr\{\inf\{(\psi - \psi_0)'\mathbf{G}_T(\psi - \psi_0)\} > 0\} = 1.$$

To this end, first note that

$$(\psi - \psi_0)'\mathbf{G}_T(\psi - \psi_0) = [\mathbf{K}_\psi(\psi - \psi_0)]'\mathcal{G}_T[\mathbf{K}_\psi(\psi - \psi_0)],$$

where $\mathbf{K}_\psi = \mathrm{diag}(T^{1/2}\mathbf{I}_k, \mathbf{I}_N)$, $\mathbf{K}_\psi(\psi - \psi_0) = \vartheta = (\mathbf{d}', (\phi - \phi_0)')'$, and $\mathcal{G}_T = \mathbf{K}_\psi^{-1}\mathbf{G}_T\mathbf{K}_\psi^{-1}$. Hence

$$\inf\{(\psi - \psi_0)'\mathbf{G}_T(\psi - \psi_0)\} \geq \delta_\vartheta^2\lambda_{\min}(\mathcal{G}_T),$$

where $\lambda_{\min}(\mathcal{G}_T)$ is the smallest eigenvalue of $\mathcal{G}_T$, and $\delta_\vartheta = (\delta_d^2 + \delta_\phi^2)^{1/2} > 0$. (Notice that for $\delta_\sigma = 0$, we have $\delta_\vartheta = \delta_\varphi > 0$. In the case where $\delta_\sigma > 0$, (A.17) will be satisfied even if $\delta_\vartheta = 0$.) As $T \to \infty$, $\lambda_{\min}(\mathcal{G}_T)$ converges weakly to $\lambda_{\min}(\mathcal{G}) > 0$, which is the smallest eigenvalue of $\mathcal{G}$, where

$$\mathcal{G} = \begin{bmatrix} \sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2}\mathbf{R}_{X_iX_i} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_1^2} q_{\xi_{10}\xi_{10}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_N^2} q_{\xi_{N0}\xi_{N0}} \end{bmatrix},$$

and $\mathbf{R}_{X_iX_i}$ and $q_{\xi_{i0}\xi_{i0}}$, $i = 1, 2, \ldots, N$, are defined in (A.1) and (A.2). But under assumptions 1, 4, and 5(a) and (b) $\mathcal{G}$ is positive definite with probability 1, and hence $\lambda_{\min}(\mathcal{G}) > 0$ with probability 1. Hence for sufficiently large T, $\inf\{(\psi - \psi_0)'\mathbf{G}_T(\psi -$

$\psi_0)\} > 0$. This completes the proof and establishes the desired consistency and superconsistency results given by (4.7).

To derive the asymptotic distribution of the MLEs of $\boldsymbol{\theta}$ and $\phi_i, i = 1, \ldots, N$, we now allow for the fact that in this case the different components of the score vector and the sample information matrix converge to their true values at different rates. Let $\mathbf{D}_\psi = \text{diag}\,(T^{-1}\mathbf{I}_k, T^{-1/2}\mathbf{I}_N)$, and define

$$\begin{aligned} \mathbf{d}_T(\boldsymbol{\psi}) &= \mathbf{D}_\psi[\partial l_T(\boldsymbol{\varphi})/\partial\boldsymbol{\psi}], \\ \mathcal{I}_T(\boldsymbol{\psi}) &= \mathbf{D}_\psi[-\partial^2 l_T(\boldsymbol{\varphi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}']\mathbf{D}_\psi. \end{aligned} \tag{A.18}$$

Then, using standard results from the unit root literature, it can be established that (see, e.g., Phillips and Durlauf 1986) that

$$\mathbf{d}_T(\boldsymbol{\psi}_0) \overset{a}{\sim} \text{MN}\{0, \mathcal{I}(\boldsymbol{\psi}_0)\}, \qquad \mathcal{I}_T(\boldsymbol{\psi}_0) \Rightarrow \mathcal{I}(\boldsymbol{\psi}_0), \tag{A.19}$$

where the $(k+N)\times(k+N)$ matrix,

$$\mathcal{I}(\boldsymbol{\psi}_0) \equiv \begin{bmatrix} \sum_{i=1}^N \frac{\phi_{i0}^2}{\sigma_{i0}^2}\,\mathbf{R}_{X_iX_i} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_{10}^2}\,q_{\xi_{10}\xi_{10}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_{N0}^2}\,q_{\xi_{N0}\xi_{N0}} \end{bmatrix}, \tag{A.20}$$

is the random information matrix, which is positive definite with probability 1.

Now using the mean-value expansion of $\partial l_T(\hat{\boldsymbol{\varphi}})/\partial\boldsymbol{\psi}$ around $\boldsymbol{\psi}_0$ and standardizing the result by $\mathbf{D}_\psi$, we have

$$\mathbf{D}_\psi^{-1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) = \mathcal{I}_T(\bar{\boldsymbol{\psi}})^{-1}\mathbf{d}_T(\boldsymbol{\psi}_0), \tag{A.21}$$

where the (i,j) element of $\partial^2 l_T(\bar{\boldsymbol{\varphi}})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'$ is evaluated at $(\bar{\varphi}_i, \bar{\varphi}_j)$ and $\bar{\varphi}_i$ is a convex combination of φ_{i0} and $\hat{\varphi}_i$.

Notice, however, that in the case of the models with unit root regressors the sample information matrix $\mathcal{I}_T(\bar{\boldsymbol{\psi}})$ need not converge weakly to $\mathcal{I}(\boldsymbol{\psi}_0)$ even when $\hat{\boldsymbol{\psi}}$ converges to $\boldsymbol{\psi}_0$. To ensure that such a convergence occurs, the sample information matrix $\mathcal{I}_T(\boldsymbol{\psi})$ should satisfy the stochastic equicontinuity condition (see, e.g., Saikkonen 1995, condition SE$_o$, p. 894). This result is summarized in the following lemma.

Lemma A.1. Under assumptions 1–4, and 5(a) and (c), and assuming that the regressors $\mathbf{x}_{it}$ in the dynamic heterogeneous panel data model (6) are $I(1)$, the sample information matrix, $\mathcal{I}_T(\boldsymbol{\psi})$, defined by (A.18), satisfy the stochastic equicontinuity condition such that

$$\sup_{\boldsymbol{\psi}_* \in B(\boldsymbol{\phi}_0,\delta)\times N_T(\boldsymbol{\theta}_0,\delta)} \|\mathcal{I}_T(\boldsymbol{\psi}_*) - \mathcal{I}_T(\boldsymbol{\psi}_0)\| \le O_p(1), \tag{A.22}$$

where $B(\boldsymbol{\phi},\delta) = \{\boldsymbol{\phi}_* \in \Theta_\phi\colon \|\boldsymbol{\phi}_* - \boldsymbol{\phi}\| < \delta\}$ and $N_T(\boldsymbol{\theta},\delta) = \{\boldsymbol{\theta}_* \in \Theta_\theta\colon T^{1/2}\|\boldsymbol{\theta}_* - \boldsymbol{\theta}\| < \delta\}$, with δ a common positive real number.

Proof. See the Appendix in the earlier version of this article (Pesaran et al. 1997).

In view of the relative orders of consistency of the MLEs established in (15) and the fact that the sample information matrix satisfies the stochastic equicontinuity condition, (A.22), it also follows that (see Saikkonen, 1995, prop. 3.2)

$$\mathcal{I}_T(\bar{\boldsymbol{\psi}}) = \mathcal{I}(\boldsymbol{\psi}_0) + o_p(1). \tag{A.23}$$

Using this result in (A.21) now yields

$$\mathbf{D}_\psi^{-1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) = \mathcal{I}_T(\boldsymbol{\psi}_0)^{-1}O_T(\boldsymbol{\psi}_0) + o_p(1). \tag{A.24}$$

Using (A.19) in (A.24) and by the continuous mapping theorem, we obtain (16).

A.3 Asymptotic Distribution of the Long-Run Coefficients: The Case of I(1) Regressors

First, note that

$$\hat{\boldsymbol{\xi}}_i = y_{i,-1} - \mathbf{X}_i\hat{\boldsymbol{\theta}} = \boldsymbol{\xi}_{i0} - \mathbf{X}_i(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{A.25}$$

and

$$\mathbf{H}_i(\Delta\mathbf{y}_i - \hat{\phi}_i\mathbf{y}_{i,-1}) = \mathbf{H}_i[-\mathbf{X}_i\boldsymbol{\theta}_0\phi_{i0} - (\hat{\phi}_i - \phi_{i0})\mathbf{y}_{i,-1} + \boldsymbol{\varepsilon}_i]. \tag{A.26}$$

Using (6) and (A.25) in (10), we have

$$\hat{\phi}_i - \phi_{i0} = (\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\hat{\boldsymbol{\xi}}_i)^{-1}[\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\mathbf{X}_i(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\phi_{i0} + \hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i]. \tag{A.27}$$

Moreover, using (A.26) in (9),

$$\begin{aligned} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = &\left\{\sum_{i=1}^N \frac{\hat{\phi}_i^2}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i\right\}^{-1} \\ &\times \left\{\sum_{i=1}^N \frac{\hat{\phi}_i}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\xi}_i(\hat{\phi}_i - \phi_{i0}) - \sum_{i=1}^N \frac{\hat{\phi}_i}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i\right\}. \end{aligned} \tag{A.28}$$

Substituting (A.27) in (A.28), and solving for $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, we obtain

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathbf{S}_{NT}^{-1}\mathbf{s}_{NT}, \tag{A.29}$$

where

$$\begin{aligned} \mathbf{S}_{NT} = &\sum_{i=1}^N \frac{\hat{\phi}_i^2}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i \\ &- \sum_{i=1}^N \frac{\hat{\phi}_i\phi_{i0}}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\xi}_{i0}(\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\hat{\boldsymbol{\xi}}_i)^{-1}\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\mathbf{X}_i \end{aligned} \tag{A.30}$$

and

$$\begin{aligned} \mathbf{s}_{NT} = &-\sum_{i=1}^N \frac{\hat{\phi}_i}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i \\ &+ \sum_{i=1}^N \frac{\hat{\phi}_i}{\hat{\sigma}_i^2}\,\mathbf{X}_i'\mathbf{H}_i\boldsymbol{\xi}_{i0}(\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\hat{\boldsymbol{\xi}}_i)^{-1}\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i. \end{aligned} \tag{A.31}$$

Using the consistency of $\hat{\boldsymbol{\theta}}$; $\hat{\phi}_1, \ldots, \hat{\phi}_N; \hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2$, it is also easily seen that

$$\begin{aligned} &T^{-1}\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\mathbf{X}_i \Rightarrow \mathbf{R}_{X_i\xi_{i0}}', \qquad T^{-1}\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\hat{\boldsymbol{\xi}}_i \Rightarrow q_{\xi_{i0}\xi_{i0}}, \\ &T^{-1/2}\hat{\boldsymbol{\xi}}_i'\mathbf{H}_i\boldsymbol{\varepsilon}_i \Rightarrow T^{-1/2}\boldsymbol{\xi}_{i0}'\mathbf{H}_i\boldsymbol{\varepsilon}_i, \end{aligned} \tag{A.32}$$

where $\mathbf{R}_{X_i\xi_{i0}}$ and $q_{\xi_{i0}\xi_{i0}}$ are defined by (A.1) and (A.2).

Using (A.1), (A.32), and the consistency of $\hat{\phi}_i$ and $\hat{\sigma}_i^2, i = 1, 2, \ldots, N$, in (A.30) and (A.31), we now have

$$T^{-2}\mathbf{S}_{NT} \Rightarrow \sum_{i=1}^N \frac{\phi_{i0}^2}{\sigma_{i0}^2}\,\mathbf{R}_{X_iX_i},$$

$$T^{-1}\mathbf{s}_{NT} \sim \text{MN}\left\{0, \sum_{i=1}^N \frac{\phi_{i0}^2}{\sigma_{i0}^2}\,\mathbf{R}_{X_iX_i}\right\},$$

where $\mathbf{R}_{X_iX_i}$ is defined by (A.1). Using these results in (A.29) and by the continuous mapping theorem, we obtain (17).

[Received August 1997. Revised December 1998.]

REFERENCES

Ahn, S. C., and Schmidt, P. (1995), "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 29–52.

Anderson, T. W., and Hsiao, C. (1981), "Estimation of Dynamic Models With Error Components," *Journal of the American Statistical Association*, 76, 598–606.

——— (1982), "Formation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Arellano, M. (1989), "A Note on the Anderson–Hsiao Estimator for Panel Data," *Economics Letters*, 31, 337–341.

Arellano, M., and Bover, O. (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–52.

Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, New York: Wiley.

Boyd, D. A. C., and Smith, R. P. (in press), "Testing for Purchasing Power Parity: Econometric Issues and an Application to Developing Countries," unpublished manuscript, *Manchester School.*

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S. (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal*, 88, 661–692.

Haque, N. U., and Montiel, P. (1989), "Consumption in Developing Countries: Tests for Liquidity Constraints and Finite Horizons," *Review of Economics and Statistics*, 71, 408–415.

Hausman, J. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

Hsiao, C. (1996), "Random Coefficient Models," in *The Economics of Panel Data: A Handbook of Theory With Applications* (2nd rev. ed.), eds. L. Mátyás and P. Sevestre, Dordrecht: Kluwer, chap. 5.

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (1999), "Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models," in *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G. S. Maddala*, eds. C. Hsiao, K. Lahiri, L-F. Lee, and M. H. Pesaran, Cambridge, U.K.: Cambridge University Press.

Hsiao, C., Sun, B. H., and Lightwood, J. (1995), "Fixed Versus Random Effects Specification for Panel Data Analysis," unpublished manuscript, University of Southern California.

Hsiao, C., and Tahmiscioglu, A. K. (1997), "A Panel Analysis of Liquidity Constraints and Firm Investment," *Journal of the American Statistical Association*, 92, 455–465.

Im, K. S., Pesaran, M. H., and Shin, Y. (1997), "Testing for Unit Roots in Heterogeneous Panels," unpublished manuscript, University of Cambridge (*http://www.econ.cam.ac.uk/faculty/pesaran*).

Keane, M., and Runkle, D. (1992), "On the Estimation of Panel Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," *Journal of Business and Economic Statistics*, 10, 1–9.

Kiviet, J. F., and Phillips, G. D. A. (1993), "Alternative Bias Approximations With a Lagged Dependent Variable," *Econometric Theory*, 9, 62–80.

Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

Maddala, G. S., and Hu, W. (1996), "The Pooling Problem," in *The Economics of Panel Data: A Handbook of Theory With Applications*, (2nd rev. ed.), eds. L. Mátyás and P. Sevestre, Dordrecht: Kluwer, chap. 13.

Mátyás, L., and Sevestre, P. (Eds.) (1996), *The Econometrics of Panel Data: A Handbook of Theory With Applications*, (2nd rev. ed.), Dordrecht: Kluwer.

Pesaran, M. H. (1997), "The Role of Economic Theory in Modelling the Long-Run," *The Economic Journal*, 107, 178–191.

Pesaran, M. H., and Pesaran, B. (1997), *Microfit 4.0: Interactive Econometric Analysis*. Oxford, U.K.: Oxford University Press.

Pesaran, M. H., and Shin, Y. (1999), "Long-Run Structural Modelling," unpublished manuscript, University of Cambridge. (*http://www.econ.cam.ac.uk/faculty/pesaran/*).

——— (1999), "An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis," in *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, ed. S. Strom, Cambridge, U.K.: Cambridge University Press.

Pesaran, M. H., Shin, Y., and Smith, R. J. (1999), "Bounds Testing Approaches to the Analysis of Long Run Relationships," Unpublished manuscript, University of Cambridge (*http://www.econ.cam.ac.uk/faculty/pesaran*).

Pesaran, M. H., Shin, Y., and Smith, R. P. (1997), "Pooled Estimation of Long-Run Relationships in Dynamic Heterogenous Panels," DAE Working Papers Amalgamated Series 9721, University of Cambridge.

Pesaran, M. H., and Smith, R. P. (1995), "Estimating Long-Run Relationships From Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68, 79–113.

Pesaran, M. H., Smith, R. P., and Akiyama, T. (1998), *Energy Demand in Asian Economies*, Oxford, U.K.: Oxford University Press.

Pesaran, M. H., Smith, R. P., and Im, K. S. (1996), "Dynamic Linear Models for Heterogeneous Panels," in *The Economics of Panel Data: A Handbook of Theory With Applications*, (2nd rev. ed.), eds. L. Mátyás and P. Sevestre, Dordrecht: Kluwer, chap. 8.

Pesaran, M. H., and Zhao, Z. (1999), "Bias Reduction in Estimating Long-Run Relationships From Dynamic Heterogeneous Panels," in *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G. S. Maddala*, eds. C. Hsiao, K. Lahiri, L-F. Lee, and M. H. Pesaran, Cambridge, U.K.: Cambridge University Press.

Phillips, P. C. B., and Durlauf, S. (1986), "Multiple Time Series With Integrated Variables," *Review of Economic Studies*, 53, 473–496.

Rogoff, K. (1996), "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34, 647–668.

Saikkonen, P. (1995), "Problems With the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems," *Econometric Theory*, 11, 888–911.

Stoker, T. M. (1993), "Empirical Approaches to the Problem of Aggregation Over Individuals," *Journal of Economic Literature*, 31, 1827–1874.

Swamy, P. A. V. B. (1970), "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38, 311–323.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Test for Aggregation Bias," *Journal of the American Statistical Association*, 57, 348–368.