

Trapani, L. & Urga, G. (2010). Micro versus macro cointegration in heterogeneous panels. *Journal of Econometrics*, 155(1), pp. 1-18. doi: 10.1016/j.jeconom.2009.07.005



**CITY UNIVERSITY  
LONDON**

[City Research Online](#)

**Original citation:** Trapani, L. & Urga, G. (2010). Micro versus macro cointegration in heterogeneous panels. *Journal of Econometrics*, 155(1), pp. 1-18. doi: 10.1016/j.jeconom.2009.07.005

**Permanent City Research Online URL:** <http://openaccess.city.ac.uk/6111/>

### **Copyright & reuse**

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

### **Versions of research**

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

### **Enquiries**

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at [publications@city.ac.uk](mailto:publications@city.ac.uk).

# Micro versus Macro Cointegration in Heterogeneous Panels

Lorenzo Trapani

(Cass Business School, UK and Università' di Bergamo, Italy)

Giovanni Urga\*

(Cass Business School, UK and Università' di Bergamo, Italy)

June 15, 2009

---

\**Corresponding author:* Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ Tel. +/44/(0)20/70408698; Fax. +/44/(0)20/70408885; g.urg@city.ac.uk; <http://www.cass.city.ac.uk/faculty/g.urg>.

# Micro versus Macro Cointegration in Heterogeneous Panels

## Abstract

We consider the issue of cross-sectional aggregation in nonstationary and heterogeneous panels where each unit cointegrates. We derive asymptotic properties of the aggregate estimate, and necessary and sufficient conditions for cointegration to hold in the aggregate relationship. We then analyze the case when cointegration does not carry through the aggregation process, and we investigate whether the violation of the formal conditions for perfect aggregation can still lead to an aggregate equation that is observationally equivalent to a cointegrated relationship. We derive a measure of the degree of noncointegration of the aggregate relationship and we explore its asymptotic properties. We propose a valid bootstrap approximation of the test. A Monte Carlo exercise evaluates size and power properties of the bootstrap test.

**J.E.L. Classification Numbers:** C12, C13, C23

**Keywords:** Heterogeneous Panels, Aggregation, Cointegration, Spurious Regression, Bootstrap.

# 1 Introduction

The assumption of the existence of a representative agent in macroeconomics has generated a huge body of literature on aggregation (see e.g. Granger 1990; Stoker, 1993; Pesaran, 2003). The main research question is of how well the aggregate relationship approximates the properties of the individual components. This question cannot be examined when only aggregate data are available. However, when data are available at disaggregate level, it is quite well known that the features of micro models may not be preserved at the macro level, and a crucial role is played by the degree of heterogeneity amongst micro units. In a series of papers, Lippi and Forni (see e.g. Lippi, 1988; Forni and Lippi 1997, 1998, 1999) show theoretically and empirically that irrespective of the approach one chooses for macroeconomic analysis, when heterogeneity across agents is allowed, the dynamic properties of aggregated equations differ from those of micro equations, thereby leading to substantially different interpretations. Basic properties of the micro models describing the panel units do not carry through aggregation, thus increasing "the difficulties involved in formulating a macro model" (Forni and Lippi, 1998). Examples are the introduction of dynamics after aggregating static micro equations and of Granger causality among aggregated variables when it is absent at the disaggregated level. This is a double-edged sword: on the one hand, in Forni and Lippi's (1998) words, "existing models which are at odds with aggregate data under the representative agent assumption could be reconciled with empirical evidence", on the other hand the exact opposite can happen and macroeconomic relationship that are supposed to be valid would not be verified by the data.

A classical example of a property that is shared by the micro equations, and that is almost always wiped out after aggregation, is cointegration. Pesaran and Smith (1995) show that aggregation of heterogeneous cointegrating equations does not imply cointegration in the aggregate relationship unless some specific conditions are satisfied. A crucial role is played by the number of units  $n$  in the panel. A

well known theoretical result (Phillips and Moon, 1999) is that when large panels are available, i.e. under the  $(n, T) \rightarrow \infty$  case, the fact that  $n \rightarrow \infty$  entails that a long-run average relationship between two nonstationary panel vectors exists even when the single units do not cointegrate. On the other hand, with fixed  $n$ , Granger (1993) considers a model where each equation is a cointegration relationship with one explanatory variable, and finds that a necessary and sufficient condition for cointegration to be maintained after aggregation is that the number of stochastic common trends that generate the nonstationary variables is equal to one. The presence of a greater number of common trends therefore leads to a spurious regression after aggregation. Gonzalo (1993) bases his analysis on a more complex multivariate model and derives a sufficient condition for cointegration to hold after aggregation. The conditions laid out by Granger (1993) and Gonzalo (1993) are very restrictive; however, the existence of cointegration at macro level is a well established result. Hence the need for a test that is capable of checking whether cointegration holds after aggregation or not.

There are important empirical implications of the ability to determine whether a macro model is observationally equivalent to a cointegration relationship. An illuminating example of the case where the information content of macro data clashes with that of micro data has recently been provided by Hsiao, Shen and Fujiki (2005). When using micro prefecture level data at an annual frequency, authors find cointegrated money demand functions in Japan. Cointegration is no longer valid when aggregated data at a quarterly frequency are used. Other potential applications that may illustrate the relevance of aggregate cointegration are Campbell and Shiller's (1987) investigation of the relation between stock prices and dividends, Vuolteenaho's (2002) exploitation of cointegrating relationships between accounting variables (book equity and market equity; dividend and market equity), and the firm-level cointegration tests between trading strategies as investigated in Gatev, Goetzmann and Rouwenhorst (2006).

*Hypotheses of interest and the main results of this paper*

In this paper we propose a test for aggregate cointegration that uses the information contained in both the aggregate and the disaggregate data. Using micro data in order to test the macro relationship also proves useful since, both under the null hypothesis of aggregate cointegration and under the alternative, cointegration holds in the micro relationships; thus, estimates derived from micro data are always  $T$ -consistent, whether cointegration holds in the aggregate relationship or not.

The null hypothesis of our testing framework is *presence* of cointegration in the aggregate relationship. Thus, the test developed here is similar in spirit to the stationarity test developed by Kwiatkowski, Phillips, Schmidt and Shin (1992) and tests for cointegration developed thereafter, e.g. Shin (1994) and Xiao (1999), and McCoskey and Kao (1998) for the case of panel data. Testing for the null of cointegration is natural in our framework, since aggregate cointegration is the hypothesis of relevance, also in light of the prior information that cointegration does hold in the micro relationships.

Building on the measure of departure from aggregate cointegration developed in Lazarova, Trapani and Urga (2007) for a simple bivariate model, in this contribution we consider a heterogeneous panel where each micro equation contains several explanatory variables and several common stochastic trends. We propose a test statistic for the null of cointegration in the aggregate relationship using the disaggregated data. We provide an estimation procedure based on Principal Components when common trends are unobservable for the case of finite  $n$ , expanding the framework in Bai (2004); thus, testing is feasible even in the presence of latent variables. The test statistic  $\widehat{D}$  is shown to be  $O_p(T^{-2})$  under the null. The test is shown to be powerful versus local alternatives of order  $O_p(T^{-1})$  and to diverge at a rate  $O_p(T^2)$  under global alternatives. Thus our test differs from those by e.g. Shin (1994) and Xiao (1999), where consistency versus global alternatives is achieved at a slower rate due to different rates of convergence of the estimators under the null and under the alter-

native and to the need to employ nonparametric estimates of the long run variances. **The advantage of our approach is that the micro relationships are cointegrated both under the null hypothesis (macro cointegration) and under the alternative (lack of macro cointegration). Thus, when constructing the test, the estimates that one employs are always (i.e. whether the null hypothesis of macro cointegration be true or not)  $T$ -consistent, which explains why under the alternative the test does not suffer from the slow rate of divergence found in the literature.** These results hold for any  $T$ -consistent estimator, and are robust to serial correlation and cross dependence. Since the test simply requires  $T$ -consistent estimators, OLS can be applied even under weak endogeneity. The asymptotic law of the test statistic is not nuisance free. Therefore, we propose a bootstrap approximation for the critical values based on the methods of sieves (see e.g. Chang, Park and Song, 2006). We prove the consistency of the procedure and Monte Carlo simulations provide evidence of good size and power properties of the testing framework. Last, though the focus of our paper is on the case of fixed  $n$ , we study the case of aggregation with large panels, i.e. for  $n \rightarrow \infty$ , investigating the conditions whereby cointegration is preserved after aggregation as  $n \rightarrow \infty$ . In this respect, our paper complements the analysis of Phillips and Moon (1999).

The paper is organized as follows. The theoretical framework is presented in Section 2, where we set up a model for heterogeneous panels, present the aggregate cointegration relationship and analyze the probabilistic structure of the ordinary least squares (OLS) estimates of the aggregate model. Section 3 presents the conditions for cointegration to carry through the aggregation process. We characterize the system's behavior when the conditions derived in the previous section are not satisfied and we develop an asymptotic theory for assessing the deviation from the case of aggregate cointegration. In Section 4 we propose a bootstrap approximation of the test. Monte Carlo simulations, reported in Section 5, evaluate size and power properties of the

bootstrap test. Section 6 concludes.

A word on notation: integrals of Brownian motions  $W(r)$  such as  $\int_0^1 W(r) dr$  are denoted as  $\int W$ ,  $\xrightarrow{p}$  denotes convergence in probability and  $\xrightarrow{d}$  denotes convergence in distribution;  $\|\cdot\|$  denotes the Euclidean norm, defined, for a matrix  $A$ , as  $\|A\| = \sqrt{\text{tr}(A'A)}$ .

## 2 Asymptotics for the Aggregate Relationship

Let us consider a system of  $n$  cointegrated micro relationships each with  $p$  explanatory variables:

$$y_{it} = \sum_{h=1}^p \beta_{hi} x_{hit} + u_{it}, \quad (1)$$

where  $t = 1, \dots, T$ , and  $i = 1, \dots, n$ . The covariates  $x_{hit}$  are I(1) processes that share  $k$  common stochastic trends:

$$x_{hit} = \alpha'_{hi} z_t + v_{hit}, \quad (2)$$

with  $z_t = [z_{1t}, \dots, z_{kt}]'$  a  $k$ -dimensional vector where

$$z_{jt} = z_{jt-1} + \epsilon_{jt},$$

with  $h = 1, \dots, p$ ,  $j = 1, \dots, k$ , and  $\alpha_{hi}$  is a  $k \times 1$  vector.

The model can also be rewritten in matrix form:

$$y_{it} = x'_{it} \beta_i + u_{it}, \quad (3)$$

$$x_{it} = \Gamma_i z_t + v_{it}, \quad (4)$$

$$z_t = z_{t-1} + \epsilon_t, \quad (5)$$

where  $x_{it} = [x_{1it}, \dots, x_{pit}]'$ ,  $\beta_i = (\beta_{1i}, \dots, \beta_{pi})'$  and  $\Gamma_i = [\alpha_{1i}, \dots, \alpha_{pi}]'$ . The matrices dimensions are respectively  $p \times 1$  and  $p \times k$ . The trend vector is assumed to initiate



at  $z_0 = 0$ .

Let  $u_t = [u_{1t}, \dots, u_{nt}]'$ ,  $v_t = [v'_{1t}, \dots, v'_{nt}]$ ,  $\varepsilon_t = [u'_t, v'_t, \epsilon'_t]'$ . We assume that the sequence of innovations satisfies the following assumption:

**Assumption 1**

(i) a functional central limit theorem (FCLT) holds for the partial sums of  $\varepsilon_t$ ,

$$S_t = \sum_{l=1}^t \varepsilon_l;$$

(ii)  $\varepsilon_t$  is independent of  $[u'_t, v'_t]'$ , and the trends  $z_t$  have a unit long-run variance,

$$\lim_{T \rightarrow \infty} \text{Var}(T^{-\frac{1}{2}} \sum_{t=1}^T \varepsilon_t) = I_k.$$

**Remarks**

**Assumption 1** summarizes the requirements on the behaviour of the error term  $\varepsilon_t$ . These are very general, and basically only require that a FCLT holds; some of the implications of the Assumption are discussed hereafter.

**A1.1** Assumption 1(i) allows  $\varepsilon_t$  to belong to a very general class of processes, among which linear processes are just a special case. In particular, time dependence is allowed for the process  $\varepsilon_t$  as long as it decays at an appropriate rate. Also, the covariance structure of  $\varepsilon_t$  could be time-varying, as long as the FCLT holds.

**A1.2** The orthonormality requirement in Assumption 1(ii) makes the trends  $z_{it}$  neutral in the model so that the behavior of the system is fully described by the coefficients  $\beta_{hi}$  and  $\alpha_{hi}$ . Therefore, the long run variance of the  $x_{it}$ s,  $\lim_{T \rightarrow \infty} T^{-1} E(x_{it} x'_{it})$ , is given by  $\Gamma_i \Gamma'_i$ . Note that Assumption 1 ensures that for  $r = [0, 1]$ ,  $T^{-1/2} \sum_{t=1}^{[Tr]} \varepsilon_t \xrightarrow{d} W_z(r)$ , where  $W_z(\cdot)$  is the  $k$ -dimensional standard Brownian motion.

**A1.3** Assumption 1 does not make any requirement as to the existence and extent of cross sectional dependence. **This is allowed**, given that our analysis is conducted for fixed  $n$ , which makes the cross-sectional dimension (and its features)

irrelevant for the purpose of asymptotic theory. **Therefore, Assumption 1 in principle allows for arbitrary contemporaneous and dynamic correlation across units. This also includes the presence of a factor structure in  $u_{it}$  and  $v_{it}$ , as long as an invariance principle holds for the partial sums of  $\varepsilon_t$ . The presence and extent of cross dependence are also discussed further in Sections 3.3.2 and 4.** Also, we do not need any restriction on the correlation between  $u_t$  and  $v_t$ , and therefore we do not need to impose weak exogeneity in the cointegration equation (3).

**A1.4** A restriction contained in Assumption 1 is that the idiosyncratic shock  $\varepsilon_t$  be statistically independent of  $\{u_{it}, v_{it}\}$ . This assumption, which is similar e.g. to Assumption D in Bai (2004), rules out a dynamic factor representation, whereby  $\{y_{it}, x_{it}\}$  would depend on  $z_t$  and lagged values of  $z_t$ . As a consequence, the structure in (3)-(5) could be described as a "static factor model" with common factors  $z_t$ . Assumption 1(ii) could be relaxed, since the main results of the paper (e.g. rate of convergence of the test statistic, power under local alternatives, etc...) hold irrespective of it. We discuss the possibility of allowing for dependence between  $\varepsilon_t$  and  $\{u_{it}, v_{it}\}$  after Proposition 3, thereby obtaining a "dynamic factor model".

## Assumption 2

(i) *the number of regressors in the cointegration equation (3),  $p$ , is not larger than the number of common trends  $k$ , i.e.  $p \leq k$ . Also,  $\text{rank}(\Gamma_i) = p$ , for  $i = 1, \dots, n$ .*

(ii) *for  $\Gamma = \sum_{i=1}^n \Gamma_i$ ,  $\text{rank}(\Gamma) = \min\{p, k\} = p$ .*

(iii)  *$k \leq n(p + 1)$ .*

(iv) *for all  $i$ , it holds that  $\|\beta_i\| < \infty$  and  $\|\Gamma_i\| < \infty$ .*

## Remarks

**Assumption 2 refers to the model representation.**

**A2.1** The lower bound on  $k$  in Assumption 2(*i*) ensures that model (3)-(5) can embed both common and/or unit specific stochastic trends. A result that follows directly from this assumption is that the  $x_{it}$  in equation (4) do not cointegrate among themselves for all  $i$ . This is a standard assumption from cointegration analysis and it is necessary to rule out the degenerate cointegration case - see Phillips (1986) for discussion.

**A2.2** Assumption 2(*ii*) requires that also the sum of the  $\Gamma_i$ s must have full rank. This condition will prove useful in the analysis of the aggregate cointegration relationship properties.

**A2.3** The upper bound  $n(p+1)$  in Assumption 2(*iii*) is necessary for the estimation of the factors and it prevents the number of unit specific factors from being too large, even though it states that their number can grow linearly with the number of units. This assumption plays a role when the  $z_t$ s are not observable - see Section 3.3.1 below. In such case, the coefficients  $\beta_i$  and  $\Gamma_i$  are estimated applying the Principal Component estimator to the  $n(p+1)$ -dimensional panel  $\{y_{it}, x_{it}\}_{i=1}^n$ . Assumption 2(*iii*) simply states that the number of factors  $z_t$  in  $\{y_{it}, x_{it}\}_{i=1}^n$  does not exceed the number of units in the panel.

**A2.4** The integrability conditions on  $\beta_i$  and  $\Gamma_i$  in Assumption 2(*iv*) are standard in the literature when the  $z_t$ s are not observable and estimation of  $\beta_i$  and  $\Gamma_i$  is required - see e.g. Assumption B in Bai (2004).

## 2.1 The Aggregate Cointegration Relationship

Aggregation of equation (2) across units leads to the equation

$$\bar{x}_{ht} = \sum_{j=1}^k a_{hj} z_{jt} + \bar{v}_{ht},$$

where  $h = 1, \dots, p$ ;  $t = 1, \dots, T$ ;  $\bar{x}_{ht} = \sum_{i=1}^n x_{hit}$ ,  $a_{hj} = \sum_{i=1}^n \alpha_{hi,j}$  with  $\alpha_{hi,j}$  being the  $j$ -th element in vector  $\alpha_{hi}$  and  $\bar{v}_{ht} = \sum_{i=1}^n v_{hit}$ . We assume there is at least one  $j$  for which  $a_{hj} \neq 0$ , so that  $\bar{x}_{ht}$  is  $I(1)$ .

For the dependent variable, cross sectional aggregation of equation (1) gives equation

$$\bar{y}_t = \sum_{j=1}^k b_j z_{jt} + \bar{s}_t,$$

where  $t = 1, \dots, T$ ,  $\bar{y}_t = \sum_{i=1}^n y_{it}$ ,  $b_j = \sum_{h=1}^p \sum_{j=1}^n \beta_{hi} \alpha_{hi,j}$  and  $\bar{s}_t = \sum_{h=1}^p \sum_{i=1}^n \beta_{hi} v_{hit} + \sum_{i=1}^n u_{it}$ . We assume there is at least one  $j$  for which  $b_j \neq 0$ , so that  $\bar{y}_t$  contains a unit root.

Let now  $\bar{x}_t = [\bar{x}_{1t}, \bar{x}_{2t}, \dots, \bar{x}_{pt}]'$  and  $b = \sum_{i=1}^n \Gamma'_i \beta_i$ . The aggregate forms of (3) and (4) can be written in vector form as

$$\bar{x}_t = \Gamma z_t + \bar{v}_t \tag{6}$$

$$\bar{y}_t = b' z_t + \bar{s}_t \tag{7}$$

where  $t = 1, \dots, T$ .

## 2.2 Asymptotics for $\hat{\beta}$

With respect to the aggregate relationship, let us consider the least-squares estimator  $\hat{\beta}$  of the slope coefficient in the linear regression of  $\bar{y}_t$  on  $\bar{x}_t$

$$\hat{\beta} = \left( \sum_{t=1}^T \bar{x}_t \bar{x}_t' \right)^{-1} \left( \sum_{t=1}^T \bar{x}_t \bar{y}_t \right).$$

We are going to evaluate the case of  $T$  large and  $n$  finite, and the case of  $T$  and  $n$  large.

### 2.2.1 The Case of $T$ Large and $n$ Finite.

In this case, when  $\bar{y}_t$  and  $\bar{x}_t$  are cointegrated, the estimator  $\hat{\beta}$  is superconsistent and converges in probability to a vector which is the true value of the aggregation coefficient,  $\beta$ . On the other hand, if the aggregate series are not cointegrated, the regression  $\bar{y}_t = \hat{\beta}'\bar{x}_t + \hat{e}_t$  is spurious and  $\hat{\beta}$  converges in distribution to a non-degenerate vector random variable.

The following proposition characterizes the limiting distribution of the estimator  $\hat{\beta}$  for large  $T$  and finite  $n$ .

**Proposition 1** *Let Assumptions 1(i) and 2(i) hold. Then, in the OLS regression of  $\bar{y}_t$  on  $\bar{x}_t$ ,  $\hat{\beta}$  converges to a non degenerate random variable  $S$ ,*

$$\hat{\beta} \xrightarrow{d} S = \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' b \right]. \quad (8)$$

**Proof.** From equations (6) and (7) and standard asymptotic results, it follows that

$$\hat{\beta} = \left[ \Gamma \sum_{t=1}^T z_t z_t' \Gamma' + o_p(1) \right]^{-1} \left[ \Gamma \sum_{t=1}^T z_t z_t' b + o_p(1) \right].$$

In addition, Assumption 1 ensures that  $T^{-2} \sum_{t=1}^T z_t z_t' \xrightarrow{d} \int W_z W_z'$ . ■

For further details, see also Park and Phillips (1988). Note that the only requirement needed for Proposition 1 to hold is that the FCLT holds for  $\varepsilon_t$ ; thus, (8) is valid for any degree of correlation (weak exogeneity and endogeneity) between  $x_{it}$  and  $u_{it}$  (and therefore between  $\bar{x}_t$  and  $e_t$ ) and also between  $\varepsilon_t$  and  $\{u_{it}, v_{it}\}$ .

As pointed out above in commenting Assumption 1, the presence of contemporaneous correlation among the panel units is not ruled out in our model. The use of OLS is a valid choice under any arbitrary level of cross sectional dependence. This is due to the fact that  $n$  is finite and therefore cross sectional dependence is neutralized by aggregation. Assumptions 2(i) and 2(ii) are needed for the  $p \times p$  term  $\Gamma \int W_z W_z' \Gamma'$  to be a nondegenerate Brownian motion - see a related discussion by Phillips (1986). Since  $p \leq k$  and  $\Gamma$  is a full rank matrix, it holds that the matrix

$\Gamma \int W_z W_z' \Gamma'$  is almost surely positive definite and the inverse  $[\Gamma \int W_z W_z' \Gamma']^{-1}$  exists almost surely. Thus, assumption 2(ii) requires that not only the individual  $x_{it}$ s, but also their aggregate  $\bar{x}_t$  does not cointegrate.

Note that Equations (1) and (2) could be extended to incorporate deterministic terms, such as constant terms

$$\begin{aligned} y_{it} &= a_{yi} + x'_{it} \beta_i + u_{it}, \\ x_{it} &= a_{xi} + \Gamma_i z_t + v_{it}. \end{aligned}$$

This would result in the aggregate relationships having a constant term as well, i.e.

$$\begin{aligned} \bar{x}_t &= \bar{a}_x + \Gamma z_t + \bar{v}_t \\ \bar{y}_t &= \bar{a}_y + \bar{a}_{xy} + b' z_t + \bar{s}_t, \end{aligned}$$

where  $\bar{a}_x = \sum_{i=1}^n a_{xi}$ ,  $\bar{a}_y = \sum_{i=1}^n a_{yi}$  and  $\bar{a}_{xy} = \sum_{i=1}^n a_{xi} a_{yi}$ . In this case, standard cointegration theory entails that Proposition 1 still holds. If a deterministic term is considered in the aggregate cointegration relationship, such as  $\bar{y}_t = \hat{\alpha} + \hat{\beta}' \bar{x}_t + \hat{e}_t$ , then (8) should be modified as

$$\hat{\beta} \xrightarrow{d} S = \left[ \Gamma \int \bar{W}_z \bar{W}_z' \Gamma' \right]^{-1} \left[ \Gamma \int \bar{W}_z \bar{W}_z' b \right],$$

where  $\bar{W}_z$  is the demeaned Brownian motion associated to the  $z_t$ s, i.e.  $\bar{W}_z(r) = W_z(r) - \int_0^1 W_z(s) ds$ .

Proposition 1 is valid for large  $T$  and finite  $n$ . In the next section we present the case of when both  $T$  and  $n$  are large.

### 2.2.2 The Case of $T$ and $n$ Large.

Though our paper is focused on the fixed  $n$  case, it is interesting to study the case  $(n, T) \rightarrow \infty$  to see where our framework fits within the large panels asymptotics

developed by Kao (1999) and Phillips and Moon (1999). Granger (1990) discusses the consequences of  $n$  being large and Granger (1993) provides an interesting characterization of  $n$  being large or small. The following proposition holds when  $T$  and  $n$  are large.<sup>1</sup>

**Proposition 2** *Let Assumptions 1 and 2(i)-(ii)-(iv) hold, and let the regression coefficients  $\beta_i$  and  $\Gamma_i$  be i.i.d. random variables across  $i$ , independent of  $\varepsilon_t$ . Let  $E(\beta_i) = \bar{\beta}$ ,  $E(\Gamma_i) = \bar{\Gamma}$  and  $E(\Gamma_i' \beta_i) = \bar{\Gamma}'(\bar{\beta} + c)$  with  $c$  a  $p \times 1$  vector. Then, as  $(n, T) \rightarrow \infty$*

$$\hat{\beta} \xrightarrow{p} \bar{\beta} + c. \quad (9)$$

**Proof.** See Appendix. ■

Proposition 2 states that, as  $(n, T) \rightarrow \infty$ ,  $\hat{\beta}$  is a consistent estimator of the long run average  $\bar{\beta}$  if and only if the  $\beta_i$ s and the  $\Gamma_i$ s are uncorrelated, i.e. if  $E[(\Gamma_i - \bar{\Gamma})'(\beta_i - \bar{\beta})] = E(\Gamma_i' \beta_i) - \bar{\Gamma}'\bar{\beta} = \bar{\Gamma}'c = 0$ . In this case, as reported in the proof, consistency is achieved at a rate  $\sqrt{n}$ , a finding in line with the large panel literature when units are described by a spurious regression - see Kao (1999) and Phillips and Moon (1999). Thus, the OLS estimate picks the long-run average relationship between  $\bar{y}$  and each of the  $\bar{x}_h$ s, regardless of the existence of a cointegration relationship. When the  $\beta_i$ s and the  $\Gamma_i$ s are correlated such that  $\bar{\Gamma}'c \neq 0$ , then  $\hat{\beta}$  is inconsistent.

Proposition 2 is valid for any degree of contemporaneous correlation, and therefore the presence of e.g. a factor structure in  $u_{it}$  and/or  $v_{it}$  is allowed for. However, statistical independence between  $\varepsilon_t$  and  $\{u_{it}, v_{it}\}$  as in Assumption 1(ii) is needed - see also the comments in remark A1.4. Note that under the more restrictive assumption of no cross-sectional dependence among units, the OLS estimator  $\hat{\beta}$  is asymptotically equivalent to the pooled-OLS estimator in Phillips and Moon (1999).

As a final remark, equation (9) has been proved using a joint limit argument. Phillips and Moon (1999) provide joint limit theory for panels with independent

---

<sup>1</sup>We wish to thank a referee whose question led to this result.

units. In this paper, instead, we deal with strong cross sectional dependence across units. Our joint limits are obtained by using a cross-sectional CLT for martingale difference sequences (MDS), using the approach developed by Kao, Trapani and Urga (2008). As in Phillips and Moon (1999), no restrictions are required on the rate of expansion between  $n$  and  $T$  as they approach infinity when proving consistency.

### 3 Aggregate Cointegration: Validity and Testing

Proposition 2 states that, for large  $n$ , the consistency (or lack thereof) of  $\widehat{\beta}$  does not depend on the existence of cointegration in the aggregate relationship. We henceforth restrict our analysis to the case of large  $T$  and finite  $n$  only. We develop an estimation theory for both aggregate and disaggregate models. We first discuss the formal requirements under which cointegration holds in the aggregate relationship  $\bar{y}_t = \widehat{\beta}' \bar{x}_t + \widehat{e}_t$ , laying out a necessary and sufficient condition in order for cointegration to be maintained after aggregation. Second, we explore the consequences of a failure of this condition to hold though cointegration can still be present in the data.

#### 3.1 Cointegration in the Aggregate Relationship

The results in this section are based on superconsistency of the OLS estimates when cointegration is present. In this case,  $\widehat{\beta} \xrightarrow{p} \beta$ . In order to have aggregate cointegration,  $S$  in equation (8) must degenerate to a vector of constants rather than a vector of random variables. Given that  $b \neq 0$  by assumption, this means that

$$\Gamma' \beta = b. \tag{10}$$



In this case,

$$\begin{aligned}
S &= \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' b \right] = \\
&= \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' \Gamma' \right] \beta = \beta,
\end{aligned} \tag{11}$$

and cointegration holds. Another consequence of superconsistency is that the linear system (10) has a unique solution

$$\beta = (\Gamma \Gamma')^{-1} \Gamma b. \tag{12}$$

Note that  $b$  must be a linear combination of the rows of  $\Gamma$  for equation (10) to admit non trivial solutions, and this holds if and only if  $\text{rank}(\Gamma') = \text{rank}(\Gamma' | b) = p$ .

Thus, the following results hold:

**Theorem 1** *Let Assumptions 1(i) and 2(i)-(ii) hold. Cointegration in the aggregate relationship  $\bar{y}_t = \hat{\beta}' \bar{x}_t + \hat{e}_t$  holds if and only if  $\text{rank}(\Gamma' | b) = p$ .*

**Corollary 1** *Let Assumptions 1(i) and 2(i)-(ii) hold. If the number of regressors in the cointegration equations (1) equals the number of stochastic trends (i.e. if  $p = k$ ), then the aggregate relationship  $\bar{y}_t = \hat{\beta}' \bar{x}_t + \hat{e}_t$  is cointegrated.*

When the number of common stochastic trends is limited, i.e. when the amount of cointegration in the single units is large enough, then aggregation does not have a completely destructive effect on cointegration in the aggregate relationship. It should be noted that when the number of common trends  $k$  is large with respect to the number of covariates  $p$ ,  $\text{rank}(\Gamma' | b)$  is more likely to be equal to  $p + 1$ , and hence aggregated cointegration is unlikely to hold.

Theorem 1 always holds when  $\Gamma$  is a  $k \times k$  matrix. Assumption 2(ii) ensures that  $\text{rank}(\Gamma) = k$  and therefore  $\text{rank}(\Gamma' | b) = k$  as well. Corollary 1 is an alternative

formulation of Theorem 1 in Gonzalo (1993) when the common trends in the disaggregate system are the same across all  $is$ . Note that Theorem 1 contains a different formulation of the conditions for aggregate cointegration with respect to Lemma 1 in Hsiao, Shen and Fujiki (2005). The difference is due to the presence of common stochastic trends in the DGP of the  $x_{it}$ s in (4), which is not assumed in Hsiao et al. (2005).

### 3.2 Measuring Departure from Cointegration

When the formal conditions for aggregate cointegration are violated, we can still have "some degree of cointegration" in the aggregate relationship if the requirements in Theorem 1 are only "mildly violated", as pointed out by Granger (1993). In what follows, we derive a statistical measure of departure from cointegration when Theorem 1 does not hold, and therefore, strictly speaking, equation  $\bar{y}_t = \hat{\beta}'\bar{x}_t + \hat{e}_t$  represents a spurious relationship. The testing framework we derive is based on

$H_0$  : presence of aggregate cointegration,

$H_A$  : spurious aggregate regression.

A natural way to address the issue of testing is to consider the statistical properties of the limiting distribution of  $\hat{\beta}$ ,  $S$ . From equation (11), we know that

$$S = \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' b \right].$$

Denote  $P = I_k - \Gamma' (\Gamma \Gamma')^{-1} \Gamma$  and  $M = \Gamma' (\Gamma \Gamma')^{-1} \Gamma$ , and writing  $b = Mb + Pb$ , equation (11) becomes

$$S = \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' Mb \right] + \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' Pb \right] \quad (13)$$

or

$$S = \beta + \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' P b \right], \quad (14)$$

using (12). To analyse the second term of the right hand side of (14), define  $W^\Gamma(r) = \Gamma W_z(r)$  and  $W^P(r) = b' P W_z(r)$ . By construction, we have

$$E [W^\Gamma W^{P'}] = \Gamma E [W_z(r) W_z'(r)] P b = \Gamma (r I_k) P b = 0.$$

Thus,  $W^\Gamma(r)$  and  $W^P(r)$  are independent. Hence the expected value of the random variable  $S$  is

$$E(S) = \beta,$$

and the variance of  $S$  is equal to

$$Var(S) = Var \left\{ \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' P b \right] \right\}.$$

Therefore, we have aggregate cointegration if the second term on the right hand side of (14),  $\left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' P b \right]$ , degenerates to a zero constant, i.e.

$$\left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' P b \right] = 0 \quad a.s.$$

This holds if and only if  $P b = 0$ , which implies that  $Var(S) = 0$  if we have aggregate cointegration, while  $Var(S) > 0$  if the aggregated relationship is not cointegrated.

Thus, for testing purposes, we can define the following indicator:

$$D = \frac{b' P b}{b' b}. \quad (15)$$

Under the null hypothesis of cointegration in the aggregate relationship  $D = 0$ , whilst  $D > 0$  under the alternative hypothesis that aggregation eliminates cointegration. Note that, given that  $M$  and  $P$  are idempotent, (15) can be rewritten

as

$$D = \sin^2(b, Mb). \quad (16)$$

From (16), the indicator  $D$  depends on the angle between the two vectors  $b$  and  $Mb$ . The smaller the angle between the two vectors, the smaller the distance from the case of aggregate cointegration. The aggregate cointegration occurs when the two vectors  $b$  and  $Mb$  are parallel. This condition is met when  $b$ , which gives the response of  $\bar{y}_t$  to the stochastic trends  $z_t$ , can be fully represented in terms of the basis associated to the column space of  $\Gamma$ , which represents the response of  $\bar{x}_t$  to the common stochastic trends. Algebraically, this means that we have cointegration when  $b$  is a linear combination of the columns of  $\Gamma$ .

The definition of  $D$  illustrates possible sources of the violation of the necessary and sufficient condition for cointegration in the aggregate relationship. When  $\text{rank}(\Gamma' | b) > p$ , cointegration is not preserved under aggregation. Nonetheless, if the stochastic trends in the error term in the aggregate relationship are relatively unimportant then  $\text{Var}(S)$  is small and the degree of departure from aggregate cointegration is not large.

### 3.3 Testing for Cointegration

The hypotheses of interest are as follows

$$\begin{aligned} H_0 : D = 0 \\ H_1 : D > 0 \end{aligned}, \quad (17)$$

where the null hypothesis  $H_0$  is the presence of cointegration in the aggregate relationship. To test the null hypothesis in (17),  $b$  and  $\Gamma$  need to be estimated.

#### 3.3.1 Estimation of $b$ and $\Gamma$

The estimation of  $b$  and  $\Gamma$  depends crucially on whether the  $z_t$ s are observable or unobservable. When the  $z_t$ s are observable, estimates of  $b$  and  $\Gamma$  can be obtained by

OLS and are given by

$$\hat{b}^{OLS} = \left( \sum_{t=1}^T z_t z_t' \right)^{-1} \left( \sum_{t=1}^T z_t \bar{y}_t \right) \quad (18)$$

$$\hat{\Gamma}^{OLS} = \left( \sum_{t=1}^T z_t z_t' \right)^{-1} \left( \sum_{t=1}^T z_t \bar{x}_t' \right). \quad (19)$$

Since equations (6) and (7) are cointegrating relationship, OLS estimators in (18)-(19) are superconsistent, i.e. letting  $\Theta = [b|\Gamma]'$  we have

$$\hat{\Theta}^{OLS} - \Theta = O_p(T^{-1}).$$

In the more likely case that the common trends  $z_t$  are not observable, another approach should be considered. Let us express model (3)-(4) as

$$y_{it} = \beta_i' \Gamma_i z_t + \beta_i v_{it} + u_{it}$$

$$x_{it} = \Gamma_i z_t + v_{it}.$$

Writing  $W_{it} = \begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix}$ ,  $\Xi_i = \begin{pmatrix} \beta_i' \Gamma_i \\ \Gamma_i \end{pmatrix}$ , and  $e_{it}^W = \begin{bmatrix} \beta_i v_{it} + u_{it} \\ v_{it} \end{bmatrix}$ ,

we have

$$W_{it} = \Xi_i z_t + e_{it}^W,$$

and by stacking  $W_{it}$ , it holds

$$W_t = \begin{bmatrix} W_{1t} \\ W_{2t} \\ \cdot \\ W_{nt} \end{bmatrix} = \begin{bmatrix} \Xi_1 \\ \Xi_2 \\ \cdot \\ \Xi_n \end{bmatrix} z_t + \begin{bmatrix} e_{1t}^W \\ e_{2t}^W \\ \cdot \\ e_{nt}^W \end{bmatrix} = \Xi z_t + e_t^W. \quad (20)$$

Consistent estimator of  $\Xi$  can be obtained by principal component. More specifically, consider the  $n(p+1) \times n(p+1)$  matrix  $\sum_{t=1}^T W_t W_t'$ . The principal component

estimator of  $\Xi$ , say  $\hat{\Xi}^{PC}$ , is given by  $\sqrt{n}$  times the  $k$  eigenvectors corresponding to the largest eigenvalues of  $\sum_{t=1}^T W_t W_t'$  subject to the normalization

$$\hat{\Xi}^{PC'} \sum_{t=1}^T W_t W_t' \hat{\Xi}^{PC} = nT^2 I_k.$$

The procedure we propose is based on Bai (2004) but extended to our case of  $n$  finite and  $T$  large. It is also known that  $\Xi_i$  and  $z_t$  are not directly identifiable but they are identifiable up to a transformation defined by a rotation matrix  $H$ . For our setup, knowing  $\Xi_i H$  is as good as knowing  $\Xi_i$ , since the test statistic  $D$  does not depend on  $H$ . For the purpose of notational simplicity, we assume  $H$  being an identity matrix in this paper. The following proposition ensures consistency of the estimates  $\hat{\Xi}^{PC}$ .

**Proposition 3** *Let Assumptions 1(i)-(ii) and 2 hold. Further, assume that  $E \|\epsilon_t\|^{4+\delta} < \infty$  for some  $\delta > 0$ ,  $E |u_{it}|^8 < \infty$ ,  $E \|v_{it}\|^8 < \infty$  and  $\sum_{t=1}^T E \|e_{it}^W e_{js}^W\| < \infty$  for all  $(i, j, s)$ . Then, as  $T \rightarrow \infty$ ,*

$$\hat{\Xi}^{PC} - \Xi = O_p(T^{-1}). \quad (21)$$

**Proof.** See Appendix. ■

Proposition 3 states that the "loadings"  $\Xi$  in (20) can be estimated consistently even for fixed  $n$ . This result has been derived by Bai (2004) for the case of nonstationary panel factor models, and it is in contrast with the stationary case whereby the order of magnitude of the loadings estimation error is  $\min(1/n, 1/\sqrt{T})$  and therefore both  $n$  and  $T$  need to be large - see Theorem 2 in Bai (2003).

**As pointed out in Remark A1.4, assuming that  $\epsilon_t$  and  $\{u_{it}, v_{it}\}$  are independent is not strictly necessary for (21) to hold, even though it simplifies the asymptotics of  $\hat{\Xi}^{PC}$ .** If the idiosyncratic shock  $\epsilon_t$  in the DGP of the  $z_t$ s were allowed to be correlated with  $\{u_{it}, v_{it}\}$ , then a static factor model like (20) would no longer be an adequate representation for  $W_t$  and (20) would have to be modified as a dynamic factor model in order to allow  $W_t$  to depend upon  $z_t$  and lagged values thereof. However, as proved in Theorem 6 in Bai (2004), the loadings

$\Xi_i$  could still be estimated consistently, even though their asymptotic law would be different to the static factor case. We refer to Bai (2004) for a comprehensive discussion of the asymptotics for nonstationary static factor models.

The principal component estimator of  $\Theta$  is given by

$$\hat{\Theta}^{PC} = \begin{pmatrix} \hat{b}^{PC} \\ \hat{\Gamma}^{PC} \end{pmatrix} = \sum_{i=1}^n \hat{\Xi}_i^{PC}. \quad (22)$$

and from Proposition 3

$$\hat{\Theta}^{PC} - \Theta = O_p(T^{-1}).$$

Therefore, even when the  $z_{it}$ s are unobservable, we have a  $T$ -consistent estimate for  $\Theta$ .

Henceforth, we shall also use the following matrix notation. Defining the  $[n(p+1)] \times (p+1)$  matrix  $F$  by stacking  $n(p+1)$ -dimensional identity matrices, i.e. as

$$F = [I_{p+1}, \dots, I_{p+1}]', \quad (23)$$

$\hat{\Theta}^{PC}$  can also be defined as  $\hat{\Theta}^{PC} = F' \hat{\Xi}^{PC}$ . Letting the  $(p+1)$ -dimensional vector  $i_b = [1, 0, \dots, 0]'$  and the  $p \times (p+1)$  matrix  $i_\Gamma = [0 | I_p]$  we also have  $\hat{b}^{PC} = \hat{\Xi}^{PC'} F i_b$  and  $\hat{\Gamma}^{PC} = i_\Gamma F' \hat{\Xi}^{PC}$ .

### 3.3.2 Testing

Let

$$\hat{D} = \frac{\hat{b}' \hat{P} \hat{b}}{\hat{b}' \hat{b}}$$

where  $\hat{P} = I_k - \hat{\Gamma}' (\hat{\Gamma} \hat{\Gamma}')^{-1} \hat{\Gamma}$  and  $\hat{b}$ ,  $\hat{\Gamma}$  are estimators of  $b$ ,  $\Gamma$ . The following theorem characterizes the rate of convergence of  $\hat{D}$  under the null hypothesis of cointegration.

**Theorem 2** *Let Assumptions 1 and 2 hold, and assume that  $\hat{b}$  and  $\hat{\Gamma}$  are  $T$ -consistent estimators of  $b$  and  $\Gamma$ . Under the null hypothesis of cointegration, we*

have  $D = 0$  and

$$\widehat{D} = O_p(T^{-2}). \quad (24)$$

**Proof.** See Appendix. ■

Theorem 2 asserts that rate of convergence of  $\widehat{D}$  is of order  $T^2$  irrespective of whether the  $z_t$ s are observable or not and of the type of estimation technique employed to derive  $\widehat{b}$  and  $\widehat{\Gamma}$ , as long as they are  $T$ -consistent estimators of  $b$  and  $\Gamma$ , e.g. the OLS or the Principal Component estimators. This result is reinforced and generalised by the following corollary.

**Corollary 2** *Under the assumptions of Theorem 2, let  $\delta > 0$  and consider two estimators  $\widehat{b}_\delta$  and  $\widehat{\Gamma}_\delta$  such that  $\widehat{b} - b = O_p(T^{-\delta})$  and  $\widehat{\Gamma} - \Gamma = O_p(T^{-\delta})$ . Then, under the null hypothesis of cointegration, we have  $D = 0$  and*

$$\widehat{D} = O_p(T^{-2\delta}).$$

**Proof.** See Appendix. ■

From Theorem 2 and Corollary 2, the rate of convergence of  $\widehat{D}$  is the square power of the rate of convergence of the estimators  $\widehat{b}$  and  $\widehat{\Gamma}$ . The intuition behind this result is that under the null, the angle between  $b$  and  $Mb$  is equal to zero. Recalling the definition of  $D$  in (16), the function  $\sin^2(\cdot)$  is an even function in a neighborhood of zero, i.e. the odd powers of its Taylor expansion have coefficients equal to zero.

**Theorem 2 and Corollary 2 hold under a broad set of circumstances, as pointed out in Remarks A1.3 and A1.4. The rate of convergence of the test statistic under the null is preserved as long as the idiosyncratic component  $\varepsilon_t$  satisfies an FCLT. Of course, this is not the case with the limiting distribution of the statistic, which does depend upon the features of  $\varepsilon_t$ . The results below illustrate this.**

When the  $z_t$ s are observable, the limiting distribution of  $\widehat{D}$  is given in the following theorem.



**Theorem 3** *Let Assumptions 1(i) and 2 hold and consider the OLS estimators of  $b$ ,  $\Gamma$ , say  $\hat{b}^{OLS}$  and  $\hat{\Gamma}^{OLS}$  respectively, defined in (18)-(19). Under the null of aggregate cointegration*

$$T^2 \hat{D} \xrightarrow{d} \frac{1}{\|b\|^2} Q' \left[ I_k - \frac{bb'}{\|b\|^2} \right] Q, \quad (25)$$

where

$$Q = (M - I_k) Q_b + \left[ \Gamma' (\Gamma \Gamma')^{-1} Q_\Gamma - \Gamma' (\Gamma \Gamma')^{-1} \Gamma Q_\Gamma' (\Gamma \Gamma')^{-1} \Gamma - \Gamma' (\Gamma \Gamma')^{-1} Q_\Gamma \Gamma' (\Gamma \Gamma')^{-1} \Gamma + Q_\Gamma' (\Gamma \Gamma')^{-1} \Gamma \right] b,$$

and  $Q_b = (\int W_z W_z')^{-1} \int W_z dW_{\bar{s}}$ ,  $Q_\Gamma = (\int W_z W_z')^{-1} (\int W_z dW_{\bar{v}}')$ , with  $W_{\bar{v}}$  and  $W_{\bar{s}}$  Brownian motion processes associated with the partial sums of the processes  $\bar{v}_t$  and  $\bar{s}_t$  in (6) and (7) respectively.

**Proof.** See Appendix. ■

The following theorem gives the limiting distribution of  $\hat{D}$  when  $z_t$ s are not observable.

**Theorem 4** *Let the assumptions of Proposition 3 hold, and consider the PC estimators of  $b$ ,  $\Gamma$ , say  $\hat{b}^{PC}$  and  $\hat{\Gamma}^{PC}$  respectively. Under the null of aggregate cointegration*

$$T^2 \hat{D} \xrightarrow{d} \frac{1}{\|b\|^2} Q^{pc'} \left[ I_k - \frac{bb'}{\|b\|^2} \right] Q^{pc}, \quad (26)$$

where

$$Q^{pc} = (M - I_k) Q_b^{pc} + \left[ \Gamma' (\Gamma \Gamma')^{-1} Q_\Gamma^{pc} - \Gamma' \Gamma Q_\Gamma^{pc'} \Gamma - \Gamma' Q_\Gamma^{pc} \Gamma' \Gamma + Q_\Gamma^{pc'} (\Gamma \Gamma')^{-1} \Gamma \right] b,$$

$Q_b^{pc} = \Pi' i_b$ ,  $Q_\Gamma^{pc} = i_\Gamma \Pi$ , and  $\Pi$  is the limiting distribution of  $\hat{\Theta}^{PC}$ , the principal component estimator of  $\Theta$  defined in (22), given in Proposition 4 below - see equation (27).

**Proof.** See Appendix. ■

The following proposition provides the limiting distribution of the principal component estimator of  $\Theta$ .

**Proposition 4** *Let  $W_e$  be the Wiener process associated to the partial sums of  $e_t^W$  in equation (20) and define  $\Omega_e = E(e_t^W e_t^{W'})$  and  $B = \int W_z W_z'$ . Then under the assumptions of Theorem 4*

$$\begin{aligned} T \left( \hat{\Theta}^{PC} - \Theta \right) &\xrightarrow{d} F' [I_{n(p+1)} - n^{-1} \Xi B \Xi'] \left( \int dW_e W_z' \right) B^{-1} \\ &- n^{-1} F' \Xi' \left( \int dW_e W_z' \right) \Xi' \\ &+ n^{-1} F' [I_{n(p+1)} - 2n^{-1} \Xi B \Xi'] \Omega_e \Xi \equiv \Pi \end{aligned} \quad (27)$$

**Proof.** See Appendix. ■

To evaluate the capability of our statistic to reject local alternatives, we consider the following sequence of local alternatives

$$H_1^l : b = \Gamma' \beta + \delta_T, \quad (28)$$

where the  $k$ -dimensional vector  $\delta_T$  is orthogonal to  $\Gamma$  and is chosen to be  $\lim_{T \rightarrow \infty} T \delta_T = \delta \neq 0$ . The orthogonality condition  $\delta_T' \Gamma = 0$  means that the response of  $\bar{y}_t$  to the stochastic trends  $z_t$  also contains a component  $\delta_T' z_t$  which cannot be explained in terms of the  $\bar{x}_t$ s, and therefore the possibility that  $\bar{y}_t$  and  $\bar{x}_t$  cointegrate is ruled out. Therefore, under the sequence of local alternatives  $H_1^l$ ,  $D > 0$ . The following theorem shows that the statistic  $\hat{D}$  has non-trivial power versus such a sequence of local alternatives.

**Theorem 5** *Let  $\hat{b}$  and  $\hat{\Gamma}$  be  $T$ -consistent estimators of  $b$  and  $\Gamma$  respectively. Under*

the alternative hypothesis  $H_1^l$ , we have

$$T^2 \widehat{D} \xrightarrow{d} \frac{1}{\|b_0\|^2} \left\{ \|\delta\|^2 + Q^{*'} \left[ I_k - \frac{b_0 b_0'}{\|b_0\|^2} \right] Q^* - 2\delta' Q^* \right\}, \quad (29)$$

where  $b_0 = \Gamma' \beta$ ,  $Q^*$  is equal to either  $Q$  or  $Q^{pc}$  depending on whether the  $z_t$ s are observable or unobservable. The definitions of  $Q$  and  $Q^{pc}$  are in Theorems 3 and 4. In either case,  $E[Q^*] = 0$ .

**Proof.** See Appendix. ■

Theorem 5 shows that the test has nontrivial power against local alternatives of order  $O(T^{-1})$ . This result too holds irrespective of whether the  $z_t$ s are observable or not as long as  $\hat{b}$  and  $\hat{\Gamma}$  are superconsistent estimators of  $b$  and  $\Gamma$ , such as e.g. the OLS or the PC estimators.

Finally, to evaluate the consistency of our test, we will study the asymptotic behaviour of  $T^2 \widehat{D}$  under the alternative hypothesis  $H_1 : D > 0$ . The following theorem shows that the test based on  $\widehat{D}$  is consistent against fixed alternatives.

**Theorem 6** *Let  $\hat{b}$ ,  $\hat{\Gamma}$  be  $T$ -consistent estimators of  $b$ ,  $\Gamma$ . Then under the alternative hypothesis  $H_1 : D > 0$  it holds that, as  $T \rightarrow \infty$*

$$\widehat{D} = D + O_p(T^{-1}), \quad (30)$$

and therefore, under  $H_1$ , the statistic  $T^2 \widehat{D} \xrightarrow{p} \infty$ .

**Proof.** See Appendix. ■

Theorem 6 shows that  $T^2 \widehat{D}$  diverges under the global alternative  $H_1$ . Consequently, the probability of rejecting the null hypothesis when the alternative  $H_1$

holds is asymptotically equal to one. This means that the test based on  $T^2\widehat{D}$  is consistent.

The rate of divergence of the test statistic is  $O_p(T^2)$ , thereby faster than existent tests for the null of cointegration based on the Lagrange Multiplier approach (see e.g. Shin, 1994, and Xiao, 1999). As already discussed in the introduction, the estimates upon which these tests are calculated have different asymptotics under the null and the alternative; in addition, non parametric estimation of long run variance is required. Our test statistic is constructed employing estimates of the parameters in the micro equations, which cointegrate under both the null and the alternative hypothesis, thereby having the same asymptotics in both cases. This ensures consistency at rate  $O_p(T^2)$ . Note that our results are robust to the cases of weak endogeneity in the micro equations, since the OLS estimator is  $T$ -consistent. **The theory developed in this section still holds if one employs other estimation techniques than OLS, although this would have implications on the limiting distribution of the test statistic.**

An ancillary result is that under  $H_1$ , when  $D$  is no longer equal to zero, the remainder term in the asymptotic expansion of  $\widehat{D}$  around  $D$  is no longer  $O_p(T^{-2})$ , but  $O_p(T^{-1})$ . An explanation of this result is that while the function  $\sin^2(\cdot)$  is an even function in a neighborhood of zero, this is not the case around other values of its argument, whence the presence of the term of order  $O_p(T^{-1})$  in the expansion of  $\widehat{D}$  around  $D \neq 0$ . Last, it could be proved, along the same lines as for Corollary 2, that when using  $T^\delta$ -consistent estimators  $\widehat{b}_\delta$  and  $\widehat{\Gamma}_\delta$ , under the alternative  $H_1 : D > 0$  it holds that  $\widehat{D} = D + O_p(T^{-\delta})$  and thus  $T^{2\delta}\widehat{D} \xrightarrow{p} \infty$ .

## 4 Bootstrap Approximation of Critical Values

In this section, we propose a bootstrap procedure to obtain critical values.

Since our model does not rule out the possibility of serial correlation in the error terms, we employ a procedure which is similar to the sieve bootstrap approach

employed by Chang, Park and Song (2006) for cointegrating regressions.

For the purposes of bootstrapping, we rewrite model (6)-(7) as follows

$$\begin{bmatrix} \bar{y}_t \\ \bar{x}_t \end{bmatrix} = \bar{W}_t = \Theta z_t + \bar{e}_t. \quad (31)$$

We propose the following bootstrap algorithm:

**Step 1.** (1.1) Estimate  $\Theta$  in equation (31) consistently, via OLS if  $z_t$ s are observable, or via principal component if  $z_t$ s are unobservable. We obtain  $\hat{\Theta} = \hat{\Theta}^{OLS}$  and  $\hat{\Theta} = \hat{\Theta}^{PC}$  respectively. Project the estimator of  $b$ ,  $\hat{b} = (\hat{b}^{OLS} \text{ or } \hat{b}^{PC})$  onto the column space of the estimated  $\Gamma$ ,  $\hat{\Gamma} = \hat{\Gamma}^{OLS}$  or  $\hat{\Gamma} = \hat{\Gamma}^{PC}$  respectively, obtaining  $\tilde{b} = (\tilde{b}^{OLS} \text{ or } \tilde{b}^{PC})$  defined as

$$\tilde{b} = \hat{\Gamma}' \left( \hat{\Gamma} \hat{\Gamma}' \right)^{-1} \hat{\Gamma} \hat{b}.$$

Let  $\tilde{\Theta}^{OLS} = [\tilde{b}^{OLS} | \hat{\Gamma}^{OLS}']'$  and  $\tilde{\Theta}^{PC} = [\tilde{b}^{PC} | \hat{\Gamma}^{PC}']'$ .

(1.2) Compute the residuals  $\hat{e}_t = \bar{W}_t - \hat{\Theta}^{OLS} z_t$  or  $\tilde{e}_t = \bar{W}_t - \hat{\Theta}^{PC} \hat{z}_t$ , where  $\hat{z}_t$  is the principal component estimator of  $z_t$ . Define  $\hat{w}_t = [\hat{e}_t', \Delta z_t']'$  and  $\tilde{w}_t = [\tilde{e}_t', \Delta \hat{z}_t']'$ .

(1.3) Compute the statistics  $\hat{D}$  as

$$\hat{D} = \frac{\hat{b}' \hat{P} \hat{b}}{\hat{b}' \hat{b}}.$$

**Step 2.** (2.1) Sieve estimation. For the case observable  $z_t$ s, compute the sieve estimates of the VAR

$$\hat{w}_t = \sum_{l=1}^q \Psi_l \hat{w}_{t-l} + \eta_{qt} \quad (32)$$

where, following Chang, Park and Song (2006), the choice of  $q$  can be done via an information criterion such as AIC or BIC. Let  $\hat{\Psi}_l$  and  $\hat{\eta}_{qt}$  denote the OLS estimates and residuals from equation (32), respectively.

(2.2) Resampling. Draw (with replacement)  $T$  values from the centered residuals

$$\left\{ \hat{\eta}_{qt} - \frac{1}{T} \sum_{t=1}^T \hat{\eta}_{qt} \right\}_{t=1}^T$$

to obtain  $\{\eta_{qt}^*\}_{t=1}^T$ .

(2.3) Construct recursively  $\hat{w}_t^*$  as

$$\hat{w}_t^* = \sum_{j=1}^q \hat{\Psi}_l \hat{w}_{t-l}^* + \eta_{qt}^*,$$

using initialization  $(\hat{w}_0^*, \dots, \hat{w}_{1-q}^*) = (\hat{w}_0, \dots, \hat{w}_{1-q})$ .

When  $z_t$ s are unobservable, steps (2.1)-(2.3) can be applied to  $\tilde{w}_t$  to obtain  $\{\tilde{\eta}_{qt}^*\}_{t=1}^T$  and  $\tilde{w}_t^*$ .

**Step 3.** (3.1) Integrate the last  $k$  elements of  $\hat{w}_t^*$  or  $\tilde{w}_t^*$  to obtain  $z_t^*$  as

$$z_t^* = z_0 + \sum_{j=1}^t \hat{w}_j^{*(z)},$$

or

$$\tilde{z}_t^* = z_0 + \sum_{j=1}^t \tilde{w}_j^{*(z)}$$

where  $\hat{w}_t^{*(z)}$  and  $\tilde{w}_t^{*(z)}$  refer to the last  $k$  elements of  $\hat{w}_t^*$  and  $\tilde{w}_t^*$  respectively.

(3.2) Generate  $\bar{W}_t^*$  as

$$\bar{W}_t^* = \tilde{\Theta}^{OLS} z_t^* + \hat{e}_t^*, \quad (33)$$

or

$$\bar{W}_t^* = \tilde{\Theta}^{PC} \tilde{z}_t^* + \tilde{e}_t^*. \quad (34)$$

(3.3) Estimate  $\Theta$  from either equation (33) or (34) using OLS. Denote the estimator as  $\Theta^*$ .

- (3.4) Compute the bootstrap counterpart of the test statistics, say  $\widehat{D}^*$ , using  $\Theta^*$ .

The resampling scheme we propose is based on sieve estimation and follows the same lines as in the approach of Chang, Park and Song (2006). Note that projecting the estimates of  $b$  onto the column space of  $\Gamma$  means that resampling is performed under the null hypothesis. As it is illustrated below, this ensures the validity of the bootstrap under the null and the alternative hypothesis.

Denote now the null limiting distribution of  $T^2\widehat{D}$  as  $Z_0$  and the bootstrap probability conditional on the sample as  $P^*$ . The form of  $Z_0$  is given by Theorems 3 and 4 for  $z_t$ s observable and unobservable, respectively. To prove that the bootstrap procedure is valid, two conditions need to be satisfied. First, we need to show that both under the null hypothesis  $H_0$  and under the local alternatives  $H_1^l$ , the conditional distribution of  $T^2\widehat{D}^*$  given  $\{\bar{W}_t\}_{t=1}^T$ , consistently estimates the limiting distribution of  $T^2\widehat{D}$ , that is

$$P^* \left[ T^2\widehat{D}^* \leq v \right] \xrightarrow{P} P \{ Z_0 \leq v \},$$

for each  $v$  which is a continuity point of the distribution function of  $T^2\widehat{D}$ . More compactly, this statement will be referred to as  $T^2\widehat{D}^* \xrightarrow{d^B} Z_0$ . Second, under the alternative hypothesis  $H_1$  the bootstrap statistic  $T^2\widehat{D}^*$  must be bounded in probability, or even possibly converge to  $Z_0$ .

Consider the following Assumption which we need to prove the bootstrap validity.

**Assumption 3**

- (i) Let  $[\bar{e}'_t, \Delta z'_t] = \Phi(L)\eta_t$  where  $\Phi(L) = \sum_{k=0}^{\infty} \Phi_k L^k$ . The sequence  $\eta_t$  is *i.i.d.* with  $E(\eta_t) = 0$ ,  $E(\eta_t \eta'_t) > 0$ , finite fourth moment and such that  $|\Phi(z)| \neq 0$  for all  $|z| \leq 1$  and  $\sum_{l=0}^{\infty} |k|^\zeta |\Phi_l| < \infty$  for some  $\zeta \geq 1$ ;
- (ii) In equation (32), let  $q \rightarrow \infty$  and  $q = o(T^{1/2})$  as  $T \rightarrow \infty$ .

## Remarks

**Assumption 3** postulates a linear process for the idiosyncratic term  $\varepsilon_t$ , which is in line with the use of sieve bootstrap.

**A3.1** Assumption 3(i) ensures the validity of an invariance principle for the partial sums of  $[\bar{e}'_t, \Delta z'_t]'$ , and it is essentially the same as in Chang, Park and Song (2006).

**A3.2** Assumption 3(ii) is required to ensure the consistency of the estimates  $\hat{\Psi}_l$ . Assumption 3 is a stronger version of Assumption 1(i), since it requires that be a stationary linear process. This is needed in order to apply sieve bootstrap and to prove its validity using the methods employed in Park (2002).

**A3.3** With respect to Remark A1.3, in this section the possibility of having a factor structure in  $\varepsilon_t$  is ruled out. This is natural, if one considers the way in which the sieve bootstrap works. At Step 2.1 of the algorithm, a consistent estimate of the parameters in the model for  $\varepsilon_t$  is required. But estimating consistently the parameters of a factor model for  $\varepsilon_t$  would not be possible, since  $n$  is fixed. Thus, sieve bootstrap should not be employed when  $\varepsilon_t$  has a factor structure. In that case, a natural choice would be block-bootstrap, in order to preserve both time and cross-sectional dependence.

The following theorem asserts the validity of the bootstrap procedure.

**Theorem 7** *Under Assumptions 1-3, we have that, under the null hypothesis  $H_0$ , the alternative hypothesis  $H_1$  and the local alternatives  $H_1^l$*

$$T^2 \hat{D}^* \xrightarrow{d^B} Z_0, \tag{35}$$



where  $Z_0$  is the null limit distribution which is  $Z_0 = \|b\|^{-2} Q' [I_k - \|b\|^{-2} (bb')] Q$  for observable  $z_t$ s and  $Z_0 = \|b\|^{-2} Q^{pc'} [I_k - \|b\|^{-2} (bb')] Q^{pc}$  for unobservable  $z_t$ s.

**Proof.** See Appendix. ■

Theorem 7 extends the sieve bootstrap algorithm proposed by Chang, Park and Song (2006) to the case of principal component estimates. The validity of our bootstrap procedure is ensured by equation (35), which shows that under the null and the local alternatives the bootstrap consistently approximates the asymptotic distribution of  $T^2 \hat{D}$  and under the alternative the bootstrap statistic  $T^2 \hat{D}^*$  has the same distribution as the null. This is a consequence of the resampling algorithm being implemented under the null hypothesis.

It is worth noting that whilst the estimation technique employed to estimate  $\hat{\Theta}$  necessarily differs (i.e. we use OLS when the  $z_t$ s are observable and principal component when  $z_t$ s are not observable), the bootstrap estimator  $\Theta^*$  is computed via OLS irrespective of the method employed to derive  $\hat{\Theta}$ .

## 5 Monte Carlo Results

In this section, we present an assessment, via a small Monte Carlo exercise, of the power and size of the bootstrap testing procedure we propose.

The data generating process for the Monte Carlo exercise is described by equations (6) and (7). We generate the  $k$  stochastic trends  $z_t$  as random walks according to Assumption 1. Let  $\bar{v}_t = [\bar{v}'_t, \bar{s}'_t]'$ , we consider the following processes for  $\bar{v}_t$ : a white noise process, an AR(1) model with autoregressive root equal to 0.75, an MA(1) process with root equal to 0.75. These choices allow to check for robustness and efficiency of our procedure under alternative error dynamics. Under the alternative hypothesis, we generate  $\bar{y}_t$  using specification (28). We also consider alternative size of  $T = \{20, 35, 50, 100, 200\}$  and of the number of trends  $k = \{2, 3, 4, 5\}$ . The number of Monte Carlo and bootstrap replications is 5000 and 1000, respectively.

The results are reported in Table 1.

**[Insert Table 1 somewhere here]**

The main finding is that the bootstrap test shows good size and power and its performance is affected by the number of trends considered.

In particular, there is a strong impact of the number of factors  $k$  on the size of the test. When the error term  $\bar{v}_t$  has no dynamics, which is the baseline case, the size decreases as  $k$  increases. This happens uniformly in  $T$ , and the size tends, asymptotically, to its nominal value. The test exhibits a good performance when the error term is white noise even for small samples. When AR(1) and MA(1) processes are present, the impact of  $k$  still leads to size decrease as the number of stochastic trends increase. Note though that now the test is oversized for small samples, especially when AR dynamics is present. This effect tends to be wiped out asymptotically, when irrespective of the error dynamics and for the large  $k$  (4, 5) cases, there is a slight undersize tendency of the test.

The power too is affected by  $k$ . Though small sample performance seems to be very good, especially in the white noise case, irrespective of  $k$ , however, for all cases, as  $k$  increases, the power slightly decreases. Nonetheless, asymptotically the power approaches one irrespective of the error dynamics and of the number of stochastic trends.

## 6 Conclusions

In nonstationary heterogeneous panels where each unit cointegrates, the aggregate relationship in general does not cointegrate unless a large number of conditions is satisfied. However, the aggregate equation may be observationally equivalent to a cointegrating relationship even when the conditions for perfect aggregation are violated. How well the aggregate relationship approximates the properties of individual components cannot be tested when only aggregate data are available. When data are

available at disaggregate level, as in the case of panels, one can test whether features of micro relationships are preserved after aggregation.

This paper addresses the issue of micro versus macro cointegration by considering nonstationary heterogeneous panels with a fixed number of units and a large number of time observations. Our results can be viewed as complementary to the analysis in Phillips and Moon (1999) of the case when  $(n, T) \rightarrow \infty$ . No restrictions are placed regarding the existence of the degree of contemporaneous correlation between units and between regressors and error terms in the cointegration regressions.

We derive the test statistic  $D = \sin^2(b, Mb)$  for the null hypothesis of cointegration, building upon the formal conditions for cointegration valid at micro level to hold after aggregation. The test is powerful against local alternatives and consistent. We propose a valid bootstrap approximation and Monte Carlo evidence suggests that the test exhibits good size and power properties.

The test under the null is of asymptotic order  $O_p(T^{-2})$ . This property has important implications for empirical applications of the test procedure. For instance, macro data may be available at monthly/quarterly frequency but micro data could be available at lower frequency (e.g. census data). In that case, the  $T^2$  convergence might be an important asset given the short length of each micro series.

Our asymptotics has been derived for panels with fixed  $n$ . Thus, it is also empirically relevant to see how our method performs in simulations in comparison with the Phillips and Moon (1999) asymptotics.

A comprehensive set of empirical applications and an extensive simulation exercise are beyond the scope of the present paper but are subject of separate studies.

## Acknowledgment

This is a new version of a paper that has circulated with the title "Cointegration versus Spurious Regression in Heterogeneous Panels". We are grateful to participants in the workshop on "Recent Developments in the Econometrics of Panel Data" (Cass

Business School, 15 March 2004), in particular Badi Baltagi, Chihwa Kao, Yoosoon Chang, and Hashem Pesaran for useful discussions; the North American Summer Meeting of the Econometrics Society (Providence, 17-20, June 2004), the Econometrics Group Seminar at Tinbergen Institute (Amsterdam, 4 February 2005), the Department of Economics/FAS and C.V. Starr Center for Applied Economics "Econometrics Seminar" (NYU, 28 October 2005) in particular Jushan Bai and Xiaohong Chen, the "Econometrics and Statistics Colloquium", Graduate School of Business Chicago (Chicago, 26 April 2006), in particular F. Bandi, A. Bester, T. Conley, R. S. Tsay, and A. Zellner, the "Staff Seminar Series 2005/2006" (Department of Economics, University of Leicester, 1 June 2006), in particular Steve Hall, the MMF Workshop on "The Use of Panel Data in Macroeconomics and Finance" (University of Leicester, 14 December 2007) in particular Kaddour Hadri and Hashem Pesaran, the Seminar@ECARES series in "Econometrics and Statistics" (ECARES at Université Libre de Bruxelles, 20 March 2008), in particular David Veredas, the "Econometrics Workshop" (Department of Economics, Michigan State University, 3 April 2008) in particular Emma Iglesias, Peter Schmidt, Timothy Vogelsang, Jeff Wooldridge, provided helpful insights. Jushan Bai, Joerg Breitung, Stepana Lazarova, and Marco Lippi read very carefully a previous version of the manuscript and provided very useful comments and suggestions which helped to greatly improve the paper. However, the usual disclaimer applies. Lorenzo Trapani acknowledges financial support from Cass Business School under the RAE Development Fund scheme and from ESRC Post-Doctoral Fellowship scheme (a.r. PTA-026-27-1107).

## References

BAI, J. (2003), "Inferential Theory for Structural Models of Large Dimensions", *Econometrica*, 71, 135-171.

BAI, J. (2004), "Estimating Cross-Section Common Stochastic Trends in Non-stationary Panel Data", *Journal of Econometrics*, 122, 137-183.

CAMPBELL, J.Y., SHILLER, R.J. (1987), "Cointegration and Tests of Present Value Models", *Journal of Political Economy*, 95, 1062-88.

CHANG, Y., PARK, J.Y. (2002), "On the Asymptotic of ADF Tests for Unit Roots", *Econometric Reviews*, 21, 431-448.

CHANG, Y., PARK, J.Y. (2003), "A Sieve Bootstrap for the Test of a Unit Root", *Journal of Time Series Analysis*, 24, 379-400.

CHANG, Y., PARK, J.Y., SONG, K. (2006), "Bootstrapping Cointegrating Regressions", *Journal of Econometrics*, 133, 703-739.

FORNI, M. AND LIPPI, M. (1997), *Aggregation and the Microfoundations of Dynamic Macroeconomics*. Oxford University Press: Oxford.

FORNI, M. AND LIPPI, M. (1998), "On the Microfoundations of Dynamic Macroeconomics", WP Series, University of Modena.

FORNI, M. AND LIPPI, M. (1999), "Aggregation of Linear Dynamic Microeconomic Models", *Journal of Mathematical Economics*, 31, 131-158.

GATEV, E, GOETZMANN, W.N., ROUWENHORST, K.G. (2006), "Pairs Trading: Performance of a Relative-Value Arbitrage Rule," *Review of Financial Studies*, 19, 797-827.

GONZALO, J. (1993), "Cointegration and Aggregation", *Ricerche Economiche*, 47, 281-291.

GRANGER, C.W.J. (1990), "Aggregation of Time Series Variables - a Survey in Disaggregation", in T. Barker and M. H. Pesaran, eds. *Econometric Modelling*, London Routledge.

GRANGER, C.W.J. (1993), "Implications of Seeing Economic Variables Through an Aggregation Window", *Ricerche Economiche*, 47, 269-279.

HSIAO, C., SHEN, Y. AND FUJIKI, H. (2005), "Aggregate vs. Disaggregate Data Analysis - A Paradox in the estimation of a Money Demand Function of Japan under the Low Interest Rate Policy", *Journal of Applied Econometrics*, 20, 579-601.

KAO, C. (1999), "Spurious Regression and Residual-Based Tests for Cointegra-

tion in Panel Data”, *Journal of Econometrics*, 90, 1-44.

KAO, TRAPANI, URGAL (2008), "Asymptotics for Panel Models with Common Shocks", mimeo. This is a reviewed version of the working paper WP-CEA-01-2006 (CEA@Cass Working Paper series):

[http://www.cass.city.ac.uk/cea/research\\_papers/Working%20Papers%202006/WP-CEA-01-2006.pdf](http://www.cass.city.ac.uk/cea/research_papers/Working%20Papers%202006/WP-CEA-01-2006.pdf)

KWIATKOWSKI, D, PHILLIPS, P.C.B, P. SCHMIDT, P., SHIN Y. (1992), "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root", *Journal of Econometrics*, 54, 159–178.

LAZAROVA, S., TRAPANI, L., URGAL, G. (2007), “Common Stochastic Trends and Aggregation in Heterogeneous Panels”, *Econometric Theory*, 23, 89-105.

LIPPI, M. (1988), “On the Dynamic Shape of Aggregated Error Correction Models”, *Journal of Economic Dynamics and Control*, 12, 561-585.

MCCOSKEY, S, KAO, C. (1998), "A Residual-Based Test of the Null of Cointegration in Panel Data", *Econometric Reviews*, 17, 57-84.

PARK, J.Y. (2002), “An Invariance Principle for Sieve Bootstrap in Time Series”, *Econometric Theory*, 18, 469-490.

PARK, J.Y. AND PHILLIPS, P.C.B. (1988), “Statistical Inference in Regressions with Integrated Processes: Part 1”, *Econometric Theory*, 4, 468-497.

PARK, J.Y. AND PHILLIPS, P.C.B., (1999), "Asymptotics for Nonlinear Transformations of Integrated Time Series", *Econometric Theory*, 15, 269, 298.

PESARAN, M. H. (2003), “Aggregation of Linear Dynamic Models: An Application to Life-Cycle Consumption Models under Habit Formation”, *Economic Modelling*, 20, 383-415.

PESARAN, M.H. AND SMITH R. (1995), “Estimating Long-Run Relationships from Dynamic Heterogeneous Panels”, *Journal of Econometrics*, 68, 79-113.

PHILLIPS, P.C.B. (1986), “Understanding Spurious Regressions in Econometrics”, *Journal of Econometrics*, 33, 311-340.

PHILLIPS, P.C.B. AND MOON, H.R. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica*, 67, 1057-1111.

SAKHANENKO, A.I. (1980), "On Unimprovable Estimates of the Rate of Convergence in Invariance Principle", in *Nonparametric Statistical Inference*, Colloquia Mathematica Societatis Janos Bolyai 32, 779-783, Budapest, Hungary.

SHIN, Y. (1994), "A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration", *Econometric Theory*, 10, 91-115.

STOKER, T. M. (1993), "Empirical Approaches to the Problem of Aggregation over Individuals", *Journal of Economic Literature*, 31, 1827-1874.

VUOLTEENAHO, T. (2002), "What Drives Firm-Level Stock Returns?", *Journal of Finance*, 57, 233-264.

XIAO, Z. (1999), "A Residual Based Test for the Null Hypothesis of Cointegration", *Economics Letters*, 64, 133-141.

# Appendix

**Proof of Proposition 2.** Assumption 2(*iv*) and the iid-ness of  $\beta_i$  and  $\Gamma_i$  entail that a LLN holds and thus  $n^{-1} \sum_{i=1}^n \Gamma_i \xrightarrow{p} \bar{\Gamma}$ ,  $n^{-1} \sum_{i=1}^n \beta_i \xrightarrow{p} \bar{\beta}$  and  $n^{-1} \sum_{i=1}^n \Gamma_i \beta_i \xrightarrow{p} \bar{\Gamma}' (\bar{\beta} + c)$ .

Consider the following Assumption:

**Assumption 1\***. Assumption 1 holds and

(*iii*)  $\varepsilon_t$  is a linear process with  $E \|\varepsilon_t\|^{2+\delta} \leq M < \infty$  for some  $\delta > 0$  and some constant  $M$ ;

(*iv*) a Beveridge-Nelson decomposition holds for  $z_t$  and  $v_{it}$  such that  $z_t = z_t^* + R_{zt}$  and  $v_{it} = v_{it}^* + R_{vit}$ , where  $z_t^*$  is a random walk with unit long run covariance matrix,  $v_{it}^*$  is an iid  $I(0)$  process whose long run variance is the same as that of  $v_{it}$ , and  $R_{zt}$  and  $R_{vit}$  are the remainders of the BN decomposition;

(*v*)  $\{u_{it}\}$  and  $\{v_{it}\}$  are independent across  $i$  and it holds that  $\sum_{t=1}^T \|\bar{v}_t\|^2 = O_p(nT)$  and  $\sum_{t=1}^T \bar{u}_t^2 = O_p(nT)$  for  $\bar{u}_t = \sum_{i=1}^n u_{it}$ .

Consider  $\hat{\beta}$ ; since  $\bar{y}_t = \bar{x}_t' \bar{\beta} + \bar{u}_t + \sum_{i=1}^n (\beta_i - \bar{\beta})' x_{it}$ , and recalling the definition of  $\bar{x}_t$ , it holds that

$$\begin{aligned} \hat{\beta} &= \bar{\beta} + \left[ \Gamma \sum_{t=1}^T z_t z_t' \Gamma' + \Gamma \sum_{t=1}^T z_t \bar{v}_t' + \sum_{t=1}^T \bar{v}_t z_t' \Gamma' + \sum_{t=1}^T \bar{v}_t \bar{v}_t' \right]^{-1} \times \\ &\quad \left[ \sum_{t=1}^T \bar{x}_t \bar{u}_t + \sum_{i=1}^n \sum_{t=1}^T \bar{x}_t x_{it}' (\beta_i - \bar{\beta}) \right]. \end{aligned} \quad (36)$$

Consider first the denominator of (36). Since  $\Gamma = O_p(n)$ , Assumption 1(*i*) ensures that  $\Gamma \sum_{t=1}^T z_t z_t' \Gamma' = O_p(n^2 T^2)$ . In order to find the order of magnitude of  $\Gamma \sum_{t=1}^T z_t \bar{v}_t'$ , we will use the joint limit theory developed in Kao, Trapani and Urga (2008), based on showing that a CLT for martingale difference sequences (MDS) holds for  $\sum_{t=1}^T z_t \bar{v}_t'$ . This involves proving that: (1)  $\sum_{t=1}^T z_t \bar{v}_t'$  is an MDS and (2) a Liapunov condition holds jointly as  $(n, T) \rightarrow \infty$ . Let  $\xi_{iT} = T^{-1} \sum_{t=1}^T z_t v_{it}'$ . Then



$\xi_{iT} = \xi_{iT}^* + R_{iT}$ , where  $\xi_{iT}^* = T^{-1} \sum_{t=1}^T z_t^* v_{it}'$  and  $R_{iT}$  is the remainder in the BN decomposition. Following similar arguments as in Phillips and Moon (1999), it can be shown that  $R_{iT} = O_p(T^{-1/2})$ ; thus,  $T \sum_{i=1}^n R_{iT} = O_p(n\sqrt{T})$ . As far as the order of magnitude of  $\xi_{iT}^*$  is concerned, let  $C$  be the  $\sigma$ -field generated by the  $z_t$ s. Then  $E[\xi_{iT}^* | C] = 0$  and, conditional on  $C$ ,  $\xi_{iT}^*$  is an independent sequence due to Assumption 1\*(v). Define  $I_i$  as the  $\sigma$ -field generated by  $C$  and by  $\{\xi_{jT}^*\}_{j=1}^i$ ; then  $\{\xi_{iT}^*, I_i\}$  is a martingale difference sequence since  $E[\xi_{iT}^* | I_{i-1}] = E[\xi_{iT}^* | C] = 0$ . Also, a Liapunov condition holds whereby  $E \|\xi_{iT}^* | C\|^{2+\delta} < \infty$  for all  $i$  since

$$\begin{aligned} E \|\xi_{iT}^* | C\|^{2+\delta} &= \frac{1}{T^{2+\delta}} E \left\| \sum_{t=1}^T z_t^* v_{it}' \middle| C \right\|^{2+\delta} \\ &\leq \frac{1}{T^{2+\delta}} \sum_{t=1}^T \|z_t^*\|^{2+\delta} E \|v_{it}^*\|^{2+\delta} \\ &\leq M \frac{1}{T^{2+\delta}} \sum_{t=1}^T \|z_t^*\|^{2+\delta}, \end{aligned}$$

which is finite in light of Assumption 1\*(iii) and of Theorem 5.2 in Park and Phillips (1999). Thus, an MDS CLT can be applied to show that  $n^{-1/2} \sum_{i=1}^n \xi_{iT}^* = O_p(1)$ . Recalling the definition of  $\xi_{iT}^*$  we finally have

$$\begin{aligned} \Gamma \sum_{t=1}^T z_t \bar{v}_t' &= \Gamma \sum_{i=1}^n \sum_{t=1}^T z_t v_{it}' \\ &= O_p(n) \left[ O_p(\sqrt{nT}) + O_p(n\sqrt{T}) \right] = O_p(n\sqrt{nT}) + O_p(n^2\sqrt{T}). \end{aligned}$$

Last,  $\sum_{t=1}^T \bar{v}_t \bar{v}_t' = O_p(nT)$  by Assumption 1\*(v). Thus, as far as the denominator of (36) is concerned, as  $(n, T) \rightarrow \infty$

$$\frac{1}{n^2 T^2} \sum_{t=1}^T \bar{x}_t \bar{x}_t' \stackrel{a.s.}{=} \Gamma \int W_z W_z' \Gamma' + o_p(1).$$

Let us now consider the numerator of (36). We have  $\sum_{t=1}^T \bar{x}_t \bar{u}_t = \Gamma \sum_{t=1}^T z_t \bar{u}_t + \sum_{t=1}^T \bar{v}_t \bar{u}_t$ . Similar arguments as above entail  $\Gamma \sum_{t=1}^T z_t \bar{u}_t = O_p(n\sqrt{nT}) + O_p(n^2\sqrt{T})$ ;

Assumption 1\* (iii) and the Cauchy-Schwartz inequality ensures that  $\sum_{t=1}^T \bar{v}_t \bar{u}_t = O_p(nT)$ . Note that we do not need to assume that  $u_{it}$  and  $v_{it}$  are uncorrelated. As far as the term  $\sum_{i=1}^n \sum_{t=1}^T \bar{x}_t x'_{it} (\beta_i - \bar{\beta})$  is concerned, it holds that

$$\begin{aligned} \sum_{i=1}^n \sum_{t=1}^T \bar{x}_t x'_{it} (\beta_i - \bar{\beta}) &= \Gamma \sum_{i=1}^n \sum_{t=1}^T z_t x'_{it} (\beta_i - \bar{\beta}) + \sum_{i=1}^n \sum_{t=1}^T \bar{v}_t x'_{it} (\beta_i - \bar{\beta}) \\ &= \Gamma \sum_{i=1}^n \sum_{t=1}^T z_t z_t \Gamma'_i (\beta_i - \bar{\beta}) + \Gamma \sum_{i=1}^n \sum_{t=1}^T z_t v'_{it} (\beta_i - \bar{\beta}) + \sum_{i=1}^n \sum_{t=1}^T \bar{v}_t x'_{it} (\beta_i - \bar{\beta}) \\ &= I + II + III, \end{aligned}$$

where the second passage follows from (4). Consider  $I$ :

$$\begin{aligned} I &= \Gamma \sum_{t=1}^T z_t z_t \sum_{i=1}^n \Gamma'_i \beta_i - \Gamma \sum_{t=1}^T z_t z_t \Gamma' \bar{\beta} \\ &= O_p(n) O_p(T^2) O_p(n) = O_p(n^2 T^2); \end{aligned}$$

by assumption, we know that  $n^{-1} \sum_{i=1}^n \Gamma'_i \beta_i \xrightarrow{p} \Gamma'(\bar{\beta} + c)$ , and therefore as  $(n, T) \rightarrow \infty$ ,  $(nT)^{-2} \times I \xrightarrow{p} \Gamma \int W_z W'_z \Gamma' c$ . Also, similar arguments as above would lead to  $II = O_p(n\sqrt{nT}) + O_p(n^2\sqrt{T})$ . Last

$$III \leq \left( \sum_{t=1}^T \|\bar{v}_t\|^2 \right)^{1/2} \left[ \sum_{t=1}^T \left\| \sum_{i=1}^n x'_{it} (\beta_i - \bar{\beta}) \right\|^2 \right]^{1/2},$$

with  $\left( \sum_{t=1}^T \|\bar{v}_t\|^2 \right)^{1/2} = O_p(\sqrt{nT})$  by Assumption 1\* (v) and

$$\begin{aligned} \sum_{t=1}^T \left\| \sum_{i=1}^n x'_{it} (\beta_i - \bar{\beta}) \right\|^2 &\leq \max_{1 \leq i \leq n} \|\beta_i - \bar{\beta}\|^2 \sum_{t=1}^T \left\| \sum_{i=1}^n x_{it} \right\|^2 \\ &= O_p(1) O_p(n^2 T), \end{aligned}$$

so that  $III$  is bounded by  $O_p(n\sqrt{nT})$ . Thus, as far as the numerator is concerned

we have, as  $(n, T) \rightarrow \infty$

$$\frac{1}{n^2 T^2} \sum_{t=1}^T \bar{x}_t \bar{s}_t \stackrel{a.s.}{=} \Gamma \int W_z W_z' \Gamma' c + o_p(1).$$

Combining this with the denominator, it holds that as  $(n, T) \rightarrow \infty$

$$\hat{\beta} \xrightarrow{p} \bar{\beta} + \left[ \Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[ \Gamma \int W_z W_z' \Gamma' \right] c = \bar{\beta} + c.$$

■

**Proof of Proposition 3.**  $T$ -consistency of the PC estimator for nonstationary panel factor models is proved in Lemma 3 in Bai (2004), where it is established that loadings can be estimated consistently irrespective of whether  $n$  be fixed or large. Lemma 3 in Bai (2004) holds here because the assumptions we make are the same as Assumptions A-E in Bai (2004). Assumption 1(*i*) and the requirement that  $E \|\epsilon_t\|^{4+\delta} < \infty$  for some  $\delta > 0$  correspond to Assumption A. Assumption 2(*iv*) corresponds to Assumption B, and it implies that  $\|\Xi_i\| < \infty$ . The requirements that  $E |u_{it}|^8 < \infty$ ,  $E \|v_{it}\|^8 < \infty$  and  $\sum_{t=1}^T E \|e_{it}^W e_{js}^W\| < \infty$  for all  $(i, j, s)$  imply Assumption C and E. Note that we do not need to assume any bounds on cross-sectional correlations given that  $n$  is fixed and finite sums of finite summands are finite. Having finite  $n$  simplifies the restrictions needed for cross-sectional and time series dependence. Finally, Bai's Assumption D is the same as Assumption 1(*ii*) in this paper. ■

**Proof of Theorem 2.** Let  $\hat{b}$  and  $\hat{\Gamma}$  be  $T$ -consistent estimators of  $b$  and  $\Gamma$  and define

$$\varepsilon_b = \hat{b} - b,$$

$$\varepsilon_\Gamma = \hat{\Gamma} - \Gamma;$$

by definition, and in light of Assumption 1(*i*),  $\varepsilon_b = O_p(T^{-1})$  and  $\varepsilon_\Gamma = O_p(T^{-1})$ . For

the sake of the notation, let also  $Mb = a$  and  $\varepsilon_a = \hat{a} - a$ . We have

$$\begin{aligned}\varepsilon_a &= \hat{a} - a = \hat{M}\hat{b} - Mb \\ &= (M + \varepsilon_M)(b + \varepsilon_b) - Mb \\ &= \varepsilon_M b + M\varepsilon_b + \varepsilon_M \varepsilon_b = O_p(T^{-1}).\end{aligned}$$

This is because we have

$$\hat{M} = \hat{\Gamma}' \left( \hat{\Gamma} \hat{\Gamma}' \right)^{-1} \hat{\Gamma},$$

and

$$\begin{aligned}\hat{\Gamma} \hat{\Gamma}' &= (\Gamma + \varepsilon_\Gamma) (\Gamma + \varepsilon_\Gamma)' = \\ &= \Gamma \Gamma' + \Gamma \varepsilon_\Gamma' + \varepsilon_\Gamma \Gamma' + \varepsilon_\Gamma \varepsilon_\Gamma' .\end{aligned}$$

Using Taylor's approximation,

$$[\Gamma \Gamma' + \Gamma \varepsilon_\Gamma' + \varepsilon_\Gamma \Gamma' + \varepsilon_\Gamma \varepsilon_\Gamma']^{-1} = (\Gamma \Gamma')^{-1} - (\Gamma \Gamma')^{-1} (\Gamma \varepsilon_\Gamma' + \varepsilon_\Gamma \Gamma' + \varepsilon_\Gamma \varepsilon_\Gamma') (\Gamma \Gamma')^{-1} + R_T,$$

where the remainder  $R_T$  is defined as  $R_T = o(\|\Gamma \varepsilon_\Gamma' + \varepsilon_\Gamma \Gamma' + \varepsilon_\Gamma \varepsilon_\Gamma'\|)$ , we have

$$\begin{aligned}\hat{\Gamma}' \left( \hat{\Gamma} \hat{\Gamma}' \right)^{-1} \hat{\Gamma} &= [\Gamma + \varepsilon_\Gamma]' \left[ (\Gamma \Gamma')^{-1} - (\Gamma \Gamma')^{-1} (\Gamma \varepsilon_\Gamma' + \varepsilon_\Gamma \Gamma' + \varepsilon_\Gamma \varepsilon_\Gamma') (\Gamma \Gamma')^{-1} \right] [\Gamma + \varepsilon_\Gamma] \\ &= \Gamma' (\Gamma \Gamma')^{-1} \Gamma + \Gamma' (\Gamma \Gamma')^{-1} \varepsilon_\Gamma - \Gamma' (\Gamma \Gamma')^{-1} \Gamma \varepsilon_\Gamma' (\Gamma \Gamma')^{-1} \Gamma - \Gamma' (\Gamma \Gamma')^{-1} \varepsilon_\Gamma \Gamma' (\Gamma \Gamma')^{-1} \Gamma \\ &\quad + \varepsilon_\Gamma' (\Gamma \Gamma')^{-1} \Gamma + O_p(T^{-2}).\end{aligned}$$

Let  $\varepsilon_M = \Gamma' (\Gamma \Gamma')^{-1} \varepsilon_\Gamma + \varepsilon_\Gamma' (\Gamma \Gamma')^{-1} \Gamma - \Gamma' (\Gamma \Gamma')^{-1} \Gamma \varepsilon_\Gamma' (\Gamma \Gamma')^{-1} \Gamma - \Gamma' (\Gamma \Gamma')^{-1} \varepsilon_\Gamma \Gamma' (\Gamma \Gamma')^{-1} \Gamma = O_p(T^{-1})$ .

We have

$$\sin^2(\hat{a}, \hat{b}) - \sin^2(a, b) = \left[ \cos(\hat{a}, \hat{b}) + \cos(a, b) \right] \left[ \cos(a, b) - \cos(\hat{a}, \hat{b}) \right].$$

Slutsky's theorem implies that  $\cos(\hat{a}, \hat{b}) = \cos(a, b) + o_p(1)$ , and under the null we have  $\cos(a, b) = 1$ , so that

$$\begin{aligned}
\sin^2(\hat{a}, \hat{b}) - \sin^2(a, b) &= [2 + o_p(1)] \left[ \cos(a, b) - \cos(\hat{a}, \hat{b}) \right] \\
&= [2 + o_p(1)] \left[ \frac{a'b}{\|a\| \|b\|} - \frac{\hat{a}'\hat{b}}{\|\hat{a}\| \|\hat{b}\|} \right] \\
&= [2 + o_p(1)] \frac{\|\hat{a}\| \|\hat{b}\| (a'b) - \|a\| \|b\| (\hat{a}'\hat{b})}{\|a\| \|b\| \|\hat{a}\| \|\hat{b}\|}. \quad (37)
\end{aligned}$$

It holds that

$$\begin{aligned}
\hat{a}'\hat{b} &= (a + \varepsilon_a)'(b + \varepsilon_b) \\
&= a'b + a'\varepsilon_b + b'\varepsilon_a + \varepsilon_a'\varepsilon_b.
\end{aligned}$$

Let now  $\varepsilon_{\|a\|} = \|\hat{a}\| - \|a\|$  and  $\varepsilon_{\|b\|} = \|\hat{b}\| - \|b\|$ . We have

$$\varepsilon_{\|a\|} = \|a\| \sqrt{1 + \frac{2a'\varepsilon_a + \varepsilon_a'\varepsilon_a}{\|a\|^2}} - \|a\|.$$

Using Taylor's expansion, we get

$$\sqrt{1 + \frac{2a'\varepsilon_a + \varepsilon_a'\varepsilon_a}{\|a\|^2}} = 1 + \frac{1}{2} \frac{2a'\varepsilon_a + \varepsilon_a'\varepsilon_a}{\|a\|^2} - \frac{1}{8} \left( \frac{2a'\varepsilon_a + \varepsilon_a'\varepsilon_a}{\|a\|^2} \right)^2 + R_T,$$

with  $R_T = o(\|2a'\varepsilon_a + \varepsilon_a'\varepsilon_a\|^2)$  so that

$$\varepsilon_{\|a\|} = \frac{a'\varepsilon_a}{\|a\|} + \frac{\varepsilon_a'\varepsilon_a}{2\|a\|} - \frac{(a'\varepsilon_a)^2}{2\|a\|^3} + O_p(T^{-3}). \quad (38)$$

Likewise,

$$\varepsilon_{\|b\|} = \frac{b'\varepsilon_b}{\|b\|} + \frac{\varepsilon_b'\varepsilon_b}{2\|b\|} - \frac{(b'\varepsilon_b)^2}{2\|b\|^3} + O_p(T^{-3}). \quad (39)$$

Under the null,  $a = b$  and  $a'b = \|a\| \|b\| = \|a\|^2$ . Therefore we may write

$$\begin{aligned} & \|\hat{a}\| \|\hat{b}\| (a'b) - \|a\| \|b\| (\hat{a}'\hat{b}) \\ = & \|a\|^2 \left[ \|\hat{a}\| \|\hat{b}\| - (\hat{a}'\hat{b}) \right], \end{aligned}$$

and

$$\begin{aligned} & \|\hat{a}\| \|\hat{b}\| - (\hat{a}'\hat{b}), \\ = & (\|a\| + \varepsilon_{\|a\|}) (\|a\| + \varepsilon_{\|b\|}) - a'b - a'\varepsilon_b - a'\varepsilon_a - \varepsilon'_a\varepsilon_b \\ = & \|a\| \varepsilon_{\|a\|} + \|a\| \varepsilon_{\|b\|} + \varepsilon_{\|a\|}\varepsilon_{\|b\|} - a'\varepsilon_b - a'\varepsilon_a - \varepsilon'_a\varepsilon_b \\ = & a'\varepsilon_a + \frac{\varepsilon'_a\varepsilon_a}{2} - \frac{(a'\varepsilon_a)^2}{2\|a\|^2} + a'\varepsilon_b + \frac{\varepsilon'_b\varepsilon_b}{2} - \frac{(a'\varepsilon_b)^2}{2\|a\|^2} + \frac{(a'\varepsilon_a)(b'\varepsilon_b)}{\|a\|^2} - a'\varepsilon_b - a'\varepsilon_a - \varepsilon'_a\varepsilon_b \\ = & \frac{1}{2} (\varepsilon_a - \varepsilon_b)' \left[ I_k - \frac{aa'}{\|a\|^2} \right] (\varepsilon_a - \varepsilon_b). \end{aligned}$$

Finally, from equation (37) it holds that, since under the null  $a = b$

$$\begin{aligned} \sin^2(\hat{a}, \hat{b}) - \sin^2(a, b) &= [2 + o_p(1)] \frac{\|\hat{a}\| \|\hat{b}\| (a'b) - \|a\| \|b\| (\hat{a}'\hat{b})}{\|a\| \|b\| \|\hat{a}\| \|\hat{b}\|} \\ &= \frac{1}{\|b\|^2} (\varepsilon_a - \varepsilon_b)' \left[ I_k - \frac{bb'}{\|b\|^2} \right] (\varepsilon_a - \varepsilon_b) + O_p(T^{-3}) \\ &= O_p(T^{-2}). \end{aligned} \tag{40}$$

■

**Proof of Corollary 2.** When  $\hat{b}_\delta - b = O_p(T^{-\delta})$  and  $\hat{\Gamma}_\delta - \Gamma = O_p(T^{-\delta})$ , it also holds that  $\varepsilon_a = O_p(T^{-\delta})$  and  $\varepsilon_b = O_p(T^{-\delta})$ , and hence

$$\begin{aligned} \varepsilon_{\|a\|} &= \frac{a'\varepsilon_a}{\|a\|} + \frac{\varepsilon'_a\varepsilon_a}{2\|a\|} - \frac{(a'\varepsilon_a)^2}{2\|a\|^3} + O_p(T^{-3\delta}), \\ \varepsilon_{\|b\|} &= \frac{b'\varepsilon_b}{\|b\|} + \frac{\varepsilon'_b\varepsilon_b}{2\|b\|} - \frac{(b'\varepsilon_b)^2}{2\|b\|^3} + O_p(T^{-3\delta}). \end{aligned}$$

Then equation (40) becomes

$$\begin{aligned}\sin^2(\hat{a}_\delta, \hat{b}_\delta) - \sin^2(a, b) &= \frac{1}{\|b\|^2} (\varepsilon_a - \varepsilon_b)' \left[ I_k - \frac{bb'}{\|b\|^2} \right] (\varepsilon_a - \varepsilon_b) + O_p(T^{-3\delta}) \\ &= O_p(T^{-2\delta}).\end{aligned}$$

■

**Proof of Theorem 3.** From equation (40) we know that under  $H_0$  asymptotically the following results holds

$$T^2 \hat{D} = \frac{1}{\|b\|^2} (\varepsilon_a - \varepsilon_b)' \left[ I_k - \frac{bb'}{\|b\|^2} \right] (\varepsilon_a - \varepsilon_b) + o_p(1). \quad (41)$$

Under  $H_0$  we know that  $aa' = bb'$ , and from equations (18)-(19) we know that

$$\begin{aligned}\varepsilon_b &= \hat{b} - b = \left( \sum_{t=1}^T z_t z_t' \right)^{-1} \left( \sum_{t=1}^T z_t \bar{s}_t \right), \\ \varepsilon_\Gamma &= \hat{\Gamma} - \Gamma = \left( \sum_{t=1}^T z_t z_t' \right)^{-1} \left( \sum_{t=1}^T z_t \bar{v}_t' \right).\end{aligned}$$

Further, we know that  $\varepsilon_a = \varepsilon_M b + M \varepsilon_b$ , with  $\hat{M} = M + \varepsilon_M$  and  $\varepsilon_M = \Gamma' (\Gamma \Gamma')^{-1} \varepsilon_\Gamma + \varepsilon_\Gamma' (\Gamma \Gamma')^{-1} \Gamma - \Gamma' (\Gamma \Gamma')^{-1} \Gamma \varepsilon_\Gamma' (\Gamma \Gamma')^{-1} \Gamma - \Gamma' (\Gamma \Gamma')^{-1} \varepsilon_\Gamma \Gamma' (\Gamma \Gamma')^{-1} \Gamma$ .

From Assumption 1 we know that

$$\begin{aligned}T \varepsilon_b &\xrightarrow{d} \left( \int W_z W_z' \right)^{-1} \int W_z dW_{\bar{s}}, \\ T \varepsilon_\Gamma &\xrightarrow{d} \left( \int W_z W_z' \right)^{-1} \int W_z dW_{\bar{v}}.\end{aligned}$$

■

**Proof of Theorem 4.** Using (41) and recalling the definitions of  $\varepsilon_a$ ,  $\varepsilon_M$  and

$\varepsilon_b$ , from equation (22) we have

$$\varepsilon_b = \hat{b}^{PC} - b = \left( \hat{\Xi}^{PC} - \Xi \right)' F i_b,$$

$$\varepsilon_\Gamma = \hat{\Gamma}^{PC} - \Gamma = i_\Gamma' F' \left( \hat{\Xi}^{PC} - \Xi \right),$$

so that

$$T \varepsilon_b \xrightarrow{d} \Pi' F i_b,$$

$$T \varepsilon_\Gamma \xrightarrow{d} i_\Gamma F' \Pi.$$

■

**Proof of Proposition 4.** The limiting distribution of  $\hat{\Theta}^{PC}$  can be computed recalling that  $\hat{\Theta}^{PC} = F' \hat{\Xi}^{PC}$  and evaluating the limiting distribution of  $\hat{\Xi}^{PC}$ . Let  $\hat{z}_t$  be the principal component estimator for  $z_t$  based upon  $\hat{\Xi}^{PC}$ . Then we know (see e.g. the proof of Lemma 3 in Bai, 2004) that  $T \left( \hat{\Xi}^{PC} - \Xi \right)$  can be decomposed as

$$\begin{aligned} T \left( \hat{\Xi}^{PC} - \Xi \right) &= \\ &= \frac{1}{T} \left[ \sum_{t=1}^T e_t^W z_t' + \sum_{t=1}^T e_t^W (\hat{z}_t - z_t)' + \Xi \sum_{t=1}^T (z_t - \hat{z}_t) \hat{z}_t' \right] \\ &\quad \left[ \frac{1}{T^2} \sum_{t=1}^T \hat{z}_t \hat{z}_t' \right]^{-1}. \end{aligned} \quad (42)$$

In the denominator of (42), we can rewrite

$$\begin{aligned} \sum_{t=1}^T \hat{z}_t \hat{z}_t' &= \sum_{t=1}^T z_t z_t' + \sum_{t=1}^T (\hat{z}_t - z_t) \hat{z}_t' + \sum_{t=1}^T \hat{z}_t (\hat{z}_t - z_t)' + \sum_{t=1}^T (\hat{z}_t - z_t) (\hat{z}_t - z_t)' \\ &= I + II + III + IV. \end{aligned}$$

We know that

$$I = O_p(T^2);$$



from Lemma B.4(ii) in Bai (2004)

$$II = III = O_p(T)$$

and from Lemma B.1 in Bai (2004)

$$IV = O_p(T),$$

Therefore the denominator of (42) is

$$T^{-2} \sum_{t=1}^T \hat{z}_t \hat{z}'_t = T^{-2} \sum_{t=1}^T z_t z'_t + O_p(T^{-1})$$

and thus

$$T^{-2} \sum_{t=1}^T \hat{z}_t \hat{z}'_t \xrightarrow{d} \int W_z W'_z = B.$$

As far as the numerator (42) is concerned, let

$$\frac{1}{T} \left[ \sum_{t=1}^T e_t^W z'_t + \sum_{t=1}^T e_t^W (\hat{z}_t - z_t)' + \Xi \sum_{t=1}^T (z_t - \hat{z}_t) \hat{z}'_t \right] = A + B + C.$$

We have that  $A \xrightarrow{d} \int dW_e W'_z$ . To study the the limiting distribution of  $B$  and  $C$ , consider the following decomposition as proposed in Bai (2004, p. 164) for the definition of  $\tilde{z}_t$ :

$$\hat{z}_t - z_t = T^{-2} \sum_{s=1}^T \tilde{z}_s \gamma_n(s, t) + T^{-2} \sum_{s=1}^T \tilde{z}_s \zeta_{st} + T^{-2} \sum_{s=1}^T \tilde{z}_s \eta_{st} + T^{-2} \sum_{s=1}^T \tilde{z}_s \xi_{st},$$

where

$$\begin{aligned}
\gamma_n(s, t) &= E(e_t^{W'} e_s^W / n) \\
\zeta_{st} &= e_t^{W'} e_s^W / n - \gamma_n(s, t) \\
\eta_{st} &= z_s' \Xi' e_t^W / n \\
\xi_{st} &= z_t' \Xi' e_s^W / n.
\end{aligned}$$

Then we have

$$\begin{aligned}
B &= T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W \tilde{z}_s' \gamma_n(s, t) + T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W \tilde{z}_s' \zeta_{st} + T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W \tilde{z}_s' \eta_{st} + \\
&\quad T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W \tilde{z}_s' \xi_{st}, \\
&= n^{-1} T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W e_t^{W'} e_s^W \tilde{z}_s' + T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W \tilde{z}_s' \eta_{st} + T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W \tilde{z}_s' \xi_{st}, \\
&= I + II + III
\end{aligned}$$

Then

$$I = n^{-1} T^{-1} \left( T^{-1} \sum_{t=1}^T e_t^W e_t^{W'} \right) \left( T^{-1} \sum_{s=1}^T e_s^W \tilde{z}_s' \right) = O_p(T^{-1});$$

$$\begin{aligned}
II &= n^{-1} T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W \tilde{z}_s' z_s' \Xi' e_t^W \\
&= n^{-1} T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W e_t^{W'} \Xi z_s \tilde{z}_s' \\
&= n^{-1} \left( T^{-1} \sum_{t=1}^T e_t^W e_t^{W'} \right) \Xi \left( T^{-2} \sum_{s=1}^T z_s \tilde{z}_s' \right) = O_p(1);
\end{aligned}$$

and

$$\begin{aligned}
III &= n^{-1}T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W \tilde{z}'_s z'_t \Xi' e_s^W \\
&= n^{-1}T^{-3} \sum_{s=1}^T \sum_{t=1}^T e_t^W z'_t \Xi' e_s^W \tilde{z}'_s \\
&= n^{-1}T^{-1} \left( T^{-1} \sum_{t=1}^T e_t^W z'_t \right) \Xi' \left( T^{-1} \sum_{s=1}^T e_s^W \tilde{z}'_s \right) = O_p(T^{-1}).
\end{aligned}$$

Therefore the only term that matters is  $II$  and thus

$$n^{-1} \left( T^{-1} \sum_{t=1}^T e_t^W e_t^{W'} \right) \Xi \left( T^{-2} \sum_{s=1}^T z_s \tilde{z}'_s \right) \xrightarrow{d} n^{-1} \Omega_e \Xi B,$$

where the distribution limit  $T^{-2} \sum_{s=1}^T z_s \tilde{z}'_s \xrightarrow{d} B$  follows from the same argument as in the proof of the denominator.

Finally, as far as the term  $C$  of the numerator is concerned we have

$$\begin{aligned}
C &= -T^{-3} \sum_{s=1}^T \sum_{t=1}^T \tilde{z}_s \tilde{z}'_t \gamma_n(s, t) - T^{-3} \sum_{s=1}^T \sum_{t=1}^T \tilde{z}_s \tilde{z}'_t \zeta_{st} - T^{-3} \sum_{s=1}^T \sum_{t=1}^T \tilde{z}_s \tilde{z}'_t \eta_{st} \\
&\quad - T^{-3} \sum_{s=1}^T \sum_{t=1}^T \tilde{z}_s \tilde{z}'_t \xi_{st} \\
&= I + II + III + IV.
\end{aligned}$$

From Lemma B.4 in Bai (2004) we have that

$$I = O_p(T^{-1})$$

$$II = O_p(T^{-1}).$$

As far as terms *III* and *IV* are concerned, we have that

$$\begin{aligned}
III &= n^{-1}T^{-3} \sum_{s=1}^T \sum_{t=1}^T \tilde{z}_s \hat{z}'_t z'_s \Xi' e_t^W \\
&= n^{-1}T^{-3} \sum_{s=1}^T \sum_{t=1}^T z_s \tilde{z}'_s \Xi' e_t^W \hat{z}'_t \\
&= n^{-1} \left( T^{-2} \sum_{s=1}^T z_s \tilde{z}'_s \right) \Xi' \left( T^{-1} \sum_{t=1}^T e_t^W \hat{z}'_t \right) = O_p(1),
\end{aligned}$$

and

$$\begin{aligned}
IV &= n^{-1}T^{-3} \sum_{s=1}^T \sum_{t=1}^T \tilde{z}_s \hat{z}'_t z'_t \Xi' e_s^W \\
&= n^{-1}T^{-3} \sum_{s=1}^T \sum_{t=1}^T \tilde{z}_s e_s^{W'} \Xi z_t \hat{z}'_t \\
&= n^{-1} \left( T^{-1} \sum_{s=1}^T \tilde{z}_s e_s^{W'} \right) \Xi' \left( T^{-1} \sum_{t=1}^T z_t \hat{z}'_t \right) = O_p(1).
\end{aligned}$$

Thus, the limiting distribution of  $C$  is determined by *III* and *IV*, and we have

$$\begin{aligned}
III &= n^{-1} \left( T^{-2} \sum_{s=1}^T z_s \tilde{z}'_s \right) \Xi' \left( T^{-1} \sum_{t=1}^T e_t^W z'_t \right) + \\
&\quad n^{-1} \left( T^{-2} \sum_{s=1}^T z_s \tilde{z}'_s \right) \Xi' \left[ T^{-1} \sum_{t=1}^T e_t^W (\hat{z}_t - z_t)' \right] \\
&\xrightarrow{d} n^{-1} B \Xi' \left[ \int dW_e W'_z + n^{-1} \Omega_e \Xi B \right],
\end{aligned}$$

and

$$IV \xrightarrow{d} n^{-1} \left[ \int W_z dW'_e + n^{-1} B \Xi' \Omega_e \right] \Xi B.$$

Combining the results above, the distribution (42) is

$$\begin{aligned}
& T \left( \hat{\Xi}^{PC} - \Xi \right) \xrightarrow{d} \left[ \int dW_e W'_z + n^{-1} \Omega_e \Xi B - n^{-1} B \Xi' \left( \int dW_e W'_z + n^{-1} \Omega_e \Xi B \right) \right. \\
& \quad \left. - n^{-1} \left( \int W_z dW'_e + n^{-1} B \Xi' \Omega_e \right) \Xi B \right] B^{-1} \\
& = \left( \int dW_e W'_z \right) B^{-1} + n^{-1} \Omega_e \Xi - n^{-1} B \Xi' \int dW_e W'_z B^{-1} \\
& \quad - n^{-2} B \Xi' \Omega_e \Xi - n^{-1} \int W_z dW'_e \Xi - n^{-2} B \Xi' \Omega_e \Xi.
\end{aligned}$$

■

**Proof of Theorem 5.** Let  $b_0 = \Gamma\beta$ . Under  $H_1^l$ ,

$$\|b\| = \|b_0\| + R_T = \|a\| + R_T,$$

where

$$R_T = \frac{1}{2} \frac{\delta'_T \delta_T}{\|a\|} = O_p(T^{-2}),$$

which follows from applying Taylor's expansion to  $\|b\|$  and that  $\delta'_T \Gamma = 0$ .

Moreover

$$\varepsilon_{\|b\|} = \varepsilon_{\|b_0\|} + \frac{\delta'_T \varepsilon_b}{\|b_0\|} - R_T b' \varepsilon_b,$$

which follows from

$$\varepsilon_{\|b_0\|} = \frac{b' \varepsilon_b}{\|b\|} + \frac{\varepsilon'_b \varepsilon_b}{2 \|b\|} - \frac{(b' \varepsilon_b)^2}{2 \|b\|^3} + O_p(T^{-3}),$$

and application of Taylor's expansion to  $\|b\|^{-1}$ .

Also, from  $\delta'_T \Gamma = 0$  it follows  $a'b = a'b_0 = \|a\|^2$ . We know from equation (37)

that

$$\sin^2(\hat{a}, \hat{b}) = \sin^2(a, b) + [2 + o_p(1)] \frac{\|\hat{a}\| \|\hat{b}\| (a'b) - \|a\| \|b\| (\hat{a}'\hat{b})}{\|\hat{a}\| \|\hat{b}\| \|a\| \|b\|},$$

with  $\|\hat{a}\| \|\hat{b}\| \|a\| \|b\| = \|a\|^2 + o_p(1)$ .

As far as  $\sin^2(a, b)$  is concerned, we have

$$\begin{aligned}
\sin^2(a, b) &= 1 - \frac{a'b}{\|a\| \|b\|} \\
&= 1 - \frac{\|a\|^2}{\|a\| (\|a\| + R_T)} \\
&= \frac{R_T}{\|a\| (\|a\| + R_T)} = O(T^{-2}).
\end{aligned} \tag{43}$$

Consider the numerator  $\|\hat{a}\| \|\hat{b}\| (a'b) - \|a\| \|b\| (\hat{a}'\hat{b})$ , we have

$$\begin{aligned}
& [\|a\| + \varepsilon_{\|a\|}] [\|b\| + \varepsilon_{\|b\|}] (a'b) - \|a\| \|b\| [a'b + a'\varepsilon_b + b'\varepsilon_a + \varepsilon'_a\varepsilon_b] \\
&= [\|a\| + \varepsilon_{\|a\|}] [\|a\| + R_T + \varepsilon_{\|b_0\|} + \|b_0\|^{-1} \delta'_T \varepsilon_b - R_T b' \varepsilon_b] \|a\|^2 - \\
& \quad \|a\| [\|a\| + R_T] [a'b + a'\varepsilon_b + b'_0\varepsilon_a + \delta'_T \varepsilon_a + \varepsilon'_a\varepsilon_b] \\
&= \frac{\|a\|^2}{2} (\varepsilon_a - \varepsilon_b)' \left[ I_k - \frac{aa'}{\|a\|^2} \right] (\varepsilon_a - \varepsilon_b) - \\
& \quad \|a\|^2 \delta'_T (\varepsilon_a - \varepsilon_b) + O_p(T^{-3}).
\end{aligned} \tag{44}$$

Combining equations (43) and (44), we finally have

$$\begin{aligned}
\sin^2(\hat{a}, \hat{b}) &= \frac{R_T}{\|a\| (\|a\| + R_T)} + \frac{1}{\|a\|^2} (\varepsilon_a - \varepsilon_b)' \left[ I_k - \frac{aa'}{\|a\|^2} \right] (\varepsilon_a - \varepsilon_b) \\
& \quad - \frac{2}{\|a\|^2} \delta'_T (\varepsilon_a - \varepsilon_b) + O_p(T^{-3}).
\end{aligned}$$

Thus, the limiting distribution of  $\hat{D} = \sin^2(\hat{a}, \hat{b})$  is

$$T^2 \hat{D} \xrightarrow{d} \frac{\|\delta\|^2}{\|a\|^2} + \frac{1}{\|a\|^2} Q^{*'} \left[ I_k - \frac{aa'}{\|a\|^2} \right] Q^* - \frac{2}{\|a\|^2} \delta' Q^*.$$

■

**Proof of Theorem 6.** We prove the Theorem, merely for the sake of the notation and with no loss of generality, by considering alternative hypotheses  $H_1$  of the form

$$H_1 : b = \Gamma' \beta + \delta,$$

where the  $k$ -dimensional vector  $\delta$  is, as in the local alternative case, orthogonal to  $\Gamma$ , i.e.  $\delta'\Gamma = 0$ . Let  $b_0 = \Gamma'\beta$  and  $k = \|\delta\| / \|a\|$ . From condition  $\delta'\Gamma = 0$ , under  $H_1$ ,  $a = b_0$  and

$$\|b\| = \|b_0\| \sqrt{1 + k^2} = \|a\| \sqrt{1 + k^2}.$$

Therefore, it holds that

$$\begin{aligned} D &= \sin^2(a, b) \\ &= 1 - \left( \frac{a'b}{\|a\| \|b\|} \right)^2 \\ &= \frac{k^2}{1 + k^2} > 0. \end{aligned} \tag{45}$$

We know that

$$\begin{aligned} \widehat{D} &= \sin^2(\widehat{a}, \widehat{b}) \\ &= \sin^2(a, b) + \left[ \cos(\widehat{a}, \widehat{b}) + \cos(a, b) \right] \left[ \cos(a, b) - \cos(\widehat{a}, \widehat{b}) \right] \\ &= \frac{k^2}{1 + k^2} + [2 \cos(a, b) + o_p(1)] \frac{\|\widehat{a}\| \|\widehat{b}\| (a'b) - \|a\| \|b\| (\widehat{a}'\widehat{b})}{\|\widehat{a}\| \|\widehat{b}\| \|a\| \|b\|}. \end{aligned}$$

From equation (45) it follows that

$$\cos(a, b) = \frac{1}{\sqrt{1 + k^2}}. \tag{46}$$

As far as the term

$$\frac{\|\widehat{a}\| \|\widehat{b}\| (a'b) - \|a\| \|b\| (\widehat{a}'\widehat{b})}{\|\widehat{a}\| \|\widehat{b}\| \|a\| \|b\|}$$

is concerned, we have, with respect to the denominator and after Slutsky's theorem

$$\begin{aligned} \|\widehat{a}\| \|\widehat{b}\| \|a\| \|b\| &= \|a\|^2 \|b\|^2 + o_p(1) \\ &= \|a\|^4 (1 + k^2) + o_p(1). \end{aligned} \tag{47}$$

As far as the numerator is concerned, we have

$$\begin{aligned}
& [\|a\| + \varepsilon_{\|a\|}] [\|b\| + \varepsilon_{\|b\|}] (a'b) - \|a\| \|b\| [a'b + a'\varepsilon_b + b'\varepsilon_a + \varepsilon'_a\varepsilon_b] \\
= & [\|b_0\| + \varepsilon_{\|a\|}] \left[ \|b_0\| \sqrt{1+k^2} + \varepsilon_{\|b\|} \right] \|b_0\|^2 - \\
& \|b_0\|^2 \sqrt{1+k^2} [\|b_0\|^2 + b'_0\varepsilon_b + b'_0\varepsilon_a + \delta'\varepsilon_a + \varepsilon'_a\varepsilon_b] \\
= & \|b_0\|^2 \left[ \|b_0\| \sqrt{1+k^2} \varepsilon_{\|a\|} + \|b_0\| \varepsilon_{\|b\|} - \sqrt{1+k^2} b'_0 (b'_0\varepsilon_b + b'_0\varepsilon_a + \delta'\varepsilon_a) + O_p(T^{-2}) \right].
\end{aligned}$$

Recalling the definitions of  $\varepsilon_{\|a\|}$  and  $\varepsilon_{\|b\|}$  given in equations (38) and (39) respectively, we have

$$\begin{aligned}
& \|b_0\|^2 \left[ \left( \frac{1}{\sqrt{1+k^2}} - \sqrt{1+k^2} \right) b'_0\varepsilon_b + \frac{\delta'\varepsilon_b}{\sqrt{1+k^2}} - \sqrt{1+k^2} \delta'\varepsilon_a \right] + O_p(T^{-2}) \\
= & O_p(T^{-1}).
\end{aligned}$$

Combining this with equations (45), (46) and (47), we obtain

$$\hat{D} = \frac{k^2}{1+k^2} + \frac{2}{\|b_0\|^2} \left[ \delta' \left( \frac{\varepsilon_b}{1+k^2} - \varepsilon_a \right) - \frac{k^2}{1+k^2} b'_0\varepsilon_b \right] + O_p(T^{-2}),$$

where

$$\frac{2}{\|b_0\|^2} \left[ \delta' \left( \frac{\varepsilon_b}{1+k^2} - \varepsilon_a \right) - \frac{k^2}{1+k^2} b'_0\varepsilon_b \right] = O_p(T^{-1}).$$

■

**Proof of Theorem 7.** To prove the theorem, consider the following preliminary result which states the distributional equivalence between the quantities  $T(\hat{\Theta}^{OLS} - \Theta)$  and  $T(\hat{\Theta}^{PC} - \Theta)$  with their bootstrap counterpart  $T(\Theta^* - \tilde{\Theta}^{OLS})$  and  $T(\Theta^* - \tilde{\Theta}^{PC})$  respectively.

**Lemma A.1** *Consider the estimators  $\hat{\Theta}^{OLS}$  and  $\hat{\Theta}^{PC}$  of  $\Theta$  and their linear transformations  $\tilde{\Theta}^{OLS}$  and  $\tilde{\Theta}^{PC}$  defined in Step (1.1) of the bootstrap algorithm. Let  $\Theta^*$  be the bootstrap estimator for  $\tilde{\Theta}^{OLS}$  and  $\tilde{\Theta}^{PC}$ , and define the limiting distribution of  $T(\hat{\Theta}^{OLS} - \Theta)$  and  $T(\hat{\Theta}^{PC} - \Theta)$  as  $Z_{\Theta}^{OLS}$  and  $Z_{\Theta}^{PC}$  respectively. Then it holds*



that

$$T \left( \Theta^* - \tilde{\Theta}^{OLS} \right) \xrightarrow{d} Z_{\Theta}^{OLS},$$

and

$$T \left( \Theta^* - \tilde{\Theta}^{OLS} \right) \xrightarrow{d} Z_{\Theta}^{PC}.$$

**Proof.** We distinguish the case of  $z_t$ s observable from that in which the  $z_t$ s are unobservable.

*The case of  $z_t$ s observable.* The proof is based on the three following steps: (1) we derive a strong approximation for the limiting distribution of the partial sums of the process  $\eta_{qt}$ ; (2) we derive the strong approximation for the bootstrap counterpart  $\eta_{qt}^*$ ; (3) we extend these results to the limiting distribution of processes  $\hat{w}_t$  and  $\hat{w}_t^*$ .

(1) Define  $S_{\eta}(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \eta_{qt}$ . Assumption 3(i) ensures that an invariance principle holds such that  $S_{\eta}(r) \xrightarrow{d} W(r)$ , where  $W(r)$  is a Brownian motion. Following Sakhanenko's (1980) and Park (2002), for some  $l > 2$  and for any  $\delta > 0$ , the following strong approximation holds

$$P \left\{ \sup_{0 \leq r \leq 1} |S_{\eta}(r) - W(r)| \geq \delta \right\} \leq T^{1-l/2} K_l \left\{ E |\eta_t|^l \right\},$$

where  $K_l$  is an absolute constant depending only on  $l$ .

(2) Define  $S_{\eta}^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \eta_{qt}^*$ . Similarly:

$$P \left\{ \sup_{0 \leq r \leq 1} |S_{\eta}^*(r) - W(r)| \geq \delta \right\} \leq T^{1-l/2} K_l \left\{ E |\eta_{qt}^*|^l \right\}.$$

Thus, from our resampling scheme

$$E |\eta_{qt}^*|^l = \frac{1}{T} \sum_{t=1}^T \left| \hat{\eta}_{qt} - \frac{1}{T} \sum_{t=1}^T \hat{\eta}_{qt} \right|^l.$$

Assumption 3(i) and the law of large numbers ensure that  $E |\eta_{qt}^*|^l < \infty$ .

Hence, as  $T \rightarrow \infty$

$$P \left\{ \sup_{0 \leq r \leq 1} |S_{\eta}^*(r) - W(r)| \geq \delta \right\} = 0$$

This proves the strong approximation is valid for the bootstrap  $\eta_{qt}^*$ .

(3) Following Chang, Park and Song (2006), the bootstrap invariance principle for  $\eta_{qt}^*$  carries over to  $w_t^*$  provided that the  $\hat{\Psi}_k$  are consistent estimators for  $\Psi_k$ . Assumption 3(ii) ensures that  $\hat{\Psi}_k$  is a consistent estimator for  $\Psi_k$ . See also Chang and Park (2002, 2003).

It holds

$$T \left( \hat{\Theta}^{OLS} - \Theta \right) \xrightarrow{d} \left( \int dB_W W_z' \right) \left( \int W_z W_z' \right)^{-1},$$

where  $B_W$  is the Brownian motion associated with the partial sums of  $\bar{e}_t$ .

Thus, it holds:

$$T \left( \Theta^* - \tilde{\Theta}^{OLS} \right) \xrightarrow{d} \left( \int dB_W W_z' \right) \left( \int W_z W_z' \right)^{-1}.$$

The use of the continuous mapping theorem leads to equation (35), under the null, for the case when  $z_t$  is observed.

*The case of  $z_t$ s unobservable.* Though this part of the proof is similar to the case where  $z_t$  is observable, however in this case the error term  $w_t$  also contains the extra component  $\Theta(z_t - \hat{z}_t)$ , which leads to different asymptotics. It is natural in this case to derive the proof directly for  $\tilde{w}_t$ .

From (34), we know that

$$\bar{W}_t^* = \tilde{\Theta}^{PC} \tilde{z}_t^* + \tilde{e}_t^*.$$

Since in this case the bootstrap estimator  $\Theta^*$  is given by

$$\Theta^* = \left[ \sum_{t=1}^T \bar{W}_t^* \tilde{z}_t^{*'} \right] \left[ \sum_{t=1}^T \tilde{z}_t^* \tilde{z}_t^{*'} \right]^{-1},$$

we have

$$\Theta^* - \tilde{\Theta}^{PC} = \left[ \sum_{t=1}^T \tilde{e}_t^* \tilde{z}_t^{*'} \right] \left[ \sum_{t=1}^T \tilde{z}_t^* \tilde{z}_t^{*'} \right]^{-1}. \quad (48)$$

(1) Define  $X_T(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \tilde{w}_t$  and  $X(r)$  the corresponding limiting distribution as  $T \rightarrow \infty$ , i.e.  $X_T(r) \xrightarrow{d} X(r)$ . Markov inequality ensures that, for any  $\delta > 0$  and some  $l > 2$

$$P \left\{ \sup_{0 \leq r \leq 1} |X_T(r) - X(r)| > \delta \right\} \leq \delta^{-l} E \left[ \sup_{0 \leq r \leq 1} |X_T(r) - X(r)|^l \right].$$

From martingale theory, we have

$$E \left[ \sup_{0 \leq r \leq 1} |X_T(r) - X(r)|^l \right] \leq c_l T \left\{ E |T^{-1/2} \tilde{w}_t|^l \right\} = T^{1-1/2l} \left\{ E |\tilde{w}_t|^l \right\},$$

where  $c_l$  is an absolute constant. Thus,

$$P \left\{ \sup_{0 \leq r \leq 1} |X_T(r) - X(r)| > \delta \right\} \leq \delta^{-l} T^{1-1/2l} \left\{ E |\tilde{w}_t|^l \right\}.$$

This result provides an assessment of the rate of convergence of  $X_T$  to its limiting distribution  $X$  and mimics the strong approximation result in Sakhanenko (1980) used by Park (2002).

(2) In the same fashion, define  $X_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \tilde{w}_t^*$ , we can write a similar result as above

$$P \left\{ \sup_{0 \leq r \leq 1} |X_T^*(r) - X(r)| > \delta \right\} \leq \delta^{-l} T^{1-1/2l} E |\tilde{w}_t^*|^l,$$

and from our resampling scheme we have

$$E |\tilde{w}_t^*|^l = \frac{1}{T} \sum_{t=1}^T \left| \tilde{w}_t - \frac{1}{T} \sum_{t=1}^T \tilde{w}_t \right|^l.$$

Give that  $\tilde{w}_t^* = [\tilde{e}_t^{*'}, \Delta \hat{z}_t^{*'}]'$ , in order to prove that  $E |\tilde{w}_t^*|^l$  is finite we need to show

that both  $E |\tilde{e}_t^*|^l$  and  $E |\Delta \hat{z}_t^*|^l$  are finite. Assumption 3(i) ensures that  $\tilde{e}_t^*$  has finite 4th moment, and therefore

$$E |\tilde{e}_t^*|^l = \frac{1}{T} \sum_{t=1}^T \left| \bar{e}_t - \frac{1}{T} \sum_{t=1}^T \bar{e}_t \right|^l$$

is finite.

As far as  $E |\Delta \hat{z}_t^*|^l$  is concerned, let us consider the quantity  $T^{-1} \sum_{t=1}^T |\Delta \check{z}_t|^l$ , where  $\Delta \check{z}_t = \Delta \hat{z}_t - T^{-1} \sum_{t=1}^T \Delta \hat{z}_t$ , and let  $\Delta \bar{z}_t = \Delta z_t - T^{-1} \sum_{t=1}^T \Delta z_t$ . Thus we have that

$$\begin{aligned} E |\Delta \hat{z}_t^*|^l &= \frac{1}{T} \sum_{t=1}^T |\Delta \check{z}_t|^l = \\ &= \frac{1}{T} \sum_{t=1}^T |\Delta \bar{z}_t + (\Delta \check{z}_t - \Delta \bar{z}_t)|^l \\ &\leq \frac{1}{T} \sum_{t=1}^T |\Delta \bar{z}_t|^l + \frac{1}{T} \sum_{t=1}^T |\Delta \check{z}_t - \Delta \bar{z}_t|^l. \end{aligned} \quad (49)$$

We have that the first term in the inequality above,  $T^{-1} \sum_{t=1}^T |\Delta \bar{z}_t|^l$ , is finite from Assumption 3(i). As far as the second term,  $T^{-1} \sum_{t=1}^T |\Delta \check{z}_t - \Delta \bar{z}_t|^l$  is concerned, we have

$$\begin{aligned} \Delta \check{z}_t - \Delta \bar{z}_t &= T^{-2} \sum_{s=1}^T \Delta \check{z}'_s \Delta e_s^{W'} \Delta e_t^W + T^{-2} \sum_{s=1}^T \Delta \check{z}'_s \Delta e_s^{W'} \Xi \Delta z_t \\ &\quad + T^{-2} \sum_{s=1}^T \Delta \check{z}'_s \Delta z'_s \Xi \Delta e_t^{W'}, \end{aligned}$$

and

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T \left| T^{-1} \sum_{s=1}^T \Delta \tilde{z}'_s \Delta e_s^{W'} \Delta e_t^W \right|^l &= \left\| T^{-1} \sum_{s=1}^T \Delta \tilde{z}'_s \Delta e_s^{W'} \right\|^l T^{-1} \sum_{t=1}^T \|\Delta e_t^W\|^l \\
&= O(T^{-l/2}), \\
\frac{1}{T} \sum_{t=1}^T \left| T^{-1} \sum_{s=1}^T \Delta \tilde{z}'_s \Delta e_s^{W'} \Xi \Delta z_t \right|^l &= \left\| T^{-1} \sum_{s=1}^T \Delta \tilde{z}'_s \Delta e_s^{W'} \right\|^l T^{-1} \sum_{t=1}^T \|\Xi \Delta z_t\|^l \\
&= O(T^{-l/2}), \\
\frac{1}{T} \sum_{t=1}^T \left| T^{-1} \sum_{s=1}^T \Delta \tilde{z}'_s \Delta z'_s \Xi \Delta e_t^{W'} \right|^l &= \left\| T^{-1} \sum_{s=1}^T \Delta \tilde{z}'_s \Delta z'_s \right\|^l T^{-1} \sum_{t=1}^T \|\Delta e_t^{W'}\|^l \\
&= O(T^{-l/2}). \tag{50}
\end{aligned}$$

Therefore, we have that  $E |\Delta \hat{z}_t^*|^l$  is finite.

From (49) and (50), the vector  $E |\tilde{w}_t^*|^l$  is finite.

As  $T \rightarrow \infty$ ,

$$P \left\{ \sup_{0 \leq r \leq 1} |X_T^*(r) - X(r)| > \delta \right\} = 0. \tag{51}$$

This result jointly with continuous mapping theorem prove that numerator in (48)

is  $T^{-2} \sum_{t=1}^T \tilde{z}_t^* \tilde{z}_t^{*'} \xrightarrow{d} \int W_z W_z'$ .

As far as the numerator in (48) is concerned, we have

$$\sum_{t=1}^T \tilde{e}_t^* \tilde{z}_t^{*'} = \sum_{t=1}^T \bar{e}_t^* \tilde{z}_t^{*'} + \sum_{t=1}^T \Theta(z_t^* - \tilde{z}_t^*) \tilde{z}_t^{*'} + o_p^*(1) \tag{52}$$

Expression (51) ensures a strong approximation result holds for the partial sums of  $\tilde{z}_t^*$ ,  $z_t^* - \tilde{z}_t^*$  and  $\bar{e}_t^*$ .

Therefore, continuous mapping theorem and consistency of the  $\hat{\Psi}_k$ s ensured by Assumption 3(ii), lead to

$$\begin{aligned}
T^{-1} \sum_{t=1}^T \Theta(z_t^* - \tilde{z}_t^*) \tilde{z}_t^{*'} &\xrightarrow{d} n^{-1} F' B \Xi' \left[ \int dW_e W_z' + n^{-1} \Omega_e \Xi B \right] \\
&+ n^{-1} F' \left[ \int W_z dW_e' + n^{-1} B \Xi' \Omega_e \right] \Xi B, \tag{53}
\end{aligned}$$

which is the same result as for  $T^{-1} \sum_{t=1}^T \Theta(z_t - \hat{z}_t) \hat{z}'_t$ .

Combining the results from equations (52) and (53), we obtain

$$\begin{aligned} & T \left( \Theta^* - \tilde{\Theta}^{PC} \right) \xrightarrow{d} \\ & F' \left[ \int dW_e W'_z + n^{-1} \Omega_e \Xi B - n^{-1} B \Xi' \left( \int dW_e W'_z + n^{-1} \Omega_e \Xi B \right) \right. \\ & \left. - n^{-1} \left( \int W_z dW'_e + n^{-1} B \Xi' \Omega_e \right) \Xi B \right] B^{-1}, \end{aligned}$$

which is the same as the distribution of  $T \left( \hat{\Theta}^{PC} - \Theta \right)$  provided in Theorem 4. Therefore, we have that  $T \left( \Theta^* - \tilde{\Theta}^{PC} \right)$  and  $T \left( \hat{\Theta}^{PC} - \Theta \right)$  are equal in distribution. QED.

Lemma A.1 ensures the distributional equivalence between  $\hat{\Theta}^{OLS}$  and  $\hat{\Theta}^{PC}$  with their bootstrap counterpart  $\Theta^*$ . Therefore, after the continuous mapping theorem, letting

$$\varepsilon_b^* = b^* - \tilde{b},$$

$$\varepsilon_\Gamma^* = \Gamma^* - \tilde{\Gamma},$$

we have

$$T \varepsilon_b^* \xrightarrow{d} \left( \int W_z W'_z \right)^{-1} \int W_z dW_{\bar{s}},$$

$$T \varepsilon_\Gamma^* \xrightarrow{d} \left( \int W_z W'_z \right)^{-1} \int W_z dW'_{\bar{v}},$$

if the  $z_t$ s are observable and

$$T \varepsilon_b^* \xrightarrow{d} \Pi' F i_b,$$

$$T \varepsilon_\Gamma^* \xrightarrow{d} i_\Gamma F' \Pi.$$

if the  $z_t$ s are unobservable, where  $F$  and  $\Pi$  are defined in equation (23) and Theorem 4 respectively.

We can now prove equation (35) by analysing the asymptotic behaviour of  $\hat{D}^*$ .

We have

$$\begin{aligned}\widehat{D}^* &= \sin^2(a^*, b^*) \\ &= \sin^2(\tilde{a}, \tilde{b}) + \left[ \cos(a^*, b^*) + \cos(\tilde{a}, \tilde{b}) \right] \frac{\|a^*\| \|b^*\| (\tilde{a}'\tilde{b}) - \|\tilde{a}\| \|\tilde{b}\| (a^{*\prime}b^*)}{\|a^*\| \|b^*\| \|\tilde{a}\| \|\tilde{b}\|}.\end{aligned}$$

Since  $a^*$  and  $b^*$  are superconsistent estimators, by Slutsky's theorem we have

$$\cos(a^*, b^*) = \cos(\tilde{a}, \tilde{b}) + o_p(1),$$

and by definition of  $\tilde{b}$  we have

$$\sin(\tilde{a}, \tilde{b}) = 0,$$

$$\cos(\tilde{a}, \tilde{b}) = 1.$$

Therefore

$$\widehat{D}^* = [2 + o_p(1)] \frac{\|\tilde{b}\|^2 \|a^*\| \|b^*\| - a^{*\prime}b^*}{\|\tilde{b}\|^4 + o_p(1)}.$$

Since

$$\begin{aligned}& \|a^*\| \|b^*\| - a^{*\prime}b^* \\ &= (\|\tilde{a}\| + \varepsilon_{\|a\|}^*) (\|\tilde{a}\| + \varepsilon_{\|b\|}^*) - \tilde{a}'\tilde{b} - \tilde{a}'\varepsilon_b^* - \tilde{a}'\varepsilon_a^* - \varepsilon_a^{*\prime}\varepsilon_b^* \\ &= \|\tilde{a}\| \varepsilon_{\|a\|}^* + \|\tilde{a}\| \varepsilon_{\|b\|}^* + \varepsilon_{\|a\|}^* \varepsilon_{\|b\|}^* - \tilde{a}'\varepsilon_b^* - \tilde{a}'\varepsilon_a^* - \varepsilon_a^{*\prime}\varepsilon_b^* \\ &= \frac{1}{2} (\varepsilon_a^* - \varepsilon_b^*)' \left[ I_k - \frac{\tilde{a}\tilde{a}'}{\|\tilde{a}\|^2} \right] (\varepsilon_a^* - \varepsilon_b^*) + O_p(T^{-3}),\end{aligned}$$

we have that

$$T^2 \widehat{D}^* = \frac{1}{\|\tilde{b}\|^2} (\varepsilon_a^* - \varepsilon_b^*)' \left[ I_k - \frac{\tilde{b}\tilde{b}'}{\|\tilde{b}\|^2} \right] (\varepsilon_a^* - \varepsilon_b^*) + o_p(1).$$

Lemma A.1 and the continuous mapping theorem ensure that equation (35) holds.

Note that since resampling was done under the null via the use of  $\tilde{b}$ , equation (35)

is valid not only under the null but also under the alternative hypothesis  $H_1$  (and under the local alternatives  $H_1^l$ ). ■



	WN	AR(1)	MA(1)	WN	AR(1)	MA(1)	WN	AR(1)	MA(1)	WN	AR(1)	MA(1)
	$k = 2$			$k = 3$			$k = 4$			$k = 5$		
<b>T</b>	$\rho = .75$		$\vartheta = .75$	$\rho = .75$		$\vartheta = .75$	$\rho = .75$		$\vartheta = .75$	$\rho = .75$		$\vartheta = .75$
<b>Size</b>												
<b>20</b>	0.097	0.405	0.133	0.096	0.224	0.130	0.063	0.235	0.205	0.044	0.205	0.126
<b>35</b>	0.092	0.296	0.095	0.080	0.183	0.098	0.053	0.202	0.107	0.035	0.186	0.078
<b>50</b>	0.088	0.213	0.079	0.065	0.133	0.072	0.046	0.151	0.074	0.038	0.149	0.057
<b>100</b>	0.083	0.107	0.061	0.063	0.078	0.044	0.042	0.073	0.035	0.027	0.071	0.028
<b>200</b>	0.081	0.058	0.051	0.052	0.044	0.034	0.033	0.033	0.020	0.028	0.032	0.018
<b>Power</b>												
<b>20</b>	0.821	0.660	0.697	0.642	0.598	0.553	0.506	0.597	0.530	0.460	0.579	0.498
<b>35</b>	0.947	0.711	0.822	0.870	0.629	0.669	0.758	0.602	0.566	0.625	0.590	0.552
<b>50</b>	0.987	0.739	0.894	0.960	0.637	0.788	0.907	0.626	0.671	0.835	0.616	0.570
<b>100</b>	1.000	0.870	0.982	0.999	0.792	0.958	0.997	0.707	0.910	0.994	0.647	0.861
<b>200</b>	1.000	0.961	0.999	1.000	0.936	0.998	1.000	0.897	0.996	1.000	0.851	0.992

Table 1: Size and power of the  $T^2\hat{D}$  statistic under several error term dynamics (WN=white noise; AR(1)=autoregressive of order one; MA(1)=moving average of order one) and k (2,3,4,5) stochastic trends. Nominal size: 5 percent.