

A New Poolability Test for Cointegrated Panels*

Joakim Westerlund[†] and Wolfgang Hess[‡]

March 22, 2007

Abstract

This paper proposes a new test of the null hypothesis that the parameters in a cointegrated panel data regression are equal across the cross-section. The asymptotic distribution of the new test statistic is derived and Monte Carlo results are provided to suggest that it performs very well in small samples. An empirical application to the monetary exchange rate model is also provided.

JEL Classification: C12; C32; C33; F31; F41.

Keywords: Panel Cointegration; Pooling; Homogeneity Testing; Monetary Exchange Rate Model.

1 Introduction

The problem of testing and estimating cointegrated panel data relationships has attracted considerable interest in the literature recently. Some of the most recent theoretical contributions within this field include Pedroni (2004) and Westerlund (2005). However, these studies have mostly been limited to applications where the sole purpose is to determine whether or not a particular panel data regression can be considered as cointegrated, and there has been relatively little work done in other areas.

One such area, which has been given surprisingly little attention in the theoretical literature, is that concerned with the testing of various homogeneity restrictions. For example, in many applications it is not only the hypothesis of cointegration itself that is of interest, but also if the individual cointegration parameters can be regarded as equal. Indeed, one of the most important

*A previous version of this paper was presented at a seminar at Lund university. The author would like to thank seminar participants, and in particular David Edgerton for many valuable comments and suggestions. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number W2006-0068:1.

[†]Corresponding author: Department of Economics, Lund University, P. O. Box 7082, S-220 07 Lund, Sweden. Telephone: +46 46 222 8670; Fax: +46 46 222 4118; E-mail address: joakim.westerlund@nek.lu.se.

[‡]Department of Economics, Lund University.

attractions of panel data is the ability to selectively pool the long-run information regarding the cointegration parameters while simultaneously permitting the short-run dynamics and fixed effects to differ between the cross-sectional units. This issue is particularly important in applied work since whenever the hypothesis of equal parameters is tested, it is almost always rejected, see for example Baltagi *et al.* (2000), Pesaran *et al.* (1999) and Rapach and Wohar (2004).

The by far most popular approach for testing restrictions of this kind is that of Mark and Sul (2003), in which the null hypothesis of equal parameters is tested versus the general heterogenous alternative by means of a simple Wald test. From a theoretical point of view, this approach is very appealing because it leads to statistics that are asymptotically chi-squared distributed. In samples of realistic size, however, tests of this kind can be misleading in the sense that they have a tendency to reject the null hypothesis too frequently, in which case the researcher is left with little or no idea of how to proceed.

Another drawback with the Wald approach is that it is only practical if the cross-sectional units are independent of each other, an assumption that is perhaps unreasonable given that many economic variables tend to exhibit strong comovement from across different units. Mark *et al.* (2005) and Moon and Perron (2004) try to alleviate this problem by resorting to seemingly unrelated regressions techniques. However, this approach is not feasible when the cross-sectional dimension N is of the same order of magnitude as the time series dimension T , since the covariance matrix of the regression errors then becomes rank deficient. In fact, for this approach to work properly, one usually requires T being substantially larger than N , a condition that is rarely fulfilled in practice.

In this paper, we try to overcome these problems by proposing a new poolability test based on the Hausman (1978) principle whereby two estimators of the cointegration parameters, one heterogenous and one pooled, are compared.¹ To handle the impact of violations of the cross-sectional independence assumption, we assume that the dependence can be modelled using a small number of factors that are common across the units of the panel. The main advantage of this assumption is that it reduces the dimension of the cross-sectional dependency, which makes the new test applicable even in situations when N is larger than T . The asymptotic null distribution of the new test statistic is derived, and verified in small samples using Monte Carlo simulations.

In the empirical part of the paper, we revisit the monetary exchange rate model, and the sample used by Mark and Sul (2001) and Rapach and Wohar (2004) to test it. Our results suggest that although there is evidence of homogeneity across a majority of the countries, the subpanels considered by Mark and Sul (2001) and Rapach and Wohar (2004) are not suitable for pooling. When the monetary model is fitted to those countries for which the poolability null was not rejected, in contrast to theory, we find no evidence to suggest that

¹See Kapetanios (2003) for a similar approach in the case of a stationary panel regression.

money is neutral or that the elasticity of income is negative. The conclusion is therefore that the monetary model must be rejected.

The remainder of this paper is organized as follows. Section 2 introduces the Hausman test, and analyzes its asymptotic and small-sample properties. Section 3 then contains the empirical application, while Section 4 concludes. Proofs of important results are relegated to the appendix.

2 The Hausman poolability test

This section introduces the new poolability test. We begin with a brief discussion of the underlying assumptions, and then we go on to analyze the test statistic, and its asymptotic and small-sample properties.

2.1 The model and assumptions

Let y_{it} be a scalar integrated variate and let x_{it} be a K dimensional vector of integrated variables. The data generating process of y_{it} is assumed to be given by the following cointegrated system of equations

$$y_{it} = \alpha_i + \beta_i x_{it} + e_{it}, \quad (1)$$

$$x_{it} = x_{it-1} + v_{it}. \quad (2)$$

Note that in this setup β_i is a $1 \times K$ vector of unknown slope parameters that are constant through time but possibly different across i . It is the homogeneity of this vector that is of particular interest in this paper. Also, in contrast to many other studies, in this paper we do not require that the cross-sectional units are independent of each other. To model deviations from this assumption, we follow Bai and Kao (2005) and assume that the error e_{it} is generated by the following factor model

$$e_{it} = \lambda_i' f_t + u_{it}, \quad (3)$$

where f_t is an k dimensional vector of unobservable common factors, λ_i is a vector of factor loadings that is conformable with f_t , and u_{it} is a scalar idiosyncratic disturbance. The rest of the assumptions may be summarized in the following way. As usual, $\|A\|$ will denote the Euclidean norm $(\text{tr}(A'A))^{1/2}$ of the matrix A .

Assumption 1. (The common component.)

- (a) λ_i is a nonrandom vector such that $\|\lambda_i\| < \infty$,
- (b) $\frac{1}{N} \sum_{i=1}^N \lambda_i \lambda_i' \rightarrow \Sigma$ as $N \rightarrow \infty$, where Σ is positive definite,
- (c) k is known.

Assumption 2. (The idiosyncratic component.)

- (a) u_{it} and v_{it} are independent across i ,
- (b) u_{it} is independent of f_t .

Assumption 3. (Invariance.) $z_{it} = (f'_t, u_{it}, v'_{it})'$ satisfies the usual invariance principle for each i .

Assumptions 1 (a) and (b) ensure the consistency of the principal components estimates of the common factors, and are standard in common factor analysis. Although Assumption 1 (a) requires that the loadings are nonrandom, this is only for simplicity. Random loadings can be permitted at the expense of some additional moment conditions. Assumption 1 (b) ensures that the common factors have a nontrivial contribution to the variation of e_{it} , which in turn ensures that the factor model is identified. For now, we also assume that k is known, but this is not necessary, as will be explained later.

Assumptions 2 (a) ensures that u_{it} and v_{it} are cross-sectionally independent, which is the same as saying that all cross-section dependence is assumed to be absorbed by the common factors.² The extent of this dependence is determined by λ_i , as can be seen by writing

$$E(e_{it}e_{jt}) = \lambda'_i E(f_t f'_t) \lambda_j \quad \text{for } i \neq j.$$

Finally, Assumption 3 guarantees that the following invariance result holds for the partial sum process constructed from z_{it} for each i

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} z_{it} \Rightarrow B_i \quad \text{as } T \rightarrow \infty,$$

where \Rightarrow signifies weak convergence and $B_i = (B'_f, B_{ui}, B'_{vi})'$ is an $k + K + 1$ dimensional vector Brownian motion that is partitioned conformably with z_{it} . If we let Ω_i denote the covariance matrix of this vector, then it is clear that B_i can be written as $\Omega_i^{1/2} W_i$, where W_i is a vector standard Brownian motion with covariance equal to the identity matrix.³ This invariance assumption is convenient for at least two reasons. Firstly, it makes the asymptotic analysis relatively straightforward. Secondly, apart from some mild regulatory conditions, it places very little restrictions on the time series properties of the data generating process as captured by Ω_i . This matrix, also known as the long-run covariance matrix of z_{it} , may be decomposed in the following fashion

$$\Omega_i = \Sigma_i + \Gamma_i + \Gamma'_i = \Lambda_i + \Gamma'_i,$$

²As pointed out by Bai and Kao (2005), v_{it} can also be permitted to have a common factor structure. In this paper, however, we focus on the case when the cross-sectional dependence originates from e_{it} only.

³For notational simplicity, in this paper the Brownian motions $W_i(r)$ and $B_i(r)$ defined on the interval $r \in [0, 1]$ are written W_i and B_i , respectively, with the measure of integration omitted.

where $\Sigma_i = E(z_{i0}z'_{i0})$ and $\Gamma_i = \sum_{j=1}^{\infty} E(z_{i0}z'_{ij})$ are the contemporaneous and lagged covariances of z_{it} , respectively. We can further partition Ω_i as

$$\Omega_i = \begin{bmatrix} \Omega_{ff} & \Omega_{fui} & \Omega_{fvi} \\ \Omega_{ufi} & \Omega_{uui} & \Omega_{uvi} \\ \Omega_{vfi} & \Omega_{vui} & \Omega_{vvi} \end{bmatrix}.$$

Here, we need to assume that Ω_{vvi} is positive definite, which ensures that x_{it} is not cointegrated in case we have multiple regressors.

2.2 The test statistic

Having described the data generating process considered in this paper, we now introduce the Hausman statistic. The hypothesis to be tested can be written as

$$H_0 : \beta_i = \beta \text{ for all } i \text{ versus } H_1 : \beta_i \neq \beta \text{ for some } i.$$

Thus, in this paper we are interested in testing whether or not the individual slopes β_i take on a common value β . In doing so, it is convenient to let $\hat{\beta}_i$ and $\hat{\beta}_N$ denote the individual and pooled least squares estimators of β , respectively. Although generally biased, as shown by Westerlund (2007), it is still possible to make valid inference based on the bias-adjusted counterparts of these estimators, which are defined in the following way

$$\hat{\beta}_i^+ = \hat{\beta}_i - \hat{b}_i \quad \text{and} \quad \hat{\beta}_N^+ = \hat{\beta}_N - \hat{b}_N,$$

where

$$\begin{aligned} \hat{b}_i &= \left(\hat{\lambda}'_i \hat{U}_{fvi} + \hat{U}_{uvi} \right) Q_i^{-1}, \quad \hat{b}_N = \sum_{i=1}^N \left(\hat{\lambda}'_i \hat{U}_{fvi} + \hat{U}_{uvi} \right) Q_N^{-1}, \\ \hat{U}_{fvi} &= \hat{\Omega}_{fvi} \hat{\Omega}_{vvi}^{-1} \left(\sum_{t=1}^T \Delta x_{it} \tilde{x}'_{it} - T \hat{\Lambda}_{vvi} \right) + T \hat{\Lambda}_{fvi}, \\ \hat{U}_{uvi} &= \hat{\Omega}_{uvi} \hat{\Omega}_{vvi}^{-1} \left(\sum_{t=1}^T \Delta x_{it} \tilde{x}'_{it} - T \hat{\Lambda}_{vvi} \right) + T \hat{\Lambda}_{uvi}. \end{aligned}$$

Note that in this definition, we have used that

$$\tilde{x}_{it} = x_{it} - \frac{1}{T} \sum_{s=1}^T x_{is}, \quad Q_i = \sum_{t=1}^T \tilde{x}_{it} \tilde{x}'_{it} \quad \text{and} \quad Q_N = \sum_{i=1}^N Q_i.$$

Also, for now it is enough to know that $\hat{\Omega}_i$ and $\hat{\Gamma}_i$ are the Newey and West (1994) estimators of Ω_i and Γ_i , respectively, while $\hat{\lambda}_i$ is the principal components estimator of λ_i . To test the hypothesis of H_0 versus H_1 , we propose using the

following test statistic, which can be thought of as a maximum Hausman (1978) type statistic

$$\hat{H}_N = \max_{1 \leq i \leq N} \hat{H}_i,$$

where

$$\hat{H}_i = T^2(\hat{\beta}_i^+ - \hat{\beta}_N^+) \left(Q_i^{-1} \left(\frac{1}{6} \hat{\Omega}_{e.vi} \hat{\Omega}_{vvi} \right) Q_i^{-1} \right)^{-1} (\hat{\beta}_i^+ - \hat{\beta}_N^+)'.$$

Here $\hat{\Omega}_{e.vi}$ is a consistent estimator of $\Omega_{e.vi} = \lambda_i' \Omega_{f.vi} \lambda_i + \Omega_{u.vi}$, the long-run variance of e_{it} conditional on v_{it} , where $\Omega_{f.vi} = \Omega_{ff} - \Omega_{fvi} \Omega_{vvi}^{-1} \Omega_{vfi}$ and $\Omega_{u.vi} = \Omega_{uu} - \Omega_{uvi} \Omega_{vvi}^{-1} \Omega_{vui}$ are the corresponding long-run variances of f_t and u_{it} , respectively, conditional on v_{it} .

Define $a_N = 2$ and

$$b_N = F^{-1} \left(1 - \frac{1}{N} \right),$$

where $F^{-1}(x)$ is the inverse of the chi-squared distribution function with K degrees of freedom evaluated at x . The asymptotic distribution of \hat{H}_N can now be stated in the following way.

Theorem 1. Under H_0 and Assumptions 1 to 3, given some x defined on the entire real line, then as $N, T \rightarrow \infty$ with $\sqrt{N}/T \rightarrow 0$

$$P \left(\frac{1}{a_N} (\hat{H}_N - b_N) \leq x \right) \rightarrow \exp(-e^{-x}).$$

Remark 1. As the theorem makes clear, the normalized statistic $(\hat{H}_N - b_N)/a_N$ has a limiting Gumbel distribution as both T and N grow. This result follows naturally from the fact that while $\hat{\beta}_N^+$ is consistent at rate \sqrt{NT} , the rate of consistency for $\hat{\beta}_i^+$ is lower, only T , which suggests that the former estimator effectively drops out in the limit. Therefore, since the asymptotic distribution of the remaining part, which only depends on $\hat{\beta}_i^+$, is chi-squared, we can apply some well-known results from the theory of extreme values to show that \hat{H}_N is indeed Gumbel distributed, see for example Embrechts *et al.* (1997). Note the requirement that \sqrt{N}/T should go to zero, which in practice means that T should be substantially larger than N .

Remark 2. As is well known from the theory of extreme values, the Gamma distribution belongs to the maximum domain of attraction of the Gumbel distribution. Since the chi-squared distribution is a special case of the Gamma distribution, it also belongs to the maximum domain of attraction of the Gumbel distribution. In other words, suppose that we have N independent chi-squared distributed random variables X_1, \dots, X_N with maximum

$$H_N = \max\{X_1, \dots, X_N\}.$$

Then there exist sequences $a_N > 0$ and b_N such that the normalized maximum $(H_N - b_N)/a_N$ has a limiting Gumbel distribution as $N \rightarrow \infty$.⁴ The normalizing constants a_N and b_N are such that if $F(x)$ belongs to the maximum domain of attraction of the Gumbel distribution, then $b_N = F^{-1}(1 - 1/N)$ and $a_N = g(b_N)$, where $g(x)$ is given as follows

$$g(x) = \int_x^{x_{max}} \left(\frac{1 - F(y)}{1 - F(x)} \right) dy$$

with $x_{max} \leq \infty$ being the right endpoint of the distribution function $F(x)$. Motivated by von Mises functions, the function $g(x)$ can be interpreted as an auxiliary function of $F(x)$. In fact, if X is a random variable then $g(x)$ is nothing but the mean excess function $g(x) = E((X - x)|X > x)$ for all $x < x_{max}$. In particular, if X is chi-squared distributed, then we have as a natural estimator

$$F^{-1} \left(1 - \frac{1}{Ne} \right) - b_N,$$

where $F^{-1}(x)$ is now the inverse of the chi-squared distribution function with K degrees of freedom evaluated at x .

Remark 3. Very relevant for practical applications is the speed at which the normalized maximum $(H_N - b_N)/a_N$ converges to its limiting distribution. Since the sample maximum H_N is the empirical version of the $(1 - 1/N)$ quantile of the underlying distribution function, the latter is an appropriate centering constant for all $N \geq 2$. For the chi-squared distribution, which has a strictly increasing distribution function $F(x)$, the quantiles correspond to the inverse of $F(x)$. Hence, setting $b_N = F^{-1}(1 - 1/N)$ works very well even if N is small. The sequence $a_N = F^{-1}(1 - 1/(Ne)) - b_N$ deserves a somewhat closer look in this context. In general, it holds that

$$a_N \rightarrow 2 \quad \text{as } N \rightarrow \infty.$$

If the chi-squared distribution considered has two degrees of freedom, then a_N will be exactly equal to two for all N . In case K is less than two, a_N will approach two from below as $N \rightarrow \infty$, while a_N will approach two from above if K is larger than two. In other words, if $a_N = 2$ is chosen as a normalizing constant, the variance of the limiting Gumbel distribution will be underestimated in small samples if $K < 2$ and it will be overestimated if $K > 2$. However, our simulation results suggest that setting $a_N = 2$ works well even in panels where N is very small.⁵

⁴It should be noted that independence of the random variables is sufficient but far from necessary. Maxima of stationary series can be dependent as long as the dependence at extreme levels is weak. Such maxima follow the same distributional limit laws as those of independent series.

⁵Note that although we are taking the maximum of a sequence of asymptotically chi-

Remark 4. Because the individual Hausman statistic \widehat{H}_i is asymptotically chi-squared, an alternative panel statistic could be defined by noting that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \widehat{H}_i - \sqrt{N}K \Rightarrow N(0, 2K) \quad \text{as } N, T \rightarrow \infty.$$

Thus, the normalized average could be used for the test of H_0 versus H_1 . However, unreported simulation results suggest that this test statistic can be quite distorted in small samples, while the asymptotic distribution of \widehat{H}_N works well even if N is small. In addition, it is not difficult to see that \widehat{H}_N should dominate the average in terms of power. For these reasons, in this paper we focus on \widehat{H}_N .

Remark 5. It is important to note that the least squares estimator is not unbiased in the data generating process considered here. The reason for this originates from the well-known endogeneity problem, but in this case it is not only the endogeneity of the regressors that matters. Indeed, as is clear by just looking at the bias formulas, there are actually two sources of bias here, U_{fv} is due to the endogeneity of the common factor f_t , while U_{uv} is due to the endogeneity of the idiosyncratic component u_{it} .

Remark 6. Theorem 1 is based on the assumption that k , the number of common factors, is known. When it is unknown, a natural approach is to treat the estimation problem as a model selection issue, and to estimate k by minimizing an information criterion. The particular estimator opted for this paper is given as follows

$$\widehat{k} = \arg \min_{0 \leq k \leq k_{max}} \log(\widehat{\sigma}^2) + k \log \left(\frac{NT}{N+T} \right) \frac{N+T}{NT},$$

where $\widehat{\sigma}^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \widehat{u}_{it}^2$ and k_{max} is a bounded integer not smaller than k . The use of this estimator is motivated in part by its popularity in studies such as Bai and Ng (2002, 2004), in part by its good performance in the simulations reported by Westerlund (2007).

The discussion up to this point relies on the availability of consistent estimates $\widehat{\Omega}_i$, $\widehat{\Gamma}_i$ and $\widehat{\lambda}_i$ of Ω_i , Γ_i and λ_i , respectively. These can be obtained using the following two-step procedure, which begins by estimating λ_i and f_t using the method of principal components. To be precise, let λ , f and \widehat{e} be $K \times N$, $T \times K$ and $T \times N$ matrices of stacked observations on λ_i , f_t and \widehat{e}_{it} , respectively, where \widehat{e}_{it} is the usual least squares residual. The principal components estimator \widehat{f} of

squared test statistics, the approach used here can in principle be applied in any panel testing situation as long as the individual statistics have a standard distribution belonging to the maximum domain of attraction of the Gumbel distribution. However, with equally large and finite N , since the degree of miscalculation of the variance depend positively on $|K - 2|$, it is clear that the Gumbel approximation will not work well if the individual test statistics are normal, which would mean passing $K \rightarrow \infty$.

f can be obtained by computing \sqrt{T} times the eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix $\widehat{e}\widehat{e}'$. The corresponding matrix of estimated loadings is given by $\widehat{\lambda} = \frac{1}{T}\widehat{f}'\widehat{e}$.

These estimates are then used in the second step to compute

$$\widehat{\Omega}_i = \widehat{\Sigma}_i + \widehat{\Gamma}_i + \widehat{\Gamma}'_i = \widehat{\Lambda}_i + \widehat{\Gamma}'_i,$$

where $\widehat{\Sigma}_i$ is the usual contemporaneous estimator of Σ_i based on \widehat{z}_{it} , which is z_{it} with \widehat{f}_t and $\widehat{u}_{it} = \widehat{e}_{it} - \widehat{\lambda}'_i\widehat{f}_t$ in place of f_t and u_{it} , respectively, while $\widehat{\Gamma}_i$ is the usual Newey and West (1994) estimator of Γ_i based on \widehat{z}_{it} and a bandwidth of M , say.

2.3 Monte Carlo simulations

In this section, we evaluate the small-sample properties of the Hausman test using Monte Carlo simulations. The simulation design used for this purpose is taken from Westerlund (2007), and consists of creating 5,000 random samples using (1) through (3) to generate the data, where

$$\begin{pmatrix} f_t \\ u_{it} \\ v_{it} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{ff} & \sigma_{fu} & \sigma_{fv} \\ \cdot & \sigma_{uu} & \sigma_{uv} \\ \cdot & \cdot & \sigma_{vv} \end{bmatrix} \right).$$

We assume that f_t and x_{it} are scalars so that the above error vector has dimension 3×1 . Two experiments depending on the assumed data generating process for this vector will be considered.

In the first experiment, we focus on the covariance matrix of the errors, which is constructed with ones along the diagonal and σ_{fu} set to zero. The off-diagonal elements σ_{fv} and σ_{uv} control the degree of the endogeneity in the data generating process, which can be of two types. The first type is generated from the idiosyncratic component u_{it} and is measured by σ_{uv} , while the second is generated from the common component f_t and is measured by σ_{fv} .

In the second experiment, we set the covariance matrix of the errors equal to identity, and focus instead on the effects of serial correlation in the idiosyncratic error u_{it} , which can be of two types

$$u_{it} = \rho u_{it-1} + w_{it} \quad \text{or} \quad u_{it} = w_{it} + \phi w_{it-1}.$$

In other words, depending on the parametrization of ρ and ϕ , the serial correlation can be of either autoregressive or moving average type.

In both experiments, we use $\alpha_i \sim U(0,1)$ for the parametrization of (1) and $\lambda_i \sim N(1,1)$ for the parametrization of (3). The slope parameter β is of course key in the simulations, and is parameterized as follows. Under the null hypothesis H_0 , we set $\beta_i = 1$ for all i . Moreover, if we let b denote the fraction of non-poolable units, then the alternative H_1 is such that $\beta_i \sim U(0,1)$ for the first bN units and $\beta_i = 1$ for the remaining $(1-b)N$ units.

In implementing the Hausman test, the maximum number of common factors considered for the information criteria is set to five, and all long-run variances and covariances are obtained using the Newey and West (1994) estimator. For brevity, only the results on the size and size-adjusted power at the nominal 5% level are reported. All computations were performed in GAUSS. The results contained in Table 1 may be summarized in the following way.

The first thing to notice is that the size accuracy is generally very good. In fact, our results suggest that the Hausman test is remarkably robust even to very high degrees of serial correlation, a valuable property that is not very common. We also see that the test is able to handle the impact of cross-sectional dependence and endogeneity in both the common and idiosyncratic component.

The fact that the new test is able to maintain good size accuracy even when N and T are comparable makes it very attractive in comparison to the seemingly unrelated regressions based Wald test of Mark *et al.* (2005), which is known to be heavily distorted unless T is substantially larger than N . In fact, based on their simulation results, for this test to be reasonable sized, T should be over 100 times larger than N , which is of course rarely satisfied in practice.

The performance of the new test under the alternative hypothesis is also quite good. As expected, we see that the power can sometimes be poor when T and b are small. However, with T equal to 100 and b larger than 0.1, power is generally very good. Another interesting observation is that the power does not seem to be affected very much by the presence of endogeneity and serial correlation.

3 The monetary exchange rate model

In this section we illustrate the empirical implementation of the Hausman test using as an example the monetary exchange rate model. We begin by a brief motivation, and then we present the results.

3.1 Motivation

The monetary model, in its usual empirical formulation, states that the nominal exchange rate of two countries should cointegrate with the relative money supply and relative output of these countries. We investigate the United States dollar exchange rate, in which case the implied cointegrating regression reads

$$e_{it} = \alpha_i + \beta_{1i}(m_t^* - m_{it}) + \beta_{2i}(y_t^* - y_{it}) + u_{it}, \quad (4)$$

where e_{it} , m_{it} and y_{it} are the logarithm of the nominal exchange rate, money supply and real income for country i at time t , respectively. Asterisks denote the United States. As usual, these variables are all assumed to be nonstationary in their levels.

Now, if the monetary model holds, then β_{1i} should be equal to one for all i while β_{2i} should be negative. In addition, the error u_{it} should be stationary, which, together with the nonstationarity of e_{it} , $(m_t^* - m_{it})$ and $(y_t^* - y_{it})$, implies that (4) should be a cointegrating relationship. The most common way of testing these implications is to use the conventional country-by-country approach. Unfortunately, the results obtained from using this testing strategy have been very mixed and far from convincing.

As a response to this, more recent studies have resorted to larger panel data sets in order to illuminate the issue. For example, Mark and Sul (2001) use quarterly observations for 18 countries between the first quarter of 1973 and the first quarter of 1997, and find support in favor of cointegration for all countries considered, which is noteworthy given the poor performance of the monetary model on an individual country basis.

However, this study suffers from a critical shortcoming that makes it difficult to interpret, and that may well have lead to an overstatement of the results. Namely, that it is based on very restrictive assumptions. This fact was recently pointed out by Rapach and Wohar (2004), who question the preference of Mark and Sul (2001) to assume that the individual slopes β_{1i} and β_{2i} are completely homogenous and prespecified in the cointegrating relationship.

Indeed, by using the same data set, Rapach and Wohar (2004) find that the restriction of homogenous slopes must be rejected, which leaves them with an intricate dilemma. On the one hand, when exploiting the additional information that comes from pooling, they obtain estimates that are much more in line with economic theory than those obtained on an individual country-by-country basis. On the other hand, since the poolability restriction is rejected, they run the risk of obtaining spurious results when pooling. Their conclusion is that, although one should be careful when interpreting the results, there are good reasons to favor the pooled estimates.

In this section, we employ our new test to reevaluate the validity of the poolability restrictions used by both Mark and Sul (2001) and Rapach and Wohar (2004), which is an important undertaking in at least two respects. Firstly, as Rapach and Wohar (2004) show in their simulations, falsely imposed homogeneity is likely to result in estimates that are biased in such a way that they will seem to fulfill the restrictions of the monetary model. Thus, pooling should really not be attempted unless there is very strong evidence in favor of doing so.

Secondly, the poolability tests used by Rapach and Wohar (2004) are not well suited for this particular application. For example, although their Wald test is robust to cross-sectional dependence, which is there by construction of the nominal exchange rate, it is based on seemingly unrelated regressions methods. Hence, for this test to work properly, T should be substantially larger than N , which is clearly not the case in the Mark and Sul (2001) data. Their likelihood ratio test is even more restrictive and requires both cross-sectional independence

and strict exogeneity of relative money and output.

The data that we use is the same as that used by Mark and Sul (2001) and Rapach and Wohar (2004). It comprises quarterly observations on nominal money supply, industrial production and nominal dollar denominated exchange rates for the recent float period from the first quarter of 1973 to the first quarter of 1997. It is comprised of 18 countries, the United Kingdom, Austria, Belgium, Denmark, France, Germany, Netherlands, Canada, Japan, Finland, Greece, Spain, Australia, Italy, Switzerland, Korea, Norway and Sweden. The data is mainly taken from the International Financial Statistics database of the International Monetary Fund. For more details, we make reference to Mark and Sul (2001).

Since the purpose is to evaluate the validity of pooling the data in the way it is done in Mark and Sul (2001) and Rapach and Wohar (2004), we will assume that the cointegration restriction is satisfied. In fact, both Mark and Sul (2001) and Rapach and Wohar (2004) test this condition, and find that it appears to be quite realistic. Thus, the error coming from making the analysis conditional upon this restriction should be relatively small.⁶

3.2 Poolability results

The Hausman test is constructed in exactly the same way as in the simulations, using the Newey and West (1994) variance estimator and a maximum of five common factors. As in Rapach and Wohar (2004), the data is grouped into five panels, the European community panel, the European monetary system panel, the Group of 6 panel, the Group of 10 panel and the full panel.⁷

The test results for each of these panels are reported in Table 2. It is seen that the asymptotic p -values based on the Gumbel distribution provide no evidence of poolability. In fact, the null hypothesis is strongly rejected for all five panels. Of course, this not only casts doubts on the test results provided by Rapach and Wohar (2004), but also on the forecasts generated by Mark and Sul (2001), which are even more restrictive in the sense that the slopes are not only assumed to be homogenous but also prespecified.

However, although the panels considered by Rapach and Wohar (2004) may not be suitable for pooling, these are obviously not the only candidates for a poolable sample. In fact, in our data there are no fewer than 262,125 subpanels with at least two countries. Thus, it seems reasonable to expect at least some of these to be homogenous enough for pooling. The problem is how to select which ones.

The approach taken in this section is very simple and consists of applying the

⁶Although this paper relies on the results of Mark and Sul (2001) and Rapach and Wohar (2004) indicating that the panel is indeed cointegrated, readers should be aware that in general one should never conduct the analysis conditionally in this way without first testing if the cointegration restriction is actually satisfied by the data at hand.

⁷See Table 2 for a complete list of the countries included in each of the four subpanels.

Hausman test in an iterative fashion, by sequentially dropping the maximizing countries in case of a rejection. In doing so, however, we face the problem of controlling the overall significance level of the test, as the asymptotic Gumbel distribution is only valid in the first step of the iteration. In particular, if we denote by \tilde{H}_N^j the normalized Hausman statistic obtained at step j , the problem is that although $P(\tilde{H}_N^j \leq x) \rightarrow \exp(-e^{-x})$ as $N, T \rightarrow \infty$ unconditionally on what happened in the previous steps, subsequent steps are applied only if the previous one ended up in a rejection. Thus, the critical values need to be modified somehow. Fortunately, this is not very difficult.

Consider for simplicity the second step, where it holds that

$$\begin{aligned} & \lim_{N, T \rightarrow \infty} P\left(\{\tilde{H}_N^2 > x\} \mid \{\tilde{H}_N^1 > x_\alpha\}\right) \\ &= \frac{\lim_{N, T \rightarrow \infty} P\left(\{\tilde{H}_N^2 > x\} \cap \{\tilde{H}_N^1 > x_\alpha\}\right)}{\lim_{N, T \rightarrow \infty} P(\tilde{H}_N^1 > x_\alpha)} \\ &= \frac{1}{\alpha} \left(\lim_{N, T \rightarrow \infty} P\left(\{\tilde{H}_N^2 > x\} \cap \{\tilde{H}_N^1 > x_\alpha\}\right) \right) \\ &= \frac{1}{\alpha} \left(\lim_{N, T \rightarrow \infty} P(\tilde{H}_N^2 > x) \right), \end{aligned}$$

where $x_\alpha < x$ is the critical value for an asymptotic α level test. By using further iterations of the same argument, it is not difficult to see that

$$\begin{aligned} & \lim_{N, T \rightarrow \infty} P\left(\{\tilde{H}_N^j > x\} \mid \{\tilde{H}_N^{j-1} > x_\alpha\}, \dots, \{\tilde{H}_N^1 > x_\alpha\}\right) \\ &= \frac{1}{\alpha^{j-1}} \left(\lim_{N, T \rightarrow \infty} P(\tilde{H}_N^j > x) \right). \end{aligned}$$

Thus, for an overall significance level of α , we have

$$\lim_{N, T \rightarrow \infty} P(\tilde{H}_N^j > x) = \alpha^j.$$

It follows that the correct critical value for an iterative α level test at step j is the upper α^j percentile from the Gumbel distribution.

The results obtained from applying this iterative scheme to the monetary exchange rate model are reported in Table 3. It is seen that if we look at the 5% level, then there is evidence of poolability for all countries but Greece and Finland. Hence, based on this evidence, it would appear as that the individual slopes β_{1i} and β_{2i} are actually quite similar across the panel. In fact, even if we look at the liberal 40% level, there are only four countries that cannot be pooled, Greece, Finland, United Kingdom and Italy.

Of course, although these results suggest that a majority of the countries can be pooled, one should keep in mind that this fully corrected iterative procedure has poor power, as the critical values to be used at each iteration increase.

For example, the 5% critical value at the first four steps are 2.97, 5.99, 8.99 and 11.98, respectively. In other words, the test becomes less powerful for each additional iteration, which is probably the main reason for why Tables 2 and 3 seem to convey conflicting results. Specifically, while Table 3 suggests that all countries but Greece and Finland can be pooled at the 5% level, Table 3 suggests that none of the five panels can be pooled, even though three of them do not include Greece or Finland. Similarly, if we look at the 10% level, then all countries but Greece, Finland and United Kingdom can be pooled according to Table 3. Yet the European monetary system panel cannot be pooled according to Table 2. Of course, considering subsamples of similar countries as in Table 2 is also problematic in the sense that it is based on conditioning, which again makes the Hausman test incorrectly sized.

These results make the issue of which countries to pool a difficult one. However, if the primary concern is to not impose any false homogeneity, then this should also be reflected in a relatively high tolerance level in terms of significance. We might for example look at the 15% level rather than the 5% level, in which case there are four non-poolable countries, Greece, Finland, United Kingdom and Italy.

Alternatively, we can iterate as before but using the asymptotic first step critical value at each step. Although this means losing control of the overall test size, it also means that the null hypothesis will be easier to reject, which reduces the risk of obtaining spurious results due to falsely imposed homogeneity. The results reported in the rightmost column of Table 3 indicate that if we remove Greece, Finland, United Kingdom and Italy, there is evidence of poolability if the 1% level is used at each step.

Thus, both strategies yield the conclusion that Greece, Finland, United Kingdom and Italy cannot be pooled, which also seems quite reasonable in the sense that it does not lead to any contradictions in comparison to those reported in Table 2. Our conclusion regarding the issue of poolability is therefore that there is evidence of homogeneity across a majority of the countries, although not across the subpanels considered by Mark and Sul (2001) and Rapach and Wohar (2004).

3.3 Estimation results

Since our results suggest that a substantial portion of the panel can in fact be pooled, it is interesting to estimate and test whether the restrictions of the monetary model hold for these countries. That is, we would like to test if the common value β_1 of β_{1i} is indeed equal to one, and if β_2 , the common value of β_{2i} , is negative.

In order to test these hypotheses, we employ the conventional least squares estimator of Kao and Chiang (2000), the bias-adjusted estimator of Westerlund (2007) and the fully modified estimator of Bai and Kao (2005). As mentioned earlier, the first of these does not permit for cross-section dependence nor en-

dogenuous regressors, and is generally biased. The latter two estimators, on the other hand, allow for these features and are therefore more widely applicable. The bias-adjusted estimator was discussed in Section 2 and requires no further presentation. The Bai and Kao (2005) estimator, which is new, can be seen as a factor augmented version of the more conventional panel fully modified estimator of Kao and Chiang (2000).

Table 4 contains the results obtained from applying these estimators to the 14 countries that were earlier found to be poolable. For comparability, we also report the results for the nine countries that lead to an acceptance of the poolability null when the 5% level is used at each step in the iterative scheme. We see that all three estimators yield very similar results, which may be summarized as follows. Firstly, the estimates of β_1 are markedly smaller than one, and the unit slope hypothesis is strongly rejected, which means that the unit slope, or monetary neutrality, restriction of the monetary model is violated. Secondly, although the results based on the least squares estimator provide some evidence of a negative β_2 , there is no such evidence if we look at the other two estimators, at least not at the 1% level. In other words, if we disregard the least squares estimator, which is biased, then there is no real evidence to suggest that the elasticity of income is negative. Thus, since there is no evidence of either monetary neutrality or a negative income elasticity, the monetary model has to be rejected.

4 Conclusions

A common problem in applied work using panel cointegration techniques is that the hypothesis of poolability is almost always rejected, even though there are strong reasons to expect the cointegration parameters to be equal. A widely held interpretation of this result is that the routinely applied Wald test performs poorly in samples of realistic size, and that it is this poor performance that causes the test to reject so often. Moreover, with the ability to pool forfeited, the panel cointegration methodology loses much of its appeal.

This paper is motivated by these concerns. In particular, a new poolability test based on the Hausman (1978) principle is proposed that is shown to have a well-defined limiting distribution and excellent small-sample properties. In our empirical application to the monetary exchange rate model, we further show how to implement the new test in an iterative fashion to determine the number of poolable units. The results suggest that although a substantial proportion of the panel can be pooled, the estimated regression does not satisfy the restrictions of the monetary model.

Appendix: Mathematical proofs

In this appendix, we derive the asymptotic distribution of the Hausman test statistic provided in Theorem 1. We do this by first showing that H_N , the infeasible test based on the known bias-adjusted estimators $\beta_i^+ = \widehat{\beta}_i - b_i$ and $\beta_N^+ = \widehat{\beta}_N - b_N$ has an asymptotic Gumbel distribution. We then show that replacing β_i^+ and β_N^+ with their estimated counterparts has no effect on this result. Before we proceed to the proof of the theorem, however, we need some auxiliary results.

Lemma A.1. Under Assumptions 1 through 3, as $T \rightarrow \infty$

$$T(\beta_i^+ - \beta) \Rightarrow N\left(0, Q_i^{-1} \left(\frac{1}{6} \Omega_{e.vi} \Omega_{vvi} \right) Q_i^{-1}\right).$$

Proof of Lemma A.1.

From the definition of β_i^+ , we have that

$$T(\beta_i^+ - \beta) = T(\widehat{\beta}_i - \beta) - Tb_i.$$

The first term on the right-hand side of this equation is simply the individual least squares estimator of β , which can be stated as

$$T(\widehat{\beta}_i - \beta) = \left(\frac{1}{T} \sum_{t=1}^T e_{it} \tilde{x}'_{it} \right) \left(\frac{1}{T^2} \sum_{t=1}^T \tilde{x}_{it} \tilde{x}'_{it} \right)^{-1}.$$

This expression, together with the definition of b_i , suggests that

$$\begin{aligned} T(\beta_i^+ - \beta) &= T(\widehat{\beta}_i - \beta) - Tb_i \\ &= \left(\frac{1}{T} \sum_{t=1}^T e_{it} \tilde{x}'_{it} - \frac{1}{T} (\lambda'_i U_{fvi} + U_{uvi}) \right) \left(\frac{1}{T^2} \sum_{t=1}^T \tilde{x}_{it} \tilde{x}'_{it} \right)^{-1}. \end{aligned} \quad (\text{A1})$$

Now, the limit of the denominator of (A1) as $T \rightarrow \infty$ is given by

$$\frac{1}{T^2} \sum_{t=1}^T \tilde{x}_{it} \tilde{x}'_{it} \Rightarrow \int_0^1 \tilde{B}_{vi} \tilde{B}'_{vi} = \Omega_{vvi}^{1/2} \left(\int_0^1 \tilde{W}_{vi} \tilde{W}'_{vi} \right) \Omega_{vvi}^{1/2},$$

while the first part of the numerator may be written

$$\frac{1}{T} \sum_{t=1}^T e_{it} \tilde{x}'_{it} = \frac{1}{T} \sum_{t=1}^T \lambda'_i f_t \tilde{x}'_{it} + \frac{1}{T} \sum_{t=1}^T u_{it} \tilde{x}'_{it}, \quad (\text{A2})$$

where we have the following results as $T \rightarrow \infty$

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T f_{it} \tilde{x}'_{it} &\Rightarrow \int_0^1 dB_f \tilde{B}'_{vi} + \Lambda_{fvi} \\
&= \int_0^1 dB_{f.vi} \tilde{B}'_{vi} + \Omega_{fvi} \Omega_{vvi}^{-1} \int_0^1 dB_{vi} \tilde{B}'_{vi} + \Lambda_{fvi}, \\
\frac{1}{T} \sum_{t=1}^T u_{it} \tilde{x}'_{it} &\Rightarrow \int_0^1 dB_{ui} \tilde{B}'_{vi} + \Lambda_{uvi} \\
&= \int_0^1 dB_{u.vi} \tilde{B}'_{vi} + \Omega_{uvi} \Omega_{vvi}^{-1} \int_0^1 dB_{vi} \tilde{B}'_{vi} + \Lambda_{uvi},
\end{aligned}$$

where $B_{f.vi}$ and $B_{u.vi}$ may be decomposed as $B_{f.vi} = B_f - \Omega_{fvi} \Omega_{vvi}^{-1} B_{vi} = \Omega_{f.vi}^{1/2} W_f$ and $B_{u.vi} = B_{ui} - \Omega_{uvi} \Omega_{vvi}^{-1} B_{vi} = \Omega_{u.vi}^{1/2} W_{ui}$, respectively. Thus, the limit of (A2) can be written as

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T e_{it} \tilde{x}'_{it} &\Rightarrow \lambda'_i \left(\int_0^1 dB_{f.vi} \tilde{B}'_{vi} \right) + \int_0^1 dB_{u.vi} \tilde{B}'_{vi} \\
&+ \lambda'_i \left(\Omega_{fvi} \Omega_{vvi}^{-1} \int_0^1 dB_{vi} \tilde{B}'_{vi} + \Lambda_{fvi} \right) \\
&+ \Omega_{uvi} \Omega_{vvi}^{-1} \int_0^1 dB_{vi} \tilde{B}'_{vi} + \Lambda_{uvi}. \tag{A3}
\end{aligned}$$

The second part of the numerator of (A1) is given by

$$\begin{aligned}
\frac{1}{T} \left(\lambda'_i U_{fvi} + U_{uvi} \right) &= \lambda'_i \left(\Omega_{fvi} \Omega_{vvi}^{-1} \left(\frac{1}{T} \sum_{t=1}^T \Delta x_{it} \tilde{x}'_{it} - \Lambda_{vvi} \right) + \Lambda_{fvi} \right) \\
&+ \Omega_{uvi} \Omega_{vvi}^{-1} \left(\frac{1}{T} \sum_{t=1}^T \Delta x_{it} \tilde{x}'_{it} - \Lambda_{vvi} \right) + \Lambda_{uvi},
\end{aligned}$$

where

$$\frac{1}{T} \sum_{t=1}^T \Delta x_{it} \tilde{x}'_{it} - \Lambda_{vvi} \Rightarrow \int_0^1 dB_{vi} \tilde{B}'_{vi}.$$

By using (A3) and this result, it is not difficult to see that the numerator of (A1) has the following limit as $T \rightarrow \infty$

$$\begin{aligned}
\frac{1}{T} \left(\sum_{t=1}^T e_{it} \tilde{x}'_{it} - \left(\lambda'_i U_{fvi} + U_{uvi} \right) \right) &\Rightarrow \lambda'_i \left(\int_0^1 dB_{f.vi} \tilde{B}'_{vi} \right) + \int_0^1 dB_{u.vi} \tilde{B}'_{vi} \\
&= U_i, \quad \text{say.} \tag{A4}
\end{aligned}$$

Now, by using the independence of dW_f , dW_{ui} and \widetilde{W}_{vi} , we get

$$E(U_i) = \lambda'_i \Omega_{f.vi}^{1/2} E \left(\int_0^1 dW_f \widetilde{W}'_{vi} \right) \Omega_{vvi}^{1/2} + \Omega_{u.vi}^{1/2} E \left(\int_0^1 dW_{ui} \widetilde{W}'_{vi} \right) \Omega_{vvi}^{1/2} = 0.$$

The variance of U_i can be computed as follows

$$\begin{aligned} E(U'_i U_i) &= E \left(\int_0^1 \widetilde{B}_{vi} dB'_{f.vi} \lambda_i \lambda'_i \int_0^1 dB_{f.vi} \widetilde{B}'_{vi} \right) \\ &+ E \left(\int_0^1 \widetilde{B}_{vi} dB_{u.vi} \int_0^1 dB_{u.vi} \widetilde{B}'_{vi} \right) \\ &= \Omega_{vvi}^{1/2} E \left(\int_0^1 \widetilde{W}_{vi} dW'_f C_i \int_0^1 dW_f \widetilde{W}'_{vi} \right) \Omega_{vvi}^{1/2} \\ &+ \Omega_{u.vi} \Omega_{vvi}^{1/2} E \left(\int_0^1 \widetilde{W}_{vi} dW_{ui} \int_0^1 dW_{ui} \widetilde{W}'_{vi} \right) \Omega_{vvi}^{1/2}, \quad (\text{A5}) \end{aligned}$$

where $C_i = \Omega_{f.vi}^{1/2} \lambda_i \lambda'_i \Omega_{f.vi}^{1/2}$. The first expectation in (A5) can be written

$$\begin{aligned} E \left(\int_0^1 \widetilde{W}_{vi} dW'_f C_i \int_0^1 dW_f \widetilde{W}'_{vi} \right) &= \int_0^1 \int_0^1 E(\widetilde{W}_{vi} dW'_f C_i dW_f \widetilde{W}'_{vi}) \\ &= \int_0^1 \int_0^1 E(dW'_f C_i dW_f) E(\widetilde{W}_{vi} \widetilde{W}'_{vi}) \\ &= \lambda'_i \Omega_{f.vi} \lambda_i \left(\int_0^1 E(\widetilde{W}_{vi} \widetilde{W}'_{vi}) \right), \quad (\text{A6}) \end{aligned}$$

where the integrals run over r and s , and the second equality holds since $dW'_f C_i dW_f$ is a scalar. The third equality follows directly from the fact that $E(dW'_f C_i dW_f) = E(dW'_f) C_i E(dW_f) = 0$ if $r \neq s$, whereas if $r = s$, then

$$\begin{aligned} E(dW'_f C_i dW_f) &= \text{tr}(E(dW'_f C_i dW_f)) = \text{tr}(E(dW_f dW'_f) C_i) \\ &= \text{tr}(I_K C_i) = \text{tr}(C_i) = \lambda'_i \Omega_{f.vi} \lambda_i. \end{aligned}$$

Also, by the moments of Brownian motion

$$\int_0^1 E(\widetilde{W}_{vi} \widetilde{W}'_{vi}) = \frac{1}{6} I_K,$$

which suggests that (A6) reduces to

$$E \left(\int_0^1 \widetilde{W}_{vi} dW'_f C_i \int_0^1 dW_f \widetilde{W}'_{vi} \right) = \lambda'_i \Omega_{f.vi} \lambda_i \left(\frac{1}{6} I_K \right).$$

Similarly, the second expectation in (A5) can be written

$$\begin{aligned}
E\left(\int_0^1 \widetilde{W}_{vi} dW_{ui} \int_0^1 dW_{ui} \widetilde{W}'_{vi}\right) &= \int_0^1 \int_0^1 E(\widetilde{W}_{vi} dW_{ui} dW_{ui} \widetilde{W}'_{vi}) \\
&= \int_0^1 \int_0^1 E(dW_{ui} dW_{ui}) E(\widetilde{W}_{vi} \widetilde{W}'_{vi}) \\
&= \int_0^1 E(\widetilde{W}_{vi} \widetilde{W}'_{vi}) = \frac{1}{6} I_K,
\end{aligned}$$

where the integrals again run over r and s .

Thus, by combining the results, we get

$$E(U'_i U_i) = \Omega_{vvi}^{1/2} \left(\lambda'_i \Omega_{f.vi} \lambda_i \left(\frac{1}{6} I_K \right) + \Omega_{u.vi} \left(\frac{1}{6} I_K \right) \right) \Omega_{vvi}^{1/2} = \frac{1}{6} \Omega_{e.vi} \Omega_{vvi},$$

which implies that as $T \rightarrow \infty$

$$\frac{1}{T} \sum_{t=1}^T e_{it} \widetilde{x}'_{it} - \frac{1}{T} (\lambda'_i U_{fvi} + U_{uvi}) \Rightarrow N\left(0, \frac{1}{6} \Omega_{e.vi} \Omega_{vvi}\right).$$

This, together with (A1), implies that the following result holds conditionally on x_{it}

$$T(\widehat{\beta}_i^+ - \beta) \Rightarrow N\left(0, Q_i^{-1} \left(\frac{1}{6} \Omega_{e.vi} \Omega_{vvi} \right) Q_i^{-1}\right).$$

This completes the proof. ■

Lemma A.2. Under Assumptions 1 through 3, as $N, T \rightarrow \infty$

$$T(\widehat{\beta}_i^+ - \widehat{\beta}_N^+) = T(\beta_i^+ - \beta_N^+) + o_p(1).$$

Proof of Lemma A.2.

Note that

$$T(\widehat{\beta}_i^+ - \widehat{\beta}_N^+) - T(\beta_i^+ - \beta_N^+) = T(\widehat{\beta}_i^+ - \beta_i^+) - T(\widehat{\beta}_N^+ - \beta_N^+).$$

The second term is $T(\widehat{\beta}_N^+ - \beta_N^+) = o_p(1)$ by Theorem 2 of Westerlund (2007). Also, note that by definition of the bias-adjusted estimator, the first term can be written as

$$\begin{aligned}
T(\widehat{\beta}_i^+ - \beta_i^+) &= T(\widehat{b}_i - b_i) \\
&= \frac{1}{T} \left((\widehat{\lambda}'_i \widehat{U}_{fvi} - \lambda'_i U_{fvi}) + (\widehat{U}_{uvi} - U_{uvi}) \right) \left(\frac{1}{T^2} \sum_{t=1}^T \widetilde{x}_{it} \widetilde{x}'_{it} \right)^{-1} \\
&= \frac{1}{T} (\widehat{\lambda}'_i \widehat{U}_{fvi} - \lambda'_i U_{fvi}) O_p(1) + \frac{1}{T} (\widehat{U}_{uvi} - U_{uvi}) O_p(1) \\
&= I + II, \quad \text{say.} \tag{A7}
\end{aligned}$$

We begin with I , which may be written as

$$I = (\hat{\lambda}_i - \lambda_i)' \left(\frac{1}{T} \hat{U}_{fvi} \right) O_p(1) + \frac{1}{T} \lambda_i' (\hat{U}_{fvi} - U_{fvi}) O_p(1).$$

From the consistency of the estimated variances it follows that

$$\frac{1}{T} \hat{U}_{fvi} = \hat{\Omega}_{fvi} \hat{\Omega}_{vvi}^{-1} \left(\frac{1}{T} \sum_{t=1}^T \Delta x_{it} \tilde{x}'_{it} - \hat{\Lambda}_{vvi} \right) + \hat{\Lambda}_{fvi} = O_p(1).$$

Also, if we let $C_{NT} = \max\{\sqrt{T}, N\}$, then it follows from Lemma 1 (c) of Bai and Ng (2004) that

$$\hat{\lambda}_i - \lambda_i = O_p(1/C_{NT}^2).$$

The second term on the right-hand side of I can be written as

$$\begin{aligned} \frac{1}{T} \lambda_i' (\hat{U}_{fvi} - U_{fvi}) &= \lambda_i' (\hat{\Omega}_{fvi} \hat{\Omega}_{vvi}^{-1} - \Omega_{fvi} \Omega_{vvi}^{-1}) \left(\frac{1}{T} \sum_{t=1}^T \Delta x_{it} \tilde{x}'_{it} \right) \\ &\quad - \lambda_i' (\hat{\Omega}_{fvi} \hat{\Omega}_{vvi}^{-1} \hat{\Lambda}_{vvi} - \Omega_{fvi} \Omega_{vvi}^{-1} \Lambda_{vvi}) \\ &\quad + \lambda_i' (\hat{\Lambda}_{fvi} - \Lambda_{fvi}). \end{aligned} \tag{A8}$$

Consider the last term on the right-hand side of this expression. By using $\|AB\| \leq \|A\| \|B\|$ and Theorem 1 of Andrews (1991), we get

$$\left\| \lambda_i' (\hat{\Lambda}_{fvi} - \Lambda_{fvi}) \right\| \leq \|\lambda_i\| \|\hat{\Lambda}_{fvi} - \Lambda_{fvi}\| = O_p(\sqrt{M/T}).$$

The remaining two terms on the right-hand side of (A8) can be evaluated in exactly the same fashion, and are both $O_p(\sqrt{M/T})$. Thus, we can show that

$$I = O_p(1/C_{NT}^2) + O_p(\sqrt{M/T}) = O_p(\sqrt{M/T}).$$

Next, consider II . Similar to (A8), we have

$$\begin{aligned} II &= \frac{1}{T} (\hat{U}_{uvi} - U_{uvi}) O_p(1) \\ &= (\hat{\Omega}_{uvi} \hat{\Omega}_{vvi}^{-1} - \Omega_{uvi} \Omega_{vvi}^{-1}) \left(\frac{1}{T} \sum_{t=1}^T \Delta x_{it} \tilde{x}'_{it} \right) O_p(1) \\ &\quad - (\hat{\Omega}_{uvi} \hat{\Omega}_{vvi}^{-1} \hat{\Lambda}_{vvi} - \Omega_{uvi} \Omega_{vvi}^{-1} \Lambda_{vvi}) O_p(1) + (\hat{\Lambda}_{uvi} - \Lambda_{uvi}) O_p(1) \\ &= O_p(\sqrt{M/T}). \end{aligned}$$

Thus, by putting everything together, we see that

$$T(\hat{\beta}_i^+ - \beta_i^+) = O_p(\sqrt{M/T}).$$

Hence, if we assume that $M/T \rightarrow 0$ as $M, N, T \rightarrow \infty$, then $\sqrt{NT}(\widehat{\beta}_i^+ - \beta_i^+) = o_p(1)$ as required for the proof. \blacksquare

Proof of Theorem 1.

To prove Theorem 1, simply note that

$$T(\beta_i^+ - \beta_N^+) = T(\beta_i^+ - \beta) - T(\beta_N^+ - \beta) = T(\beta_i^+ - \beta) + O_p(N^{-1/2}).$$

where we have used Theorem 1 of Westerlund (2007), which states that as $N, T \rightarrow \infty$ with $\sqrt{N}/T \rightarrow 0$

$$\sqrt{NT}(\beta_N^+ - \beta) = O_p(1).$$

For the first term on the right-hand side, we use Lemma A.1, from which it follows that as $T \rightarrow \infty$

$$T(\beta_i^+ - \beta_N^+) \Rightarrow N \left(0, Q_i^{-1} \left(\frac{1}{6} \Omega_{e.vi} \Omega_{vvi} \right) Q_i^{-1} \right),$$

which in turn implies that

$$H_i = T^2(\beta_i^+ - \beta_N^+) \left(Q_i^{-1} \left(\frac{1}{6} \Omega_{e.vi} \Omega_{vvi} \right) Q_i^{-1} \right)^{-1} (\beta_i^+ - \beta_N^+)' \Rightarrow \chi^2(K).$$

We can now apply Lemma A.2 to show that

$$\widehat{H}_i \Rightarrow \chi^2(K). \tag{A9}$$

The result in (A9) implies that as $N, T \rightarrow \infty$ with $\sqrt{N}/T \rightarrow 0$

$$P \left(\frac{1}{a_N} (\widehat{H}_N - b_N) \leq x \right) \rightarrow \exp(-e^{-x}),$$

where $-\infty < x < \infty$. The quantities a_N and b_N are given by $a_N = 2$ and $b_N = F^{-1}(1 - 1/N)$, see Embrechts *et al.* (1997). Note that according to the Poisson limit theorem, we have that the limiting statement

$$\lim_{N \rightarrow \infty} P \left(\frac{1}{a_N} (\widehat{H}_N - b_N) \leq x \right) = \lim_{N \rightarrow \infty} F^N(a_N x + b_N) = \exp(-e^{-x}),$$

is equivalent to

$$\lim_{N \rightarrow \infty} N \left(1 - F(a_N x + b_N) \right) = e^{-x}.$$

In the special case where $F(x)$ is the chi-squared distribution function with two degrees of freedom, then this statement reduces to

$$\begin{aligned}
 N(1 - F(2x + b_N)) &= N\left(1 - \int_0^{(2x+b_N)/2} e^{-t} dt\right) \\
 &= N\left(e^{-(2x+b_N)/2}\right) \\
 &= N(1 - F(b_N)) e^{-x} \\
 &= e^{-x}.
 \end{aligned}$$

Hence, with K equal to two, it follows that for any $N \geq 2$

$$P\left(\frac{1}{a_N}(\widehat{H}_N - b_N) \leq x\right) = \exp(-e^{-x}),$$

as required for the proof. ■

Table 1: Size and size-adjusted power on the 5% level.

σ_{fv}	σ_{uv}	N	T	H_1 fraction					
				0	0.1	0.2	0.3	0.4	0.5
0	0	10	50	3.2	15.9	22.3	30.5	36.4	40.0
		20	50	3.4	17.9	25.6	31.6	38.9	42.0
		10	100	4.3	30.5	48.6	62.2	70.3	77.3
		20	100	4.5	39.0	58.7	73.0	80.1	85.5
-0.5	0	10	50	1.5	24.3	38.0	46.6	54.5	55.6
		20	50	1.4	29.4	44.3	54.8	62.1	65.3
		10	100	2.0	39.9	63.5	74.6	81.7	86.3
		20	100	2.9	48.3	70.2	82.2	86.8	91.3
0	-0.5	10	50	3.0	15.4	24.3	30.7	35.8	39.1
		20	50	3.1	17.9	26.2	32.6	39.5	44.0
		10	100	4.0	31.4	50.3	63.9	72.5	77.1
		20	100	4.8	39.0	59.0	73.3	80.3	85.8
-0.5	-0.5	10	50	1.2	27.5	43.4	53.8	62.8	64.7
		20	50	0.7	36.5	51.6	63.0	66.3	72.0
		10	100	0.9	50.1	71.8	82.0	89.3	91.6
		20	100	1.2	57.1	76.7	87.0	93.6	95.5

Continued overleaf

Table 2: Continued.

ρ	ϕ	N	T	H_1 fraction					
				0	0.1	0.2	0.3	0.4	0.5
0.5	0	10	50	3.2	13.6	21.6	28.7	32.5	36.4
		20	50	4.4	13.2	20.5	28.1	29.1	31.6
		10	100	4.8	27.8	43.7	56.9	67.5	73.0
		20	100	7.3	31.6	48.3	60.6	69.5	76.2
0	-0.5	10	50	2.9	16.4	26.8	33.3	38.6	44.8
		20	50	3.8	20.1	29.1	37.3	42.2	42.6
		10	100	4.0	33.3	54.8	66.5	74.7	81.8
		20	100	4.7	41.0	58.1	73.5	81.2	85.1
0.5	-0.5	10	50	2.9	17.6	27.9	33.2	41.0	45.4
		20	50	2.7	22.4	33.5	40.1	45.5	50.3
		10	100	4.8	30.4	48.8	62.4	72.9	76.9
		20	100	4.9	38.9	59.1	71.8	80.3	85.0

Notes: The parameter σ_{fv} measures the common endogeneity, σ_{uv} measures the idiosyncratic endogeneity, ρ measures the autoregressive serial correlation and ϕ measures the moving average serial correlation. In the top panel, there is no serial correlation, while in the bottom panel, there is no endogeneity. The numbers in the table represent the size if the H_1 fraction is zero and the size-adjusted power if the H_1 fraction is greater than zero.

Table 2: Poolability tests of the monetary model.

Panel	Value	p -value
European community ^a	32.583	0.000
European monetary system ^b	17.823	0.000
Group of 6 ^c	18.795	0.000
Group of 10 ^d	33.132	0.000
All 18	43.331	0.000

Notes: The p -values are based on the asymptotic Gumbel distribution.

^aBelgium, Denmark, France, Germany, Greece, Italy, Netherlands, Spain and the United Kingdom.

^bBelgium, Denmark, France, Germany, Italy, Netherlands, Spain.

^cCanada, France, Germany, Italy, Japan and the United Kingdom.

^dBelgium, Canada, France, Germany, Italy, Japan, Netherlands, Sweden Switzerland, and the United Kingdom.

Table 3: Iterative poolability tests of the monetary model.

No.	Country	Value	p -value ^a	p -value ^b
1	Greece	40.580	0.000	0.000
2	Finland	10.765	0.005	0.000
3	United Kingdom	8.410	0.061	0.000
4	Italy	7.653	0.148	0.000
5	Canada	4.162	0.434	0.015
6	Belgium	3.918	0.520	0.020
7	Australia	3.788	0.581	0.022
8	Austria	3.718	0.627	0.024
9	Norway	3.492	0.677	0.030
10	Switzerland	2.858	0.749	0.056
11	Germany	2.139	0.819	0.111
12	Denmark	2.033	0.840	0.123
13	Japan	1.778	0.867	0.155
14	Spain	1.807	0.874	0.151
15	Korea	0.776	0.936	0.369
16	Netherland	0.697	0.943	0.392
17	France	0.282	0.963	0.530
18	Sweden	0.281	0.965	0.530

Notes: The values reported in the table are the maximum test statistics at each iteration.

^aThe p -values are corrected for iterations so as to maintain the overall significance level of the test.

^bThe p -values are not corrected for iterations.

Table 4: Pooled estimates of the monetary model.

Value	Poolable at the 1% level			Poolable at the 5% level		
	LS	BA	FM	LS	BA	FM
β_1	0.345	0.359	0.505	0.397	0.412	0.621
SE	3.399	31.959	31.959	2.947	33.783	33.783
p -value ^c	0.000	0.000	0.000	0.000	0.000	0.002
β_2	-0.129	-0.186	-0.128	-0.216	-0.266	-0.285
SE	3.759	38.402	38.402	3.288	40.777	40.777
p -value ^d	0.000	0.044	0.120	0.000	0.033	0.024

Notes: LS refers to the least squares estimator of Kao and Chiang (2000), BA refers to the bias-adjusted estimator of Westerlund (2007) and FM refers to the fully modified estimator of Bai and Kao (2005). The number of poolable countries is 14 at the 1% level and nine at the 5% level, see Table 4.

^aThe p -values are for the null hypothesis of a unit slope against the double-sided alternative.

^bThe p -values are for the null hypothesis of a zero slope against the left-sided alternative.

References

- Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, Vol. 59, pp. 817-858.
- Bai, J., and C. Kao (2005). On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence. In Baltagi, B. (Ed.), *Contributions to Economic Analysis*, Elsevier, Amsterdam.
- Bai, J., and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, Vol. 70, pp. 191-221.
- Bai, J., and S. Ng (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica*, Vol. 72, pp. 1127-1177.
- Baltagi, B. H., J. M. Griffin and W. Xiong (2000). To Pool or Not to Pool: Homogeneous Versus Heterogeneous Estimators Applied to Cigarette Demand. *Review of Economics and Statistics*, Vol. 82, pp. 117-126.
- Embrechts, P., C. Klüppelberg, T. Mikosch (1997). *Modelling extremal events for insurance and finance*. Berlin, Springer.
- Hausman, J. (1978). Specification Tests in Econometrics. *Econometrica*, Vol. 46, pp. 1251-1271.
- Kao, C., and M. H. Chiang (2000). On the Estimation and Inference of a Cointegrated Regression in Panel Data. *Advances in Econometrics*, Vol. 15, pp. 179-222.
- Kapetanios, G. (2003). Determining the Poolability of Individual Series in Panel Datasets. University of London Queen Mary Economics Working Paper No. 499.
- MacKinnon, J. G. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests, *Journal of Applied Econometrics*. Vol. 11, pp. 601-618.
- Mark, N. C., M. Ogaki and D. Sul (2005). Dynamic Seemingly Unrelated Cointegrating Regression. *Review of Economic Studies*, Vol. 72, pp. 797-820.
- Mark, N. C., and D. Sul (2001). Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Post-Bretton Woods Panel. *Journal of International Economics*, Vol. 53, pp. 29-52.
- Mark, N. C., and D. Sul (2003). Cointegration Vector Estimation by Panel DOLS and the Long-Run Money Demand. *Oxford Bulletin of Economics and Statistics*, Vol. 65, pp. 655-680.

- Moon, H. R., and B. Perron (2004). Efficient Estimation of the SUR Cointegration Regression Model and Testing for Purchasing Power Parity. *Econometric Reviews*, Vol. 23, pp. 293-323.
- Newey, W., and K. West (1994). Autocovariance Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, Vol. 61, pp. 613-653.
- Pedroni, P. (2004). Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis. *Econometric Theory*, Vol. 20, pp. 597-625.
- Pesaran, M. H., Y. Shin and R. Smith (1999). Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association*, Vol. 94, pp. 621-634.
- Rapach, D. E., and M. E. Wohar (2004). Testing the Monetary Model of Exchange Rate Determination: A Closer Look at Panels. *Journal of International Money and Finance*, Vol. 23, pp. 867-895.
- Westerlund, J. (2005). New Simple Tests for Panel Cointegration. *Econometric Reviews*, Vol. 24, pp. 297-316.
- Westerlund, J. (2007). Estimating Cointegrated Panels with Common Factors and the Forward Rate Unbiasedness Hypothesis. *Journal of Financial Econometrics*, forthcoming.