CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Micro versus Macro Cointegration in Heterogeneous Panels

Lorenzo Trapani and Giovanni Urga

CEA@Cass Working Paper Series

WP–CEA–13-2007

Micro versus Macro Cointegration in Heterogeneous Panels

Lorenzo Trapani

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

Giovanni Urga*

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

October 18, 2007

*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.urga@city.ac.uk; http://www.cass.city.ac.uk/faculty/g.urga.

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. 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 microequations and of Granger causality among aggregated variables when it is absent in 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. On the other hand, Phillips and Moon (1999) found that with $n \to$

∞ (and T large) a long-run average relationship between two nonstationary panel vectors exists even when the single units do not cointegrate. 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, and Forni and Lippi (1998) argue that they correspond to a zero Lebesgue measure set in the model parameter space. Therefore, formally speaking, cointegration in the aggregate relationship should be almost surely never found, implying that macroeconomic models never cointegrate. Of course, this conclusion clashes with macroeconomic theory and reality. The existence of cointegration at macro level is a well established result. Therefore, though formally aggregate cointegration may not hold, the macro relationship could be observationally equivalent to, and hence properly described by, a long run relationship. There are important empirical implications of the ability to determine whether a macro model is observationally equivalent to a cointegration relationship. An illuminating example has recently been provided by Hsiao, Shen and Fujiki (2005). When using micro prefecture level data at a annual frequency, authors find cointegrated money demand functions in Japan. Cointegration is not longer valid when aggregated data at a quarterly frequency are used.

In this paper, inspired by Lazarova, Trapani and Urga (2007) where authors offer a simple descriptive measure of the departure from aggregate cointegration for a simple bivariate model, we consider a heterogeneous panel where each micro equation contains several explanatory variables, p, and several common stochastic trends, k (≥ 1). We develop a test statistic for the null hypothesis of aggregate cointegration. The main novel contributions of the paper are twofold. First, we provide an estimation procedure when common trends are unobservable for the case of finite n, expanding the framework in Bai (2004); this makes testing feasible even in the presence of latent variables. Second, we propose a sieve bootstrapping algorithm to get critical values, proving the consistency of the procedure and Monte Carlo simulations provide evidence of size and power properties of the testing framework.

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 the 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. A Monte Carlo simulation, reported in Section 5, evaluates 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.

2 Asymptotics for the Aggregate Relationship

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

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

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

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

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

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

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

The model can also be rewritten in matrix form:

$$y_{it} = x'_{it}\beta_i + u_{it}, \tag{3}$$

$$x_{it} = \Gamma_i z_t + v_{it}, \tag{4}$$

$$z_t = z_{t-1} + \epsilon_t, \tag{5}$$

where $x_{it} = [x_{1it}, ..., x_{pit}]'$, $\beta_i = \left(\beta_{1i}, ..., \beta_{pi}\right)'$ and $\Gamma_i = [\alpha_{1i}, ..., \alpha_{pi}]'$. The matrices dimensions are respectively $p \times 1$ and $p \times k$. The trend vector is assumed to initiate at $z_0 = 0$.

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

Assumption 1

(i) *a functional central limit theorem holds for the partial sums of* ε_t, $S_t = \sum_{l=1}^{t} \varepsilon_l$;

(ii) ϵ_t *is independent of* u_t *and* v_t *and the trends* z_t *have a unit long-run variance,* $\lim_{T\to\infty} Var(T^{-\frac{1}{2}} \sum_{t=1}^{T} \epsilon_t) = I_k$.

Assumption 1 summarizes the requirements on the behaviour of the error term ε_t. Assumption 1(i) allows ε_t to belong to a very general class of processes. In particular, time dependence is allowed for the process ε_t as long as it decays at an appropriate rate. 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 β_{hi} and α_{hi}. Therefore, the long run variance of the x_{it}s, $\lim_{T\to\infty} T^{-1}E\left(x_{it}x_{it}'\right)$, is given by $\Gamma_i\Gamma_i'$. Note that Assumption 1 ensures that for $r=[0,1]$, $T^{-1/2}\sum_{t=1}^{[Tr]}\epsilon_t \xrightarrow{d} W_z(r)$, where $W_z(\cdot)$ is the k-dimensional standard Brownian motion. Further, Assumption 1 does not make any requirements on the existence and extent of cross sectional dependence; we therefore allow for arbitrary contemporaneous and dynamic correlation across units (as long as the mixing conditions for the central limit theorem for functional spaces hold). 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).

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, rank* $(\Gamma_i) = p$, *for* $i = 1, ..., n$.

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

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

Assumption 2 refers to the model representation. 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}s 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. Conversely, the upper bound $n(p+1)$ in Assumption 2(iii), which is necessary for the factors identification and estimation, 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.

Assumption 2(ii) requires that also the sum of the Γ_is must have full rank. This condition will prove useful in the analysis of the aggregate cointegration relationship properties.

2.1 The Aggregate Cointegration Relationship

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

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

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

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

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

where $t = 1, ..., T$, $\overline{y}_t = \sum_{i=1}^{n} y_{it}$, $b_j = \sum_{h=1}^{p} \sum_{j=1}^{n} \beta_{hi} \alpha_{hi,j}$ 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 $\overline{y}_t$ contains a unit root.

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

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

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

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

2.2 Asymptotics for $\widehat{\beta}$

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

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

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

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

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

$$\widehat{\beta} \xrightarrow{d} S = \left[\Gamma \int W_z W_z' \Gamma' \right]^{-1} \left[\Gamma \int W_z W_z' b \right]. \tag{8}$$

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

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

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 Proposition 1 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.

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 Γ 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 $\left[\Gamma \int W_z W_z' \Gamma'\right]^{-1}$ exists almost surely. Thus, assumption 2(ii) requires that not only the individual x_{it}s, but also their aggregate $\overline{x}_t$ 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.

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

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 $\overline{y}_t = \widehat{\alpha} + \widehat{\beta}' \overline{x}_t + \widehat{e}_t$, then (8) should be modified as

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

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

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.

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.

Proposition 2 *Let the regression coefficients β_i and Γ_i be i.i.d. random variables across i with means $\bar{\beta}$ and $\bar{\Gamma}$ and uncorrelated with each other. Then, as $(n,T) \to \infty$,*

$$\widehat{\beta} \xrightarrow{p} \bar{\beta}. \tag{9}$$

Proof. See Appendix. ■

Proposition 2 is valid for any degree of contemporaneous correlation. The OLS estimate picks the average relationship between $\bar{y}$ and each of the $\bar{x}_h$s, regardless of the existence of a cointegration relationship. A similar result is in Phillips and Moon (1999).

Note that under the more restrictive assumption of no cross-sectional dependence among units it can be shown that the OLS estimator $\widehat{\beta}$, defined in this case as $\widehat{\beta} = \left[\sum_{t=1}^{T}\left(\sum_i x_{it}\right)\left(\sum_i x'_{it}\right)\right]^{-1}\left[\sum_{t=1}^{T}\left(\sum_i x_{it}\right)\left(\sum_i y_{it}\right)\right]$, is asymptotically equivalent to the pooled-OLS estimator in Phillips and Moon (1999).

3 Aggregate Cointegration: Validity and Testing

Given that, for large n, $\widehat{\beta}$ is consistent regardless of the existence of a cointegration 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 $\overline{y}_t = \widehat{\beta}' \overline{x}_t + \widehat{e}_t$, 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

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

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

Thus, the following results hold:

Theorem 1 *Cointegration in the aggregate relationship* $\overline{y}_t = \widehat{\beta}' \overline{x}_t + \widehat{e}_t$ *always holds if and only if* $rank\,(\Gamma' \mid b) = p$.

Corollary 1 *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* $\overline{y}_t = \widehat{\beta}' \overline{x}_t + \widehat{e}_t$ *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, $rank\,(\Gamma' \mid b)$ is more likely to be equal to $p+1$, and hence aggregated cointegration is unlikely to hold.

Theorem 1 always holds when Γ is a $k \times k$ matrix. Assumption 2*(ii)* ensures that $rank\,(\Gamma) = k$ and therefore $rank\,(\Gamma' \mid b) = k$ 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.

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 $\overline{y}_t = \widehat{\beta}' \overline{x}_t + \widehat{e}_t$ represents a spurious relationship. The testing framework we derive is based on

$$\begin{aligned} H_0 &: \text{ presence of aggregate cointegration,} \\ H_A &: \text{ spurious aggregate regression.} \end{aligned}$$

A natural way to address the issue of testing is to consider the statistical properties of the limiting distribution of $\widehat{\beta}$, 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

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

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

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

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

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

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

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

This holds if and only if $Pb = 0$, 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:

$$D = \frac{b'Pb}{b'b}. \tag{15}$$

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

$$D = \sin^2(b, Mb). \tag{16}$$

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 Γ, 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 Γ.

The definition of D illustrates possible sources of the violation of the necessary and sufficient condition for cointegration in the aggregate relationship. When $rank(\Gamma' \mid b) > p$, cointegration is not preserved under aggregation. Nonetheless, if these trends are relatively unimportant then $Var(S)$ 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

$$\begin{array}{l} H_0 : D = 0 \\ H_1 : D > 0 \end{array}, \tag{17}$$

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

3.3.1 Estimation of b and Γ

The estimation of b and Γ depends crucially on whether the z_ts are observable or unobservable. When z_ts are observable, OLS estimators of b and Γ can be obtained by OLS and are given by

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

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

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

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

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

$$\begin{aligned} y_{it} &= \beta_i' \Gamma_i z_t + \beta_i v_{it} + u_{it} \\ x_{it} &= \Gamma_i z_t + v_{it}. \end{aligned}$$

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

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

Consistent estimator of Ξ 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 Ξ, 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

$$\hat{\Xi}^{PC\prime} \sum_{t=1}^T W_t W_t \hat{\Xi}^{PC} = nT^2 I_k.$$

The procedure we propose is based on Bai (2004) but extended to our case of n finite and T larger. It is also known that the solution to the above minimization problem is not unique, i.e., Ξ_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 Ξ_i. For the purpose of notational simplicity, we shall assume H being an identity matrix in this paper. The following proposition ensures consistency of the estimates $\hat{\Xi}^{PC}$.

Proposition 3 *If Assumption 1 is valid, as $T \to \infty$,*

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

Proof. See Lemma 3 in Bai (2004). ■

The principal component estimator of Θ is given by

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

and from Proposition 3

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

Therefore, even when the z_ts are unobservable, we have a T-consistent estimate for Θ.

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

$$\mathcal{F}=\left[I_{k+1},...,I_{k+1}\right]', \tag{23}$$

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

3.3.2 Testing

Let

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

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

Theorem 2 *Let $\hat{b}$, $\hat{\Gamma}$ be superconsistent estimators of b, Γ. Under the null hypothesis of cointegration, we have $D=0$ and*

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

Proof. See Appendix. ■

Theorem 2 asserts that rate of convergence of $\widehat{D}$ is of order T^2 irrespective of whether the z_ts are observable or not and of the type of estimation technique employed to derive $\hat{b}$ and $\hat{\Gamma}$, as long as they are superconsistent estimators of b and Γ. This result is reinforced by the following corollary.

Corollary 2 *For any $\delta > 0$ such that $\hat{b} - b = O_p\left(T^{-\delta}\right)$ and $\hat{\Gamma} - \Gamma = O_p\left(T^{-\delta}\right)$, under the null hypothesis of cointegration, we have $D = 0$ and*

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

Proof. See Appendix. ■

From Theorem 2 and Corollary 2 it is clear that the rate of convergence of $\widehat{D}$ is the square power of the rate of convergence of $\hat{b}$ and $\hat{\Gamma}$. This faster convergence arises from $\sin^2(b, Mb)$ being an even function in a neighborhood of zero.

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

Theorem 3 *Let $\hat{b}$, $\hat{\Gamma}$ be the OLS estimators of b, Γ in (18)-(19). Under the null of aggregate cointegration*

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

where

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

and $Q_b = \left(\int W_z W_z'\right)^{-1} \int W_z dW_{\bar{s}}$, $Q_\Gamma = \left(\int W_z W_z'\right)^{-1} \left(\int W_z dW_{\bar{v}}'\right)$, 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 $\widehat{D}$ when z_ts are not observable.

Theorem 4 *Let $\hat{b}$, $\hat{\Gamma}$ be the OLS estimators of b, Γ in (18)-(19) and $\hat{\Theta}^{PC}$ the principal component estimator of Θ in (22). Under the null of aggregate cointegration*

$$T^2\widehat{D} \xrightarrow{d} \frac{1}{\|b\|^2} Q^{pc\prime} \left[I_k - \frac{bb'}{\|b\|^2} \right] Q^{pc}, \tag{26}$$

where

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

$Q_b^{pc} = \Pi' i_b$, $Q_\Gamma^{pc} = i_\Gamma \Pi$, and Π is the limiting distribution of $\hat{\Theta}^{PC}$ given in Proposition 4 below.

Proof. See Appendix. ■

The following proposition provides the limiting distribution of the principal component estimator of Θ.

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

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

Proof. See Appendix. ■

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

$$H_1^l : b = \Gamma'\beta + \delta_T, \tag{28}$$

where the k-dimensional vector δ_T is orthogonal to Γ and is chosen to be $\lim_{T\to\infty} T\delta_T = \delta \neq 0$. 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 $\widehat{D}$ has non-trivial power versus such a sequence of local alternatives.

Theorem 5 *Let $\hat{b}$ and $\hat{\Gamma}$ be superconsistent estimators of b and Γ respectively. Under the alternative hypothesis H_1^l, we have*

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

where $b_0 = \Gamma'\beta$, Q^* is equal to either Q or Q^{pc} depending on whether the z_ts 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_ts are observable or not as long as $\hat{b}$ and $\hat{\Gamma}$ are superconsistent estimators of b and Γ.

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 global alternatives.

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

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

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.

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 $O_p\left(T^{-2}\right)$, but $O_p\left(T^{-1}\right)$. 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, and hence the presence of the term of order $O_p\left(T^{-1}\right)$ in the expansion of $\widehat{D}$ around $D \neq 0$.

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

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

We propose the following bootstrap algorithm:

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

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

Let

$$\tilde{\Theta} = \left[\tilde{b}|\hat{\Gamma}'\right]' \tag{32}$$

and

$$\tilde{\Theta} = \left[\tilde{b}|\hat{\Gamma}'\right]'. \tag{33}$$

(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 $\widehat{D}$ as

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

Step 2. (2.1) Sieve estimation. For the case observable z_ts, compute the sieve estimates of the VAR

$$\hat{w}_t = \sum_{l=1}^{q} \Psi_l \hat{w}_{t-l} + \eta_{qt} \tag{34}$$

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 (34), 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 $\left\{\eta_{qt}^{*}\right\}_{t=1}^{T}$.

(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 $\left(\hat{w}_0^*, ..., \hat{w}_{1-q}^*\right) = (\hat{w}_0, ..., \hat{w}_{1-q})$.

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

$$\bar{W}_t^* = \tilde{\Theta}^{OLS} z_t^* + \hat{e}_t^*, \tag{35}$$

or

$$\bar{W}_t^* = \tilde{\Theta}^{PC} \tilde{z}_t^* + \tilde{e}_t^*. \tag{36}$$

(3.3) Estimate Θ from either equation (35) or (36) using OLS. Denote the estimator as Θ^*.

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

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 Γ 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_ts 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 $\left\{\bar{W}_t\right\}_{t=1}^T$, consistently estimates the limiting distribution of $T^2\widehat{D}$, that is

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

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 $\Phi(L) = \sum_{k=0}^{\infty}\Phi_l L^l$. The sequence η_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 $\sum_{l=0}^{\infty}|k|^b|\Phi_l| < \infty$ for some $b \geq 1$;

(ii) In equation *(34), let $q \to \infty$ and $q = o\left(T^{1/2}\right)$ as $T \to \infty$.*

Assumption 3*(i)* ensures that both central limit theorem and invariance principle hold, and it is essentially the same as in Chang, Park and Song (2006). Assumption 3*(ii)* is required to ensure the consistency of the estimates $\hat{\Psi}_l$.

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

$$T^2 \widehat{D}^* \xrightarrow{d^B} Z_0, \tag{37}$$

where Z_0 *is the null limit distribution which is* $Z_0 = \|b\|^{-2} Q' \left[I_k - \|b\|^{-2} (bb')\right] Q$ *for observable* z_t*s and* $Z_0 = \|b\|^{-2} Q^{pc\prime} \left[I_k - \|b\|^{-2} (bb')\right] Q^{pc}$ *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 (37), which shows that under the null and the local alternatives the bootstrap consistently approximates the asymptotic distribution of $T^2\widehat{D}$ and under the alternative the bootstrap statistic $T^2\widehat{D}^*$ 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_ts are observable and principal component when z_ts are not observable), the bootstrap estimator Θ^* 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 $\bar{\nu}_t = [\bar{v}_t', \bar{s}_t]'$, we consider the following processes for $\bar{\nu}_t$: 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 $\overline{y}_t$ 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 $\bar{\nu}_t$ 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 heterogenous 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 $(n, T) \to \infty$. 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, data may be available at monthly/quarterly frequency but micro-level data is 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 particulat 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, 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. (2004), "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data", *Journal of Econometrics*, 122, 137-183.

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.

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.

Lazarova, S., Trapani, L., Urga, 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.

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.

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.

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

Appendix

Proof of Proposition 2. By assumption, the regression coefficients β_i and Γ_i are *i.i.d.* random variables across i with mean $\bar{\beta}$ and $\bar{\Gamma}$ respectively, and uncorrelated with each other. Hence, the weak law of large number ensures that $n^{-1}\sum_{i=1}^{n}\Gamma_i \xrightarrow{p} \bar{\Gamma}$, $n^{-1}\sum_{i=1}^{n}\beta_i \xrightarrow{p} \bar{\beta}$ and $n^{-1}\sum_{i=1}^{n}\Gamma_i\beta_i \xrightarrow{p} \bar{\Gamma}\bar{\beta}$. The distribution limit of $\widehat{\beta}_{n,T}$ for large T

$$\widehat{\beta}_{n,T} \xrightarrow{d} S = \left[\Gamma \int W_z W_z' \Gamma'\right]^{-1}\left[\Gamma \int W_z W_z' b\right],$$

can also be written as

$$S = \left[\sum \Gamma_i \int W_z W_z' \sum \Gamma_i'\right]^{-1}\left[\sum \Gamma_i \int W_z W_z' \sum \Gamma_i'\beta_i\right],$$

where the sums are for i from 1 to n. Then the weak law of large number on β_i and Γ_i ensure that, as $n \to \infty$

$$\left[\sum \Gamma_i \int W_z W_z' \sum \Gamma_i'\right]^{-1}\left[\sum \Gamma_i \int W_z W_z' \sum \Gamma_i'\beta_i\right]$$

$$\to_p \left[\bar{\Gamma} \int W_z W_z' \bar{\Gamma}'\right]^{-1}\left[\bar{\Gamma} \int W_z W_z' \bar{\Gamma}'\bar{\beta}\right] = \bar{\beta}.$$

Note that equation (9) has been derived using a sequential limit argument. Extension to the joint limit case can be obtained following Phillips and Moon (1999). ■

Proof of Theorem 2. Let $\hat{b}$ and $\hat{\Gamma}$ be superconsistent estimators of b and Γ and define

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

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

by definition, $\varepsilon_b = O_p\left(T^{-1}\right)$ and $\varepsilon_\Gamma = O_p\left(T^{-1}\right)$. For the sake of the notation, let also $Mb = a$ and $\varepsilon_a = \hat{a} - a$. We have

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

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,

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

so that

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

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

We have

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

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

$$\begin{aligned}
\sin^2\left(\hat{a},\hat{b}\right) - \sin^2(a,b) &= [2 + o_p(1)]\left[\cos(a,b) - \cos\left(\hat{a},\hat{b}\right)\right] \\
&= [2 + o_p(1)]\left[\frac{a'b}{\|a\|\,\|b\|} - \frac{\hat{a}'\hat{b}}{\|\hat{a}\|\,\left\|\hat{b}\right\|}\right] \\
&= [2 + o_p(1)]\,\frac{\|\hat{a}\|\,\left\|\hat{b}\right\|(a'b) - \|a\|\,\|b\|\left(\hat{a}'\hat{b}\right)}{\|a\|\,\|b\|\,\|\hat{a}\|\,\left\|\hat{b}\right\|}. \qquad (38)
\end{aligned}$$

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 $\varepsilon_{\|b\|} = \left\|\hat{b}\right\| - \|b\|$. We have

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

Using Taylor's expansion, we get

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

so that

$$\varepsilon_{\|a\|} = \frac{a'\varepsilon_a}{\|a\|} + \frac{\varepsilon_a'\varepsilon_a}{2\|a\|} - \frac{(a'\varepsilon_a)^2}{2\|a\|^3} + O_p\left(T^{-3}\right). \qquad (39)$$

Likewise,

$$\varepsilon_{\|b\|} = \frac{b'\varepsilon_b}{\|b\|} + \frac{\varepsilon_b'\varepsilon_b}{2\|b\|} - \frac{(b'\varepsilon_b)^2}{2\|b\|^3} + O_p\left(T^{-3}\right). \qquad (40)$$

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

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

and

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

Finally, from equation (38) it holds that, since under the null $a = b$

$$
\begin{aligned}
\sin^2\left(\hat{a}, \hat{b}\right) - \sin^2(a, b) &= [2 + o_p(1)] \frac{\|\hat{a}\| \left\|\hat{b}\right\| (a'b) - \|a\| \|b\| \left(\hat{a}'\hat{b}\right)}{\|a\| \|b\| \|\hat{a}\| \left\|\hat{b}\right\|} \\
&= \frac{1}{\|b\|^2}(\varepsilon_a - \varepsilon_b)'\left[I_k - \frac{bb'}{\|b\|^2}\right](\varepsilon_a - \varepsilon_b) + O_p\left(T^{-3}\right) \\
&= O_p\left(T^{-2}\right). \qquad (41)
\end{aligned}
$$

■

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

$$
\varepsilon_{\|a\|} = \frac{a'\varepsilon_a}{\|a\|} + \frac{\varepsilon_a'\varepsilon_a}{2\|a\|} - \frac{(a'\varepsilon_a)^2}{2\|a\|^3} + O_p\left(T^{-3\delta}\right),
$$

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

Then equation (41) becomes

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

■

Proof of Theorem 3. From equation (41) we know that under H_0 asymptotically the following results holds

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

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 $\varepsilon_M = \Gamma'(\Gamma\Gamma')^{-1}\varepsilon_\Gamma + \varepsilon_\Gamma'(\Gamma\Gamma')^{-1}\Gamma - \Gamma\Gamma'\varepsilon_\Gamma'\Gamma - \Gamma'\varepsilon_\Gamma\Gamma\Gamma'$.

From Assumption 1 we know that

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

$$T\varepsilon_\Gamma \xrightarrow{d} \left(\int W_z W_z'\right)^{-1}\int W_z dW_{\bar{v}}'.$$

■

Proof of Theorem 4. After equation (41) we have.

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

and under H_0 we have $aa' = bb'$. Also, it holds that $\varepsilon_a = \varepsilon_M b + M\varepsilon_b$, with $\hat{M} = M + \varepsilon_M$ and $\varepsilon_M = \Gamma'(\Gamma\Gamma')^{-1}\varepsilon_\Gamma + \varepsilon_\Gamma'(\Gamma\Gamma')^{-1}\Gamma - \Gamma'\Gamma\varepsilon_\Gamma'\Gamma - \Gamma'\varepsilon_\Gamma\Gamma'\Gamma$. 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

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

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

■

Proof of Proposition 4. The limiting distribution of $\hat{\Theta}^{PC}$ can be computed recalling that $\hat{\Theta}^{PC} = \digamma'\hat{\Xi}^{PC}$ 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

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

In the denominator of (42), we can rewrite

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

We know that

$$I = O_p\left(T^2\right);$$

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

$$II = III = O_p\left(T\right)$$

and from Lemma B.1 in Bai (2004)

$$IV = O_p\left(T\right),$$

Therefore the denominator of (42) is

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

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

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

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 $\tilde{z}_t$:

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

where

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

Then we have

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

Then

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

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

and

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

Therefore the only term that matters is II and thus

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

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

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

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

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

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

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

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

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

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

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

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

■

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

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

where

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

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

Moreover

$$\varepsilon_{\|b\|} = \varepsilon_{\|b0\|} + \frac{\delta_T'\varepsilon_b}{\|b_0\|} - R_T b'\varepsilon_b,$$

which follows from

$$\varepsilon_{\|b0\|} = \frac{b'\varepsilon_b}{\|b\|} + \frac{\varepsilon_b'\varepsilon_b}{2\|b\|} - \frac{(b'\varepsilon_b)^2}{2\|b\|^3} + O_p\left(T^{-3}\right),$$

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 (38) that

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

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

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

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

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

$$\begin{aligned}
& \left[\|a\| + \varepsilon_{\|a\|}\right] \left[\|b\| + \varepsilon_{\|b\|}\right] (a'b) - \|a\| \|b\| \left[a'b + a'\varepsilon_b + b'\varepsilon_a + \varepsilon_a'\varepsilon_b\right] \\
= & \left[\|a\| + \varepsilon_{\|a\|}\right] \left[\|a\| + R_T + \varepsilon_{\|b0\|} + \|b_0\|^{-1} \delta_T'\varepsilon_b - R_T b'\varepsilon_b\right] \|a\|^2 - \\
& \|a\| \left[\|a\| + R_T\right] \left[a'b + a'\varepsilon_b + b_0'\varepsilon_a + \delta_T'\varepsilon_a + \varepsilon_a'\varepsilon_b\right] \\
= & \frac{\|a\|^2}{2} (\varepsilon_a - \varepsilon_b)' \left[I_k - \frac{aa'}{\|a\|^2}\right] (\varepsilon_a - \varepsilon_b) - \\
& \|a\|^2 \delta_T' (\varepsilon_a - \varepsilon_b) + O_p\left(T^{-3}\right). \qquad (44)
\end{aligned}$$

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

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

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

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

■

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 δ is, as in the local alternative case, orthogonal to Γ, 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

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

From equation (45) it follows that

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

As far as the term

$$
\frac{\|\hat{a}\|\left\|\hat{b}\right\|(a'b)-\|a\|\,\|b\|\left(\hat{a}'\hat{b}\right)}{\|\hat{a}\|\left\|\hat{b}\right\|\|a\|\,\|b\|}
$$

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

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

As far as the numerator is concerned, we have

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

Recalling the definitions of $\varepsilon_{\|a\|}$ and $\varepsilon_{\|b\|}$ given in equations (39) and (40) respectively, we have

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

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

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

where

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

■

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

Lemma A.1 *Consider the estimators $\hat{\Theta}^{OLS}$ and $\hat{\Theta}^{PC}$ of Θ and their linear transformations $\tilde{\Theta}^{OLS}$ and $\tilde{\Theta}^{PC}$ defined in equations (32) and (33) respectively. Let Θ^* be the bootstrap estimator for $\tilde{\Theta}^{OLS}$ and $\tilde{\Theta}^{PC}$, and define the limiting distribution of $T\left(\hat{\Theta}^{OLS} - \Theta\right)$ and $T\left(\hat{\Theta}^{PC} - \Theta\right)$ as Z_Θ^{OLS} and Z_Θ^{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_ts observable from that in which the z_ts 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 η_{qt}; (2) we derive the strong approximation for the bootstrap counterpart η_{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

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

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:

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

Thus, from our resampling scheme

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

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

Hence, as $T \to \infty$

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

This proves the strong approximation is valid for the bootstrap η_{qt}^*.

(3) Following Chang, Park and Song (2006), the bootstrap invariance principle for η_{qt}^* carries over to w_t^* provided that the $\hat{\Psi}_k$ are consistent estimators for Ψ_k. Assumption 3(ii) ensures that $\hat{\Psi}_k$ is a consistent estimator for Ψ_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 (37), under the null, for the case when z_t is observed.

The case of z_ts 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 $\Theta\left(z_t-\hat{z}_t\right)$, which leads to different asymptotics. It is natural in this case to derive the proof directly for $\tilde{w}_t$.

From (36), we know that

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

Since in this case the bootstrap estimator Θ^* is given by

$$\Theta^*=\left[\sum_{t=1}^T\bar{W}_t^*\tilde{z}_t^{*\prime}\right]\left[\sum_{t=1}^T\tilde{z}_t^*\tilde{z}_t^{*\prime}\right]^{-1},$$

we have

$$\Theta^{*}-\tilde{\Theta}^{PC}=\left[\sum_{t=1}^{T}\tilde{e}_{t}^{*}\tilde{z}_{t}^{*\prime}\right]\left[\sum_{t=1}^{T}\tilde{z}_{t}^{*}\tilde{z}_{t}^{*\prime}\right]^{-1}. \tag{48}$$

(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

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

where c_l is an absolute constant. Thus,

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

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

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

and from our resampling scheme we have

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

Give that $\tilde{w}_t^{*}=[\tilde{e}_t^{*\prime},\Delta\hat{z}_t^{*\prime}]^{\prime}$, in order to prove that $E\left|\tilde{w}_t^{*}\right|^{l}$ is finite we need to show that both $E\left|\tilde{e}_t^*\right|^l$ and $E\left|\Delta\hat{z}_t^*\right|^l$ are finite. Assumption 3*(i)* ensures that $\tilde{e}_t^*$ has finite $4th$ moment, and therefore

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

is finite.

As far as $E\left|\Delta\hat{z}_t^*\right|^l$ is concerned, let us consider the quantity $T^{-1}\sum_{t=1}^{T}\left|\Delta\check{z}_t\right|^l$, 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

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

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

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

and

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

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

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

As $T\to\infty$,

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

This result jointly with continuous mapping theorem prove that numerator in (48) is $T^{-2}\sum_{t=1}^{T}\tilde{z}_t^*\tilde{z}_t^{*\prime}\xrightarrow{d}\int W_zW_z'$.

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

$$\sum_{t=1}^{T}\tilde{e}_t^*\tilde{z}_t^{*\prime}=\sum_{t=1}^{T}\bar{e}_t^*\tilde{z}_t^{*\prime}+\sum_{t=1}^{T}\Theta\left(z_t^*-\tilde{z}_t^*\right)\tilde{z}_t^{*\prime}+o_p^*(1) \qquad (52)$$

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

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

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

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

$$\begin{aligned}
&T\left(\Theta^*-\tilde{\Theta}^{PC}\right)\xrightarrow{d}\\
&\digamma'\left[\int dW_eW_z'+n^{-1}\Omega_e\Xi B-n^{-1}B\Xi'\left(\int dW_eW_z'+n^{-1}\Omega_e\Xi B\right)\right.\\
&\left.-n^{-1}\left(\int W_zdW_e'+n^{-1}B\Xi'\Omega_e\right)\Xi B\right]B^{-1},
\end{aligned}$$

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 Θ^*. Therefore, after the continuous mapping theorem, letting

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

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

we have

$$T\varepsilon_b^*\xrightarrow{d}\left(\int W_zW_z'\right)^{-1}\int W_zdW_{\bar{s}},$$

$$T\varepsilon_\Gamma^*\xrightarrow{d}\left(\int W_zW_z'\right)^{-1}\int W_zdW_{\bar{v}}',$$

if the z_ts are observable and

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

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

if the z_ts are unobservable, where $\digamma$ and Π are defined in equation (23) and Theorem 4 respectively.

We can now prove equation (37) by analysing the asymptotic behaviour of $\widehat{D}^*$.

We have

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

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

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

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

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

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

Therefore

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

Since

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

we have that

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

Lemma A.1 and the continuous mapping theorem ensure that equation (37) holds. Note that since resampling was done under the null via the use of $\tilde{b}$, equation (37) 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)
		$k=2$			$k=3$			$k=4$			$k=5$	
T		$\rho=.75$	$\vartheta=.75$		$\rho=.75$	$\vartheta=.75$		$\rho=.75$	$\vartheta=.75$		$\rho=.75$	$\vartheta=.75$
Size												
20	0.097	0.405	0.133	0.096	0.224	0.130	0.063	0.235	0.205	0.044	0.205	0.126
35	0.092	0.296	0.095	0.080	0.183	0.098	0.053	0.202	0.107	0.035	0.186	0.078
50	0.088	0.213	0.079	0.065	0.133	0.072	0.046	0.151	0.074	0.038	0.149	0.057
100	0.083	0.107	0.061	0.063	0.078	0.044	0.042	0.073	0.035	0.027	0.071	0.028
200	0.081	0.058	0.051	0.052	0.044	0.034	0.033	0.033	0.020	0.028	0.032	0.018
Power												
20	0.821	0.660	0.697	0.642	0.598	0.553	0.506	0.597	0.530	0.460	0.579	0.498
35	0.947	0.711	0.822	0.870	0.629	0.669	0.758	0.602	0.566	0.625	0.590	0.552
50	0.987	0.739	0.894	0.960	0.637	0.788	0.907	0.626	0.671	0.835	0.616	0.570
100	1.000	0.870	0.982	0.999	0.792	0.958	0.997	0.707	0.910	0.994	0.647	0.861
200	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.