June, 1996

FULLY MODIFIED OLS FOR HETEROGENEOUS COINTEGRATED PANELS AND THE CASE OF PURCHASING POWER PARITY *

Peter Pedroni

Indiana University

mailing address: Economics, Indiana University
Bloomington, IN 47405
(812) 855-7925

email: PPEDRONI@INDIANA.EDU

Abstract: Using fully modified OLS principles, methods are developed for estimating and testing hypotheses for cointegrating vectors in heterogeneous panels which lead to asymptotically unbiased and nuisance parameter free standard distributions in the presence of idiosyncratic dynamics and fixed effects. Small sample properties are also investigated by Monte Carlo simulation under a variety of scenarios for the error processes. Finally, by way of illustration, the estimators are employed to address an empirical puzzle which has developed in recent panel studies of the purchasing power parity hypothesis for the post Bretton Woods data.

Keywords: Time Series, Panel Cointegration, Fully Modified OLS, Purchasing Power Parity
JEL Classifications: C22, C23, F31

* I thank especially Bob Cumby and Norman Swanson for helpful comments on earlier versions and Maria Arbatskaya for research assistance. The paper has also benefitted from presentations at the 1996 North American Econometric Society Summer Meetings, the Midwest International Economics Meetings, and workshop seminars at Rice University-University of Houston, Southern Methodist University and Washington University.

FULLY MODIFIED OLS FOR HETEROGENEOUS COINTEGRATED PANELS AND THE CASE OF PURCHASING POWER PARITY

I. Introduction

Methods for nonstationary time series panels, including unit root and cointegration tests, have been gaining increased acceptance in recent empirical research. The extension of conventional nonstationary methods 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, nonstationary panel methods provide an opportunity for researchers to exploit some of the attractive theoretical properties of nonstationary regressions while addressing in a natural and obvious 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 realistic small samples. By allowing data to be pooled in the cross sectional dimension, nonstationary panel methods have the potential to improve upon these small sample limitations. Conversely, the use of nonstationary time series asymptotics provides an opportunity to make panel methods more amenable to pooling aggregate level data. Conventional panel methods that were designed to employ short spans of stationary data have traditionally been of more limited use in areas of study that rely on aggregate data, in large part because these methods typically require that any of the associated dynamics be homogeneous among individual members of the panel. By contrast, superconsistent rates of convergence can be exploited in nonstationary panel methods to allow for practical implementations that permit relatively unrestricted heterogeneous dynamics among differing members of the panel.

Initial methodological work on nonstationary 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 and Lin (1993) derived distributions under more general conditions that allow for heterogeneous fixed effects and time trends. More recently, Im, Pesaran and 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 generally involves relationships within multivariate systems. Thus, Pedroni (1995a) studied the properties of spurious regressions and tests for cointegration in heterogeneous panels and derived appropriate distributions for these cases. These allow one to test for the presence of long run equilibria in multi-variate panels while permitting the dynamics and even the long run cointegrating relationships to be heterogeneous across individual members. Recent applications of these panel tests for cointegration include Canzoneri, Cumby and Diba (1996) to productivity and real exchange rates, Obstfeld and Taylor (1996) to international capital mobility, Pedroni (1995b) to endogenous growth theory and Taylor (1996) to historical episodes of purchasing power parity.

While these methods allow one to test for the presence of unit roots and cointegration in nonstationary panels, it would also useful to be able to test hypotheses about the cointegrating vectors in such panels. Pedroni (1995a) showed that the asymptotics and corresponding critical values associated with tests for cointegration depended in a key way on whether or not the cointegrated vectors could be assumed to be homogeneous across individual members of the panel. In other situations, economic theories may suggest a particular value of the cointegrating vector in panels which one would like to compare with available data, as in the purchasing power parity example that we consider. The purpose of this paper, therefore, is to investigate the properties of various estimators for such cointegrating vectors in panels with heterogeneous dynamics and to propose feasible statistics that can be used to make reliable inferences about the cointegrating vectors.

It is well known for the conventional single equation case that although ordinary least squares estimates of cointegrating vectors are superconsistent, the corresponding distributions are asymptotically biased and dependent on nuisance parameters associated with the serial correlation properties of the data. These difficulties are no less likely to persist for panels, and are likely to be further complicated by potential heterogeneity in the dynamics. Indeed, in related work, Kao and Chen (1995) document the extent of this bias in a panel of cointegrated time series for the case in which the dynamics are homogeneous. They also investigate the possibility of adjusting for this bias directly using a least squares dummy variable estimator, but find that the bias adjusted LSDV estimator performs no better than the unadjusted estimator in finite samples. Herein lies the challenge for panel methods to be applicable to making inferences regarding cointegrating vectors in realistic finite samples. As the cross section dimension increases, there is the potential for systematic second order biases that are associated with the poor performance of estimators designed for large time series samples to be compounded as they are averaged over the cross sectional dimension to the extent that they become substantial. Therefore, an important additional objective of this paper is to provide feasible statistics that minimize this problem in panels with modest length in the time series dimension and to demonstrate their successful use in an empirical application.

Specifically, we develop in this paper a pragmatic approach based on the fully modified OLS principles that were first employed by Phillips and Hansen (1990) to deal with the problems of asymptotic bias and nuisance parameter dependency associated with cointegrating vector estimates in the conventional single equation case. Fully modified techniques have since proven successful in treating a host of issues relating to these difficulties and continue to gain further appreciation following recent work by Phillips (1995), which demonstrates the attractive features of such an approach in the context of nonstationary VARs with unknown cointegrating rank. In this paper, we find that an appropriately modified FMOLS estimator performs relatively well for the purposes of making inferences in cointegrated panels with heterogeneous dynamics as the cross sectional dimension grows large even for panels with relatively short time series dimensions. In particular, we show that an adjustment term which captures the potential contribution of heterogeneity from the dynamics of the panel works to produce minimal size distortions in these panel FMOLS statistics.

Finally, by way of illustration, we also provide a simple example of how the methods developed here can be successfully applied toward resolving an empirical puzzle that has developed in the recent purchasing power parity literature which employs nonstationary panel methods to test for unit root and cointegrating relationships. The PPP hypothesis has often been a popular proving ground for new time series methods, and the current situation is no exception as recent developments in these methods have sparked renewed interest in the PPP question. In general, long run PPP has been relatively easy to evidence for exchange rates and price series that span long periods of time, but has been considerably more elusive for the relatively shorter spans of data available since the breakdown of the Bretton-Woods system in 1973, despite the fact that a considerable body of literature predicts that the nominal exchange rate regime should not matter for long run equilibrium real exchange rates. This has left open the question as to whether this might not simply be a consequence of the inherently low power of the tests that had been previously employed. Consequently, studies such as Frankel and Rose (1995), Oh (1994), Papell (1995), and Wei and Parsley (1995) and Wu (1996) have in one way or another examined the issue of whether purchasing power parity appears to hold under the recent float on the basis of more powerful panel unit root tests of the real exchange rate. The results have been somewhat mixed, however, and Papell (1995) demonstrates in a series of simulation exercises that most of the results of these studies are quite sensitive to the inclusion of different subsets of the panel.

These panel unit root studies implicitly assume a homogeneous unit value for the implied cointegrating vector between nominal exchange rates and aggregate price ratios. But many authors have argued that although there may be a tendency for these two variables to move together in equilibrium over long periods, the relationship need not necessarily be one for one under this more general interpretation of PPP. For example, Taylor (1988) offers the possibility that transportation costs or measurement errors may induce a non unit coefficient, while Patel (1990) suggests that differences in price indices between countries may be responsible. Fisher and Park (1991) consider the possibility that differential productivity shocks over time could produce a non unit coefficient, even though the nominal variables would continue to move together in equilibrium over long periods of time. In the context of panels, it is quite natural to imagine that if these factors play a role in the data, they are also just as likely to be of varying significance across differing countries, so that one should also be prepared for the possibility of heterogeneous cointegrating relationships. Using the methods developed in Pedroni (1995a), studies by Canzoneri, Cumby and Diba (1996), Obstfeld and Taylor (1996), Pedroni (1995a) and Taylor (1996) all find strong support for this weaker version of PPP with heterogeneous slope coefficients, and in some cases (Taylor, 1996) clearly reject stationarity of the real exchange rate.

These results for weak PPP do not by themselves exclude the stronger version of course, but they do suggest a fairly obvious interpretation for the mixed findings in tests of strong PPP based on panel unit root tests of the real exchange rate. Specifically, if the implicit maintained hypothesis that the cointegrating vector is homogeneous and equal to one for all countries is violated, even very slightly for only a small subset of countries, then because this mixes a few integrated series in with the majority of stationary ones, this is likely to lead to an inability to reject the null of a unit root for the panel at large. Obviously, one way to resolve this issue is to test the hypothesis regarding the cointegrating vector directly. In doing so, we find this scenario to be quite plausible since we find that the hypothesis of a homogeneous cointegrating vector equal to one is robustly rejected when the methods of this paper are employed. Thus, although the data support the notion of weak PPP, the methods developed in this paper are able to confidently reject the more restrictive notion of strong PPP for a panel of countries in the Post-Bretton Woods period.

The remainder of the paper is structured as follows. In the next section, we introduce the econometric models of interest for heterogeneous cointegrated panels. We then present a number of new asymptotic results for estimators designed to produce asymptotically unbiased and nuisance parameter free distributions around the true values for the cointegrating vectors. In section III we consider the finite sample properties of these estimators and propose a feasible FMOLS statistic that performs relatively well in finite panels with heterogeneous dynamics. Finally, in section IV we investigate the consequences of applying these new methods to the purchasing power parity question for a panel of 20 to 25 countries with post Bretton Woods data.

II. Asymptotic Results for Fully Modified OLS in Heterogeneous Cointegrated Panels

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

$$\begin{aligned} y_{it} &= \alpha_{1i} + \beta x_{it} + \mu_{it} \\ x_{it} &= \alpha_{2i} + x_{it-1} + \epsilon_{it} \end{aligned} \qquad \textbf{(1)}$$

where the vector error process $\xi_{it} = (\mu_{it}, \epsilon_{it})'$ 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 terms α_{1i}, α_{2i} allow either variable to exhibit idiosyncratic nonzero drifts for individual members of the panel. 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. For simplicity, we will refer to x_i as univariate, although each of the results of this study generalize in an obvious and straightforward manner to the vector case, unless otherwise indicated.

In order to develop 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 (1995a) regarding the degree of dependency across both these dimensions. For the time series dimension, we will assume that the conditions of the multivariate functional central limit theorems used in Phillips and Durlauf (1986) and Park and 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 for the partial sum* $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \xi_{it} \to B_i(r, \Omega_i)$ *holds as* $T \to \infty$ *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 quite different for different members of the panel. The asymptotic covariance matrix, Ω_i varies by cross section, and is given by $\Omega_i \equiv \lim_{T \to \infty} E\,[\,T^{-1}(\Sigma_{t=1}^{T}\xi_{it})\,(\Sigma_{t=1}^{T}\xi_{it}^{\prime})\,]$. The off diagonal term ω_{21i} captures the endogenous feedback effect between y_{it} and x_{it}, which is also permitted to be idiosyncratic to individual members of the panel.

For the cross sectional dimension, by contrast, we will employ the standard panel data assumption of independence. Hence:

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 NxT is given as* $I_N \otimes \Omega_i > 0$, *which is block diagonal positive definite 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 nondiagonal NxN matrix of Ω_{ij} elements, and then premultiplying the errors by this matrix in order to achieve cross sectional independence. Needless to say, 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.

Next, we consider the properties of a number of statistics that might be used for a cointegrated panel as described by (1) under these assumptions regarding the time series and cross dimensional dependencies in the data. The first statistic to consider is a standard panel OLS estimator of the cointegrating relationship. It is well known that the distribution of the single equation OLS estimator is asymptotically biased and 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.

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*

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

Then,

a) The estimator converges to the true value at rate $T\sqrt{N}$*, but in general the distribution under the null for such an estimator will be asymptotically biased and 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 distribution of $\hat{\beta}_{NT}$ *or its associated t-statistic.*

As the proof of proposition 1.1 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 perhaps 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 and Chen (1995) document this bias for a panel of cointegrated time series for the 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. For example, Phillips and Hansen (1990) propose a semi-parametric fully modified OLS estimator which eliminates the influence of this feedback, and Park (1992) proposes a closely related canonical cointegrating regression approach. Johansen (1988, 1991) instead employs a parametric full information maximum likelihood approach which jointly estimates these dynamics. While the case with dynamic panels is in general further complicated by the possibility of cross sectional heterogeneity of this feedback effect, once these effects are accounted for, it will be possible to construct estimators that are asymptotically unbiased along the lines of these conventional time series approaches.

While in principle it is possible to use a systems based approach along the lines of the Johansen procedure, in practice such systems are likely to quickly become infeasible as the number of cross sectional observations grow large, as one typically encounters in multi-country macro panels. Therefore, we use a more conventional panel data approach by allowing both the time series and cross sectional dimensions to grow large in the spirit of Levin and Lin (1993) and Pedroni (1995a), and then use a fully modified approach to treat the endogenous feedback effect nonparametrically in the spirit of Phillips and Hansen (1990). This has the further advantage of producing very convenient statistics that are asymptotically unbiased, free of nuisance parameters and normally distributed. The following proposition establishes an important preliminary result.

Proposition 1.2 (Asymptotic Distribution of the 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 $\hat{L}_i$ *is the lower triangular decomposition of a consistent estimator of the idiosyncratic* *asymptotic covariance matrix* $\Omega_i = \Omega_i^o + \Gamma_i + \Gamma_i^{\prime}$ *with* $\hat{L}_i$ *normalized such that* $\hat{L}_{i22} = \hat{\Omega}_{i22}^{-1/2}$*, and where* μ_{it}^* *is given by* $\mu_{it}^* = \mu_{it} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} \Delta x_{it}$ *and the serial correlation adjustment parameter* $\hat{\gamma}_i$ *is given by* $\hat{\gamma}_i \equiv \hat{\Gamma}_{21i} + \hat{\Omega}_{21i}^o - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} (\hat{\Gamma}_{22i} + \hat{\Omega}_{22i}^o)$. *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*

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

under the null as $T \rightarrow \infty$ *and* $N \rightarrow \infty$.

Thus, the distribution is free of any nuisance parameters associated with the idiosyncratic serial correlation pattern 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 and 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 depend 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 $\hat{L}_{2ii}$, 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 $\sum_{t=1}^{T} (x_{it} - \bar{x}_i) y_{it}^* - T\hat{\gamma}_i$.

For the fixed effects model, we also require the usual constraint that $N/T \rightarrow 0$ whenever we use estimated means in place of true means to ensure that this does not affect the asymptotic distribution. 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 $T\sqrt{N}(\hat{\beta}_{NT}^* - \beta)$. 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. Nonetheless, 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. 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 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,*

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

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.

As the proof 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.

Notice, however, that in contrast to the conventional single equation case studied by Phillips and 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 paradoxical 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 $\sum_{t=1}^{T} (x_{it} - \overline{x}_i)^2$. 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 the next section we consider various strategies for specifying these statistics in terms of observables and study the small sample properties of the resulting feasible statistics.

Before preceding to the small sample properties, we first consider one additional asymptotic result that will be of use. Recently Im, Pesaran and Smith (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. Pedroni (1995a) finds in fact that for the case in which residuals are used to test for cointegration, this covariance is approximately -1.326 with the numerator and 0.243 with the denominator for the standard model, and -0.238 and 0.026 respectively for the fixed effects model, and adjusts the computation of the asymptotic distributions accordingly. 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. The need for a group mean estimator is further diminished for estimation of cointegrated panels under a fully modified OLS approach given that this approach renders the numerator asymptotically centered around zero, so that any asymptotic covariance between summed terms, of the numerator and denominator, for example, do not play a role in the asymptotic distribution of $T\sqrt{N}(\hat{\beta}^*_{NT} - \beta)$ or $t_{\hat{\beta}^*_{it}}$. Nevertheless, it is also interesting to consider the possibility of a fully modified OLS group mean statistic in the present context.

Proposition 1.3 (Asymptotic Distribution of the Panel FMOLS Group Mean Estimator). *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,*

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

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, where* $\hat{L}_i$ *is the lower triangular decomposition of a consistent estimator of the idiosyncratic asymptotic covariance matrix* $\Omega_i = \Omega^o_i + \Gamma_i + \Gamma'_i$ *with* $\hat{L}_i$ *normalized such that* $\hat{L}_{i22} = \hat{\Omega}^{-1/2}_{i22}$ *, and where* y^*_{it} *is given by* $y^*_{it} = (y_{it} - \bar{y}_i) - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} \Delta x_{it}$ *and the serial correlation adjustment parameter* $\hat{\gamma}_i$ *is given by* $\hat{\gamma}_i \equiv \hat{\Gamma}_{21i} + \hat{\Omega}^o_{21i} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}} (\hat{\Gamma}_{22i} + \hat{\Omega}^o_{22i})$.

Notice 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, the common panel OLS based fully modified t-statistic and the group mean fully modified OLS t-statistic, 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 the paper 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 panel FMOLS estimators that were developed in the previous section and propose 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. As we will see, the feasible panel FMOLS statistics and the group mean FMOLS statistic perform differently under varying situations, each providing a comparative advantage under certain circumstances.

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.

Indeed, a similar principle holds true for the panel FMOLS group mean statistic defined in proposition 1.3 . In this case, it is true that the individual cross sectional fully modified OLS estimators are each asymptotically unbiased prior to averaging for the panel. However, in small samples, the fully modified OLS statistic is well known to exhibit considerable bias under certain circumstances. Phillips and Hansen (1990), for example, report biases for samples with T=50 ranging from -0.267 and -1.102 depending on the dynamic properties of the data. For the panel FMOLS group mean statistic, this effect is likely to be exacerbated for the t-statistic, which effectively acts to sum these biases over the N dimension in such cases. Therefore, although the superconsistency properties make the panel OLS and panel group mean estimators suitable for point estimates, the potential for asymptotic or small sample bias with respect to the time series dimension leads to more serious problems for inferences when these effects are accumulated in the N dimension.

Another possibility might 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 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 proposition makes this relationship precise.

Proposition 2.1 (Feasible 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*

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

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

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 equivalent to the ones defined in proposition 1.2 and corollary 1.2 .*

This makes it difficult to construct a feasible point estimator along these lines, since any such estimator would in general depend on the true value of the parameter that it is intended to estimate. On the other hand, this does not necessarily prohibit the usefulness of the estimator in proposition 2.1 for the purposes of inference regarding cointegrating relationships in such panels. For the purposes of a feasible point estimator, a simple panel OLS or group mean FMOLS estimator is more than adequate. What is lacking is a reasonable estimator that can be used to make valid inferences regarding a particular null hypothesis regarding the true value of β. In this case, proposition 2.1 can be used to construct the panel FMOLS statistics $T\sqrt{N}(\hat{\beta}^*_{NT} - \beta)$ and $t_{\hat{\beta}^*_{NT}}$, which turn out to be relatively well behaved for heterogeneous panels.

Before proceeding to the Monte Carlo simulation results that compare the small sample properties of these various statistics, consider again the precise form of modification entailed in the panel version of the FMOLS estimator. The modification differs from the standard single series in two ways. 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 $\hat{L}_{11i}$ and $\hat{L}_{22i}$ 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 $\hat{L}_{11i}$ and $\hat{L}_{22i}$ 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.

Consequently, each of these statistics is likely to perform differently under varying circumstances in small samples. Fortunately, as we will see, it is often the case that under those circumstances where one of the statistics is less likely to perform well in very small samples, the others tend to do well. Furthermore, as we demonstrate in the purchasing power parity example of the next section, for a typical real data example, we can expect to be reasonably confident about our choice of statistics.

We next study these properties in a series of Monte Carlo simulations. To facilitate an ease of comparison along these lines of interest, we consider as a starting point a few Monte

Carlo simulations analogous to the ones studied in Phillips and Loretan (1991) and Phillips and Hansen (1990) based on their original work on FMOLS estimators for conventional single equations. Thus, following these authors, we consider the following dynamics for the cointegrated panel (1) under assumptions 1.1 and 1.2, for which we model the vector error process $\xi_{it} = (\mu_{it}, \epsilon_{it})'$ in terms of a vector moving average process

$$\xi_{it} = \eta_{it} - \theta_i \eta_{it-1} \quad ; \qquad \eta_{it} \sim \; i.i.d. \; N(0, \Psi_i) \tag{2}$$

where the elements of θ_i and Ψ_i are permitted to vary according to the experiment. 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.

We consider first as a prototypical benchmark case an experiment which captures much of the richness of the error process studied in Phillips and Loretan (1991) and yet also permits considerable heterogeneity among individual members of the panel. In particular, Phillips and Loretan (1991), following Phillips and 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 element θ_{21i} is particularly interesting since a nonzero 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 the following uniform distributions which are centered around the parameter values set by Phillips and 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: $\theta_{11i} \sim U(-0.1, 0.7)$, $\theta_{12i} \sim U(0.0, 0.8)$, $\theta_{21i} \sim U(0.0, 0.8)$, $\theta_{22i} \sim U(0.2, 1.0)$ 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, and will consider negative values for the diagonals of θ_i in subsequent experiments. Finally, since we are also interested in the consequences of heterogeneous fixed effects, we allow the regression intercept parameter α_{1i} to be drawn from $\alpha_{1i} \sim U(2.0, 4.0)$. We initially consider a panel of dimension $N = 50$, $T = 100$ to correspond to the modest sized panels that are likely to be of greatest interest for these methods. Other dimensions are also examined in subsequent experiments.

The results are reported in Table I, where we denote the feasible panel FMOLS statistics constructed from estimated residuals as $\tilde{\beta}^*_{NT}$ and the feasible panel FMOLS statistics constructed with the panel adjustment term for y^*_{it} as in proposition 2.1 as $\hat{\beta}^*_{NT}$. We denote the group mean FMOLS statistics as $\bar{\beta}^*_{NT}$, and the simple panel OLS statistics as $\hat{\beta}_{NT}$. The asymptotic covariances were estimated individually for each member *i* of the cross section using the Newey-West estimator with 5 lags, and all results are based on 10,000 draws for each panel. The first two rows of the table report the actual sizes for the 5% and 10% nominal sizes of the test statistics, and the next two rows report the bias and standard error of the same statistic. As we would expect, the bias is considerable for the standard panel OLS estimator and its t-ratio, both of which we know to be biased even asymptotically. In panels, this effect is made worse by the fact that the bias is aggregated over the cross sectional dimension. This has the consequence

Table I. Monte Carlo Simulations for $N = 50$, $T = 100$

Fixed Effects Model with Heterogeneous Parameters, θ_i, Ψ_i

	Panel OLS		Residual FM		Adjusted FM		Group Mean	
	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	t_{β^*}
5 % size	0.720	0.620	0.880	0.921	0.147	0.174	0.558	0.240
10 % size	0.781	0.701	0.944	0.964	0.230	0.263	0.622	0.329
Bias	3.030	2.414	-6.986	-3.041	-2.006	-0.873	-1.944	-1.112
Std Error	1.835	1.443	1.964	0.761	2.618	1.134	2.661	1.168

Notes: Based on 10,000 replications of cointegrated system (1) with errors ξ_{it} given by the vector MA process (2) with parameters $= 2\,0, \alpha_i \sim (2\,0, .0$, $\Psi_{11} = \Psi_i = .0$, $_{21i}$ $U(0. \ , 0.85$, *and* θ_i *centered around* $_{11i}$ $0.$, $\theta_i = .4$ $\theta_{21} = 0\,4$, $_{22i}$ $0.$ *such that* $\theta_{11} \sim U\ {-0}\,1, .7$, $\theta_{12} \sim U\ 0. \ , 0\,8)$, $_{21i}$ $U(.0\ 0. \ )$, $\theta_i \sim (0\,2, .0$. *The lag truncation for the Newey-West kernel estimator of* Ω_i *was set at* $_i =$. *All distributions for* β_{NT} β *have been standardized by* $T\sqrt{\ }$.

of driving the actual size of the nominal 10% critical value to approximately 78% for the standardized estimator and to 70% for the associated t-ratio for a sample of dimension N=50, T=100. Thus, if we attempt to use the standard OLS estimator to make inferences in these type of panels, we will almost always tend to reject the null, even when it is true. On the other hand, notice that for the purposes of a point estimator, this bias is still relatively small. Dividing the bias of the standardized statistic by $T\sqrt{N}$ gives 0.004 as the bias of the panel OLS estimator, β_{NT}, for this sample. As N grows larger, the size distortion for the standardized statistics will grow, but the bias for the point estimator will further diminish. We can expect this bias to be associated predominately with two features of any panel; (1) the extent to which strict exogeneity of the regressors is violated and (2) the degree of heterogeneity of the associated dynamics.

Next, consider the group mean panel FMOLS statistics. We know that in contrast to the panel OLS statistic, these will be unbiased asymptotically and invariant to the dynamics of the data. On the other hand, we also know from the Monte Carlo studies of Phillips and Hansen and others, that the conventional single equation FMOLS estimator can be substantially biased in small samples. For the group mean panel FMOLS estimator, this small sample bias associated with the small T dimension has the potential to become worse as the number of cross sections grow large. On this basis, we can expect the problem to diminish as the time series dimension becomes sufficiently large to eliminate the small sample bias. For the Monte Carlo simulation reported in Table I, we see that the bias is -1.112 for the group mean t-statistic, which amounts to a 33% size for the nominal 10% critical value. Still, for the purposes of a point estimate, the superconsistency of this estimator again accounts for a relatively small bias, giving -0.002 as an even smaller bias than the simple panel OLS estimator. These biases for the point estimates are notably smaller than the small sample biases reported in Phillips and Hansen for either the single equation OLS or the single equation FMOLS. This stems from the fact that the rate of convergence is $\sqrt{N}$ times faster for the panel.

Next, consider the properties of the feasible FMOLS estimators associated with proposition 1.2 that are based on estimated residuals. For Table I, our intuition proves to be correct regarding the small sample properties of these estimators. In this example the t-statistic has a bias of -3.041 and an actual size of 96% at the 10% nominal level, which is even worse than the panel OLS t-statistic. For this statistic, the reasoning behind the bias is again similar, though the mechanism is somewhat different than the others. Specifically, it is the fact that the individual member's estimator is asymptotically biased, and this has the consequence of creating a systematic relationship between the estimated residuals and the regressors, which does not exist for the true residuals. Thus, even though the individual member's estimator is superconsistent, as will be the estimate of the residuals, they do have a second order bias that produces this relationship with the regressors in the cross product of the numerator. Even though this effect is second order for a single member, when it is accumulated over all of the members of the cross section, it has the potential to becomes first order. Table I reports the results for the case in which the individual member's residuals are estimated using OLS, but the effects are quantitatively very similar when the individual member residuals are estimated using a single equation FMOLS estimator, since the small sample biases are very similar for T=100.

Finally, we consider the consequences of constructing the feasible panel FMOLS estimator in accordance with the result in proposition 2.1. This is equivalent to imposing the null for the coefficient of interest β and estimating the other parameters. For the fixed effects model, the intercepts are unknown, and must be implicitly estimated. This is accomplished by demeaning the data in the modification for y_{it}^*. Thus, it is not the case that proposition 2.1 imposes known residuals, and of course the estimated residuals may still be biased in small samples. However, in contrast to the previous estimator, the form of this bias does not create the same type of systematic correlation between the residuals and the regressors in this case, and is equivalent to empirically demeaning the stationary cointegrating relationship, which generally amounts to a much smaller bias effect. Thus, by including the heterogeneous panel FMOLS adjustment term $\frac{\hat{L}_{11i} - \hat{L}_{22i}}{\hat{L}_{22i}}\beta(x_{it} - \overline{x}_i)$, the effect is to create an estimator whose finite sample bias is relatively small in most cases even when the N dimension is comparable to the T dimension of the panel and there is considerable persistence and feedback in the data. The resulting bias is -0.873 for the t-statistic and the actual size at the 10% level in this case is approximately 26%, which is an improvement over the group mean statistic for this example.

Table I.B of the appendix reports the consequences of varying both the time series and cross sectional dimensions of the panel for this experiment. In particular, we consider combinations of N and T, where N takes on values of 20 and 50 and where T takes on values of 20 and 100. It is interesting to note that for this experiment, the size properties of the feasible statistic based on proposition 2.1 remain fairly stable. There is slight rise in the size distortion as N increases relative to T, but this is expected, since we require the ratio N/T to be small in order to allow estimated means in the fixed effects model. In the extreme case, when N=50, T=20, this statistic does about as well as the group mean statistic, but in all other cases it slightly outperforms it. Most remarkable, however, is that in this experiment the properties of the feasible statistic based on estimated residuals and the panel OLS start out reasonable for very small N, but rapidly deteriorate as N grows larger. In subsequent experiments, we will show that even with very small N dimensions, as expected from proposition 2.1, the properties of the panel OLS will rapidly deteriorate when the values of L_{11i} and differ more substantially from one another, as turns out to be the case with the PPP example. Given the relative stability of the feasible FMOLS based on proposition 2.1 and the group mean FMOLS t-statistics, these clearly appear to be the preferred choice for data sets that correspond to the characteristics of this experiment.

Before proceeding to the purchasing power parity example, we first consider a number of other experiments, which help to illustrate the circumstances under which these two statistics are likely to prove more fragile in small samples. In particular it is well known that semi-parametric estimators such as FMOLS do not perform as well in the presence of negative moving average coefficients. Thus, in Table II.B we report the properties of these statistics under various scenarios with negative MA coefficients. As it turns out, however, these panel FMOLS statistics still do remarkably well in the presence of negative MA coefficients under a variety of circumstances. The first table reports results for the a simulation with similar parameters as in Table I, except that θ_{12i} and θ_{21i} are constructed to include some negative coefficients. Specifically, they are drawn from uniform distributions with $\theta_{12i} \sim U(-0.4, 0.4)$ and $\theta_{21i} \sim U(-0.4, 0.4)$. For this case, all of the t-statistics do fairly well, and in fact considerably better than the conventional single equation FMOLS statistics.

As it turns out, even if there are negative moving average coefficients present for some members of the panel, as long as there are a sufficient number of others with positive coefficients, the statistics still do very well. This turns out to be a characteristic of the purchasing power parity data as well in the next section, and one might arguably claim that this property will tend to characterize many real world panel data sets. The situation in which some of these statistics run into more trouble for small samples is when a much greater proportion of the individual members of the panel have large negative moving average coefficients. The third experiment in the table, makes this clear. In this case the values for θ_{21i} are constructed to be negative for all members of the panel. All other parameters are the same as in the first case, but θ_{21i} is drawn from $\theta_{21i} \sim U(-0.8, 0.0)$. In this case, the bias for the group mean t-statistic and the feasible FMOLS t-statistic of proposition 2.1 rise substantially, to 3.006 and 3.804 respectively and the size distortions become large. Interestingly, the situation is much better in the second experiment in which both θ_{12i} and θ_{21i} are all constructed to be negative. In this case the group mean and adjusted FMOLS of proposition 2.1 do just as well as in the first experiment. On the other hand, in the fourth experiment, with θ_{12i} all positive and θ_{21i} all negative, they again exhibit substantial bias. Notice however, that in all of these cases the feasible FMOLS statistics based on estimated residuals do fairly well and are quite stable. Thus, in summary, it can be said that all of the statistics do fairly well in small samples even with negative moving average coefficients as long as large negative moving average coefficients are not too pervasive among too many members of the panel. In the event that this is the case, then the feasible panel FMOLS statistic based on estimated residuals reliably outperforms the others. In the next section, we consider the properties of these statistics in the context of a data set that is used for the purchasing power parity example, and then apply the statistics to test the hypothesis that strong PPP holds for a panel of countries for the post Bretton Woods period.

IV. The Case of Purchasing Power Parity

In general, the hypothesis of long run purchasing power parity indicates that nominal exchange rates and aggregate price ratios should move together over long periods of time. Under what is termed "strong" purchasing power parity, this relationship is expected to be such that the variables move one for one in the long run, so that the cointegrating vector is equal to one in a bivariate relationship. Alternatively, under what is termed "weak" purchasing power parity, these nominal variables may tend to move together in the long run, so that there exists a cointegrating relationship, but they need not move directly one for one. This is likely to be the case under a number of different scenarios. For example, although there may be an equilibrium mechanism by indices, or the presence of transportation costs, or other types of real disturbances may lead to a nonunitary relationship. See for example Fisher and Park (1991), Patel (1990) or Taylor (1988)

and must be estimated since it will generally be unknown. In this case, PPP is often tested in

cointegration.

variables is unity implies that one can also interpret this as meaning that the real exchange is

hypothesis using raw panel unit root tests with mixed results. Using the methods developed in

Pedroni (1995a) and Taylor (1996) all find strong support for this weaker version of PPP with

if the true cointegrating vector is not homogeneous as assumed in a raw panel unit root test of the

we test this hypothesis of strong PPP in a more direct fashion by testing directly by means of the

Thus, the parameter of interest is β

$$s = \alpha + \beta_{it} + _{it} \quad \textbf{(3)}$$

where $_{it}$ p is the log aggregate price ratio in terms of the CPI between the two countries. All data are from the IFS, and are for the post Bretton-Woods period 1974-1993, and thus include 20 years of annual and monthly time series observations for 20 to 25 countries. The results are contained in Table II below. The truncation lag K_i for the Newey-West estimator of the asymptotic covariance matrix was permitted to vary by individual country for the single equation estimators as well as the panel estimators, and was chosen according to the data determined selection scheme suggested in Newey and West (1994).

Whereas the data from a few of the countries are able to reject the hypothesis that β is equal to one, for the most part, even though the individual point estimates are often far from one, the single equation tests are unable to reject this hypothesis. When the data are pooled together, on the other hand, the panel OLS estimate of the β ranges depending on the frequency of the data and whether aggregate time dummy effects are included, from 1.04 to 1.09, which might be considered fairly close to one. On the other hand, when we use the adjusted feasible panel FMOLS t-ratio statistic of proposition 2.1, we see that by pooling the data we are able to sharpen our inferences considerably, and in fact conclude with well over 90% confidence that this value is statistically different than 1.0, based on the asymptotic distribution of this statistic. To check against the possibility that these results are driven primarily by the strong rejections of a few outlier countries, we computed the pooled statistics with common time dummies again after omitting those three countries which individually showed the strongest ability to reject null. For the annual data, the three strongest rejections are France, India and Pakistan. Since Pakistan is not present in the full panel of monthly data, Chile was substituted as the next strongest rejection among available countries for the monthly data. While the nominal values for the corresponding statistics diminished after these outliers were omitted, the panel statistics were still able to reject

Table II. *Individual and Panel FMOLS Tests for Strong PPP for Post Bretton Woods Period based on IFS annual and monthly (in parentheses) data, 1974-1993.*

Country	$\hat{\alpha}_i$ *(intercept)*	$\hat{\beta}_{iOLS}$ *(slope)*	$\hat{\beta}^*_{iFM}$ *(slope)*	$t_{\hat{\beta}^*_{iFM}}$ *(t-ratio)*	K_i *(lags)*
Belgium	-3.63 (-3.66)	0.13 (0.47)	-0.69 (0.32)	-2.15 (-1.73)	1 (4)
Denmark	-1.98 (-2.01)	1.23 (1.42)	1.58 (1.56)	0.73 (1.91)	1 (3)
France	-1.89 (-1.92)	1.64 (1.71)	2.22 (2.03)	3.94 (5.80)	1 (5)
Germany	-0.67 (-0.70)	0.72 (0.70)	0.92 (0.79)	-0.22 (-1.68)	1 (3)
Ireland	0.35 (----)	0.75 (---)	1.08 (---)	0.40 (---)	2 (-)
Italy	-7.22 (-7.22)	0.83 (0.88)	1.08 (0.97)	0.55 (-0.46)	1 (4)
Netherlands	-0.79 (-0.82)	0.69 (0.69)	0.60 (0.69)	-1.03 (-2.40)	1 (2)
Sweden	-1.81 (-1.80)	1.23 (1.25)	1.07 (1.22)	0.15 (1.15)	1 (4)
Switzerland	-0.51 (-0.54)	1.01 (1.17)	1.16 (1.27)	0.43 (1.96)	2 (4)
U.K.	0.53 (-0.53)	0.63 (0.69)	0.60 (0.68)	-1.02 (-2.90)	1 (3)
Canada	-0.20 (-0.20)	1.29 (1.43)	1.09 (1.42)	0.15 (2.13)	1 (3)
Japan	-5.15 (-5.19)	1.88 (1.85)	1.75 (1.76)	2.25 (4.36)	2 (6)
Greece	-4.60 (-4.57)	1.02 (1.03)	0.91 (1.02)	-1.19 (0.52)	1 (4)
Iceland	-3.50 (----)	0.99 (---)	1.04 (---)	1.21 (---)	1 (-)
Portugal	-4.78 (-4.77)	0.99 (1.02)	1.08 (1.05)	0.97 (1.33)	2 (3)
Spain	-4.74 (-4.74)	0.83 (0.86)	0.98 (0.93)	-0.09 (-0.83)	7 (4)
Turkey	-6.06 (-5.93)	1.11 (1.09)	1.10 (1.10)	2.19 (5.20)	1 (7)
Australia	-0.10 (----)	1.44 (---)	1.43 (---)	1.69 (---)	2 (-)
N. Zealand	-0.41 (-0.38)	0.90 (1.19)	0.88 (1.11)	-0.68 (3.02)	1 (3)
S. Africa	-0.41 (----)	1.16 (---)	1.00 (---)	0.00 (---)	1 (-)
Chile	-4.93 (-4.84)	1.04 (1.18)	1.11 (1.23)	1.39 (9.05)	1 (8)
Mexico	1.33 (1.45)	1.03 (1.04)	0.99 (1.03)	-0.29 (3.20)	1 (5)
India	-2.40 (-2.37)	2.23 (2.12)	2.04 (2.08)	6.99 (7.90)	1 (4)
Pakistan	-6.57 (----)	0.94 (---)	0.86 (---)	-0.93 (---)	1 (-)
Korea	-2.63 (-6.56)	2.89 (0.97)	2.93 (0.92)	4.68 (-1.17)	1 (5)
Panel (25/20)		1.05 (1.09)	1.15 (1.16)	1.87 (6.37)	4.02 (8.13)
w. T dums		1.04 (1.09)	1.04 (1.12)	2.65 (14.50)	3.42 (15.89)
Panel (22/17)		1.05 (1.07)	1.06 (1.10)	1.66 (3.84)	3.76 (8.77)

Notes: Estimated regression is (3) in the text. The null hypothesis for the t-ratio is $H_o : \beta_i = 1.0$. *The last row of the table reports panel estimates excluding the following outliers: France, India, and Pakistan for annual data, and France, India and Chile for monthly data (see text for discussion). Bottom of last column contains panel group mean t-statistics.*

with at least 90% confidence in all cases.

As we have seen in the previous section, however, the small sample properties can often vary considerably from the asymptotic distributions depending on the dimensionality of the panel as well as the degree of heterogeneity and complexity of the serial correlation properties of the data. Based on the Monte Carlo simulations reported in the previous section, we might expect the adjusted panel FMOLS statistic of proposition 2.1 and possibly the group mean statistic to be relatively well behaved in the type of panel data that we use. In order to confirm this, a simulation was conducted based on the first order vector moving average representation of the differenced data for this panel. The results are reported in Table III for simulations based on the parameterizations of both the annual and monthly data. The heterogeneous intercepts are also based on estimates from the data, and the value of is set to $\beta = .0$ under the null.

As anticipated, the panel OLS estimator is poorly behaved. This is not surprising based on the predictions of proposition 2.1. The estimated parameters for the MA coefficients (not reported), show that there is both a substantial degree of heterogeneity in the dynamics as well a sizeable difference in the relative volatility between the left hand and right hand side variables, in which case we know that the panel OLS estimator will be ill behaved. The residual based panel FMOLS estimator also does less than ideal in these simulations. The systematic relationship between the regressors and the residuals that is created by the small sample bias of the individual member residual estimation causes the standard error to be too low, giving 0.32 for the monthly and 0.85 for the annual. This leads to the potential for the statistic to seriously under reject. Finally, we note that the t-statistics for the feasible FMOLS statistic $\hat{\beta}^*_{NT}$ from proposition 2.1 and the panel FMOLS group mean statistic $t_{\bar{\beta}^*_{NT}}$ from proposition 1.3 both do well in the annual and especially the monthly data based simulations. The monthly based Monte Carlo results are virtually identical for the two, producing biases of only -0.06 and -0.03 respectively, and actual sizes of 13.8% and 13.5% respectively. For the annual data, the two statistics differ somewhat in

Table III. PPP Data Based Simulations

Annual Based, N=25, T=20

	Panel OLS		Residual FM		Adjusted FM		Group Mean	
			$\tilde{\beta}^*_{NT}-\beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT}-\beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT}-\beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.774	0.000	0.183	0.247	0.192	0.239	0.949	0.237
10 % size	0.809	0.000	0.305	0.381	0.273	0.323	0.957	0.321
Bias	-1.391	-0.004	3.094	1.389	-1.178	-0.536	5.872	0.082
Std Error	7.800	0.025	1.949	0.851	3.503	1.573	32.353	1.653

Monthly Based, N=20, T=246

	Panel OLS		Residual FM		Adjusted FM		Group Mean	
	$\hat{\beta}_{NT}-\beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT}-\beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT}-\beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT}-\beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.911	0.000	0.017	0.003	0.078	0.074	0.938	0.077
10 % size	0.925	0.000	0.075	0.038	0.140	0.138	0.948	0.135
Bias	-8.635	-0.001	2.602	1.070	-0.146	-0.063	-2.223	-0.031
Std Error	14.587	0.002	0.941	0.320	2.679	1.106	25.683	1.109

Notes: Heterogeneous parameters are based on the idiosyncratic first order vector MA representations for the differenced PPP data. Intercepts are based on estimated values and the null is set at $\beta = 1.0$. *The number of lags is set to3 for the annual and 7 for the monthly data. Simulations are based on 10,000 replications each. All statistics are defined as in Table I.*

that $t_{\hat{\beta}^*_{NT}}$ has a small downward bias of -0.536 while $t_{\bar{\beta}^*_{NT}}$ has a minimal upward bias of 0.082. This explains the differences reported for the test results in Table II, where the group mean statistics are consistently somewhat larger in value than the feasible panel FMOLS statistics. For the annual data, the small downward bias implies that the reported values for $t_{\hat{\beta}^*_{NT}}$ represent a potentially stronger rejection than one would believe based on the asymptotic distribution, and for the monthly data, with minimal biases, critical values based on the asymptotic distributions turn out to be very representative of the small sample values. Thus, the data based simulations further confirm the strength of the rejection of the strong purchasing power parity hypothesis which calls for a common unit cointegrating vector.

V. Concluding Remarks

We have explored in this paper 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 unbiased and free of nuisance parameters. Furthermore, Monte Carlos simulations have been used to show how various feasible statistics constructed on the basis of these principles will tend to behave under varying scenarios for the error processes. Finally, by applying these methods to a panel of IFS data for the post Bretton Woods periods we find evidence to support the idea that although weak PPP appears to hold, the stronger form of PPP does not, and results from raw panel unit root tests on the real exchange rate are likely to be misleading.

Needless to say, much additional theoretical and simulation work remains to be done in the area of nonstationary panels, and the PPP example is intended only as a simple illustration of how these methods can be expected to prove fruitful in applied areas of interest. Even for the PPP example, there are many other facets still to be explored, such as the consequences of relaxing the symmetry restriction of the bivariate regression form for the aggregate price ratio. While the asymptotic distributions for each of the t-statistics in this study will not be affected by the number of left hand side regressors, the small sample properties are likely to depend on this as well. Other important extensions for this line of research include the possibility of developing efficient methods that account for richer and more direct cross sectional dependencies and even intra-panel cointegrating relationships that hold across different members of the panel which are likely to be of interest in applications such as the PPP example. Finally, while the PPP result is in itself interesting, the panel FMOLS techniques developed in this study should prove to have wide applicability to any of a number of different applications involving cointegrated panels. Based on an initial version of this paper, Canzoneri, Cumby and Diba (1996) have already demonstrated the successful application of these methods to an empirical test of the Balassa-Samuelson Hypothesis for a panel of OECD countries. It will be interesting to see if these panel FMOLS methods fair equally well in other scenarios.

REFERENCES

Canzoneri M., R. Cumby and B. Diba (1996) "Relative Labor Productivity and the Real Exchange Rate in the Long Run: Evidence for a Panel of OECD Countries," manuscript, Georgetown University.

Fisher E. and J. Park (1991) "Testing Purchasing Power Parity Under the Null Hypothesis of Cointegration," *The Economic Journal*, 101, 1476-84.

Frankel, J. and A. Rose (1995) "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries," Forthcoming, *Journal of International Economics*.

Im, K. , Pesaran, H. and Y. Shin (1995) "Testing for Unit Roots in Heterogeneous Panels," manuscript, University of Cambridge.

Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica* 59, 1551-80.

Kao, C. and B. Chen (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,” manuscript, Syracuse University.

Levin and Lin (1993) "Unit Root Tests in Panel Data; Asymptotic and Finite-sample Properties," U.C. San Diego Working Paper.

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

Obstfeld M. and A. Taylor (1996) "International Capital-Market Integration over the Long Run: The Great Depression as a Watershed," manuscript, U.C. Berkeley.

Oh, K. (1994) "Purchasing Power Parity and Unit Root Tests Using Panel Data," forthcoming *Journal of International Money and Finance*.

Papell, D. (1995) "Searching for Stationarity: Purchasing Power Parity Under the Current Float," manuscript, University of Houston.

Park, J. and M. Ogaki (1991) "Seemingly Unrelated Canonical Cointegrating Regressions," The Rochester Center for Economic Research, Working Paper No. 280.

Park, J. and P. Phillips (1988) "Statistical Inference in Regressions with Integrated Processes: Part I," *Econometric Theory*, 4, 468-497.

Patel, J. (1990) "Purchasing Power Parity as a Long Run Relation," *Journal of Applied Econometrics*, 5, 367-79.

Pedroni, P. (1995a) "Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis," Indiana University Working Papers in Economics, No. 95-013.

Pedroni, P. (1995b) "Human Capital, Endogenous Growth, and Cointegration for Multi-Country Panels," manuscript, Indiana Unversity.

Phillips, P. (1995) "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1023-1078.

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

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

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

Park, J. (1992) "Canonical Cointegrating Regressions," *Econometrica* 60, 119-143.

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," manuscript, Northwestern University.

Taylor, M. (1988) "An Empirical Examination of Long-Run Purchasing Power Parity Using Cointegration Techniques," *Applied Economics*, 20, 1369-82.

Wei, S. and D. Parsley (1995) "Purchasing Power Dis-Parity During the Floating Rate Period: Exchange Rate Volatility, Trade Barriors and Other Culprits," manuscript, Harvard University.

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

MATHEMATICAL APPENDIX

Proposition 1.1: *We establish notation here which will be used throughout the remainder of the appendix. Let* $Z_{it} = \alpha_i + Z_{it-1} + \xi_{it}$ *where* $\alpha_i = (\alpha_{1i}, \alpha_{2i})'$ *and* $\xi_{it} = (\mu_{it}, \epsilon_{it})'$. *Then by virtue of assumption 1.1 and the functional central limit theorem,*

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

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

for all i, where $\tilde{Z}_{it} = Z_{it} - \bar{Z}_i$ *refers to the demeaned discrete time process and* $\bar{B}(r,\Omega_i)$ *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 Cholesky of* Ω_i *and* $\tilde{W}(r) = (W_1(r) - \int_0^1 W_1(r)dr,\, W_2(r) - \int_0^1 W_2(r)dr)'$ *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*

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

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

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

the first term of which can be decomposed as

$$\left(\int_{r=0}^{1} \tilde{B}(r,\Omega_i)\, d\tilde{B}(r,\Omega_i)^{\prime} \right)_{21} = L_{11i}L_{22i}\left(\int W_{2i}dW_{1i} - W_{2i}(1)\int W_{2i}\right) + L_{21i}L_{22i}\left(\int W_{2i}dW_{2i} - W_{1i}(1)\int W_{2i}\right) \quad \textbf{(A5)}$$

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

$$E\left[L_{21i}L_{22i}\left(\int W_{2i}dW_{2i} - W_{1i}(1)\int W_{2i}\right)\right] = \tfrac{1}{2}L_{21i}L_{22i} \quad \textbf{(A6)}$$

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* $L_{21i} = \Gamma_{21i} = \Omega^{o}_{21i} = 0$*, 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 nontrivial data dependencies and a failure of the conditions for the appropriate central limit theorems if they are not properly purged from the data prior to summing over N.*

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

$$T\sqrt{N}(\hat{\beta}^{*}_{NT} - \beta) = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left((0,1)(T^{-1}\sum_{t=1}^{T}\tilde{Z}_{it}\xi_{it}^{\prime})(1,\frac{\hat{L}_{21i}}{\hat{L}_{22i}})^{\prime} - \hat{\gamma}_i\right)}{\frac{1}{N}\sum_{i=1}^{N}\hat{L}_{22i}^{-2}\left((T^{-2}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}^{\prime})_{22}\right)} \quad \textbf{(A7)}$$

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

$$\left(\int_{r=0}^{1} \tilde{B}(r,\Omega_i)\, d\tilde{B}(r,\Omega_i)^{\prime} \right)_{21} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}}\left(\int_{r=0}^{1} \tilde{B}(r,\Omega_i)\, d\tilde{B}(r,\Omega_i)^{\prime} \right)_{22} + \Gamma_{21i} + \Omega^{o}_{21i} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}}(\Gamma_{22i} + \Omega^{o}_{22i}) \quad \textbf{(A8)}$$

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

$$\left(\int_{r=0}^{1} \tilde{B}(r,\Omega_i)\, d\tilde{B}(r,\Omega_i)' \right)_{21} = L_{11i}L_{22i}\left(\int W_{2i}dW_{1i} - W_{2i}(1)\int W_{2i}\right) + L_{21i}L_{22i}\left(\int W_{2i}dW_{2i} - W_{1i}(1)\int W_{2i}\right) \quad \textbf{(A9)}$$

and

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

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

$$\hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left((0,1)(T^{-1}\sum_{t=1}^{T} \tilde{Z}_{it}\xi_{it}')(1,\frac{\hat{L}_{21i}}{\hat{L}_{22i}})' - \hat{\gamma}_i \right) \to \int_0^1 W_{2i}(r)\, dW_{1i}(r) - W_{1i}(1)\int_0^1 W_{2i}(r)dr \quad \textbf{(A11)}$$

where the mean and variance of this expression are given by

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

$$E\left[(\int W_{2i}\, dW_{1i})^2 - 2\, W_{1i}(1)\int W_{2i}dr \int W_{2i}\, dW_{1i} + W_{1i}(1)^2(\int W_{2i}dr)^2\right] = \frac{1}{2} - 2\left(\frac{1}{3}\right) + \frac{1}{3} = \frac{1}{6} \quad \textbf{(A13)}$$

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,*

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

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

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

Thus,

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

which has finite variance, and a mean given by

$$E\left[\int_0^1 W_{2i}(r)^2 dr - \left(\int_0^1 W_{2i}(r)dr\right)^2\right] = \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \quad \textbf{(A17)}$$

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,*

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

as $N\rightarrow\infty$*. Thus, by iterated weak covergence 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* $T^{-1/2}\bar{y}_i \rightarrow \int_0^1 W_{1i}(r)dr$ *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 (A12) remains zero, but the variance in (A13) become ½ and the mean in (A17) also becomes ½. 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:*

$$t_{\hat{\beta}^*_{NT}} = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{L}_{11i}^{-1}\hat{L}_{22i}^{-1}\left((0,1)(T^{-1}\sum_{t=1}^{T}\tilde{Z}_{it}\xi_{it}^{\prime})(1,\frac{\hat{L}_{21i}}{\hat{L}_{22i}})^{\prime} - \hat{\gamma}_i\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\hat{L}_{22i}^{-2}\left((T^{-2}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}^{\prime})_{22}\right)}} \tag{A19}$$

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*

$$E\left[(\int W_{2i}dW_{1i})^2 - 2W_{1i}(1)\int W_{2i}\int W_{2i}dW_{1i} + W_{1i}(1)^2(\int W_{2i})^2\right] = E\left[\int W_{2i}^2 - \left(\int W_{2i}\right)^2\right] \tag{A20}$$

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:*

$$\bar{t}_{\hat{\beta}^*_{NT}} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{L}_{11i}^{-2}\left((0,1)(T^{-1}\sum_{t=1}^{T}\tilde{Z}_{it}\xi_{it}^{\prime})(1,\frac{\hat{L}_{21i}}{\hat{L}_{22i}})^{\prime} - \hat{\gamma}_i\right)\left((T^{-2}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}^{\prime})_{22}\right)^{-1/2} \tag{A21}$$

Then the first bracketed term converges to

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

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

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

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*

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

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.*

Table I.B: Fixed Effects Model with Heterogeneous Dynamics

Benchline Model Centered Around: $\theta_{11i} = 0.3$, $\theta_{12i} = 0.4$, $\theta_{21i} = 0.4$, $\theta_{22i} = 0.6$

$N = 50$, $T = 100$

	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.720	0.620	0.880	0.921	0.147	0.174	0.558	0.240
10 % size	0.781	0.701	0.944	0.964	0.230	0.263	0.622	0.329
Bias	3.030	2.414	-6.986	-3.041	-2.006	-0.873	-1.944	-1.112
Std Error	1.835	1.443	1.964	0.761	2.618	1.134	2.661	1.168

$N = 50$, $T = 20$

	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.682	0.535	0.772	0.874	0.270	0.344	0.568	0.332
10 % size	0.741	0.630	0.880	0.933	0.359	0.431	0.627	0.423
Bias	2.814	2.077	-6.206	-2.924	-2.576	-1.214	-1.597	-1.069
Std Error	1.872	1.374	1.933	0.856	3.427	1.606	2.963	1.661

$N = 20$, $T = 100$

	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.339	0.153	0.081	0.083	0.215	0.234	0.720	0.322
10 % size	0.419	0.231	0.170	0.189	0.301	0.328	0.767	0.422
Bias	0.395	0.262	2.704	1.153	2.608	1.111	3.482	1.408
Std Error	2.014	1.344	1.425	0.570	2.743	1.154	2.871	1.169

$N = 20$, $T = 20$

	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.345	0.131	0.180	0.264	0.224	0.286	0.688	0.410
10 % size	0.428	0.203	0.301	0.399	0.309	0.376	0.738	0.494
Bias	0.179	0.119	3.137	1.445	1.875	0.861	3.046	1.484
Std Error	2.075	1.282	1.899	0.835	3.500	1.609	3.289	1.650

Notes: Based on 10,000 replications of cointegrated system (1) with errors ξ_{it} given by the vector MA process (2) with parameters $\beta = 2.0$, $\alpha_{1i} \sim U(2.0, 4.0)$, $\Psi_{11i} = \Psi_{22i} = 1.0$, $\Psi_{21i} \sim U(-0.85, 0.85)$, $\theta_{11i} \sim U(-0.1, 0.7)$, $\theta_{12i} \sim U(0.0, 0.8)$, $\theta_{21i} \sim U(0.0, 0.8)$, $\theta_{22i} \sim U(0.2, 1.0)$. The lag truncation for the Newey-West kernel estimator of Ω_i was set at $K_i = 5$ when T=100 and $K_i = 3$ when T=20 . All distributions for $\beta_{NT} - \beta$ have been standardized by $T\sqrt{N}$.

Table II.B: Fixed Effects Model with Heterogeneous Dynamics

Consequences of Negative Moving Average Coefficients, N=50,T=100

Centered Around: $\theta_{11i} = 0.3$, $\theta_{12i} = 0.0$, $\theta_{21i} = 0.0$, $\theta_{22i} = 0.6$

	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.294	0.141	0.034	0.045	0.145	0.171	0.561	0.191
10 % size	0.380	0.220	0.097	0.130	0.225	0.256	0.622	0.278
Bias	0.309	0.219	-2.365	-1.033	-1.818	-0.796	-1.846	-0.860
Std Error	1.854	1.317	1.285	0.545	2.699	1.175	2.831	1.203

Centered Around: $\theta_{11i} = 0.3$, $\theta_{12i} = -0.4$, $\theta_{21i} = -0.4$, $\theta_{22i} = 0.6$

	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.603	0.485	0.022	0.029	0.130	0.152	0.531	0.153
10 % size	0.673	0.574	0.065	0.084	0.208	0.239	0.599	0.230
Bias	-2.416	-1.878	1.766	0.768	-1.751	-0.765	-1.599	-0.675
Std Error	1.855	1.431	1.484	0.636	2.609	1.133	2.698	1.170

Centered Around: $\theta_{11i} = 0.3$, $\theta_{12i} = 0.0$, $\theta_{21i} = -0.4$, $\theta_{22i} = 0.6$

	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.600	0.434	0.075	0.102	0.773	0.811	0.991	0.938
10 % size	0.670	0.531	0.198	0.247	0.850	0.875	0.994	0.964
Bias	2.427	1.744	2.926	1.274	6.912	3.006	8.747	3.805
Std Error	1.857	1.321	1.302	0.544	2.869	1.188	2.984	1.211

Centered Around: $\theta_{11i} = 0.3$, $\theta_{12i} = 0.4$, $\theta_{21i} = -0.4$, $\theta_{22i} = 0.6$

	$\hat{\beta}_{NT} - \beta$	$t_{\hat{\beta}_{NT}}$	$\tilde{\beta}^*_{NT} - \beta$	$t_{\tilde{\beta}^*_{NT}}$	$\hat{\beta}^*_{NT} - \beta$	$t_{\hat{\beta}^*_{NT}}$	$\bar{\beta}^*_{NT} - \beta$	$t_{\bar{\beta}^*_{NT}}$
5 % size	0.996	0.992	0.214	0.278	0.999	1.000	1.000	1.000
10 % size	0.998	0.996	0.431	0.516	1.000	1.000	1.000	1.000
Bias	7.270	5.496	3.816	1.662	14.376	6.255	19.093	7.651
Std Error	2.147	1.507	1.261	0.513	3.342	1.234	3.629	1.238

Notes: Based on 10,000 replications of cointegrated system (1) with errors ξ_{it} given by the vector MA process (2) with parameters $\beta = 2.0$, $\alpha_{1i} \sim U(2.0, 4.0)$, $\Psi_{11i} = \Psi_{22i} = 1.0$, $\Psi_{21i} \sim U(-0.85, 0.85)$. The elements of θ_i were drawn from uniform distributions centered around the indicated values with deviations of 0.4 in either direction. The lag truncation for the Newey-West kernel estimator of Ω_i was set at $K_i = 5$. All distributions for $\beta_{NT} - \beta$ have been standardized by $T\sqrt{N}$.