FULLY MODIFIED OLS FOR HETEROGENEOUS COINTEGRATED PANELS

Peter Pedroni

ABSTRACT

This chapter uses fully modified OLS principles to develop new methods for estimating and testing hypotheses for cointegrating vectors in dynamic panels in a manner that is consistent with the degree of cross sectional heterogeneity that has been permitted in recent panel unit root and panel cointegration studies. The asymptotic properties of various estimators are compared based on pooling along the 'within' and 'between' dimensions of the panel. By using Monte Carlo simulations to study the small sample properties, the group mean estimator is shown to behave well even in relatively small samples under a variety of scenarios.

I. INTRODUCTION

In this chapter we develop methods for estimating and testing hypotheses for cointegrating vectors in dynamic time series panels. In particular we propose methods based on fully modified OLS principles which are able to accommodate considerable heterogeneity across individual members of the panel. Indeed, one important advantage to working with a cointegrated panel approach of this type is that it allows researchers to selectively pool the long run information contained in the panel while permitting the short run dynamics

Nonstationary Panels, Panel Cointegration and Dynamic Panels, Volume 15, pages 93–130.

ISBN: 0–7623–0688–2

and fixed effects to be heterogeneous among different members of the panel. An important convenience of the fully modified approach that we propose here is that in addition to producing asymptotically unbiased estimators, it also produces nuisance parameter free standard normal distributions. In this way, inferences can be made regarding common long run relationships which are asymptotically invariant to the considerable degree of short run heterogeneity that is prevalent in the dynamics typically associated with panels that are composed of aggregate national data.

A. Non-Stationary Panels and Heterogeneity

Methods for non-stationary time series panels, including unit root and cointegration tests, have been gaining increased acceptance in a number of areas of empirical research. Early examples include Canzoneri, Cumby & Diba (1996), Chinn & Johnson (1996), Chinn (1997), Evans & Karras (1996), Neusser & Kugler (1998), Obstfeld & Taylor (1996), Oh (1996), Papell (1997), Pedroni (1996b), Taylor (1996) and Wu (1996), with many more since. These studies have for the most part been limited to applications which simply ask whether or not particular series appear to contain unit roots or are cointegrated. In many applications, however, it is also of interest to ask whether or not common cointegrating vectors take on particular values. In this case, it would be helpful to have a technique that allows one to test such hypothesis about the cointegrating vectors in a manner that is consistent with the very general degree of cross sectional heterogeneity that is permitted in such panel unit root and panel cointegration tests.

In general, the extension of conventional non-stationary methods such as unit root and cointegration tests to panels with both cross section and time series dimensions holds considerable promise for empirical research considering the abundance of data which is available in this form. In particular, such methods provide an opportunity for researchers to exploit some of the attractive theoretical properties of non-stationary regressions while addressing in a natural and direct manner the small sample problems that have in the past often hindered the practical success of these methods. For example, it is well known that superconsistent rates of convergence associated with many of these methods can provide empirical researchers with an opportunity to circumvent more traditional exogeneity requirements in time series regressions. Yet the low power of many of the associated statistics has often impeded the ability to take full advantage of these properties in small samples. By allowing data to be pooled in the cross sectional dimension, non-stationary panel methods have the potential to improve upon these small sample limitations. Conversely, the use of non-stationary time series asymptotics provides an opportunity to make panel methods more amenable to pooling aggregate level data by allowing researchers to selectively pool the long run information contained in the panel, while allowing the short run dynamics to be heterogeneous among different members of the panel.

Initial methodological work on non-stationary panels focused on testing for unit roots in univariate panels. Quah (1994) derived standard normal asymptotic distributions for testing unit roots in homogeneous panels as both the time series and cross sectional dimensions grow large. Levin & Lin (1993) derived distributions under more general conditions that allow for heterogeneous fixed effects and time trends. More recently, Im, Pesaran & Shin (1995) study the small sample properties of unit root tests in panels with heterogeneous dynamics and propose alternative tests based on group mean statistics. In practice however, empirical work often involves relationships within multivariate systems. Toward this end, Pedroni (1993, 1995) studies the properties of spurious regressions and residual based tests for the null of no cointegration in dynamic heterogeneous panels. This chapter continues this line of research by proposing a convenient method for estimating and testing hypotheses about common cointegrating vectors in a manner that is consistent with the degree of heterogeneity permitted in these panel unit root and panel cointegration studies.

In particular, we address here two key sources of cross member heterogeneity that are particularly important in dealing with dynamic cointegrated panels. One such source of heterogeneity manifests itself in the familiar fixed effects form. These reflect differences in mean levels among the variables of different individual members of the panel and we model these by including individual specific intercepts. The second key source of heterogeneity in such panels comes from differences in the way that individuals respond to short run deviations from equilibrium cointegrating vectors that develop in response to stochastic disturbances. In keeping with earlier panel unit root and panel cointegration papers, we model this form of heterogeneity by allowing the associated serial correlation properties of the error processes to vary across individual members of the panel.

B. Related Literature

Since the original version of this paper, Pedroni (1996a),[1] many more papers have contributed to our understanding of hypothesis testing in cointegrating panels. For example, Kao & Chiang (1997) extended their original paper on the least squares dummy variable model in cointegrated panels, Kao & Chen

(1995), to include a comparison of the small sample properties of a dynamic OLS estimator with other estimators including a FMOLS estimator similar to Pedroni (1996a). Specifically, Kao & Chiang (1997) demonstrated that a panel dynamic OLS estimator has the same asymptotic distribution as the type of panel FMOLS estimator derived in Pedroni (1996a) and showed that the small sample size distortions for such an estimator were often smaller than certain forms of the panel FMOLS estimator. The asymptotic theory in these earlier papers were generally based on sequential limit arguments (allowing the sample sizes T and N to grow large sequentially), whereas Phillips & Moon (1999) subsequently provided a rigorous and more general study of the limit theory in non-stationary panel regressions under joint convergence (allowing T and N to grow large concurrently). Phillips & Moon (1999) also provided a set of regularity conditions under which convergence in sequential limits implies convergence in joint limits, and considered these properties in the context of a FMOLS estimator, although they do not specifically address the small sample properties of feasible versions of the estimators. More recently, Mark & Sul (1999) also study a similar form of the panel dynamic OLS estimator first proposed by Kao & Chiang (1997). They compare the small sample properties of a weighted versus unweighted version of the estimator and find that the unweighted version generally exhibits smaller size distortion than the weighted version.

In this chapter we report new small sample results for the group mean panel FMOLS estimator that was originally proposed in Pedroni (1996a). An advantage of the group mean estimator over the other pooled panel FMOLS estimators proposed in the Pedroni (1996a) is that the t-statistic for this estimator allows for a more flexible alternative hypothesis. This is because the group mean estimator is based on the so called ‘between dimension’ of the panel, while the pooled estimators are based on the ‘within dimension’ of the panel. Accordingly, the group mean panel FMOLS provides a consistent test of a common value for the cointegrating vector under the null hypothesis against values of the cointegrating vector that need not be common under the alternative hypothesis, while the pooled within dimension estimators do not. Furthermore, as Pesaran & Smith (1995) argue in the context of OLS regressions, when the true slope coefficients are heterogeneous, group mean estimators provide consistent point estimates of the sample mean of the heterogeneous cointegrating vectors, while pooled within dimension estimators do not. Rather, as Phillips & Moon (1999) demonstrate, when the true cointegrating vectors are heterogeneous, pooled within dimension estimators provide consistent point estimates of the average regression coefficient, not the sample mean of the cointegrating vectors. Both of these features of the group mean estimator are often important in practical applications.

Finally, the implementation of the feasible form of the between dimension group mean estimator also has advantages over the other estimators in the presence of heterogeneity of the residual dynamics around the cointegrating vector. As was demonstrated in Pedroni (1996a), in the presence of such heterogeneity, the pooled panel FMOLS estimator requires a correction term that depends on the true cointegrating vector. For a specific null value for a cointegrating vector, the *t*-statistic is well defined, but of course this is of little use per se when one would like to estimate the cointegrating vector. One solution is to obtain a preliminary estimate of the cointegrating vector using OLS. However, although the OLS estimator is superconsistent, it still contains a second order bias in the presence of endogeneity, which is not eliminated asymptotically. Accordingly, this bias leads to size distortion, which is not necessarily eliminated even when the sample size grows large in the panel dimension. Consequently, this type of approach based on a first stage OLS estimate was not recommended in Pedroni (1996a), and it is not surprising that Monte Carlo simulations have shown large size distortions for such estimators. Even when the null hypothesis was imposed without using an OLS estimator, the size distortions for this type of estimator were large as reported in Pedroni (1996a). Similarly, Kao & Chiang (1997) also found large size distortions for such estimators when OLS estimates were used in the first stage for the correction term. By contrast, the feasible version of the between dimension group mean based estimator does not suffer from these difficulties, even in the presence of heterogeneous dynamics. As we will see, the size distortions for this estimator are minimal, even in panels of relatively modest dimensions.

The remainder of the chapter is structured as follows. In Section 2, we introduce the econometric models of interest for heterogeneous cointegrated panels. We then present a number of theoretical results for estimators designed to be asymptotically unbiased and to provide nuisance parameter free asymptotic distributions which are standard normal when applied to heterogeneous cointegrated panels and can be used to test hypotheses regarding common cointegrating vectors in such panels. In Section 3 we study the small sample properties of these estimators and propose feasible FMOLS statistics that performs relatively well in realistic panels with heterogeneous dynamics. In Section 4 we enumerate the algorithm used to construct these statistics and briefly describe a few examples of their uses. Finally, in Section 5 we offer conclusions and discuss a number of related issues in the ongoing research on estimation and inference in cointegrated panels.

II. ASYMPTOTIC RESULTS FOR FULLY MODIFIED OLS IN HETEROGENEOUS COINTEGRATED PANELS

In this section we study asymptotic properties of cointegrating regressions in dynamic panels with common cointegrating vectors and suggest how a fully modified OLS estimator can be constructed to deal with complications introduced by the presence of parameter heterogeneity in the dynamics and fixed effects across individual members. We begin, however, by discussing the basic form of a cointegrating regression in such panels and the problems associated with unmodified OLS estimators.

A. Cointegrating Regressions in Heterogeneous Panels

Consider the following cointegrated system for a panel of $i = 1, \ldots, N$ members,

$$\begin{aligned} y_{it} &= \alpha_i + \beta x_{it} + \mu_{it} \\ x_{it} &= x_{it-1} + \varepsilon_{it} \end{aligned} \tag{1}$$

where the vector error process $\xi_{it} = (\mu_{it}, \varepsilon_{it})'$ is stationary with asymptotic covariance matrix Ω_{i}. Thus, the variables x_i, y_i are said to cointegrate for each member of the panel, with cointegrating vector β if y_{it} is integrated of order one. The term α_i allows the cointegrating relationship to include member specific fixed effects. In keeping with the cointegration literature, we do not require exogeneity of the regressors. As usual, x_i can in general be an m dimensional vector of regressors, which are not cointegrated with each other. In this case, we partition $\xi_{it} = (\mu_{it}, \varepsilon'_{it})$ so that the first element is a scalar series and the second element is an m dimensional vector of the differences in the regressors $\varepsilon_{it} = x_{it} - x_{it-1} = \Delta x_{it}$, so that when we construct

$$\Omega_i = \begin{bmatrix} \Omega_{11i} & \Omega'_{21i} \\ \Omega_{21i} & \Omega_{22i} \end{bmatrix} \tag{2}$$

then Ω_{11i} is the scalar long run variance of the residual μ_{it}, and Ω_{22i} is the $m \times m$ long run covariance among the ε_{it}, and Ω_{21i} is an $m \times 1$ vector that gives the long run covariance between the residual μ_{it} and each of the ε_{it}. However, for simplicity and convenience of notation, we will refer to x_i as univariate in the remainder of this chapter. Each of the results of this study generalize in an obvious and straightforward manner to the vector case, unless otherwise indicated.[2]

In order to explore the asymptotic properties of estimators as both the cross sectional dimension, N, and the time series dimension, T, grow large, we will make assumptions similar in spirit to Pedroni (1995) regarding the degree of dependency across both these dimensions. In particular, for the time series dimension, we will assume that the conditions of the multivariate functional central limit theorems used in Phillips & Durlauf (1986) and Park & Phillips (1988), hold for each member of the panel as the time series dimension grows large. Thus, we have

Assumption 1.1 (invariance principle): *The process ξ_{it} satisfies a multivariate functional central limit theorem such that the convergence as $T \rightarrow \infty$ for the partial sum $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_{it} \rightarrow B_i(r, \Omega_i)$ holds for any given member, i, of the panel, where $B_i(r, \Omega_i)$ is Brownian motion defined over the real interval $r \in [0,1]$, with asymptotic covariance Ω_i.*

This assumption indicates that the multivariate functional central limit theorem, or invariance principle, holds over time for any given member of the panel. This places very little restriction on the temporal dependency and heterogeneity of the error process, and encompasses for example a broad class of stationary ARMA processes. It also allows the serial correlation structure to be different for individual members of the panel. Specifically, the asymptotic covariance matrix, Ω_i varies across individual members, and is given by $\Omega_i \equiv \lim_{T\rightarrow\infty} E[T^{-1}(\Sigma_{t=1}^{T}\xi_{it})(\Sigma_{t=1}^{T}\xi_{it})]$, which can also be decomposed as $\Omega_i = \Omega_i^o + \Gamma_i + \Gamma_i$, where Ω_i^o is the contemporaneous covariance and Γ_i is a weighted sum of autocovariances. The off-diagonal terms of these individual Ω_{21i} matrices capture the endogenous feedback effect between y_{it} and x_{it}, which is also permitted to vary across individual members of the panel. For several of the estimators that we propose, it will be convenient to work with a triangularization of this asymptotic covariance matrix. Specifically, we will refer to this lower triangular matrix of Ω_i as L_i, whose elements are related as follows

$$L_{11i} = (\Omega_{11i} - \Omega_{21i}^2/\Omega_{22i})^{1/2},\ L_{12i} = 0,\ L_{21i} = \Omega_{21i}/\Omega_{22i}^{1/2},\ L_{22i} = \Omega_{22i}^{1/2} \quad (3)$$

Estimation of the asymptotic covariance matrix can be based on any one of a number of consistent kernel estimators such as the Newey & West (1987) estimator.

Next, for the cross sectional dimension, we will employ the standard panel data assumption of independence. Hence we have:

Assumption 1.2 (cross sectional independence): *The individual processes are assumed to be independent cross sectionally, so that $E[\xi_{it}, \xi_{jt}] = 0$ for all $i \neq j$.*

More generally, the asymptotic covariance matrix for a panel of dimension $N \times T$ *is block diagonal with the ith diagonal block given by the asymptotic covariance for member i.*

This type of assumption is typical of our panel data approach, and we will be using this condition in the formal derivation of the asymptotic distribution of our panel cointegration statistics. For panels that exhibit common disturbances that are shared across individual members, it will be convenient to capture this form of cross sectional dependency by the use of a common time dummy, which is a fairly standard panel data technique. For panels with even richer cross sectional dependencies, one might think of estimating a full non-diagonal $N \times N$ matrix of Ω_{ij} elements, and then premultiplying the errors by this matrix in order to achieve cross sectional independence. This would require the time series dimension to grow much more quickly than the cross sectional dimension, and in most cases one hopes that a common time dummy will suffice.

While the derivation of most of the asymptotic results of this chapter are relegated to the mathematical appendix, it is worth discussing briefly here how we intend to make use of assumptions 1.1 and 1.2 in providing asymptotic distributions for the panel statistics that we consider in the next two subsections. In particular, we will employ here simple and somewhat informal sequential limit arguments by first evaluating the limits as the T dimension grows large for each member of the panel in accordance with assumption 1.1 and then evaluating the sums of these statistics as the N dimension grows large under the independence assumption of 1.2.[3] In this manner, as N grows large we obtain standard distributions as we average the random functionals for each member that are obtained in the initial step as a consequence of letting T grow large. Consequently, we view the restriction that first $T \to \infty$ and then $N \to \infty$ as a relatively strong restriction that ensures these conditions, and it is possible that in many circumstances a weaker set of restrictions that allow N and T to grow large concurrently, but with restrictions on the relative rates of growth might deliver similar results. In general, for heterogeneous error processes, such restrictions on the rate of growth of N relative to T can be expected to depend in part on the rate of convergence of the particular kernel estimators used to eliminate the nuisance parameters, and we can expect that our iterative $T \to \infty$ and then $N \to \infty$ requirements proxy for the fact that in practice our asymptotic approximations will be more accurate in panels with relatively large T dimensions as compared to the N dimension. Alternatively, under a more pragmatic interpretation, one can simply think of letting $T \to \infty$ for fixed N reflect the fact that typically for the panels in which we are interested, it is the time series dimension which can be expected to grow in actuality rather than the cross sectional dimension, which is in practice fixed. Thus, $T \to \infty$ is in a sense the true asymptotic feature in which we are interested, and this leads to statistics which are characterized as sums of i.i.d. Brownian motion functionals. For practical purposes, however, we would like to be able to characterize these statistics for the general case in which N is large, and in this case we take $N \to \infty$ as a convenient benchmark for which to characterize the distribution, provided that we understand $T \to \infty$ to be the dominant asymptotic feature of the data.

B. Asymptotic Properties of Panel OLS

Next, we consider the properties of a number of statistics that might be used for a cointegrated panel as described by (1) under assumptions 1.1 and 1.2 regarding the time series and cross dimensional dependencies in the data. The first statistic that we examine is a standard panel OLS estimator of the cointegrating relationship. It is well known that the conventional single equation OLS estimator for the cointegrating vector is asymptotically biased and that its standardized distribution is dependent on nuisance parameters associated with the serial correlation structure of the data, and there is no reason to believe that this would be otherwise for the panel OLS estimator. The following proposition confirms this suspicion.[4]

Proposition 1.1 (Asymptotic Bias of the Panel OLS Estimator). *Consider a standard panel OLS estimator for the coefficient* β *of panel (1), under assumptions 1.1 and 1.2, given as*

$$\hat{\beta}_{NT} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2 \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)$$

where $\bar{x}_i$ *and* $\bar{y}_i$ *refer to the individual specific means. Then,*

(a) The estimator is asymptotically biased and its asymptotic distribution will be dependent on nuisance parameters associated with the dynamics of the underlying processes.

(b) Only for the special case in which the regressors are strictly exogenous and the dynamics are homogeneous across members of the panel can valid inferences be made from the standardized distribution of $\hat{\beta}_{NT}$ *or its associated t-statistic.*

As the proof of proposition 1.1 given in the appendix makes clear, the source of the problem stems from the endogeneity of the regressors under the usual assumptions regarding cointegrated systems. While an exogeneity assumption is common in many treatments of cross sectional panels, for dynamic cointegrated panels such strict exogeneity is by most standards not acceptable. It is stronger than the standard exogeneity assumption for static panels, as it implies the absence of any dynamic feedback from the regressors at all frequencies. Clearly, the problem of asymptotic bias and data dependency from the endogenous feedback effect can no less be expected to diminish in the context of such panels, and Kao & Chen (1995) document this bias for a panel of cointegrated time series for the special case in which the dynamics are homogeneous. For the conventional time series case, a number of methods have been devised to deal with the consequences of such endogenous feedback effects, and in what follows we develop an approach for cointegrated panels based on fully modified OLS principles similar in spirit to those used by Phillips & Hanson (1990).

C. Pooled Fully Modified OLS Estimators for Heterogeneous Panels

Phillips & Hansen (1990) proposed a semi-parametric correction to the OLS estimator which eliminates the second order bias induced by the endogeneity of the regressors. The same principle can also be applied to the panel OLS estimator that we have explored in the previous subsection. The key difference in constructing our estimator for the panel data case will be to account for the heterogeneity that is present in the fixed effects as well as in the short run dynamics. These features lead us to modify the form of the standard single equation fully modified OLS estimator. We will also find that the presence of fixed effects has the potential to alter the asymptotic distributions in a non-trivial manner.

The following proposition establishes an important preliminary result which facilitates intuition for the role of heterogeneity and the consequences of dealing with both temporal and cross sectional dimensions for fully modified OLS estimators.

Proposition 1.2 (Asymptotic Distribution of the Pooled Panel FMOLS Estimator). *Consider a panel FMOLS estimator for the coefficient* β *of panel (1) given by*

$$\hat{\beta}^*_{NT} - \beta = \left(\sum_{i=1}^{N} \hat{L}_{22i}^{-2} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2 \right)^{-1} \sum_{i=1}^{N} \hat{L}_{11i}^{-1} \hat{L}_{22i}^{-1} \left(\sum_{t=1}^{T} (x_{it} - \bar{x}_i)\mu^*_{it} - T\hat{\gamma}_i \right)$$

where

$$\mu^*_{it} = \mu_{it} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} \Delta x_{it},\ \hat{\gamma}_i \equiv \hat{\Gamma}_{21i} + \hat{\Omega}^o_{21i} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} (\hat{\Gamma}_{22i} + \hat{\Omega}^o_{22i})$$

and $\hat{L}_i$ is a lower triangular decomposition of $\hat{\Omega}_i$ as defined in (2) above. Then, under assumptions 1.1 and 1.2, the estimator $\hat{\beta}^*_{NT}$ converges to the true value at rate $T\sqrt{N}$, and is distributed as

$$T\sqrt{N}(\hat{\beta}^*_{NT} - \beta) \rightarrow N(0, v) \ where\ v = \begin{matrix} 2\ iff\ \bar{x}_i = \bar{y}_i = 0 \\ 6\ else \end{matrix}$$

as $T \rightarrow \infty$ and $N \rightarrow \infty$.

As the proposition indicates, when proper modifications are made to the estimator, the corresponding asymptotic distribution will be free of the nuisance parameters associated with any member specific serial correlation patterns in the data. Notice also that this fully modified panel OLS estimator is asymptotically unbiased for both the standard case without intercepts as well as the fixed effects model with heterogeneous intercepts. The only difference is in the size of the variance, which is equal to 2 in the standard case, and 6 in the case with heterogeneous intercepts, both for x_{it} univariate. More generally, when x_{it} is an m-dimensional vector, the specific values for v will also be a function of the dimension m. The associated t-statistics, however, will not depend on the specific values for v, as we shall see.

The fact that this estimator is distributed normally, rather than in terms of unit root asymptotics as in Phillips & Hansen (1990), derives from the fact that these unit root distributions are being averaged over the cross sectional dimension. Specifically, this averaging process produces normal distributions whose variance depends only on the moments of the underlying Brownian motion functionals that describe the properties of the integrated variables. This is achieved by constructing the estimator in a way that isolates the idiosyncratic components of the underlying Wiener processes to produce sums of standard and independently distributed Brownian motion whose moments can be computed algebraically, as the proof of the proposition makes clear. The estimators $\hat{L}_{11i}$ and $\hat{L}_{22i}$, which correspond to the long run standard errors of conditional process μ_{it}, and the marginal process Δx_{it} respectively, act to purge the contribution of these idiosyncratic elements to the endogenous feedback and serial correlation adjusted statistic $\sum_{t=1}^{T} (x_{it} - \bar{x}_i) y^*_{it} - T\hat{\gamma}_i$.

The fact that the variance is larger for the fixed effects model in which heterogeneous intercepts are included stems from the fact that in the presence of unit roots, the variation from the cross terms of the sample averages $\bar{x}_i$ and

$\bar{y}_i$ grows large over time at the same rate T, so that their effect is not eliminated asymptotically from the distribution of $T\sqrt{N}(\hat{\beta}^*_{NT} - \beta)$.[5] However, since the contribution to the variance is computable analytically as in the proof of proposition 1.2, this in itself poses no difficulties for inference. Nevertheless, upon consideration of these expressions, it also becomes apparent that there should exist a metric which can directly adjust for this effect in the distribution and consequently render the distribution standard normal. In fact, as the following proposition indicates, it is possible to construct a *t*-statistic from this fully modified panel OLS estimator whose distribution will be invariant to this effect.

Corollary 1.2 (Asymptotic Distribution of the Pooled Panel FMOLS t-statistic). *Consider the following t-statistic for the FMOLS panel estimator of* β *as defined in proposition 1.2 above. Then under the same assumptions as in proposition 1.2, the statistic is standard normal,*

$$t_{\hat{\beta}^*_{NT}} = (\hat{\beta}^*_{NT} - \beta)\left(\sum_{i=1}^{N} \hat{L}_{22i}^{-2} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2\right)^{-1/2} \rightarrow N(0, 1)$$

as $T \rightarrow \infty$ *and* $N \rightarrow \infty$ *for both the standard model without intercepts as well as the fixed effects model with heterogeneous estimated intercepts.*

Again, as the derivation in the appendix makes apparent, because the numerator of the fully modified estimator $\hat{\beta}^*_{NT}$ is a sum of mixture normals with zero mean whose variance depends only on the properties of the Brownian motion functionals associated with the quadratic $\sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2$, the *t*-statistic constructed using this expression will be asymptotically standard normal. This is regardless of the value of v associated with the distribution of $T\sqrt{N}(\hat{\beta}^*_{NT} - \beta)$ and so will also not depend on the dimensionality of x_{it} in the general vector case.

Note, however, that in contrast to the conventional single equation case studied by Phillips & Hansen (1990), in order to ensure that the distribution of this t-statistic is free of nuisance parameters when applied to heterogeneous panels, the usual asymptotic variance estimator of the denominator is replaced with the estimator $\hat{L}_{22i}^{-2}$. By construction, this corresponds to an estimator of the asymptotic variance of the differences for the regressors and can be estimated accordingly. This is in contrast to the t-statistic for the conventional single equation fully modified OLS, which uses an estimator for the conditional asymptotic variance from the residuals of the cointegrating regression. This distinction may appear puzzling at first, but it stems from the fact that in heterogeneous panels the contribution from the conditional variance of the residuals is idiosyncratic to the cross sectional member, and must be adjusted for directly in the construction of the numerator of the $\hat{\beta}^*_{NT}$ estimator itself before averaging over cross sections. Thus, the conditional variance has already been implicitly accounted for in the construction of $\hat{\beta}^*_{NT}$, and all that is required is that the variance from the marginal process Δx_{it} be purged from the quadratic $\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2$. Finally, note that proposition 1.2 and its corollary 1.2 have been specified in terms of a transformation, μ^*_{it}, of the true residuals. In Section 3 we will consider various strategies for specifying these statistics in terms of observables and consider the small sample properties of the resulting feasible statistics.

D. A Group Mean Fully Modified OLS t-Statistic

Before preceding to the small sample properties, we first consider one additional asymptotic result that will be of use. Recently Im, Pesaran & Shin (1995) have proposed using a group mean statistic to test for unit roots in panel data. They note that under certain circumstances, panel unit root tests may suffer from the fact that the pooled variance estimators need not necessarily be asymptotically independent of the pooled numerator and denominator terms of the fixed effects estimator. Notice, however, that the fully modified panel OLS statistics in proposition 1.2 and corollary 1.2 here have been constructed without the use of a pooled variance estimator. Rather, the statistics of the numerator and denominator have been purged of any influence from the nuisance parameters prior to summing over N. Furthermore, since asymptotically the distribution for the numerator is centered around zero, the covariance between the summed terms of the numerator and denominator also do not play a role in the asymptotic distribution of $T\sqrt{N}(\hat{\beta}^*_{NT}-\beta)$ or $t_{\hat{\beta}^*_{it}}$ as they would otherwise.

Nevertheless, it is also interesting to consider the possibility of a fully modified OLS group mean statistic in the present context. In particular, the group mean *t*-statistic is useful because it allows one to entertain a somewhat broader class of hypotheses under the alternative. Specifically, we can think of the distinction as follows. The *t*-statistic for the true panel estimator as described in corollary 1.2 can be used to test the null hypothesis $H_o: \beta_i = \beta_o$ *for all i* versus the alternate hypothesis $H_a: \beta_i = \beta_a \neq \beta_o$ *for all i* where β_o is the hypothesized common value for β under the null, and β_a is some alternative value for β which is also common to all members of the panel. By contrast, the group mean fully modified *t*-statistic can be used to test the null hypothesis $H_o:\beta_i=\beta_o$ for all i versus the alternate hypothesis $H_a:\beta_i\neq\beta_o$ for all i, so that the values for β are not necessarily constrained to be homogeneous across different members under the alternative hypothesis.

The following proposition gives the precise form of the panel fully modified OLS t-statistic that we propose and gives its asymptotic distributions.

Proposition 1.3 (Asymptotic Distribution of the Panel FMOLS Group Mean t-Statistic). *Consider the following group mean FMOLS t-statistic for* β *of the cointegrated panel (1). Then under assumptions 1.1 and 1.2, the statistic is standard normal, and*

$$\bar{t}_{\hat{\beta}^*_{NT}}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{L}_{11i}^{-1}\left(\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2\right)^{-1/2}\left(\sum_{t=1}^{T}(x_{it}-\bar{x}_i)y^*_{it}-T\hat{\gamma}_i\right)\rightarrow N(0,1)$$

where

$$y^*_{it}=(y_{it}-\bar{y}_i)-\frac{\hat{L}_{21i}}{\hat{L}_{22i}}\Delta x_{it},\ \hat{\gamma}_i\equiv\hat{\Gamma}_{21i}+\hat{\Omega}^o_{21i}-\frac{\hat{L}_{21i}}{\hat{L}_{22i}}(\hat{\Gamma}_{22i}+\hat{\Omega}^o_{22i})$$

and $\hat{L}_i$ *is a lower triangular decomposition of* $\hat{\Omega}_i$ *as defined in (2) above, as* $T\rightarrow\infty$ *and* $N\rightarrow\infty$ *for both the standard model without intercepts as well as the fixed effects model with heterogeneous intercepts.*

Note that the asymptotic distribution of this group mean statistic is also invariant to whether or not the standard model without intercepts or the fixed effects model with heterogeneous intercepts has been estimated. Just as with the previous *t*-statistic of corollary 1.2, the asymptotic distribution of this panel group mean *t*-statistic will also be independent of the dimensionality of x_{it} for the more general vector case. Thus, we have presented two different types of *t*-statistics, a pooled panel OLS based fully modified *t*-statistic based on the 'within' dimension of the panel, and a group mean fully modified OLS *t*-statistic based on the 'between' dimension of the panel, both of which are asymptotically unbiased, free of nuisance parameters, and invariant to whether or not idiosyncratic fixed effects have been estimated. Furthermore, we have characterized the asymptotic distribution of the fully modified panel OLS estimator itself, which is also asymptotically unbiased and free of nuisance parameters, although in this case one should be aware that while the distribution will be a centered normal, the variance will depend on whether heterogeneous intercepts have been estimated and on the dimensionality of the vector of regressors. In the remainder of this chapter we investigate the small sample properties of feasible statistics associated with these asymptotic results and consider their application to the purchasing power parity question.

III. SMALL SAMPLE PROPERTIES OF FEASIBLE PANEL FULLY MODIFIED OLS STATISTICS

In this section we investigate the small sample properties of the pooled and group mean panel FMOLS estimators that were developed in the previous section. We discuss two alternative feasible estimators associated with the panel FMOLS estimators of proposition 1.2 and its t-statistic, which were defined only in terms of the true residuals. While these estimators perform reasonably well in idealized situations, more generally, size distortions for these estimators have the potential to be fairly large in small samples, as was reported in the earlier version of this paper. By contrast, we find that the group mean test statistics do very well and exhibit relatively little size distortion even in relatively small panels even in the presence of substantial cross sectional heterogeniety of the error process associated with the dynamics around the cointegrating vector. Consequently, after discussing some of the basic properties of the feasible versions of the pooled estimators and the associated difficulties for small samples, we focus here on reporting the small sample properties of the group mean test statistics, which are found to do extremely well provided that the time series dimension is not smaller than the cross sectional dimension.

A. General Properties of the Feasible Estimators

First, before reporting the results for the between dimension group mean test statistic, we discuss the general properties of various feasible forms of the within dimension pooled panel fully modified OLS and consider the consequences of these properties in small samples. One obvious candidate for a feasible estimator based on proposition 1.2 would be to simply construct the statistic in terms of estimated residuals, which can be obtained from the initial N single equation OLS regressions associated with the cointegrating regression for (1). Since the single equation OLS estimator is superconsistent, one might hope that this produces a reasonably well behaved statistic for the panel FMOLS estimator. The potential problem with this reasoning stems from the fact that although the OLS regression is superconsistent it is also asymptotically biased in general. While this is a second order effect for the conventional single series estimator, for panels, as N grows large, the effect has the potential to become first order.

Another possibility might appear to be to construct the feasible panel FMOLS estimator for proposition 1.2 in terms of the original data series $y^*_{it} = (y_{it} - \bar{y}_i) - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} \Delta x_{it}$ along the lines of how it is often done for the conventional single series case. However, this turns out to be correct only in very specialized cases. More generally, for heterogeneous panels, this will introduce an asymptotic bias which depends on the true value of the cointegrating relationship and the relative volatility of the series involved in the regression. The following makes this relationship precise.

Proposition 2.1 (Regarding Feasible Pooled Panel FMOLS) *Under the conditions of proposition 1.2 and corollary 1.2, consider the panel FMOLS estimator for the coefficient* β *of panel (1) given by*

$$\hat{\beta}^*_{NT} = \left(\sum_{i=1}^{N} \hat{L}^{-2}_{22i} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2 \right)^{-1} \sum_{i=1}^{N} \hat{L}^{-1}_{11i} \hat{L}^{-1}_{22i} \left(\sum_{t=1}^{T} (x_{it} - \bar{x}_i) y^*_{it} - T\hat{\gamma}_i \right)$$

where

$$y^*_{it} = (y_{it} - \bar{y}_i) - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} \Delta x_{it} + \frac{\hat{L}_{11i} - \hat{L}_{22i}}{\hat{L}_{22i}} \beta(x_{it} - \bar{x}_i)$$

and $\hat{L}_i$ *and* $\hat{\gamma}_i$ *are defined as before. Then the statistics* $T\sqrt{N}\,(\hat{\beta}^*_{NT} - \beta)$ *and* $t_{\hat{\beta}^*_{NT}}$ *constructed from this estimator are numerically equivalent to the ones defined in proposition 1.2 and corollary 1.2.*

This proposition shows why it is difficult to construct a reliable point estimator based on the naive FMOLS estimator simply by using a transformation of y^*_{it} analogous to the single equation case. Indeed, as the proposition makes explicit, such an estimator would in general depend on the true value of the parameter that it is intended to estimate, except in very specialized cases, which we discuss below. On the other hand, this does not necessarily prohibit the usefulness of an estimator based on proposition 2.1 for the purposes of testing a particular hypothesis about a cointegrating relationship in heterogeneous panels. By using the hypothesized null value for β in the expression for y^*_{it}, proposition 2.1 can at least in principle be employed to construct a feasible FMOLS statistics to test the null hypothesis that $\beta_i = \beta$ for all i. However, as was reported in Pedroni (1996a), even in this case the small sample performance of the statistic is often subject to relatively large size distortion.

Proposition 2.1 also provides us with an opportunity to examine the consequences of ignoring heterogeneity associated with the serial correlation dynamics for the error process for this type of estimator. In particular, we notice that the modification involved in this estimator relative to the convential time series fully modified OLS estimator differs in two respects. First, it includes the estimators $\hat{L}_{11i}$ and $\hat{L}_{22i}$ that premultiply the numerator and denominator terms to control for the idiosyncratic serial correlation properties of individual cross sectional members prior to summing over N. Secondly, and more importantly, it includes in the transformation of the dependent variable y_{it}^* an additional term $\frac{\hat{L}_{11i} - \hat{L}_{22i}}{\hat{L}_{22i}} \beta(x_{it} - \bar{x}_i)$. This term is eliminated only in two special cases: (1) The elements L_{11i} and L_{22i} are identical for all members of the panel, and do not need to be indexed by i. This corresponds to the case in which the serial correlation structure of the data is homogeneous for all members of the panel. (2) The elements L_{11i} and L_{22i} are perhaps heterogeneous across members of the panel, but for each panel $L_{11i} = L_{22i}$. This corresponds to the case in which asymptotic variances of the dependent and independent variables are the same. Conversely, the effect of this term increases as (1) the dynamics become more heterogeneous for the panel, and (2) as the relative volatility becomes more different between the variables x_{it} and y_{it} for any individual members of the panel. For most panels of interest, these are likely to be important practical considerations. On the other hand, if the data are known to be relatively homogeneous or simple in its serial correlation structure, the imprecise estimation of these elements will decrease the attractiveness of this type of estimator relative to one that implicitly imposes these known restrictions.

B. Monte Carlo Simulation Results

We now study small sample properties in a series of Monte Carlo simulations. Given the difficulties associated with the feasible versions of the within dimension pooled panel fully modified OLS estimators discussed in the previous subsection based on proposition 2.1, it is not surprising that these tend to exhibit relatively large size distortions in certain scenarios, as reported in the Pedroni (1996a). Kao & Chiang (1997) subsequently also confirmed the poor small sample properties of the within dimension pooled panel fully modified estimator based on a version in which a first stage OLS estimate was used for the adjustment term. Indeed, such results should not be surprising given that the first stage OLS estimator introduces a second order bias in the presence of endogeneity, which is not eliminated asymptotically. Consequently, this bias leads to size distortion for the panel which is not necessarily eliminated even when the sample size grows large. By contrast, the feasible version of the between dimension group mean estimator does not require such an adjustment term even in the presence of heterogeneous serial correlation dynamics, and does not suffer from the same size distortion.[6] Consequently, we focus here on reporting the small sample Monte Carlo results for the between dimension group mean estimator and refer readers to Pedroni (1996a) for simulation results for the feasible versions of the within dimension pooled estimators.

To facilitate comparison with the conventional time series literature, we use as a starting point a few Monte Carlo simulations analogous to the ones studied in Phillips & Loretan (1991) and Phillips & Hansen (1990) based on their original work on FMOLS estimators for conventional time series. Following these studies, we model the errors for the data generating process in terms of a vector MA(1) process and consider the consequences of varying certain key parameters. In particular, for the purposes of the Monte Carlo simulations, we model our data generating process for the cointegrated panel (1) under assumptions 1.1 and 1.2 as

$$y_{it} = \alpha_i + \beta x_{it} + \mu_{it}$$

$$x_{it} = x_{it-1} + \varepsilon_{it}$$

$i = 1, \ldots, N$, $t = 1, \ldots, T$, for which we model the vector error process $\xi_{it} = (\mu_{it}, \varepsilon_{it})$ in terms of a vector moving average process given by

$$\xi_{it} = \eta_{it} - \theta_i \eta_{it-1};\ \eta_{it} \sim i.i.d.\ N(0, \Psi_i) \tag{3}$$

where θ_i is a 2×2 coefficient matrix and Ψ_i is a 2×2 contemporaneous covariance matrix. In order to accommodate the potentially heterogeneous nature of these dynamics among different members of the panel, we have indexed these parameters by the subscript i. We will then allow these parameters to be drawn from uniform distributions according to the particular experiment. Likewise, for each of the experiments we draw the fixed effects α_i from a uniform distribution, such that $\alpha_i \sim U(2.0, 4.0)$.

We consider first as a benchmark case an experiment which captures much of the richness of the error process studied in Phillips & Loretan (1991) and yet also permits considerable heterogeneity among individual members of the panel. In their study, Phillips & Loretan (1991), following Phillips & Hansen (1990), fix the following parameters $\theta_{11i} = 0.3$, $\theta_{12i} = 0.4$, $\theta_{22i} = 0.6$, $\Psi_{11i} = \Psi_{22i} = 1.0$, $\beta = 2.0$ and then permit θ_{21i} and Ψ_{21i} to vary. The coefficient θ_{21i} is particularly interesting since a non-zero value for this parameter reflects an absence of even weak exogeneity for the regressors in the cointegrating regression associated with (1), and is captured by the term L_{21i} in the panel FMOLS statistics. For our heterogeneous panel, we therefore set $\Psi_{11i} = \Psi_{22i} = 1.0$, $\beta = 2.0$ and draw the remaining parameters from uniform distributions which are centered around the parameter values set by Phillips &

Loretan (1991), but deviate by up to 0.4 in either direction for the elements of θ_i and by up to 0.85 in either direction for Ψ_{21i}. Thus, in our first experiment, the parameters are drawn as follows: $\theta_{11i} \sim U(-0.1, 0.7)$, $\theta_{12i} \sim (0.0, 0.8)$, $\theta_{21i} \sim U(0.0, 0.8)$, $\theta_{22i} \sim U(0.2, 1.0)$ and $\Psi_{21i} \sim U(-0.85, 0.85)$. This specification achieves considerable heterogeneity across individual members and also allows the key parameters θ_{21i} and Ψ_{21i} to span the set of values considered in Phillips and Loretan's study. In this first experiment we restrict the values of θ_{21i} to span only the positive set of values considered in Phillips and Loretan for this parameter. In several cases Phillips and Loretan found negative values for θ_{21i} to be particularly problematic in terms of size distortion for many of the conventional test statistics applied to pure time series, and in our subsequent experiments we also consider the consequences of drawing negative values for this coefficient. In each case, the asymptotic covariances were estimated individually for each member *i* of the cross section using the Newey-West (1987) estimator. In setting the lag length for the band width, we employ the data dependent scheme recommended in Newey & West (1994), which is to set the lag truncation to the nearest integer given by $K = 4\left(\frac{T}{100}\right)^{2/9}$, where T is the number of sample observations over time. Since we consider small sample results for panels ranging in dimension from $T = 10$ to $T = 100$ by increments of 10, this implies that the lag truncation ranges from 2 to 4. For the cross sectional dimension, we consider small sample results for $N = 10$, $N = 20$ and $N = 30$ for each of these values of T.

Results for the first experiment, with $\theta_{21i} \sim U($ 0.0, 0.8 are reported in Table I of Appendix B. The first column of results reports the bias of the point estimator and the second column reports the associated standard error of the sampling distribution. Clearly, the biases are small at –0.058 even in extreme cases when both the N and T dimensions are as small as $N = 10$, $T = 10$ and become minuscule as the T dimension grows larger. At $N = 10$, $T = 30$ the bias is already down to –0.009, and at $T = 100$ it goes to –0.001. This should be anticipated, since the estimators are superconsistent and converge at rate $T\sqrt{N}$, so that even for relatively small dimensions the estimators are extremely precise. Furthermore, the Monte Carlo simulations confirm that the bias is reduced more quickly with respect to growth in the T dimension than with respect to growth in the N dimension. For example, the biases are much smaller for $T = 30$, $N = 10$ than for $T = 10$, $N = 30$ for all of the experiments. The standard errors in column two confirm that the sampling variance around these biases are also very small. Similar results continue to hold in subsequent experiments with negative moving average coefficients, regardless of the data generating process for the serial correlation processes. Consequently, the first thing to note is that these estimators are extremely accurate even in panels with very heterogeneous serial correlation dynamics, fixed effects and endogenous regressors.

Of course these findings on bias should not come as a surprise given the superconsistency results presented in the previous section. Instead, a more central concern for the purposes of inference are the small sample properties of the associated *t*-statistic and the possibility for size distortion. For this, we consider the performance of the small sample sizes of the test under the null hypothesis for various nominal sizes based on the asymptotic distribution. Specifically, the last two columns report the Monte Carlo small sample results for the nominal 5% and 10% *p*-values respectively for a two sided test of the null hypothesis $\beta = 2.0$. As a general rule, we find that the size distortions in these small samples are remarkably small provided that the time series dimension, T, is not smaller than the cross sectional dimension, N. The reason for this condition stems primarily as a consequence of the estimation of the fixed effects. The number of fixed effects, α_i, grows with the N dimension of the panel. On the other hand, each of these N fixed effects are estimated consistently as T grows large, so that $\hat{\alpha}_i - \alpha_i$ goes to zero only as T grows large. Accordingly, we require T to grow faster than N in order to eliminate this effect asymptotically for the panel. As a practical consequence, small sample size distortion tends to be high when N is large relative to T, and decreases as T becomes large relative to N, which can be anticipated in any fixed effects model. As we can see from the results in Table I, in cases when N exceeds T, the size distortions are large, with actual sizes exceeding 30 and 40% when $T = 10$ and N grows from 10 to 20 and 30. This represents an unattractive scenario, since in this case, the tests are likely to report rejections of the null hypothesis when in fact it is not warranted. However, these represent extreme cases, as the techniques are designed to deal with the opposite case, where the T dimension is reasonably large relative to the N dimension. In these cases, even when the T dimension is only slightly larger than the N dimension, and even in cases where it is comparable, we find that the size distortion is remarkably small. For example, in the results reported in Table I we find that with $N = 20$, $T = 40$ the size of the nominal 5% and 10% tests becomes 4.5% and 9.3% respectively. Similarly, for $N = 10$, $T = 30$ the sizes for the Monte Carlo sample become 6.1% and 11% respectively, and for $N = 30$, $T = 60$, they become 4.7% and 9.6%. As the T dimension grows even larger for a fixed N dimension, the tests tend to become slightly undersized, with the actual size becoming slightly smaller than the nominal size. In this case the small sample tests actually become slightly more conservative than one would anticipate based on the asymptotic critical values.

Next, we consider the case in which the values for θ_{21i} span negative numbers, and for the experiment reported in Table II of Appendix B we draw this coefficient from $\theta_{21i} \sim U(-0.8, 0.0)$. Large negative values for moving average coefficients are well known to create size distortion for such estimators, and we anticipate this to be a case in which we have higher small sample distortion. It is interesting to note that in this case the biases for the point estimate become slightly positive, although as mentioned before, they continue to be very small. The small sample size distortions follow the same pattern in that they tend to be largest when T is small relative to N and decrease as T grows larger. In this case, as anticipated, they tend to be higher than for the case in which θ_{21i} spans only positive values. However, the values still fall within a fairly reasonable range considering that we are dealing with all negative values for θ_{21i}. For example, with $N = 10$, $T = 100$ we have values of 6.3% and 12% for the 5% and 10% nominal sizes respectively. For $N = 20$, $T = 100$ they become 9% and 15.6% respectively. These are still remarkably small compared to the size distortions reported in Phillips & Loretan (1991) for the conventional time series case.

Finally, we ran a third experiment in which we allowed the values for θ_{21i} to span both positive and negative values so that we draw the values from $\theta_{21i} \sim U(-0.4, 0.4)$. We consider this to be a fairly realistic case, and this corresponds closely to the range of moving average coefficients that were estimated in the purchasing power parity study contained in the Pedroni (1996a). We find the group mean estimator and test statistic to perform very well in this situation. The Monte Carlo simulation results for this case are reported in Table III of Appendix B. Whereas the biases for the case with large positive values of θ_{21i} in Table I were negative, and for the case with large negative values in Table II were positive, here we find the biases to be positive and often even smaller in absolute value than either of the first two cases. Most importantly, we find the size distortions for the t-statistic to be much smaller here than in the case where we have exclusively negative values for θ_{21i}. For example, with $N = 30$, and T as small as $T = 60$, we find the nominal 5% and 10% sizes to be 5.4% and 10.5%. Again, generally the small sample sizes for the test are quite close to the asymptotic nominal sizes provided that the T dimension is not smaller than the N dimension. Consequently, it appears to be the case that even when some members of the panel exhibit negative moving average coefficients, as long as other members exhibit positive values, the distortions tend to be averaged out so that the small sample sizes for the group mean statistic stay very close to the asymptotic sizes. Thus, we conclude that in general when the T dimension is not smaller than the N dimension, the asymptotic normality result appears to provide a very good benchmark for the sampling distribution under the null hypothesis, even in relatively small samples with heterogeneous serial correlation dynamics.

Finally, although power is generally not a concern for such panel tests, since the power is generally quite high, it is worth mentioning the small sample power properties of the group mean estimator. Specifically, we experimented by checking the small sample power of the test against the alternative hypothesis by generating the 10,000 draws for the DGP associated with case 3 above with $\beta = 1.9$. For the test of the null hypothesis that $\beta = 2.0$ against the alternative hypothesis that $\beta = 1.9$, we found that the power for the 10% p-value test reached 100% for $N = 10$ when T was 40 or more (or 98.2% when $T = 30$) and reached 100% for $N = 20$ when T was 30 or more, and for $N = 30$ the power reached 100% already when T was 20 or more. Consequently, considering the high power and the relatively small size distortion, we find the small sample properties of the estimator and associated t-statistic to be extremely well behaved in the cases for which it was designed.

IV. ESTIMATION ALGORITHM AND SOME EXAMPLES OF APPLICATIONS[7]

In this section we describe the algorithm for computing the panel FMOLS estimators and their associated test statistics and then discuss a few examples of their use. In summary, we can compute any one the desired statistics by performing the following steps:

1. *Estimate the panel regression and collect the residuals*. Specifically one should estimate the desired panel cointegration regression, making sure to include any desired intercepts, or common time dummies in the regression, and then collect the residuals $\hat{\mu}_{i,t}$ for each of the members of the panel. If the slopes are homogeneous, the common time dummy effects can be eliminated more simply by first demeaning the data over the time dimension prior to estimating the regression. Thus, construct $y_{it} - \bar{y}_t$, $x_{it} - \bar{x}_t$ for each variable, where $\bar{y}_t = N^{-1} \sum_{i=1}^{N} y_{it}$, $\bar{x}_t = N^{-1} \sum_{i=1}^{N} x_{it}$ prior to estimating the regression, and prior to the following steps.
2. *Estimate the long run covariances and autocovariances of the errors*. Use the estimated residuals from part (1) plus the differences of each of the regressors to construct a vector error series $\xi_{it} = (\mu_{it}, \varepsilon'_{it})'$. Note that the second element is a vector of dimension m, where m corresponds to the number of regressors. Now use any long run covariance matrix estimator, such as the Newey-West (1987) estimator to estimate the elements of the

long run covariance Ω_i and the autocovariances Γ_i. This can be done by applying the estimator to the entire $m+1$ vector $\xi_{it} = (\mu_{it}, \varepsilon'_{it})'$ to produce an $(m+1) \times (m+1)$ long run covariance matrix and autocovariances matrix. The elements of Ω_i and Γ_i then correspond to partitions of the $(m+1) \times (m+1)$ long run covariance matrix and autocovariance matrix respectively. Specifically, the far upper right scalar element of the $(m+1) \times (m+1)$ long run covariance matrix corresponds to Ω_{11i}. The lower $m \times m$ partition corresponds to Ω_{22i}, which is an $m \times m$ matrix representing the long run covariance among the regressors, and the remaining m elements in the column below the far upper right scalar element correspond to Ω_{21i}. Since the covariance matrix is symmetric, $\Omega_{12i} = \Omega_{21i}$. The same mapping corresponds the partitions of the $(m+1) \times (m+1)$ autocovariance matrix and the elements of Γ_i, except that unlike Ω_i, the autocovariance matrix Γ_i is not symmetric, so $\Gamma_{12i} \neq \Gamma'_{21i}$, and these elements must be extracted from the corresponding column and row partitions separately. Once Ω_i has been constructed, apply a Cholesky style triangularization to obtain the elements of the matrix L_i. Finally, we will use an estimate of the standard contemporaneous covariance matrix, Ω_i^o, for the elements of $\xi_{it} = (\mu_{it}, \varepsilon'_{it})'$, similarly partitioned.

3. *Construct the estimator.* Now we have all of the pieces required to construct the estimators. Each estimator uses a serial correlation correction term, γ_i, which can be constructed from the pieces obtained in part (2) above, as

$$\hat{\gamma}_i \equiv \hat{\Gamma}_{21i} + \hat{\Omega}^o_{21i} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} (\hat{\Gamma}_{22i} + \hat{\Omega}^o_{22i})$$

Next, using the elements of L_i, the expression for $y^*_{it} = (y_{it} - \bar{y}_i) - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} \Delta x_{it}$ can be constructed from the original data. Then the final step is to construct the cross product terms between y^*_{it} and $(x_{it} - x_i)$. This is sufficient now to compute either the point estimators or the associated t-statistics for any of the statistics.

It is worth noting two points here. The difference between the panel 'within' dimension estimators and the group mean 'between' dimension estimators is in the way in which the cross product terms are computed. For the 'within' dimension statistics, the cross product terms are computed by summing over the T and N dimensions separately for the numerator and the denominator. For the group mean 'between' dimension statistics, the cross product terms are computed by summing over the T dimension for the numerator and denominator separately, and then summing over the N dimension for the entire ratio. Consequently, the first point to note is that the algorithm as applied to the group mean estimator describes the same steps that one would take if one were estimating N different conventional FMOLS estimators and then taking the average of these. The same is true for the group mean t-statistic. Thus, if one already has a routine to estimate the conventional time series FMOLS estimator, then the group mean panel FMOLS estimator is extremely simple and convenient to estimate. The second point to note is that for the panel FMOLS 'within' dimension estimator we have used the estimates of Ω_i, Γ_i, Ω_i^o and γ_i to compute the weighted panel variances. But it is equally feasible to compute the unweighted panel variances by first averaging the values Ω_i, Γ_i, Ω_i^o before applying the transformations. Whether or not the two different treatments has much consequence for the estimate is likely to depend on how heterogeneous the values of Ω_i are across individual members.

Next, we briefly describe a few examples of the use of these panel FMOLS estimators. One obvious application is to the exchange rate literature, and in particular the purchasing power parity literature. Long run *absolute* or strong purchasing power parity predicts that nominal exchange rates and aggregate price ratios among countries should be cointegrated with a unit cointegrating vector, so that the real exchange rate is stationary. However, panel unit root tests based on Levin & Lin (1993) have generally found mixed results. See for example Oh (1996) and Papell (1997) and Wu (1996) among others. On the other hand, panel cointegration tests based on Pedroni (1995, 1997a) have generally rejected the null of no cointegration. See for example Canzoneri, Cumby & Diba (1996), Chinn (1997) and Taylor (1996) among others for these. By contrast, long run *relative* or weak purchasing power parity simply predicts that the nominal exchange rate and aggregate price ratios will be cointegrated, though not necessarily with a unit cointegrating vector. The panel FMOLS estimators presented in this paper are an obvious way to distinguish between these two hypothesis, and Pedroni (1996a, 1999) uses these panel FMOLS estimators to show that only the relative, weak form of purchasing power parity holds for a panel of post Bretton Woods period floating exchange rates. The latter paper contrasts results for both a parametric group mean DOLS estimator and nonparametric group mean FMOLS estimator for the weak purchasing power parity test. In a similar spirit, Alexius & Nilson (2000), Canzoneri, Cumby & Diba (1996), Chinn (1997) apply these panel FMOLS tests to test the Samulson-Balassa hypothesis that long run movement of real exchange rates are driven by differences in long run relative productivities among countries.

Other examples of the use of these panel FMOLS tests have been to the growth literature. Neusser & Kugler (1998) use the tests to investigate the connection between financial development and growth. Kao, Chiang & Chen (1999) use a panel FMOLS estimator and compare it to a panel DOLS estimator to investigate the connection between research and development expenditure and growth. Keller & Pedroni (1999) use the group mean panel estimator presented in this chapter to study the mechanism by which imported R&D impacts growth at the industry level and demonstrate the attractiveness of the more flexible form of the group mean estimator. Canning & Pedroni (1999) use the same group mean panel FMOLS test as a first step estimator to construct a test for the direction of long run causality between public infrastructure and long run growth. Finally Pedroni & Wen (2000) make use of the group mean panel FMOLS estimator as a first step estimator in an overlapping generations model to identify the position of the U.S., Japanese and European economies relative to the golden rule, and the extent to which social security transfer programs can move economies closer to this position.

This is just a brief summary of the application of these estimators to two literatures, the exchange rate and growth literatures. Needless to say, many potential applications exist beyond these two literatures.

V. DISCUSSION OF FURTHER RESEARCH AND CONCLUDING REMARKS

We have explored in this chapter methods for testing and making inferences about cointegrating vectors in heterogeneous panels based on fully modified OLS principles. When properly constructed to take account of potential heterogeneity in the idiosyncratic dynamics and fixed effects associated with such panels, the asymptotic distributions for these estimators can be made to be centered around the true value and will be free of nuisance parameters. Furthermore, based on Monte Carlos simulations we have shown that in particular the *t*-statistic constructed from the between dimension group mean estimator performs very well in that in exhibits relatively little small sample size distortion. To date, the techniques developed in this study have been employed successfully in a number of applications, and it will be interesting to see if the panel FMOLS methods developed in this paper fare equally well in other scenarios.

The area of research and application of nonstationary panel methods is rapidly expanding, and we take this opportunity to remark on a few further issues of current and future research as they relate to the subject of this chapter. As we have already discussed, the between dimension group mean estimator has an advantage over the within dimension pooled estimators presented in this chapter in that it permits a more flexible alternative hypothesis that allows for heterogeneity of the cointegrating vector. In many cases it is not known a priori whether heterogeneity of the cointegrating vector can be ruled out, and it would be particularly nice to test the null hypothesis that the cointegrating vectors are heterogeneous in such panels with heterogeneous dynamics. In this context, Pedroni (1998) provides a technique that allows one to test such a null hypothesis against the alternative hypothesis that they are homogeneous and demonstrates how the technique can be used to test whether convergence in the Solow growth model occurs to a distinct versus common steady states for the Summers and Heston data set.

Another important issue that is often raised for these types of panels pertains to the assumption of cross sectional independence as per assumption 1.2 in this chapter. The standard approach is to use common time dummies, which in many cases is sufficient to deal with cross sectional dependence. However, in some cases, common time dummies may not be sufficient, particularly when the cross sectional dependence is not limited to contemporaneous effects and is dynamic in nature. Pedroni (1997b) proposes an asymptotic covariance weighted GLS approach to deal with such dynamic cross sectional dependence for the case in which the time series dimension is considerably larger than the cross sectional dimension, and applies the panel fully modified form of the test to the purchasing power parity hypothesis using monthly OECD exchange rate data. It is interesting to note, however, that for this particular application, taking account of such cross sectional dependencies does not appear to impact the conclusions and it is possible that in many cases cross sectional dependence does not play as large a role as one might anticipate once common time dummies have been included, although this remains an open question.

Another important issue is parameteric versus non-parametric estimation of nuisance parameters. Clearly, any of the estimators presented here can be implemented by taking care of the nuisance parameter effects either nonparameterically using kernel estimators, or parametrically, as for example using dynamic OLS corrections. Generally speaking, non-parametric estimation tends to be more robust, since one does not need to assume a specific parametric form. On the other hand, since non-parametric estimation relies on fewer assumptions, it generally requires more data than parametric estimation. Consequently, for conventional time series tests, when data is limited it is often worth making specific parameteric assumptions. For panels, on the other hand, the greater abundance of data suggests an opportunity to take advantage of the greater robustness of nonparametric methods, though ultimately the choice may simply be a matter of taste. The Monte Carlo simulation results provided here demonstrate that even in the presence of considerable heterogeneity, non-parametric correction methods do very well for the group mean estimator and the corresponding t-statistic.

NOTES

1. The results in section 2 first appeared in Pedroni (1996a). The Indiana University working paper series is available at http://www.indiana.edu/ iuecon/workpaps/

2. In fact the computer program which accompanies this paper also allows one to implement these tests for any arbitrary number of regressors. It is available upon request from the author at ppedroni@indiana.edu

3. See Phillips & Moon (1999) for a recent formal study of the regularity conditions required for the use of sequential limit theory in panel data and a set of conditions under which sequential limits imply joint limits, including the case in which the long run variances differ among members of the panel.

4. These results are for the OLS estimator when the variables are cointegrated. A related stream of the literature studies the properties of the panel OLS estimator when the variables are not cointegrated and the regression is spurious. See for example Entorf (1997), Kao (1999), Phillips & Moon (1999) and Pedroni (1993, 1997a) on spurious regression in nonstationary panels.

5. A separate issue pertains to differences between the sample averages and the true population means. Since we are treating the asymptotics sequentially, this difference goes to zero as T grows large prior to averaging over N, and thus does not impact the limiting distribution. Otherwise, more generally we would require that the ratio N/T goes to zero as N and T grow large in order to ensure that these differences do not impact the limiting distribution. We return to this point in the discussion of the small sample properties in section 3.2.

6. Of course this is not to say that all within dimension estimators will necessarily suffer from this particular form of size distortion, and it is likely that some forms of the pooled FMOLS estimator will be better behaved than others. Nevertheless, given the other attractive features of the between dimension group mean estimator, we focus here on reporting the very attractive small sample properties of this estimator.

7. I am grateful to an anonymous referee for suggesting this section.

ACKNOWLEDGMENTS

I thank especially Bob Cumby, Bruce Hansen, Roger Moon, Peter Phillips, Norman Swanson and Pravin Trivedi and two anonymous referees for helpful comments and suggestions on various earlier versions, and Maria Arbatskaya for research assistance. The paper has also benefitted from presentations at the June 1996 North American Econometric Society Summer Meetings, the April 1996 Midwest International Economics Meetings, and workshop seminars at Rice University-University of Houston, Southern Methodist University, The Federal Reserve Bank of Kansas City, U. C. Santa Cruz and Washington University. The current version of the paper was completed while I was a visitor at the Department of Economics at Cornell University, and I thank the members of the Department for their generous hospitality. A computer program

which implements these tests is available upon request from the author at ppedroni@indiana.edu

REFERENCES

Alexius, A., & Nilson, J. (2000). Real Exchange Rates and Fundamentals: Evidence from 15 OECD Countries. *Open Economies Review,* forthcoming.

Canning, D., & Pedroni, P. (1999). *Infrastructure and Long Run Economic Growth*. CAE Working paper, No. 99–09, Cornell University.

Canzoneri M., Cumby, R., & Diba, B. (1996). *Relative Labor Productivity and the Real Exchange Rate in the Long Run: Evidence for a Panel of OECD Countries*. NBER Working paper No. 5676.

Chinn, M. (1997). *Sectoral Productivity, Government Spending and Real Exchange Rates: Empirical Evidence for OECD Countries*. NBER Working paper No. 6017.

Chinn, M., & Johnson, L. (1996). *Real Exchange Rate Levels, Productivity and Demand Shocks: Evidence from a Panel of 14 Countries*. NBER Working paper No. 5709.

Entorf, H. (1997). Random Walks and Drifts: Nonsense Regression and Spurious Fixed-Effect Estimation'. *Journal of Econometrics*, *80*, 287–96.

Evans, P., & Karras, G. (1996). Convergence Revisited. *Journal of Monetary Economics*, *37*, 249–265.

Im, K., Pesaran, H., & Shin, Y. (1995). *Testing for Unit Roots in Heterogeneous Panels*. Working paper, Department of Economics, University of Cambridge.

Kao, C. (1999). Spurious Regression and Residual-Based Tests for Cointegration in Panel Data'. *Journal of Econometrics*, *90*, 1–44.

Kao, C., & Chen, B. (1995). *On the Estimation and Inference of a Cointegrated Regression in Panel Data When the Cross-section and Time-series Dimensions Are Comparable in Magnitude*. Working paper, Department of Economics, Syracuse University.

Kao, C., & Chiang, M. (1997). *On the Estimation and Inference of a Cointegrated Regression In Panel Data*. Working paper, Department of Economics, Syracuse University.

Kao, C., Chiang, M., & Chen, B. (1999). International R&D Spillovers: An Application of Estimation and Inference in Panel Cointegration. *Oxford Bulletin of Economics and Statistics*, *61*(4), 691–709.

Keller, W., & Pedroni, P. (1999). *Does Trade Affect Growth? Estimating R&D Driven Models of Trade and Growth at the Industry Level*. Working paper, Department of Economics, Indiana University and University of Texas.

Levin, & Lin (1993). *Unit Root Tests in Panel Data; Asymptotic and Finite-sample Properties*. Working paper, Department of Economic, U. C. San Diego.

Mark, N., & Sul, D. (1999). *A Computationally Simple Cointegration Vector Estimator for Panel Data*. Working paper, Department of Economics, Ohio State University.

Neusser, K., & Kugler, M. (1998). Manufacturing Growth and Financial Development: Evidence from OECD Countries. *Review of Economics and Statistics*, *80*, 638–646.

Newey, W., & West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Coariance Matrix. *Econometrica*, *55*, 703–708.

Newey, W., & West, K. (1994). Autocovariance Lag Selection in Covariance Matrix Estimation'. *Review of Economic Studies*, *61*, 631–653.

Obstfeld M., & Taylor, A. (1996). *International Capital-Market Integration over the Long Run: The Great Depression as a Watershed*. Working paper, Department of Economics, U. C. Berkeley.

Oh, K. (1996). Purchasing Power Parity and Unit Root Tests Using Panel Data'. *Journal of International Money and Finance*, *15*, 405–418.

Papell, D. (1997). Searching for Stationarity: Purchasing Power Parity Under the Current Float'. *Journal of International Economics*, *43*, 313–32.

Pedroni, P. (1993). Panel Cointegration. Chapter 2 in *Panel Cointegration, Endogenous Growth And Business Cycles in Open Economies*, Columbia University Dissertation, Ann Arbor, MI: UMI Publishers.

Pedroni, P. (1995). *Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests, With an Application to the PPP Hypothesis*. Working paper, Department of Economics, No. 95–013, Indiana University.

Pedroni, P. (1996a). *Fully Modified OLS for Heterogeneous Cointegrated Panels and the Case of Purchasing Power Parity.* Working paper No. 96–020, Department of Economics, Indiana University.

Pedroni, P. (1996b). *Human Capital, Endogenous Growth, & Cointegration for Multi-Country Panels*. Working paper, Department of Economics, Indiana University.

Pedroni, P. (1997a). *Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests, With an Application to the PPP Hypothesis; New Results*. Working paper, Department of Economics, Indiana University.

Pedroni, P. (1997b). *On the Role of Cross Sectional Dependency in Dynamic Panel Unit Root and Panel Cointegration Exchange Rate Studies*. Working paper, Department of Economics, Indiana University.

Pedroni, P. (1998). *Testing for Convergence to Common Steady States in Nonstationary Heterogeneous Panels*. Working paper, Department of Economics, Indiana University.

Pedroni, P. (1999). *Purchasing Power Parity Tests in Cointegrated Panels*. Working paper, Department of Economics, Indiana University.

Pedroni, P., & Wen, Y. (2000). *Government and Dynamic Efficiency.* Working paper, Department of Economics, Cornell University and Indiana University.

Pesaran, H., & Smith, R. (1995). Estimating Long Run Relationships from Dynamic Heterogeneous Panels. *Journal of Econometrics*, *68,* 79–114.

Phillips, P., & Durlauf, S. (1986). Multiple Time Series Regressions with Integrated Processes'. *Review of Economic Studies*, *53*, 473–495.

Phillips, P., & Hansen, B. (1990). Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies*, *57*, 99–125.

Phillips, P., & Loretan, M. (1991). Estimating Long-run Economic Equilibria. *Review of Economic Studies*, *58*, 407–436.

Phillips, P., & Moon, H. (1999). Linear Regression Limit Theory for Nonstationary Panel Data'. *Econometrica, 67*, 1057–1112.

Quah, D. (1994). Exploiting Cross-Section Variation for Unit Root Inference in Dynamic Data'. *Economics Letters*, *44*, 9–19.

Taylor, A. (1996). *International Capital Mobility in History: Purchasing Power Parity in the Long-Run*. NBER Working paper No. 5742.

Wu, Y. (1996). Are Real Exchange Rates Nonstationary? Evidence from a Panel-Data Test. *Journal of Money Credit and Banking*, *28*, 54–63.

MATHEMATICAL APPENDIX A

Proposition 1.1: *We establish notation here which will be used throughout the remainder of the appendix. Let* $Z_{it} = Z_{it-1} + _{it}$ *where* $\xi_{it} = (\mu_{it}, \varepsilon_{it})'$. *Then by virtue of assumption 1.1 and the functional central limit theorem,*

$$T^{-1} \sum_{t=1}^{T} \tilde{Z}_{it}\xi'_{it} \rightarrow \int_{r=0}^{1} \tilde{B}(r, \Omega_i)\, d\tilde{B}(r, \Omega_i)' + \Gamma_i + \Omega_i^o \qquad \text{(A1)}$$

$$T^{-2} \sum_{t=1}^{T} \tilde{Z}_{it}\tilde{Z}'_{it} \rightarrow \int_{r=0}^{1} \tilde{B}(r, \Omega_i)\tilde{B}(r, \Omega_i)'\, dr \qquad \text{(A2)}$$

for all i, where $\tilde{Z}_{it} = Z_{it} - \bar{Z}_i$ *refers to the demeaned discrete time process and* $\tilde{B}(r, \Omega_i)$ *is demeaned vector Brownian motion with asymptotic covariance* Ω_i. *This vector can be decomposed as* $\tilde{B}(r, \Omega_i) = L'_i \tilde{W}_i(r)$ *where* $L_i = \Omega_i^{1/2}$ *is the lower triangular decomposition of* Ω_i *and* $\tilde{W}(r) = \left(W_1(r) - \int_0^1 W_1(r)\ dr,\ W_2(r) - \int_0^1 W_2(r)\ dr \right)'$ *is a vector of demeaned standard Brownian motion, with* W_{1i} *independent of* W_{2i}. *Under the null hypothesis, the statistic can be written in these terms as*

$$T\sqrt{N}(\hat{\beta}_{NT} - \beta) = \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(T^{-1} \sum_{t=1}^{T} \tilde{Z}_{it}\xi'_{it} \right)_{21}}{\frac{1}{N} \sum_{i=1}^{N} \left(T^{-2} \sum_{t=1}^{T} \tilde{Z}_{it}\tilde{Z}'_{it} \right)_{22}} \qquad \text{(A3)}$$

Based on (A1), as $T \rightarrow \infty$, *the bracketed term of the numerator converges to*

$$\left(\int_{r=0}^{1} \tilde{B}(r, \Omega_i)\, d\tilde{B}(r, \Omega_i)' \right)_{21} + \Gamma_{21i} + \Omega_{21i}^o \qquad \text{(A4)}$$

the first term of which can be decomposed as

$$\left(\int_{r=0}^{1} \tilde{B}(r, \Omega_i)\, d\tilde{B}(r, \Omega_i)' \right)_{21} = L_{11i}L_{22i} \left(\int W_{2i}\, dW_{1i} - W_{1i}(1) \int W_{2i} \right)$$

$$+ L_{21i}L_{22i} \left(\int W_{2i}\, dW_{2i} - W_{2i}(1) \int W_{2i} \right) \qquad \text{(A5)}$$

In order for the distribution of the estimator to be unbiased, it will be necessary that the expected value of the expression in (A4) be zero. But although the expected value of the first bracketed term in (A5) is zero, the expected value of the second bracketed term is given as

$$E\left[L_{21i}L_{22i}\left(\int W_{2i}\,dW_{2i} - W_{2i}(1)\int W_{2i}\right)\right] = \frac{1}{2}L_{21i}L_{22i} \qquad \text{(A6)}$$

Thus, given that the asymptotic covariance matrix, Ω_i, must have positive diagonals, the expected value of the expression (A4) will be zero only if $L_{21i} = \Gamma_{21i} = \Omega_{21i}^{o=0}$, which corresponds to strict exogeneity of regressors for all members of the panel. Finally, even if such strict exogeneity does hold, the variance of the numerator will still be influenced by the parameters L_{11i}, L_{22i} which reflect the idiosyncratic serial correlation patterns in the individual cross sectional members. Unless these are homogeneous across members of the panel, they will lead to non-trivial data dependencies in the asymptotic distribution.

Proposition 1.2: *Continuing with the same notation as above, the fully modified statistic can be written under the null hypothesis as*

$$T\sqrt{N}(\hat{\beta}^*_{NT} - \beta) = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left((0,1)\left(T^{-1}\sum_{t=1}^{T}\tilde{Z}_{it}\xi_{it}'\right)\left(1, \frac{\hat{L}_{21i}}{\hat{L}_{22i}}\right)' - \hat{\gamma}_i\right)}{\frac{1}{N}\sum_{i=1}^{N}\hat{L}_{22i}^{-2}\left(\left(T^{-2}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'\right)_{22}\right)} \qquad \text{(A7)}$$

Thus, based on (A1), as $T \to \infty$, the bracketed term of the numerator converges to

$$\left(\int_{r=0}^{1}\tilde{B}(r,\Omega_i)\,d\tilde{B}(r,\Omega_i)'\right)_{21} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}}\left(\int_{r=0}^{1}\tilde{B}(r,\Omega_i)\,d\tilde{B}(r,\Omega_i)'\right)_{22}$$

$$+\Gamma_{21i} + \Omega_{21i}^{o} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}}(\Gamma_{22i} + \Omega_{22i}^{o}) \qquad \text{(A8)}$$

which can be decomposed into the elements of $\tilde{W}_i$ such that

$$\left(\int_{r=0}^{1} \tilde{B}(r,\Omega_i)\, d\tilde{B}(r,\Omega_i)'\right)_{21} = L_{11i}L_{22i}\left(\int W_{2i}\, dW_{1i} - W_{1i}(1)\int W_{2i}\right)$$

$$+ L_{21i}L_{22i}\left(\int W_{2i}\, dW_{2i} - W_{2i}(1)\int W_{2i}\right) \qquad (A9)$$

$$\left(\int_{r=0}^{1} \tilde{B}(r,\Omega_i)\, d\tilde{B}(r,\Omega_i)'\right)_{22} = L_{22i}^{2}\left(\int W_{2i} dW_{2i} - W_{2i}(1)\int W_{2i}\right) \quad (A10)$$

where the index r has been omitted for notational simplicity. Thus, if a consistent estimator of Ω_i is employed, so that $\hat{\Omega}_i \rightarrow \Omega_i$ and consequently $\hat{L}_i \rightarrow L_i$ and $\hat{\gamma}_i \rightarrow \gamma$, then

$$\hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left((0,1)(T^{-1}\sum_{t=1}^{T} \tilde{Z}_{it}\xi_{it}')\left(1, \frac{\hat{L}_{21i}}{\hat{L}_{22i}}\right)' - \hat{\gamma}_i\right)$$

$$\rightarrow \int_0^1 W_{2i}(r)\, dW_{1i}(r) - W_{1i}(1)\int_0^1 W_{2i}(r)\, dr \qquad (A11)$$

where the mean and variance of this expression are given by

$$E\left[\int W_{2i} dW_{1i} - W_{1i}(1)\int W_{2i} dr\right] = 0 \qquad (A12)$$

$$E\left[\left(\int W_{2i}\, dW_{1i}\right)^2 - 2W_{1i}(1)\int W_{2i} dr \int W_{2i} dW_{1i} + W_{1i}(1)^2\left(\int W_{2i} dr\right)^2\right]$$

$$= \frac{1}{2} - 2\left(\frac{1}{3}\right) + \frac{1}{3} = \frac{1}{6} \qquad (A13)$$

respectively. Now that this expression has been rendered void of any idiosyncratic components associated with the original $\tilde{B}(r, \Omega_i)$, then by virtue of assumption 1.2 and a standard central limit theorem argument,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\int_0^1 W_{2i}(r)\, dW_{1i}(r) - W_{1i}(1)\int_0^1 W_{2i}(r)\, dr\right) \rightarrow N(0, 1/6) \quad (A14)$$

as $N \rightarrow \infty$. *Next, consider the bracketed term of the denominator of (A3), which based on (A1), as* $T \rightarrow \infty$, *converges to*

$$\left(\int_{r=0}^{1} \tilde{B}(r, \Omega_i)\tilde{B}(r, \Omega_i)'\right)_{22} = L_{22i}^2\left(\int_0^1 W_{2i}(r)^2\, dr - \left(\int_0^1 W_{2i}(r)\, dr\right)^2\right) \quad \text{(A15)}$$

Thus,

$$\hat{L}_{22i}^{-2}\left((T^{-2}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}')\right)_{22} \rightarrow \int_0^1 W_{2i}(r)^2 dr - \left(\int_0^1 W_{2i}(r)\, dr\right)^2 \quad \text{(A16)}$$

which has finite variance, and a mean given by

$$E\left[\int_0^1 W_{2i}(r)^2 dr - \left(\int_0^1 W_{2i}(r)\, dr\right)^2\right] = \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \quad \text{(A18)}$$

Again, since this expression has been rendered void of any idiosyncratic components associated with the original $\tilde{B}(r, \Omega_i)$, *then by virtue of assumption 1.2 and a standard law of large numbers argument,*

$$\frac{1}{N}\sum_{i=1}^{N}\left(\int_0^1 W_{2i}(r)^2\, dr - \left(\int_0^1 W_{2i}(r)\, dr\right)^2\right) \rightarrow \frac{1}{6} \quad \text{(A18)}$$

as $N \rightarrow \infty$. *Thus, by iterated weak convergence and an application of the continuous mapping theorem,* $T\sqrt{N}(\hat{\beta}^*_{NT} - \beta) \rightarrow N(0, 6)$ *for this case where heterogeneous intercepts have been estimated. Next, recognizing that* $T^{-1/2}\bar{y}_{i\cdot} \rightarrow \int_0^1 W_{1i}(r)\, dr$ *and* $T^{-1/2}\bar{x}_i \rightarrow \int_0^1 W_{2i}(r)\, dr$ *as* $T \rightarrow \infty$, *and setting* $\int W_{1i} = \int W_{2i} = 0$ *for the case where* $\bar{y}_i = \bar{x}_{i=0}$ *gives as a special case of* (A13) *and* (A17) *the results for the distribution in the case with no estimated* *intercepts. In this case the mean given by* (*A*12) *remains zero, but the variance in* (*A*13) *become* $\frac{1}{2}$ *and the mean in* (*A*17) *also becomes* $\frac{1}{2}$. *Thus,* $T\sqrt{N}(\hat{\beta}^{*}_{NT}-\beta)\rightarrow N(0,2)$ *for this case.*

Corollary 1.2: *In terms of earlier notation, the statistic can be rewritten as:*

$$t_{\hat{\beta}^{*}_{NT}}=\frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left((0,1)\left(T^{-1}\sum_{t=1}^{T}\tilde{Z}_{it}\xi_{it}'\right)\left(1,\frac{\hat{L}_{21i}}{\hat{L}_{22i}}\right)'-\hat{\gamma}_i\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\hat{L}_{22i}^{-2}\left(\left(T^{-2}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'\right)_{22}\right)}} \tag{A19}$$

where the numerator converges to the same expression as in proposition 1.2, and the root term of the denominator converges to the same value as in proposition 1.2. Since the distribution of the numerator is centered around zero, the asymptotic distribution of $t_{\hat{\beta}^{*}_{NT}}$ *will simply be the distribution of the numerator divided by the square root of this value from the denominator. Since*

$$E\left[\left(\int W_{2i}\,dW_{1i}\right)^2-2W_{1i}(1)\int W_{2i}\int W_{2i}\,dW_{1i}+W_{1i}(1)^2\left(\int W_{2i}\right)^2\right]$$

$$=E\left[\int W_{2i}^2-\left(\int W_{2i}\right)^2\right] \tag{A20}$$

by (A13) and (A17) regardless of whether or not $\int W_{1i}$, $\int W_{2i}$ *are set to zero, then* $t_{\hat{\beta}^{*}_{NT}}\rightarrow N(0,1)$ *irrespective of whether* $\bar{x}_i$, $\bar{y}_i$ *are estimated or not.*

Proposition 1.3: *Write the statistic as:*

$$\bar{t}_{\hat{\beta}^{*}_{NT}}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{L}_{11i}^{-2}\left((0,1)\left(T^{-1}\sum_{t=1}^{T}\tilde{Z}_{it}\xi_{it}'\right)\left(1,\frac{\hat{L}_{21i}}{\hat{L}_{22i}}\right)'-\hat{\gamma}_i\right)$$

$$\times\left((T^{-2}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}')_{22}\right)^{-1/2} \tag{A21}$$

Then the first bracketed term converges to

$$L_{11i}L_{22i}\left(\int_0^1 W_{2i}(r)\,dW_{1i}(r) - W_{1i}(1)\int_0^1 W_{2i}(r)\,dr\right)$$

$$\sim N\left(0,\, L_{11i}L_{22i}\int_0^1 W_{2i}(r)^2\,dr - \left(\int_0^1 W_{2i}(r)\,dr\right)^2\right) \tag{A22}$$

by virtue of the independence of $W_{21i}(r)$ and $dW_{1i}(r)$. Since the second bracketed term converges to

$$L_{22i}\left(\int_0^1 W_{2i}(r)^2\,dr - \left(\int_0^1 W_{2i}(r)\,dr\right)^2\right)^{-1/2} \tag{A23}$$

then, taken together, for $\hat{L}_i \rightarrow L_i$, (A21) becomes a standardized sum of i.i.d. standard normals regardless of whether or not $\int W_{1i}, \int W_{2i}$ are set to zero, and thus $\bar{t}_{\hat{\beta}^*_{NT}} \rightarrow N(0, 1)$ by a standard central limit theorem argument irrespective of whether $\bar{x}_i, \bar{y}_i$ are estimated or not.

Proposition 2.1: *Insert the expression for y^*_{it} into the numerator and use $y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + \mu_{it}$ to give*

$$\hat{\beta}^*_{NT} = \frac{\sum_{i=1}^{N} \hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left(\sum_{t=1}^{T}(x_{it}-\bar{x}_i)\left(\mu_{it} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}}\Delta x_{it}\right) - T\hat{\gamma}_i\right)}{\sum_{i=1}^{N}\hat{L}_{22i}^{2}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2} + \frac{\sum_{i=1}^{N}\hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left(1 + \frac{\hat{L}_{11i} - \hat{L}_{22i}}{\hat{L}_{22i}}\right)\beta\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}{\sum_{i=1}^{N}\hat{L}_{22i}^{-2}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2} \tag{A24}$$

Since $\hat{L}_{22i}^{-2} = \hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left(1 + \frac{\hat{L}_{11i} - \hat{L}_{22i}}{\hat{L}_{22i}}\right)$, the last term in (A24) reduces to β, thereby giving the desired result.

APPENDIX B

Table I. Small Sample Performance of Group Mean Panel FMOLS with Heterogeneous Dynamics

Case 1: $\theta_{21i} \sim (0.0, 0.8)$

N	*T*	bias	std error	5% size	10% size
10	10	–0.058	0.115	0.282	0.362
	20	–0.018	0.047	0.084	0.145
	30	–0.009	0.029	0.061	0.110
	40	–0.006	0.020	0.035	0.076
	50	–0.004	0.016	0.027	0.062
	60	–0.003	0.012	0.020	0.049
	70	–0.002	0.010	0.016	0.044
	80	–0.002	0.009	0.014	0.040
	90	–0.002	0.008	0.014	0.038
	100	–0.001	0.007	0.014	0.037
20	10	–0.034	0.079	0.291	0.378
	20	–0.012	0.033	0.100	0.166
	30	–0.006	0.020	0.076	0.132
	40	–0.004	0.014	0.045	0.093
	50	–0.003	0.011	0.039	0.081
	60	–0.003	0.009	0.028	0.066
	70	–0.002	0.007	0.026	0.059
	80	–0.002	0.006	0.021	0.055
	90	–0.002	0.006	0.020	0.050
	100	–0.001	0.005	0.018	0.052
30	10	–0.049	0.061	0.386	0.470
	20	–0.017	0.025	0.156	0.234
	30	–0.009	0.015	0.107	0.177
	40	–0.006	0.011	0.072	0.133
	50	–0.004	0.008	0.059	0.118
	60	–0.003	0.007	0.047	0.096
	70	–0.003	0.006	0.039	0.086
	80	–0.002	0.005	0.035	0.073
	90	–0.002	0.004	0.032	0.077
	100	–0.002	0.004	0.030	0.076

Notes: Based on 10,000 independent draws of the cointegrated system (1)–(3), with $\beta = 2.0$, $\alpha_{1i} \sim U(2.0, 4.0)$, $\Psi_{11i} = \Psi_{22i} = 1.0$, $\Psi_{21i} \sim U(-0.85, 0.85)$ and $\theta_{11i} \sim U(-0.1, 0.7)$, $\theta_{12i} \sim U(0.0, 0.8)$, $\theta_{21i} \sim U(0.0, 0.8)$, $\theta_{22i} \sim U(0.2, 1.0)$.

Table II. Small Sample Performance of Group Mean Panel FMOLS with Heterogeneous Dynamics

Case 2: $\theta_{21i} \sim U(-0.8, 0.0)$

N	*T*	bias	std error	5% size	10% size
10	10	0.082	0.132	0.422	0.498
	20	0.041	0.058	0.234	0.324
	30	0.025	0.037	0.187	0.268
	40	0.016	0.027	0.137	0.213
	50	0.012	0.021	0.115	0.185
	60	0.009	0.017	0.091	0.155
	70	0.007	0.014	0.087	0.151
	80	0.006	0.012	0.078	0.140
	90	0.005	0.011	0.072	0.135
	100	0.005	0.010	0.063	0.120
20	10	0.093	0.092	0.581	0.648
	20	0.043	0.042	0.352	0.447
	30	0.026	0.027	0.265	0.361
	40	0.017	0.020	0.205	0.294
	50	0.012	0.015	0.158	0.242
	60	0.009	0.012	0.130	0.211
	70	0.007	0.010	0.117	0.194
	80	0.006	0.009	0.109	0.181
	90	0.005	0.008	0.103	0.170
	100	0.004	0.007	0.090	0.156
30	10	0.070	0.071	0.563	0.630
	20	0.033	0.032	0.339	0.433
	30	0.020	0.020	0.259	0.352
	40	0.013	0.015	0.196	0.289
	50	0.009	0.011	0.152	0.236
	60	0.007	0.009	0.131	0.211
	70	0.006	0.008	0.113	0.190
	80	0.005	0.007	0.103	0.175
	90	0.004	0.006	0.096	0.164
	100	0.003	0.005	0.087	0.156

Notes: Based on 10,000 independent draws of the cointegrated system (1)–(3), with $\beta = 2.0$, $\alpha_{1i} \sim U(2.0, 4.0)$, $\Psi_{11i} = \Psi_{22i} = 1.0$, $\Psi_{21i} \sim U(-0.85, 0.85)$ and $\theta_{11i} \sim U(-0.1, 0.7)$, $\theta_{12i} \sim U(-0.8, 0.0)$, $\theta_{21i} \sim U(-0.8, 0.0)$, $\theta_{22i} \sim U(0.2, 1.0)$.

Table III. Small Sample Performance of Group Mean Panel FMOLS with Heterogeneous Dynamics

Case 3: $\theta_{21i} \sim U(-0.4, 0.4)$

N	*T*	bias	std error	5% size	10% size
10	10	0.009	0.129	0.284	0.367
	20	0.011	0.052	0.113	0.179
	30	0.008	0.033	0.086	0.150
	40	0.005	0.023	0.058	0.113
	50	0.004	0.018	0.048	0.093
	60	0.003	0.014	0.039	0.083
	70	0.002	0.012	0.037	0.077
	80	0.002	0.011	0.031	0.072
	90	0.002	0.009	0.029	0.068
	100	0.001	0.008	0.028	0.062
20	10	0.028	0.090	0.346	0.430
	20	0.014	0.037	0.145	0.222
	30	0.009	0.024	0.106	0.179
	40	0.006	0.017	0.077	0.138
	50	0.004	0.013	0.060	0.114
	60	0.003	0.010	0.048	0.093
	70	0.002	0.009	0.040	0.085
	80	0.002	0.008	0.037	0.083
	90	0.001	0.007	0.035	0.079
	100	0.001	0.006	0.035	0.078
30	10	0.008	0.069	0.317	0.402
	20	0.006	0.028	0.122	0.194
	30	0.004	0.018	0.095	0.155
	40	0.003	0.013	0.068	0.122
	50	0.002	0.010	0.054	0.105
	60	0.001	0.008	0.044	0.088
	70	0.001	0.007	0.038	0.082
	80	0.001	0.006	0.036	0.076
	90	0.001	0.005	0.033	0.073
	100	0.001	0.005	0.036	0.074

Notes: Based on 10,000 independent draws of the cointegrated system (1)–(3), with $\beta = 2.0$, $\alpha_{1i} \sim U(2.0, 4.0)$, $\Psi_{11i} = \Psi_{22i} = 1.0$, $\Psi_{21i} \sim U(-0.85, 0.85)$ and $\theta_{11i} \sim U(-0.1, 0.7)$, $\theta_{12i} \sim U(-0.4, 0.4)$, $\theta_{21i} \sim U(-0.4, 0.4)$, $\theta_{22i} \sim U(0.2, 1.0)$.