A parametric approach to the estimation of cointegration vectors in panel data

Jörg Breitung[*]
Humboldt University Berlin
Institute of Statistics and Econometrics
Spandauer Strasse 1
10178 Berlin, Germany

January 14, 2002

Abstract

In this paper a parametric framework for estimation and inference in cointegrated panel data models is considered that is based on a cointegrated VAR(p) model. A convenient two-step estimator is suggested where in the first step all individual specific parameters are estimated, whereas in the second step the long-run parameters are estimated from a pooled least-squares regression. The two-step estimator and related test procedures can easily be modified to account for contemporaneously correlated errors, a feature that is often encountered in multi-country studies. Monte Carlo simulations suggest that the two-step estimator and related test procedures outperform semiparametric alternatives such as the FM-OLS approach, especially if the number of time periods is small.

1 Introduction

Since the papers by Levin and Lin (1992, 1993) and Pedroni (1995) it has become popular to investigate long-run relationships by applying cointegration techniques to multi-country data. The attractive feature of such an analysis is that employing panel data with a substantial number of cross section units (countries) takes advantage of a much richer data source than using pure time series data.

An important problem with the analysis of panel data is its ability to capture heterogeneity due to country specific characteristics. The usual panel cointegration framework therefore assumes that the mean (or trend) and the short-run dynamics may differ across countries, whereas the long-run relationship is the same for all countries. The reason for assuming a homogenous long-run relationship is that the underlying economic principles that are employed to establish the long-run equilibrium (for example the purchasing power parity) should apply similarly in all economies, whereas the adjustment process towards the long-run equilibrium may differ due to behavioral and institutional characteristics.

Another important feature of the panel data model considered here is a possible contemporaneous correlation among cross section units. In many country studies this cross section correlation cannot be captured by a time-specific random effect (e.g. O'Connell 1998). Thus, to allows for arbitrary contemporaneous correlation among the errors, recent work employ simulation techniques to mimic the cross-correlation pattern among the errors (e.g. Chang 2001, Wu and Wu 2001).

Pedroni (1995, 2000) and Phillips and Moon (1999) suggest an asymptotically efficient estimation procedure that is based on the "fully-modified OLS" (FM-OLS) approach suggested by Phillips and Hansen (1990). This method employs kernel estimators of the nuisance parameters that affect the asymptotic distribution of the OLS estimator. In order to achieve asymptotic efficiency, the FM-OLS estimator accounts for a possible endogeneity of the regressors and serial correlation of the errors. Although this nonparametric approach is a very elegant way to deal with nuisance parameters, it may be problematical especially in fairly small samples. Furthermore, it is well known that nonparametric estimators may have poor properties in special cases, for example if the process has a moving average polynomial with a

root close to the unit circle (e.g. Schwert 1989, Perron and Ng 1996).

Another problem is that it is based on a single equation approach. Therefore, it is assumed that there is only a single cointegration relationship. Furthermore, the normalization of the cointegration vectors requires that the dependent variable enters the cointegration relationship. This assumption is however questionable if the cointegration vector is unknown (e.g. Boswijk 1996; Saikkonen, 1999).

For these reasons, a parametric approach may be a promising alternative, in particular, for panels with a small number of time periods. In this paper a vector error correction model (VECM) is employed to represent the dynamics of the system. Our framework can be seen as a panel analog of Johansen's cointegrated vector autoregression, where the short-run parameters are allowed to vary across countries and the long-run parameters are homogenous. Unfortunately, in such a setup the ML estimator cannot be computed from solving a simple eigenvalue problem as in Johansen (1988). Instead, in section 2 we adopt a two-step estimation procedure that was suggested by Ahn and Reinsel (1990) and Engle and Yoo (1991) for the usual time series model. As in Levin and Lin (1993) the individual specific parameters are estimated in a first step, whereas in a second step the common long-run parameters are estimated from a pooled regression. The resulting estimator is asymptotically efficient and normally distributed. Furthermore, a number of test procedures that are based on the two-step approach is considered in section 4 and extensions to more general models are addressed in section 5. The results of a couple of Monte Carlo simulations presented in section 6 suggest that the two-step estimator performs better than the FM-OLS estimator in typical sample sizes. Some conclusions and suggestions for future work can be found in section 7.

Finally, a word on the notational conventions applied in this paper. A standard Brownian motion is written as $W_i(a)$. Although there are different Brownian motions for different cross section units i, we sometimes drop the index i for convenience. This has no consequences for the final results since they depend on the *expectation* of the stochastic functionals. Furthermore, if there is no risk of misunderstanding, we drop the limits and the argument a (or da). For example, the term $\int_0^1 aW_i(a)\mathrm{d}a$ will be economically written as $\int aW$. As usual $[b]$ is used to indicate the integer part of b.

2 ML estimimation

For the ease of exposition we first consider a cointegrated VAR(1) model with the VECM presentation

$$\Delta y_{it} = \alpha_i \beta' y_{i,t-1} + \varepsilon_{it} , \quad t = 0, 1, \dots, T; \ i = 1, \dots, N , \qquad (1)$$

where ε_{it} is an k-dimensional white noise error vector with $E(\varepsilon_{it}) = 0$ and positive definite covariance matrix $\Sigma_i = E(\varepsilon_{it}\varepsilon'_{it})$. Furthermore, it is assumed that the number of time periods is the same for all cross section units (balanced panel). Various extensions of this model will be considered in Section 5.

In this specification the cointegration vectors β are the same for all cross section units, whereas the "loading matrix" α_i is allowed to vary across i. A similar setup is considered by Pesaran et al. (1999), Pedroni (1995, 2000) and Phillips and Moon (1999). Assuming normally distributed errors, we can concentrate the log-likelihood function with respect to the individual specific parameters $\alpha_1, \dots, \alpha_N$ and $\Sigma_1, \dots, \Sigma_N$ yielding

$$L_c(\beta) = c_0 - \sum_{i=1}^{N} \frac{T}{2} \log |\widehat{\Sigma}_i(\beta)| , \qquad (2)$$

where c_0 is some constant and

$$\widehat{\Sigma}_i(\beta) = T^{-1} \sum_{t=1}^{T} \widetilde{\varepsilon}_{it}(\beta)\widetilde{\varepsilon}_{it}(\beta)'$$

$$\widetilde{\varepsilon}_{it}(\beta) = \left[\Delta y_{it} - \left(\sum_{t=1}^{T} \Delta y_{it} y'_{i,t-1}\beta \right) \left(\sum_{t=1}^{T} \beta' y_{i,t-1} y'_{i,t-1}\beta \right)^{-1} \beta' y_{i,t-1} \right] .$$

The problem with this criterion function is that it cannot be maximized by solving a simple eigenvalue problem. In the pure time series case with $N = 1$, the maximization of $L_c(\beta)$ is equivalent to maximizing $|\widehat{\Sigma}|$, which leads to a simple eigenvalue problem. For $N > 1$, however, we have to maximize the expression $\prod_{i=1}^{N} |\widehat{\Sigma}_i(\beta)|$, which cannot be solved by a simple eigenvalue problem.

Nevertheless, it is possible to maximize $L_c(\beta)$ in (2) by using numerical techniques. It is well known that r^2 restrictions are required to identify the cointegration vectors. Following Johansen (1995, p. 72) the cointegration

3

vectors can be normalized as $\beta_c = \beta(c'\beta)^{-1}$. An important special case of this normalization is obtained by letting $c = [I, 0]'$, so that $\beta = [I, -B]'$, where B is a $r \times (k - r)$ parameter matrix. Such a normalization is used by Ahn and Reinsel (1990) and Phillips (1991, 1995).

The ML estimator is obtained by maximizing the log-likelihood function (2) subject to the normalization restrictions (cf. Pesaran et al. 1999). In practice this approach may become computationally burdensome and for a small number of time periods, problems with the convergence of the Gauss-Newton algorithm may occur. In the following section a simple two-step approach is suggested to obtain an estimator that is asymptotically equivalent to the ML estimator.

3   The two-step estimator

Since the ML procedure is computationally burdensome it is preferable to employ a simple two-step[1] estimation procedure that has the same large sample properties as the ML procedure. Engle and Yoo (1991) have shown that the information matrix of the Gaussian likelihood is asymptotically block diagonal with respect to the "short-run parameters" (α_i, Σ_i) and the matrix of cointegration vectors β. Therefore, the matrix β can be estimated conditional on some consistent initial estimator of α_i and Σ_i $(i = 1, \ldots, N)$.

To motivate the two-step estimator, consider the transformed VECM model

$$
\begin{aligned}
\gamma_i' \Delta y_{it} &= \gamma_i' \alpha_i \beta' y_{i,t-1} + \gamma_i' \varepsilon_{it} \\
z_{it} &= \beta' y_{i,t-1} + v_{it} ,
\end{aligned}
\tag{3}
$$

where $z_{it} = (\gamma_i'\alpha_i)^{-1}\gamma_i'\Delta y_{it}$, $v_{it} = (\gamma_i'\alpha_i)^{-1}\gamma_i'\varepsilon_{it}$ and γ_i is a $k \times r$ matrix with $rk(\gamma_i'\alpha_i) = r$. From

$$
\Sigma_v = E(v_{it}v_{it}') = (\gamma_i'\alpha_i)^{-1}\gamma_i'\Sigma_i\gamma_i(\alpha_i'\gamma_i)^{-1}
\tag{4}
$$

it follows that $\Sigma_v - (\alpha_i'\Sigma^{-1}\alpha_i)^{-1}$ is positive semi-definite and, therefore, the optimal choice of the transformation is $\gamma_i' = \alpha_i'\Sigma_i^{-1}$. The resulting estimator is asymptotically equivalent to the Gaussian ML estimator (cf. Reinsel 1993, p. 170).

---

[1]This estimator is called a "three-step estimator" by Engle and Yoo (1991). Here we follow Reinsel (1993, p. 170f) and refer to it as the two-step estimation procedure.

A consistent estimator (as $T \to \infty$) of α_i can be obtained from estimating separate models for all N cross sections. If $r = 1$, one may use the two-step estimator suggested by Engle and Granger (1987), whereas for $r > 1$ the ML estimator of Johansen (1988, 1991) can be used. However, in the latter case it is important to re-normalize the cointegration vectors so that they do not depend on individual specific parameters. Let $\widehat{\beta}_i^{ML}$ denote the ML estimator of the cointegration matrix suggested by Johansen (1988, 1991). The estimator is normalized such that $\widehat{\beta}_i^{ML\prime} S_{11,i} \widehat{\beta}_i^{ML} = I_r$, where $S_{11,i} = \sum_{t=1}^{T} y_{i,t-1} y_{i,t-1}'$. Since the distribution of $S_{11,i}$ depends on α_i and Σ_i, the ML estimator applies an individual specific normalization. To obtain the same normalization for all cointegration matrices $\widehat{\beta}_1^{ML}, \ldots, \widehat{\beta}_N^{ML}$ one may apply the normalization $\widehat{\beta}_{c,i}^{ML} = \widehat{\beta}_i^{ML}(\widehat{\beta}_{i,1}^{ML})^{-1} = [I, -\widehat{B}_i^{ML}]'$, where $\widehat{B}_i^{ML\prime} = -\widehat{\beta}_{i,2}^{ML}(\widehat{\beta}_{i,1}^{ML})^{-1}$ and $\widehat{\beta}_{i,1}^{ML}$ $(\widehat{\beta}_{i,2}^{ML})$ denotes the upper (lower) $r \times r$ $(n - r \times r)$ block of $\widehat{\beta}_i^{ML}$.

A problem with such a normalization is that $\beta_{i,1}^{ML}$ needs not to be invertible and, thus, the normalization may be invalid (see Boswijk 1996 and Saikkonen 1999). To avoid such problems an estimator can be used that is based on an eigenvalue problem not depending on nuisance parameters. Such an estimator is obtained by solving the eigenvalue problem $|\lambda_i I - S_{11,i}| = 0$. The eigenvectors corresponding to the r smallest eigenvalues are called the Principal Component (PC) estimator of the cointegration vectors (e.g. Harris, 1997). The estimated cointegration matrices $\widehat{\beta}_i^{PC}$ are normalized as $\widehat{\beta}_i^{PC\prime} \widehat{\beta}_i^{PC} = I_r$ and, thus, the normalization does not depend on individual specific parameters.

At the first estimation stage, the restriction that the cointegration vectors are the same for all cross section units is ignored, but this does not affect the asymptotic properties of the estimator. For the asymptotic properties of the two-step estimator it is only required that the parameters are estimated consistently as $T \to \infty$.

At the second stage, the system is transformed such that the cointegration matrix β can be estimated by ordinary least-squares of the *pooled* regression

$$\widehat{z}_{it} = \beta' y_{i,t-1} + \widehat{v}_{it} \quad i = 1, \ldots, N; \ t = 1, \ldots, T, \tag{5}$$

where $\widehat{z}_{it} = (\widehat{\alpha}_i' \widehat{\Sigma}_i^{-1} \widehat{\alpha}_i)^{-1} \widehat{\alpha}_i' \widehat{\Sigma}_i^{-1} \Delta y_{it}$ and $\widehat{v}_{it}$ is defined analogously.

If the cointegration vectors are normalized as $\beta = [I, -B]'$, then the

regression is rewritten as

$$\widehat{z}_{it}^{+} = B y_{i,t-1}^{(2)} + \widehat{v}_{it} \; , \tag{6}$$

where $\hat{z}_{it}^{+} = y_{i,t-1}^{(1)} - (\widehat{\alpha}_i' \widehat{\Sigma}_i^{-1} \widehat{\alpha}_i)^{-1} \widehat{\alpha}_i' \widehat{\Sigma}_i^{-1} \Delta y_{it}$, $y_{it} = [y_{it}^{(1)'}, y_{it}^{(2)'}]'$ and $y_{it}^{(1)}$ $(y_{it}^{(2)})$ are $r \times 1$ $(k-r \times 1)$ subvectors of y_{it}. It is interesting to note that $\hat{z}_{it}^{+}$ adopts an endogeneity correction similar as the estimator of Phillips and Moon (1999). The important difference is, however, that the latter approach employs a nonparametric estimate of the endogeneity effect, whereas $\hat{z}_{it}^{+}$ is based on a parametric endogeneity correction based on a VAR(p) model.

Based on a sequential limit theory, the following theorem states that the two-step estimator has a normal limiting distribution.

Theorem 1: *Let y_{it} be generated as in (1) and $\widehat{B}_{2S}$ denotes the least-squares estimator of B in the regression (6). Furthermore ε_{it} and ε_{jt} are independent for $i \neq j$. If $T \to \infty$ is followed by $N \to \infty$ we have*

$$T\sqrt{N} vec(\widehat{B}_{2S} - B) \xrightarrow{d} N(0, \Omega_2^{-1} \otimes \Sigma_v) \; ,$$

where Σ_v is defined in (4),

$$\Omega_2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \beta_{\perp,2}(\alpha_{i,\perp}' \beta_{\perp})^{-1} \alpha_{i,\perp}' \Sigma_i \alpha_{i,\perp} (\beta_{\perp}' \alpha_{i,\perp})^{-1} \beta_{\perp,2}' \; ,$$

$\alpha_{i,\perp}$ and $\beta_{\perp}$ are orthogonal complements of α_i and β and $\beta_{\perp,2}$ is the lower $(n-r) \times r$ block of $\beta_{\perp}$.

From this theorem it follows that the long-run parameters are asymptotically normally distributed and, therefore, the usual tests on the cointegration parameters involve the usual limiting distributions. In particular, the second-step regression (6) can be treated as an ordinary regression equation, that is, the nonstationarity of the regressors and the fact that $\widehat{z}_{it}^{+}$ is estimated can be ignored. Furthermore, it is interesting to note that for finite N and $T \to \infty$, the estimator are *mixed normal*, that is, normally distributed with a stochastic covariance matrix. Therefore, the normal limiting distribution is expected to yield a reliable approximation even if N is small.

4   Inference

In practice, the number of cointegration relationships is often unknown. It is therefore interesting to test hypotheses on the cointegration rank. Larsson, Lyhagen and Lothgren (2001) suggest a "LR-bar" statistic that is based on the standardized mean of the cross section LR statistics for the hypothesis $H_0 : r = r_0$ against the alternative $H_A : r > r_0$. This test statistic assumes that the cointegration vectors are different across i (heterogenous cointegration), whereas our framework assumes that the cointegration vectors are the same for all cross section units (homogenous cointegration). To improve the power of the test in the case $\beta_i = \beta$ for all i, the homogeneity assumption can be imposed.

Following Saikkonen (1999) a simple test procedure is constructed, where the restriction of a homogeneous cointegration relationship can easily be imposed. To nest the null and the alternative hypotheses we write

$$\Delta y_{it} = \alpha_i \beta' y_{i,t-1} + \gamma_i \beta'_\perp y_{i,t-1} + \varepsilon^*_{it} , \tag{7}$$

where γ_i is a $k \times (k - r)$ matrix with full column rank. Under the null hypothesis it is assumed that $\gamma_i = 0$ yielding (1), whereas under the alternative γ_i is unrestricted so that the matrix

$$\Pi_i = [\alpha_i \beta', \gamma_i \beta'_\perp]$$

has full rank for at least one $i \in \{1, \ldots, N\}$. Pre-multiplying (7) with the orthogonal complement $\alpha'_{i,\perp}$ yields

$$u_{it} = \delta_i w_{i,t-1} + e_{it} , \tag{8}$$

where $u_{it} = \alpha'_{i,\perp} \Delta y_{it}$, $\delta_i = \alpha'_{i,\perp} \gamma_i$, $w_{it} = \beta'_\perp y_{it}$, and $e_{it} = \alpha'_{i,\perp} \varepsilon^*_{it}$. To test the hypothesis $r = r_0$ the equation (8) is estimated by ordinary least-squares and a LR, Wald or LM statistic can be constructed to test the hypothesis $\delta_i = 0$ for all i.

In practice the matrices $\alpha_{i,\perp}$ and $\beta_\perp$ are unknown and must be replaced by consistent estimators. This can be done by computing orthogonal complements of the estimates of α_i from the first step and the estimate of β from the second step of the estimation procedure proposed in section 3. The following theorem states, that the limiting null distribution of the test statistic is similar to the one derived by Lyhagen et al. (2001).

Theorem 2: *Let y_t be generated by a VAR(1) process with EC representation (1) and $0 \leq r < k-1$. Furthermore ε_{it} and ε_{jt} are independent for $i \neq j$. Let $\widehat{\alpha}_{i,\perp}$ and $\widehat{\beta}_\perp$ be $\sqrt{T}$-consistent estimates for some orthogonal complements of $\alpha_{i,\perp}$ and $\beta_\perp$, respectively. If $T \to \infty$ is followed by $N \to \infty$ we have*

$$\frac{\sqrt{N}(\overline{\lambda}_r - \mu_r)}{\sigma_r} \xrightarrow{d} N(0,1) \ ,$$

where

$$
\begin{aligned}
\mu_r &= E[\lambda_i(r)] \\
&= E \ tr\left[\int dW_{k-r}W'_{k-r}\left(\int W_{k-r}W'_{k-r}\right)^{-1}\int W_{k-r}dW'_{k-r}\right] \\
\sigma_r^2 &= E tr[\lambda_i(r) - \mu_r]^2 \\
&= var\left\{tr\left[\int dW_{k-r}W'_{k-r}\left(\int W_{k-r}W'_{k-r}\right)^{-1}\int W_{k-r}dW'_{k-r}\right]\right\} \ ,
\end{aligned}
$$

$\overline{\lambda}(r) = N^{-1}\sum_{i=1}^N \lambda_i(r)$ and $\lambda_i(r)$ denotes the LR, Wald or LM statistic of the hypothesis $\delta_i = 0$ in the regression

$$\widehat{u}_{it} = \delta'_i \widehat{w}_{i,t-1} + e_{it} \ , \quad t = 1, \ldots, T \ , \tag{9}$$

where $\widehat{u}_{it} = \widehat{\alpha}'_{i,\perp}\Delta y_{it}$ and $\widehat{w}_{it} = \widehat{\beta}'_\perp y_{it}$.

A convenient (Wald type) test statistic of the null hypothesis is

$$\lambda_i^W(r) = T \ tr\left[\sum_{t=1}^T v_{it}w'_{i,t-1}\left(\sum_{t=1}^T w_{i,t-1}w'_{i,t-1}\right)^{-1}\sum_{t=1}^T w_{i,t-1}v'_{it}\left(\sum_{t=1}^T v_{it}v'_{it}\right)^{-1}\right] \ .$$

The values μ_r and σ_r^2 are computed by Lyhagen et al. (2001) for the model without deterministic terms.

Hypotheses on the cointegration parameters can be tested by using a LR statistic. Following Johansen and Juselius (1994) we consider the following class of linear hypotheses on the cointegration vectors:

$$H_0: \quad \beta = [\beta_1, \ldots, \beta_r] = [\Phi_1\theta_1, \ldots, \Phi_r\theta_r] \ , \tag{10}$$

where Φ_j is a known $k \times q_j$ matrix with $1 \leq q_j \leq k - r$ and θ_j is a $q_j \times 1$ vector for $j = 1, \ldots, r$. Note that for the identification of the cointegration vectors r normalization restrictions are required so that the maximal number

of free parameters is $k - r$ for all cointegration vectors. As demonstrated by Johansen (1995, p. 75), any linear hypothesis of the form $R_j \beta_j = r_j$ can be rewritten as in (10). Inserting the hypothesis in (5) yields a set of r equations

$$\hat{z}_{j,it} = \theta_j'(\Phi_j' y_{i,t-1}) + v_{j,it}^* , \tag{11}$$

where $\hat{z}_{j,it}$ is the j'th element of $\hat{z}_{it}$. Accordingly, under the alternative the parameter vector θ_j can be estimated by a least squares regressions of $\widehat{z}_{j,it}$ on $(\Phi_j' y_{i,t-1})$. Since the system equations for $j = 1, \ldots, r$ do no longer involve the same set of regressors, the SUR system should be estimated by GLS in order to achieve asymptotic efficiency. It is interesting to note that no "switching-algorithm" needs to be applied as in Johansen and Juselius (1994).

Let $\widetilde{v}_{it}$ and $\widetilde{v}_{it}^* = [\widetilde{v}_{1,it}^*, \ldots, \widetilde{v}_{r,it}^*]'$ denote the residual vectors of the unrestricted regression (5) and the restricted regression (11), respectively. A test statistic that is asymptotically equivalent to the ML statistic is

$$\xi_{LR} = NT \left(\log \left| \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{v}_{it} \widetilde{v}_{it}' \right| - \log \left| \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{v}_{it}^* \widetilde{v}_{it}^{*'} \right| . \right) \tag{12}$$

The asymptotic properties of such a test are considered in the following theorem.

Theorem 3: *Let y_{it} be generated as in (1). Furthermore ε_{it} and ε_{jt} are independent for $i \neq j$. Under the null hypothesis (10) the test statistic ξ_{LR} defined in (10) is asymptotically χ^2 distributed with $r(k-r) - \sum_{j=1}^{r} q_j$ degrees of freedom.*

Alternatively, a Wald test procedure can be applied that is based on the results of Theorem 1.

5   Extensions

So far we have considered the cointegrated VAR(1) model with $E(\varepsilon_{it}) = 0$ for all i and t. Although such a limitation is convenient for expositional purposes, it is of course too restrictive for practical applications. A more realistic model is the cointegrated VAR(p) model with individual specific

short-run dynamics and deterministic terms:

$$\Delta y_{it} = \Psi_i d_t + \alpha_i \beta' y_{i,t-1} + \sum_{j=1}^{p-1} \Gamma_{i,j} \Delta y_{i,t-j} + \varepsilon_{it} \ ,$$

where d_t is a vector of deterministic variables (such as a constant, trend or dummy variables) and Ψ_i is a $k \times k$ matrix of coefficients.

If Ψ_i, $\Gamma_{i,1}, \ldots, \Gamma_{i,p-1}$ are unrestricted matrices, they can be "partialled out" from the likelihood function (cf. Johansen 1988). Let $\Delta \widetilde{y}_{it}$ ($\widetilde{y}_{i,t-1}$) denote the residual vectors from a least squares regression of Δy_{it} ($y_{i,t-1}$) on $\Delta y_{i,t-1}, \ldots, \Delta y_{i,t-p+1}$ and d_t. The two-step estimator of the long-run parameters is obtained from the regression

$$\widetilde{z}_{it}^+ = B \widetilde{y}_{i,t-1}^{(2)} + \widetilde{v}_{it} \quad i = 1, \ldots, N; \ t = 1, \ldots, T, \tag{13}$$

where $\widetilde{z}_{it}^+ = \widetilde{y}_{i,t-1}^{(1)} - (\widehat{\alpha}_i' \widehat{\Sigma}_i^{-1} \widehat{\alpha}_i)^{-1} \widehat{\alpha}_i' \widehat{\Sigma}_i^{-1} \Delta \widetilde{y}_{it}$ and v_{it} is defined analogously. The asymptotic distribution of the two-step estimator $\widehat{B}_{2S}$ resulting from (13) is the same as in Theorem 1.

As in the usual time series case with $N = 1$ the asymptotic distribution of the cointegration rank statistic are affected by the deterministic terms. For example, if d_t is a constant so that (8) includes a constant, then the Brownian motions $W_{k-r}(a)$ in Theorem 2 are replaced by $\overline{W}_{k-r}(a) = W_{k-r}(a) - \int_0^1 W_{k-r}(a) \mathrm{d}a$. If d_t represents a polynomial in time, then the asymptotic expressions can be derived by using the results of Ouliaris et al. (1989). Appendix B contains the respective values of μ_r and σ_r^2 for a model with a constant and a linear trend.

An important problem of multi-country panel data sets is the apparent contemporaneous correlation among the errors (e.g. O'Connell 1998, Wu and Wu 2001). For panel unit root tests simulation techniques are applied to control for such correlation among the errors. For the FM-OLS approach, however, cross-section correlation imply more fundamental problems that have not been resolved yet.[2] Since the second step of the parametric approach is based on an ordinary least-square regression, it is straightforward to account for possible contemporaneous correlation. First, one may use a feasible GLS procedure to estimate the set of seemingly unrelated regression

---

[2]Phillips and Moon (1999, p. 1092) state that "... when there are strong correlations in a cross section (as there will be in the face of global shocks) we can expect failures in the strong laws and central limit theory arising from the nonergodicity."

(SUR) equations. Such a procedure requires that the number of time series observations (T) is substantially larger than the number of cross section units (N). In typical country studies, however, N and T are of the same order of magnitude. In such cases the inverse of the estimated covariance matrix may behave poorly and, therefore, a robust estimator in the spirit of Newey and West (1987) is preferable. Let $b = vec(B')$ and

$$\widetilde{X}_{it} = (I_r \otimes \widetilde{y}_{i,t-1}^{(2)}{}') ,$$

so that the second step regression model can be written as

$$\widetilde{z}_{it}^+ = \widetilde{X}_{it}b + \widetilde{v}_{it} .$$

Furthermore, we stack the cross section observations and define

$$\widetilde{z}_t^+ = \begin{bmatrix} \widetilde{z}_{1t}^+ \\ \vdots \\ \widetilde{z}_{Nt}^+ \end{bmatrix} , \quad \widetilde{X}_t = \begin{bmatrix} \widetilde{X}_{1t} \\ \vdots \\ \widetilde{X}_{Nt} \end{bmatrix} , \quad \widetilde{v}_t = \begin{bmatrix} \widetilde{v}_{1t} \\ \vdots \\ \widetilde{v}_{Nt} \end{bmatrix}$$

so that the regression can be written as

$$\widetilde{z}_t^+ = \widetilde{X}_t b + \widetilde{v}_t . \tag{14}$$

In this regression the error vector $\widetilde{v}_t$ is assumed to be (asymptotically) uncorrelated with $\widetilde{v}_s$ for $t \neq s$. The asymptotic covariance matrix of the least squares estimator of $b = vec(B)$ is given by

$$\lim_{T,N \to \infty} \left[\sum_{t=1}^{T} E(\widetilde{X}_t'\widetilde{X}_t) \right]^{-1} \left[\sum_{t=1}^{T} E(\widetilde{X}_t'v_t v_t'\widetilde{X}_t) \right] \left[\sum_{t=1}^{T} E(\widetilde{X}_t'\widetilde{X}_t) \right]^{-1} .$$

Therefore, a consistent estimator for the covariance matrix of the least squares estimator of b can be constructed as

$$\widetilde{V}_b = \left[\sum_{t=1}^{T} \widetilde{X}_t'\widetilde{X}_t \right]^{-1} \left[\sum_{t=1}^{T} \widetilde{X}_t'\widehat{v}_t\widehat{v}_t'\widetilde{X}_t \right] \left[\sum_{t=1}^{T} \widetilde{X}_t'\widetilde{X}_t \right]^{-1} , \tag{15}$$

where $\widehat{v}_t$ denotes the residual vector from the regression (14). This approach is similar to the robust estimator of the covariance matrix suggest by Arellano (1987). However Arellano's estimator assumes that the errors are contemporaneously but not serially uncorrelated. Our robust estimator therefore results from interchanging the role of i and t. Since the cointegration tests suggested in Theorems 2 and 3 are based on similar least-squares regressions, analog procedures can be used for the Wald test of the cointegration rank.

6 Small Sample Properties

In this section, the small sample properties of alternative estimators are studied by means of Monte Carlo simulations. The data are generated by the two-dimensional cointegrated VAR(1) model with error correction representation

$$\Delta y_{it} \;=\; \mu_i - \begin{bmatrix} \phi \\ \phi \end{bmatrix} \begin{bmatrix} 1 & b \end{bmatrix} y_{i,t-1} + \varepsilon_{it} \;,$$

where $\varepsilon_{it} \sim i.i.N.(0, I)$ and the individual effects are generated as $\mu_i \sim i.i.U(0,1)$. If $\phi > 0$ then y_t is cointegrated with cointegration rank $r = 0$. To save space, only the results for $b = 1$ and $\phi = 0.1$ are presented. The results for other values of the parameters are qualitatively similar.[3]

The comparison includes the FM-OLS estimator with individual specific constant and short-run dynamics proposed by Pedroni (1995, 2000) and Phillips and Moon (1999), the dynamic OLS (DOLS) estimator suggested by Kao and Chiang (2000), where the length of the lags and leads is two, the (inefficient) OLS estimator and the two-step estimator suggested in Section 3. The FM-OLS and DOLS estimators are computed by using the GAUSS library NPT 1.1 developed by Chiang and Kao (2000). The bias and root mean-square errors (RMSE) for various sample sizes that are typical for empirical work using country studies are reported in Table 1.

It is well known that the bias in the OLS estimator of the cointegration parameters is $O(T^{-1})$. As can be seen from the results reported in Table 1, the bias of the OLS estimator is severe if the number of time periods is small. The nonparametric bias correction of the FM-OLS estimator seems to be insufficient in short time series as it reduces the bias only marginally. For $T = 100$ the bias reduction is more effective. The DOLS estimator removes the bias by including future and past values of $\Delta y_{it}^{(2)}$. The results displayed in Table 1 suggest that this approach performs slightly worse than the FM-OLS estimator.

The bias of the two-stage estimator is much smaller in absolute value. For $T = 30$ the two-step estimator is nearly unbiased, whereas the FM-OLS and the DOLS estimator still possess a severe negative bias. For the root mean square error (RMSE) the conclusions are similar. If T is small, then

---

[3]Additional simulation results and the GAUSS codes can be found on the homepage of the authors (http://ise.wiwi.hu-berlin.de/~joerg/pancoint.html).

the RMSE of the two-stage estimator is less than half of the RMSE for the FM-OLS or DOLS estimator.

Next, we compare the small sample properties of the tests of the cointegration rank. The LR-bar test suggested by Lyhagen et al. (2001) is denoted by "LLL" and the regression based test for the cointegration rank based on (8) is labelled as "REG". The asymptotic values for $T, N \to \infty$ as reported in Appendix B are used. From the simulation results displayed in Table 2(a) it turns out that for small values of T, the LLL test tends to be very conservative, whereas the REG test performs much better in small samples.

To investigate the power of the test statistics we compute the rejection rates under the local alternative

$$\Pi = \alpha\beta' = -0.1 \begin{bmatrix} 1 & 1 \\ 1 & (1 - \nabla_{NT}) \end{bmatrix} \tag{16}$$

where $\nabla_{NT} = 10/(T\sqrt{N})$. Note that if $\nabla_{NT} \neq 0$, then $rk(\Pi) = 2$ and, therefore, the system is stationary. Such sequence of local alternatives is considered in order to make the power comparable for varying N and T. Furthermore, such alternatives allow the study of the power against alternatives that comes close to the null hypothesis of interest ("near-stationary" alternatives), which seems to be a relevant situation in empirical practice. Table 2(b) presents the size adjusted local power of the the LLL and the REG test. It turns out that in small samples the REG test is much more powerful against local alternatives even if the size bias of the tests is accounted for. Furthermore it is interesting to note that both tests seem to converge to roughly the same limiting power. For small T, however, the local power of the REG test is much higher than the respective power of the LLL test.

Finally, we study the performance of the robust estimator of the standard errors of the parameters. To this end we compute the rejection frequencies of a t-test for the hypothesis that $\beta = [1, 1]'$ in (16). Since the robust estimator (15) is consistent under contemporaneous correlation and heteroskedastic errors, the empirical size of a t-test based on the robust standard errors should approach the nominal size for sufficient sample sizes. The respective estimator is called "2S-HAC". To generate contemporaneously correlated errors, the matrix of contemporaneous errors $E_t = [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]'$ is multiplied by the $k \times k$ matrix Q such that transformed errors $\tilde{\varepsilon}_{it}$ result from the rows of the matrix $\widetilde{E}_t = QE_t$. In our simulations, the elements of the matrix Q are generated by independent draws of $U(0, 10)$ distributed random variable.

First we consider the case of uncorrelated errors, that is, $Q = I$. It turns out that for small T all tests have a tendency to reject the null hypothesis $b = 1$ too often.[4] The size bias is most severe for the DOLS procedure. If the errors are generated with a contemporaneous correlation using $Q \neq I$, then the usual t-statistics based on the assumption of i.i.d. errors (2S-OLS) rejects the null hypothesis much too often. The use of robust standard errors as in (15) reduces the size bias of the test drastically although a moderate tendency to over-reject the null hypothesis remains if T is small. In contrast, the standard errors of the DOLS procedure are seriously over-estimated if the errors are contemporaneously correlated. Consequently, the test based on the DOLS procedure has a severe (negative) size bias.

7   Conclusions

In this paper, a parametric approach for estimation and inference in cointegrated panel data models is suggested. Following Ahn and Reinsel (1990) and Engle and Yoo (1991), an asymptotically efficient estimator is proposed, where all individual specific short-run parameters are estimated in the first step and the long-run parameters are estimated from a pooled regression in a second step. A test procedure is suggested that allows to test the number of cointegrating relationships and a likelihood ratio statistic is proposed that allows to test hypotheses on the long-run parameters. Monte Carlo simulations demonstrate that the parametric approach is much more effective in reducing the small sample bias than the FM-OLS of Pedroni (1995, 2000) and Phillips and Moon (1999) or the DOLS estimator suggested by Kao and Chiang (2000). Furthermore, the estimated standard errors of the two-step estimator can easily be adjusted to account for heteroskedasticity and contemporaneous correlation of the errors, a feature that is often encountered in cross-country studies.

---

[4]The standard errors of the FM-OLS estimators computed by the NPT 1.1 program produces implausibly small values yielding an empirical size close to one.

Appendix A

Proof of Theorem 1

First assume that α_i and Σ_i are known so that $z_{it} = (\alpha_i'\Sigma_i^{-1}\alpha_i)^{-1}\alpha_i'\Sigma_i^{-1}\Delta y_{it}$ and $z_{it}^+ = z_{it} - y_{i,t-1}^{(1)}$.

From Johansen (1991) it is known that in a cointegrated VAR(1) model we have

$$y_{it} = \beta_\perp(\alpha_{i,\perp}'\beta_\perp)^{-1}\alpha_{i,\perp}' \sum_{s=1}^{t} \varepsilon_{is} + u_{it} \ ,$$

where u_{it} is (asymptotically) stationary. From $T^{-1/2}\sum_{s=1}^{t}\varepsilon_{is} = T^{-1/2}\sum_{t=1}^{[aT]}\varepsilon_{it} \Rightarrow B_i(a)$, where $B_i(a) \equiv B_i$ is a Brownian motion with covariance matrix Σ_i. It follows that

$$T^{-2}\sum_{t=1}^{T} y_{it}y_{it}' \ \Rightarrow \ \beta_\perp(\alpha_{i,\perp}'\beta_\perp)^{-1}\alpha_{i,\perp}' \left(\int B_i B_i'\right) \alpha_{i,\perp}(\beta_\perp'\alpha_{i,\perp})^{-1}\beta_\perp' \ .$$

Using $E(\int B_i B_i') = (1/2)\Sigma_i$ we obtain

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^{(2)}y_{i,t-1}^{(2)}{}' \ \xrightarrow{p} \ \frac{1}{2}\Omega_2 \ .$$

Furthermore,

$$y_{i,t-1}^{(2)}v_{it}' \ = \ \beta_{2,\perp}(\alpha_{i,\perp}'\beta_\perp)^{-1}\alpha_{i,\perp}' \left(\sum_{s=1}^{t-1}\varepsilon_{is}v_{it}'\right) \ ,$$

where $v_{it}' = (\alpha_i'\Sigma_i^{-1}\alpha_i)^{-1}\alpha_i'\Sigma_i^{-1}\varepsilon_{it}$. Since $\varepsilon_{i,t-j}$ is independent of v_{it} for $j \geq 1$ we have

$$\lim_{T,N\to\infty}\frac{1}{NT^2}E\left[\sum_{i=1}^{N}\sum_{t=1}^{T}vec(y_{i,t-1}^{(2)}v_{it}')vec(y_{i,t-1}^{(2)}v_{it}')'\right]$$

$$= \lim_{T,N\to\infty}E\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}^{(2)}y_{i,t-1}^{(2)}{}'\right] \otimes \left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}v_{it}'\right]$$

$$= (1/2)\Omega_2 \otimes \Sigma_v$$

and, therefore,

$$\frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}vec(y_{i,t-1}^{(2)}v_{it}') \ \Rightarrow \ N(0,\frac{1}{2}\Omega_2 \otimes \Sigma_v) \ .$$

15

It follows that

$$
T\sqrt{N}vec(\widehat{B}-B) = T\sqrt{N}vec\left\{\left[\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}^{(2)}y_{i,t-1}^{(2)}{}'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}^{(2)}v_{it}'\right]\right\}
$$

$$
\xrightarrow{d} N(0,\Omega_2^{-1}\otimes\Sigma_v) .
$$

Finally, it is easy to verify that if $\widehat{\alpha}_i-\alpha_i = O_p(T^{-1/2})$ and $\widehat{\Sigma}_i-\Sigma_i = O_p(T^{-1/2})$ we have

$$
\frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}^{(2)}\widehat{v}_{it}' = \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}^{(2)}v_{it}' + o_p(1)
$$

and, thus, replacing α_i and Σ_i by a consistent estimator does not affect the asymptotic distribution.

Proof of Theorem 2

First, assume that α_i and β (and therefore $\alpha_{i,\perp}$ and $\beta_\perp$) are known. The vector of regressors results as

$$
w_{it} = \beta_\perp' y_{i,t-1} = \beta_\perp'\beta_\perp(\alpha_{i,\perp}'\beta_\perp)^{-1}\alpha_{i,\perp}\sum_{s=1}^{t-1}\varepsilon_{is} + o_p(T^{1/2})
$$

$$
= \Gamma S_{i,t-1} + o_p(T^{1/2}) ,
$$

where $S_{it} = \sum_{s=1}^{t}\alpha_{i,\perp}'\varepsilon_{is}$ and $\Gamma = \beta_\perp'\beta_\perp(\alpha_{i,\perp}'\beta_\perp)^{-1}$. Accordingly, under the null hypothesis $\alpha_{i,\perp}'\Delta y_{it} = \alpha_{i,\perp}'\varepsilon_{it}$ the respective Wald statistic can be written as

$$
\lambda_i^w(r) = tr\left\{\left(\widetilde{\Sigma}^{-1/2}\sum_{t=1}^{T}\Delta S_{it}S_{i,t-1}'\Gamma'\right)\left(\sum_{t=1}^{T}\Gamma S_{i,t-1}S_{i,t-1}'\Gamma'\right)^{-1}\right.
$$

$$
\left.\times\left(\sum_{t=1}^{T}\Gamma S_{i,t-1}\Delta S_{it}'\right)\widetilde{\Sigma}^{-1/2}\right\} + o_p(1)
$$

$$
= tr\left\{\widetilde{\Sigma}^{-1/2}\left(\sum_{t=1}^{T}\Delta S_{it}S_{i,t-1}'\right)\left(\sum_{t=1}^{T}S_{i,t-1}S_{i,t-1}'\right)^{-1}\right.
$$

$$
\left.\times\left(\sum_{t=1}^{T}S_{i,t-1}\Delta S_{it}'\right)\widetilde{\Sigma}^{-1/2}\right\} + o_p(1) ,
$$

where $\widetilde{\Sigma}^{-1/2}$ is a symmetric matrix with the property $\widetilde{\Sigma}^{-1/2}\widetilde{\Sigma}^{-1/2} = \alpha_{i,\perp}'\Sigma\alpha_{i,\perp}$. It remains to show that the limiting distribution is not affected if $\beta_\perp$ and

16

$\alpha_{i,\perp}$ are replaced by estimates with $\widehat{\beta}_\perp = \beta_\perp + O_p(T^{-1/2})$ and $\widehat{\alpha}_{i,\perp} = \alpha_{i,\perp} + O_p(T^{-1/2})$. It follows that $\widehat{\beta}'_\perp y_{i,t-1} = \Gamma S_{i,t-1} + O_p(T^{1/2})$ and

$$T^{-1} \sum_{t=1}^{T} \widehat{\alpha}'_{i,\perp} \Delta y_{it} y'_{i,t-1} \widehat{\beta}_\perp = T^{-1} \sum_{t=1}^{T} \alpha'_{i,\perp} \Delta y_{it} y'_{i,t-1} \beta_\perp + o_p(1) \ .$$

Consequently, replacing $\beta_\perp$ and $\alpha_{i,\perp}$ by consistent estimates does not change the asymptotic distribution.

Proof of Theorem 3

Since the null distributions of the LR and the Wald statistics are asymptotically identical (e.g. Engle 1984), we first consider the Wald statistic of the null hypothesis. To this end we write the second step regression as

$$\widehat{z}_{j,it} = \theta'_j(\Phi'_j y_{i,t-1}) + \varphi'_j(\Psi'_j y_{i,t-1}) + v^*_{it} \quad j = 1, \ldots, r \ ,$$

where Ψ_j is a $k \times (k-r-q_j)$ dimensional matrix such that the matrix $[\Phi_j, \Psi_j]$ has full column rank. The null hypothesis is equivalent to $\varphi_j = 0$. The set of equations can be written as

$$\widehat{z}_{it} = \begin{bmatrix} (y'_{i,t-1}\Phi_1) & 0 & . & 0 & (y'_{i,t-1}\Psi_1) & 0 & . & 0 \\ . & . & . & . & & . & . & \\ 0 & 0 & . & (y'_{i,t-1}\Phi_r) & 0 & 0 & . & (y'_{i,t-1}\Psi_r) \end{bmatrix} \begin{bmatrix} \theta_1 \\ . \\ \theta_r \\ \varphi_1 \\ . \\ \varphi_r \end{bmatrix} + v^*_{it}$$

From Theorem 1 it follows that the vector $[\theta'_1, \ldots, \theta'_r, \varphi'_1, \ldots, \varphi'_r]'$ is asymptotically normal and, therefore, $W_\varphi = \widehat{\varphi}' Var(\widehat{\varphi})^{-1} \widehat{\varphi}$ is asymptotically χ^2 distributed with $r(n-r) - \sum_{j=1}^{r} q_j$ degrees of freedom, where $\widehat{\varphi}$ is the least-squares estimate of $\varphi = [\varphi'_1, \ldots, \varphi'_r]'$.

Appendix B

For the model with a constant or a linear time trend, the Brownian motions in Theorem 2 are replaced by the expressions

$$\overline{W}_{k-r} = W_{k-r} - \int W_{k-r}$$

for the model with a constant and

$$\widetilde{W}_{k-r} = W_{k-r} - \int W_{k-r} - 12a \int aW_{k-r}$$

for the model with a time trend. To estimate μ_r and σ_r^2 used in Theorem 2 for the model with a constant term and a linear time trend, the Brownian motions W_{k-r} are replaced by a $(k-r)$ dimensional vector of Gaussian random walks with $T = 500$. The mean and variances are computed from 20,000 replications of the stochastic expressions.

Table B.1: Asymptotic values of μ_r and σ_r^2

	constant		linear trend	
sig. lev.	μ_r	σ_r^2	μ_r	σ_r^2
$k - r = 1$	3.051	6.826	5.301	10.94
$k - r = 2$	9.990	18.46	14.35	26.02
$k - r = 3$	20.88	35.03	27.31	45.79
$k - r = 4$	35.67	57.49	44.13	70.82
$k - r = 5$	54.33	86.00	64.71	101.9
$k - r = 6$	76.94	119.7	89.16	136.9

References

Ahn, S.K. and G.C. Reinsel (1990), Estimation for partially nonstationary multivariate autoregressive Models, Journal of the American Statistical Association, 85, 813–823.

Arellano, M. (1987), Computing Robust Standard Errors for Within-groups Estimators, Oxford Bulletin of Economics and Statistics, 49, 431-434.

Boswijk, H.P. (1996), Testing Identifiability of Cointegration Vectors, Journal of Business & Economic Statistics, 14, 153–160.

Chang, Y. (2001), Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency, *forthcoming in:* Journal of Econometrics.

Engle, R.F. and B.S. Yoo (1991), Cointegrated Economic Time Series: An Overview with New Results, in: Fomby, T.B. and G.F. Rhodes (eds.), Advances in Econometrics: Co-Integration, Spurious Regression, and Unit Roots, Greenwich: JAI Press, 237–266.

Engle, R.F (1984), Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics, Ch. 13 in: Z. Griliches and M.D. Intriligator (eds), Handbook of Econometrics, Vol. II, Amsterdam: North-Holland.

Harris, D. (1997), Principal Components Analysis of Cointegrated Time Series, Econometric Theory, 13, 529–557.

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, 1551–1580.

Johansen, S. (1995), Likelihood-based Inference in Cointegrated Vector Autoregressive Models, Oxford: University Press.

Johansen, S. and K. Juselius (1994), Identification of the Long-Run and the Short-Run Structure: An application to the ISLM model, Journal of Econometrics, 63, 7–36.

Kao, C. and M.-H. Chiang (2000), On the Estimation and Inference of a Cointegrated Regression in Panel Data, in: B. Baltagi (ed.), Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics, Vol. 15, Amsterdam: JAI Press, 161–178.

Larsson, R., J. Lyhagen and M. Lothgren (2001), Likelihood-based Cointegration Tests in Heterogenous Panels, Econometrics Journal, 4, 109–142.

Levin, A. and C.-F. Lin (1992), Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties, University of California San Diego Working Paper.

Levin, A. and C.-F. Lin (1993), Unit Root Tests in Panel Data: New Results, University of California San Diego Working Paper.

Newey W.K. and K.D. West (1987), A Simple, Positive Semi-definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 703–708.

O'Connell, P.G.J. (1998), The Overvaluation of Purchasing Power Parity, Journal of International Economics, 44, 1–20.

Ouliaris, S., J.Y. Park and P.C.B. Phillips (1989), Testing for a Unit Root in the Presence of a Maintained Trend, in: B. Raj (ed.), Advances in Econometrics and Modelling, Amsterdam: Kluwer, 7–28.

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

Pedroni, P. (2000), Fully Modified OLS for Heterogenous Cointegrated Panels, in: B. Baltagi (ed.), Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics, Vol. 15, Amsterdam: JAI Press, 93–130.

Perron, P. and S. Ng (1996), Useful Modifications to Some Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties, Review of Economic Studies, 63, 435-463.

Pesaran, M.H., Y. Shin and R. Smith (1999), Pooled Mean Group Estimation of Dynamic Heterogeneous Panels, Journal of the American Statistical Association, 1999, 94, 621–634.

Phillips, P.C.B. (1991), Optimal Inference in Co-integrated Systems, Econometrica, 59, 282–306.

Phillips, P.C.B. (1995), Fully Modified Least Squares and Vector Autoregressions, Econometrica, 1023–1078.

Phillips, P.C.B. and B.E. Hansen (1990), Statistical Inference in Instrumental Variable Regression with I(1) Processes, Review of Economic Studies, 57, 99–125.

Phillips, P.C.B. and H.R. Moon (1999), Linear Regression Limit Theory for Nonstationary Panel Data, Econometrica, 67, 1057–1111.

Reinsel, G.C. (1993), Elements of Multivariate Time Series Analysis, New York: Springer.

Saikkonen, P. (1999), Testing the Normalization and Overidentification of Cointegrating Vectors in Vector Autoregressive Processes, Econometric Reviews, 18, 235–257.

Schwert, G.W. (1989), Tests for Unit Roots: A Monte Carlo Investigation, Journal of Business and Economic Statistics, 7, 147–159.

Wu, J.-L. and S. Wu (2001), Is Purchasing Power Parity Overvalued, Journal of Money, Credit, and Banking, 33, 804–812

Table 1: Estimation bias of various cointegration estimators

	Bias			
	two-stage	FM-OLS	DOLS	OLS
$T =$		$N = 10$		
15	−0.1363	−0.4280	−0.4294	−0.4824
15	−0.1304	−0.6487	−0.6487	−0.7476
20	−0.0900	−0.5675	−0.5675	−0.6811
30	−0.0300	−0.4578	−0.4578	−0.5759
50	0.0124	−0.3233	−0.3233	−0.4390
100	0.0233	−0.1842	−0.1842	−0.2732
$T =$		$N = 20$		
15	−0.1316	−0.6439	−0.6439	−0.7524
20	−0.0884	−0.5654	−0.5654	−0.6832
30	−0.0327	−0.4515	−0.4515	−0.5752
50	0.0091	−0.3158	−0.3158	−0.4362
100	0.0198	−0.1766	−0.1766	−0.2688
	RMSE			
	two-stage	FM-OLS	DOLS	OLS
$T =$		$N = 10$		
15	0.2100	0.6806	0.6806	0.7651
20	0.1723	0.5999	0.5999	0.6983
30	0.1227	0.4874	0.4874	0.5926
50	0.0883	0.3485	0.3485	0.4541
100	0.0650	0.2016	0.2016	0.2849
$T =$		$N = 20$		
15	0.1746	0.6604	0.6604	0.7611
20	0.1355	0.5813	0.5813	0.6921
30	0.0909	0.4664	0.4664	0.5840
50	0.0633	0.3284	0.3284	0.4438
100	0.0461	0.1851	0.1851	0.2749

Note: The entries of the Table report the estimated bias and root mean squared error (RMSE) of the cointegration parameter b based on 5000 replication of the model (16). "two-step" indicates the two-step estimator suggested in Section 3, "FM-OLS" denotes the Fully-modified panel cointegration estimator suggested by Pedroni (1996), "DOLS" is the dynamic OLS estimator of Kao and Chiang (2000), and "OLS" indicates the ordinary least-squares estimator of the pooled model, where the first variable is regressed on the second variable.

Table 2: Sizes and local powers of tests of the cointegration rank

	(a) Empirical sizes			
	$N = 10$		$N = 20$	
T	REG	LLL	REG	LLL
20	0.012	0.000	0.010	0.000
30	0.033	0.000	0.033	0.000
50	0.053	0.004	0.050	0.002
100	0.056	0.030	0.068	0.033
	(b) Local powers (size adjusted)			
	$N = 10$		$N = 20$	
T	REG	LLL	REG	LLL
20	0.242	0.150	0.240	0.143
30	0.221	0.150	0.227	0.148
50	0.186	0.141	0.206	0.153
100	0.164	0.146	0.169	0.166

Note: Rejection frequency for tests of the null hypothesis $r = 1$. "REG" indicates the regression based test suggested in Theorem 2 and "LLL" denotes the LR-bar statistic suggested by Lyhagen et al. (2001). The local power is computed by simulating the data under the local alternative $\nabla_{NT} = -10/(T\sqrt{N})$. The critical values are used that yield tests with an exact size of 0.05 under the null hypothesis (size adjusted power). 5000 replications are used to compute the rejection frequencies.

Table 3: Sizes of the t-test

	$N = 10$			$N = 20$		
	2S-OLS	2S-HAC	DOLS	2S-OLS	2S-HAC	DOLS
	Uncorrelated errors					
20	0.098	0.097	0.157	0.140	0.115	0.162
30	0.057	0.052	0.124	0.057	0.046	0.115
50	0.044	0.041	0.081	0.039	0.036	0.078
100	0.058	0.064	0.045	0.072	0.069	0.045
	Contemporaneously correlated errors					
20	0.292	0.085	0.010	0.403	0.083	0.008
30	0.274	0.069	0.009	0.397	0.070	0.005
50	0.308	0.070	0.006	0.420	0.069	0.004
100	0.355	0.073	0.003	0.479	0.066	0.002

Note: Rejection frequency for tests of the null hypothesis $b = 0$ in model (16). "2S-OLS" indicates rejection frequencies of a t-test where the errors are assumed to be i.i.d., whereas "2S-HAC" indicates a t-test based on robust standard errors computed from (15). The column "DOLS" reports rejection frequencies of the DOLS estimator assuming contemporaneously uncorrelated errors. The nominal size is of the tests is 0.05 and 5000 replications are used to compute the rejection frequencies.