View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

# Cointegration versus Spurious Regression in Heterogeneous Panels

Lorenzo Trapani
Cass Business School and Bergamo University

Giovanni Urga*
Cass Business School

This Version: 27 January 2004

**Corresponding author*: Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ Tel.+/44/(0)20/70408698; Fax.+/44/(0)20/70408885; e-mail: g.urga@city.ac.uk. http://www.staff.city.ac.uk/~giourga. We wish to thank Stepana Lazarova for useful discussions and comments. The usual disclaimer applies. Lorenzo Trapani wishes to thank Marie Curie Training Site for financial support (grant N. HPMT-CT-2001-00330).

## Cointegration versus Spurious Regression in Heterogeneous Panels


### Abstract

We consider the issue of cross sectional aggregation in nonstationary, heterogeneous panels where each unit cointegrates. We first derive the asymptotic properties of the aggregate estimate, and a necessary and sufficient condition for cointegration to hold in the aggregate relationship. We also develop an estimation and testing framework to verify whether the condition is met. Secondly, we analyze the case when cointegration doesn't carry through the aggregation process, investigating whether a mild violation can still lead to an aggregate estimator that summarizes the micro relationships reasonably well. We derive the asymptotic measure of the degree of non cointegration of the aggregated estimate and we provide estimation and testing procedures. A Monte Carlo exercise evaluates the small sample properties of the estimator.



J.E.L. Classification Numbers: C12, C13, C23
Keywords: Aggregation, Cointegration, Heterogeneous Panel, Monte Carlo Simulation.

# 1 INTRODUCTION

The effects of cross-sectional aggregation in panel data model have been explored by several contributions in the econometric literature. It is well known that most economic theories are based on microeconomic foundations, in that they are specified as panels where each equation represents a single agent, household or firm. Anyway, most often data are available only as aggregates. One may therefore wonder whether the models based on aggregated micro relationships can still provide a good summary of the properties shared by each equation of the panel.

Introducing cointegration within this framework is quite natural, in the light that it may be viewed as a property that each equation of a nonstationary panel data model shares. Previous contributions (see e.g. Pesaran and Smith, 1995) have proved that when the variables in the panel data model are integrated, cointegration in the micro level does not imply cointegration in the aggregate relationship unless some specific conditions on the micro relationships are satisfied. Spurious regression occurs when cointegration fails to hold in the aggregate relationship, which therefore becomes meaningless. It is well known that a sufficient condition for cointegration to carry through the aggregation process is the panel to be homogeneous, which would match the representative agent requirement. Anyway things get more complicated for the heterogeneous panel case.

This aggregation issue has been investigated in a few contributions, which provide conditions for cointegration to hold in the aggregates given that it holds in the micro relationships. 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. Having a greater number of common trends therefore leads to a spurious regression after aggregation. Gonzalo (1993) bases his analysis on a more complicated, multivariate model, and obtain a sufficient condition for cointegration to hold after cointegration. According to his findings, this can be maintained when there is enough cointegration in the model, which - like in Granger's (1993) analysis - happens when there is a sufficiently small number of trends to drive the system. Ghose (1995) considers a single equation framework and investigates the issue of aggregating a subset of regressor in a single variable without damaging the consistency of the estimates of the parameters of interest. Granger (1993) also investigates the case when the formal conditions for cointegration to be maintained fail to hold. He considers an example in which only few common stochastic trends are shared across all the micro series of the model, and other trends are shared by only small groups of the series. In this case the coefficients of the shared common trends in the aggregate regression are shown to be higher than the coefficients for the non-shared common trends by an order of magnitude. Therefore, removing the large trends from the aggregate regression by establishing a cointegration relationship leaves "small" $I(1)$ elements in the residuals that may not be found by standard tests applied to relatively small samples. These result suggests that when the system is described by a sufficiently low number of dominant components, and the conditions for cointegration are only "mildly" violated, then the aggregate relationship "approximately cointegrates".

In the light of Granger's contribution, which shows how data behaviour can be different from the formal conditions laid out for the model, we consider a heterogeneous panel data model where each equation contains several independent variables, say $p$, and common stochastic trends, say $k$. As a preliminary step, we provide a necessary and sufficient condition for cointegration to hold in the aggregates. Second, on the basis of these formal conditions, we develop a measure to assess the degree of departure from cointegration when these conditions don't hold, and employ it to test whether the departure from the case of perfect cointegration leads to a completely spurious relationship, or to a hybrid case where the presence of cointegration is not made insignificant by the spurious element. This strategy is aimed at formalizing the results obtained by Granger (1993).

This paper is organized as follows. The theoretical framework is presented in Section 2, where we set up a model for panel data. We first present the aggregate relationship and the properties of its estimate (2.1). To develop the following statistical framework, we analyze the probabilistic structure of the OLS estimates for both the disaggregate and the aggregate models. Section 3 presents the conditions for the cointegration to carry through the aggregation process, and a testing theory is developed and illustrated via a numerical example (3.1). In Section 4 we characterize the system's behavior when the conditions derived in the previous section are not satisfied; we derive a measure to assess the deviation from the case of non perfect cointegration and apply it to a numerical example (4.1). Moreover, as the results given in Section 2 for the aggregate estimation are only asymptotic, in Section 5 we present a Monte-Carlo examination of a small-sample distribution of the aggregate estimate. Section 6 concludes.

# 2 BASIC MODEL AND ASSUMPTIONS

Consider the following data generating process:

$$
\begin{aligned}
y_{it} &= \sum_{h=1}^{p} \beta_{hi} x_{hit} + u_{it}, && (1)\\
x_{hit} &= \alpha'_{hi} z_t + v_{hit}, && (2)\\
z_{jt} &= z_{jt-1} + \epsilon_{jt}. && (3)
\end{aligned}
$$

where $i = 1, ...n$, $j = 1, ..., k$, $h = 1, .., p$ and $\alpha_{hi}$ is a $k \times 1$ vector. The common trends representation implied by equation (2) follows from Granger (1993). The following notation will be employed henceforth: $q \equiv n(p+1) + k$, $B_h \equiv diag\{\beta_{h1}, ...\beta_{hn}\}$, $A_h \equiv [\alpha_{h1}, ...\alpha_{hn}]'$, $\beta_i \equiv (\beta_{1i}, ..., \beta_{pi})'$ and $\Gamma_i \equiv [\alpha_{1i}, ..., \alpha_{pi}]'$. The matrices dimensions are respectively: $n \times n$, $n \times k$, $p \times 1$ and $p \times k$. With this notation, we can consider two compact forms for model (1)-(3):

$$
\begin{aligned}
y_t &= \sum_{h=1}^{p} B_h x_{ht} + u_t, && (4)\\
x_{ht} &= A_h z_t + v_{ht}, && (5)\\
z_t &= z_{t-1} + \epsilon_t && (6)
\end{aligned}
$$

and

$$
\begin{aligned}
y_{it} &= x'_{it} \beta_i + u_{it}, && (7)\\
x_{it} &= \Gamma_i z_t + v_{it}, && (8)\\
z_t &= z_{t-1} + \epsilon_t. && (9)
\end{aligned}
$$

Let $\varepsilon_t \equiv \left[u'_t, v'_{1t}, , ..., v'_{pt}, \epsilon'_t\right]'$ and consider the vector of partial sums be $S_t \equiv \sum_{i=1}^{t} \varepsilon_i$ with $S_0 = 0$; we assume that the sequence of innovations satisfies the following assumptions:

### Assumption 2.1

(2.1a) *$\varepsilon_t$ follows an invertible $MA(\infty)$ process: $\varepsilon_t = \Phi(L)\eta_t = \sum_{j=0}^{\infty} \Phi_j \eta_{t-j}$, where $\eta_t$ is a zero mean iid process with finite fourth moment and $E(\eta_t \eta'_t) = I_q$*[1]*;*

[1] The spherical covariance requirement implies that the $MA(\infty)$ representation is not the one for the fundamental innovations, since $\Phi_0 \neq I_q$.

(2.1b) *the sequence* $\{j\Phi_j\}_{j=0}^{\infty}$ *is absolutely summable;*

(2.1c) *the components of* $\epsilon_t$ *are iid and the trends* $z_t$ *have unit long-run variance:* $\lim_{T\to\infty} T^{-1}E(z_t z_t') = I_k$.

Assumptions (2.1a) and (2.1b) are needed for the central limit theorem for the functional spaces to be valid[2]. The orthonormality requirement (2.1c) 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 $A_h$. Let now $\Sigma \equiv \lim_{T\to\infty} E\left(S_T S_T'\right)$. Then:

$$\Sigma = 2\pi H_\varepsilon\left(0\right) = \Phi\left(1\right)\left[\Phi\left(1\right)\right]' = \Omega + \Lambda + \Lambda',$$

where

1. $H_\varepsilon\left(\omega\right)$ is $\varepsilon_t$ spectral density matrix at density $\omega$;

2. under assumption (2.1c), we write

$$\Phi\left(1\right) = \begin{bmatrix} \Phi_{11} & 0 & 0 \\ 0 & \Phi_{22} & 0 \\ 0 & 0 & I_k \end{bmatrix} \tag{10}$$

   with $\Phi_{ij} = \Phi_{ij}\left(1\right)$. Also, $H_\varepsilon(0) = \Phi(1)[\Phi(1)]'$;

3. $\Omega \equiv \sum_{j=0}^{\infty} \Phi_j \Phi_j'$ and $\Lambda \equiv \sum_{j=2}^{\infty} E[\eta_1 \eta_j'] = \sum_{j=2}^{\infty} \sum_{i=0}^{\infty} \Phi_{j+1}\Phi_i'$.

The following Lemma holds:

**Lemma 2.1** *Let* $W^*$ *be a q-dimensional standard Brownian motion, partitioned as* $(\mathtt{W}^*)' = [W_y'^*, W_x'^*, W_z'^*]'$ *where the three vectors are of order* $n$, $np$ *and* $k$ *respectively. Then:*

1. $T^{-\frac{1}{2}} S_T \Rightarrow \Phi(1) W^*(1)$;
2. $T^{-1}\sum_{t=1}^{T} S_{t-1}\eta_t' \Rightarrow \Phi(1) \int W^*\left(dW^*\right)' \Phi'(1) + \Lambda$;
3. $T^{-2}\sum_{t=1}^{T} S_{t-1}S_{t-1}' \Rightarrow \Phi(1) \int W^*\left(W^*\right)' \Phi'(1)$.

[2] Notice that the usual central limit theorem framework also requires $\Sigma$ to be positive definite. According to equation (2), this will never happen, since the $x_{hit}$s cointegrate among themselves. See Lazarova *et al.* (2003) for details.

Lemma 2.1 considers the convergence of the $I(1)$ process $S_T$ to a linear combination of normally distributed variables (central limit theorem), and of products of the process with itself and with its increments according to a standard framework. Notice that the last $k$ elements of $\Phi(1)W^*$ are standard Brownian motions, which will be referred to as $W_z$. Henceforth, we will also denote the first $n$ elements of $\Phi(1)W^*$ as $W_1$, and the next $np$ ones as $W_2$. Therefore, $W_{2i}$, $i = 1, ..., n$ will denote the Brownian motion associated with the disturbance of each $x_{it}$ data generating process.

After describing the general framework, we now consider the issues that arise from aggregating the equations in model (1)-(3).

## 2.1 The aggregate cointegration relationship

When we aggregate the regressors across units, we obtain

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

where $\overline{x}_{ht} \equiv \sum_{i=1}^{n} x_{hit}$, $a_{hj} \equiv \sum_{i=1}^{n} \alpha_{hi,j}$ with $\alpha_{hi,j}$ being the $j$-th element in vector $\alpha_{hi}$ and $\overline{v}_{ht} \equiv \sum_{i=1}^{n} v_{hit}$. 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 we have:

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

where $\overline{y}_t \equiv \sum_{i=1}^{n} y_{it}$, $b_j \equiv \sum_{h=1}^{p} \sum_{j=1}^{n} \beta_{hi} \alpha_{hi,j}$ and $\bar{s}_t \equiv \sum_{h=1}^{p} \sum_{i=1}^{n} \beta_{hi} v_{hit} + \sum_{i=1}^{n} u_{it}$. Let $\bar{x}_t = [\bar{x}_{1t}, \bar{x}_{2t}, ..., \bar{x}_{pt}]'$, $\Gamma \equiv \sum_{i=1}^{n} \Gamma_i$ and $b \equiv \sum_{i=1}^{n} \Gamma_i' \beta_i$; then, in vector form, the aggregate cointegration relationship may be written as

$$\overline{x}_t = \Gamma z_t + \bar{v}_t \qquad (11)$$

$$\overline{y}_t = b' z_t + \bar{s}_t \qquad (12)$$

We again assume there is at least one $j$ for which $b_j \neq 0$, so that $\overline{y}_t$ contains a unit root. We consider the least-squares regression equation

$$\overline{y}_t = \widehat{\beta}' \overline{x}_t + \widehat{e}_t, \qquad (13)$$

where $\widehat{\beta}$ is the OLS estimator, defined as $\widehat{\beta} = \left( \sum_{t=1}^{T} \overline{x}_t \overline{x}_t' \right)^{-1} \left( \sum_{t=1}^{T} \overline{x}_t \overline{y}_t \right)$.

When $\overline{y}_t$ and $\overline{x}_t$ are cointegrated, $\widehat{\beta}$ is superconsistent and converges in probability to a vector which is the true value of the aggregation coefficient.

In case the aggregate series are not cointegrated, the regression is spurious and $\widehat{\beta}$ converges in distribution to a non-degenerate vector random variable. Consider the following assumption:

**Assumption 2.2**

(2.2a) $\forall i$, *rank* $(\Gamma_i) = p$;

(2.2b) *The number of regressors in the cointegration equations (1) is not greater than the number of common trends, i.e.* $p \leq k$ *for any* $i$. *Moreover,* *rank* $(\Gamma) = \min\{p, k\} = p$.

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

**Proposition 1** *If* $y_t$ *and* $x_t$ *are generated by (1)-(3) where the innovation sequence* $\{\varepsilon_t\}_1^\infty$ *satisfies assumption 1, and if assumption (2.2b) holds, then in the OLS regression of* $\overline{y}_t$ *on* $\overline{x}_t$ $\widehat{\beta}$ *converges to a non degenerate random variable*

$$\widehat{\beta} \Rightarrow S \equiv \left[\Gamma \int W_z W_z' \Gamma'\right]^{-1} \left[\Gamma \int W_z W_z' b\right], \tag{14}$$

*where* $\int W_z W_z' = \int_0^1 W_z(r) W_z'(r)\, dr$ *and where "*$\Rightarrow$*" denotes weak convergence of the associated probability measures as* $T \to \infty$.[3]

**Remarks**

(a) The proof is the same as in Park and Phillips (1988). Assumption (2.2b) is needed for the $p \times p$ term $\left[\Gamma \int W_z W_z' \Gamma'\right]$ not to be a degenerate Brownian motion - see a related discussion by Phillips (1986). Being $p \leq k$ and $\Gamma$ a full rank matrix, it holds that $\left[\Gamma \int W_z W_z' \Gamma'\right]$ is almost surely positive definite and $\left[\Gamma \int W_z W_z' \Gamma'\right]^{-1}$ exists almost surely;

(b) Notice that the OLS estimator $\widehat{\beta}$ converges to a weighted average of $\beta_i$ coefficients where weights are given by the $\Gamma_i$ coefficients. This finding is consistent with the analysis of Gonzalo (1993). Hall, Lazarova and Urga (1999) highlight this case when providing a counterexample to the general statement of Pesaran and Smith (1995) that the aggregate relationship does not cointegrate even if the individual unit do cointegrate;

[3]The proof of this limiting representation can be found in Park and Phillips (1988, 1989).

(c) Assumption (2.2a) is needed for the following Lemma, which rules out the degenerate cointegration regression case:

**Lemma 2.2** *If assumption (2.2a) and (2.2b) hold, then the* $x_{it}$ *in equation (8) don't cointegrate among themselves for any* $i$.

**Proof** See Appendix I.

The results obtained so far are valid for large $T$. In many applications, anyway, also the number of units $n$ can be large. It is therefore worth investigating the limit behavior of $\hat{\beta}$ when $n$ is large, according to the framework provided by Phillips and Moon (1999). Consider the following preliminary assumptions:

**Assumption 2.3** *With respect to model (7)-(9), the regression coefficients* $\beta_i$ *and* $\Gamma_i$ *are iid random variables across* $i$ *with mean* $\bar{\beta}$ *and* $\bar{\Gamma}$ *respectively, and are assumed to be incorrelated with each other.*

**Assumption 2.4** *Let* $\hat{\beta}_{n,T}$ *be the finite* $T$*, finite* $n$ *estimator. We assume that*

$$\lim_{n,T} \sup P\left\{\left\|\hat{\beta}_{n,T} - S\right\| > \varepsilon\right\} = 0 \quad \forall\varepsilon > 0.$$

**Lemma 2.3** *Given assumptions 2.3-2.4, the following results hold*

$$\frac{1}{n}\sum_{i=1}^{n}\Gamma_i \rightarrow_p \bar{\Gamma}$$

$$\frac{1}{n}\sum_{i=1}^{n}\beta_i \rightarrow_p \bar{\beta}$$

$$\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\beta_i \rightarrow_p \bar{\Gamma}\bar{\beta}.$$

The first two equalities in Lemma 2.3 hold by weak law of large numbers, and the last by incorrelatedness; proof is straightforward and is omitted. With these assumptions and Lemma 2.3, the following Proposition holds:

**Proposition 2** *Let assumption 2.3 and 2.4 be valid and consider Phillips and Moon's (1999) notation. Then:*

$$\hat{\beta}_{n,T} \rightarrow_p \bar{\beta} \quad for \quad (T, n \rightarrow \infty)_{seq}$$

$$\hat{\beta}_{n,T} \rightarrow_p \bar{\beta} \quad for \quad (T, n \rightarrow \infty)$$

**Proof** See Appendix I.

**Remarks**

(a) Assumption 2.4 is needed for Lemma 6(a) in Phillips and Moon (1999) to hold in order to ensure the equivalence between joint and sequential limit.

(b) The results in Proposition 2 refer to the probability limit. They state that as $n$ increases, the OLS estimate converges to the average relationship between $\bar{y}$ and each of the $\bar{x}_h$s, regardless of the existence of a cointegrating relationship between the aggregated variables.

(c) According to Phillips and Moon (1999), the joint probability limit also implies the convergence to the same value for all monotonic diagonal paths $(n, T(n)) \rightarrow \infty$.

(d) An interesting characterization of $n$ being large or small is provided in an example in Granger (1993). Granger (1990) points out how the two cases may lead to different results.

Given that, for large $n$, $\hat{\beta}$ will be superconsistent regardless of the existence of a cointegration relationship, henceforth we will restrict our analysis to the case of finite $n$ only. In the next section, we will develop an estimation theory framework for both the aggregate and the disaggregate model.

## 2.2 Probability structure for OLS estimation

The estimation method we consider in this section is OLS, and the results we obtain are a mere application of Lemma 2.2. OLS estimation will be considered for both models (11)-(12) and (1)-(3). Henceforth, let $\Phi_{11,i}$ be the $i$-th row in $\Phi_{11}$, and $\Phi_{22,i}$ is the $i$-th block of $p$ rows in $\Phi_{22}$, where $\Phi_{11}$ and $\Phi_{22}$ are given in equation (10). Also, let $\Phi_{1j}$, $\Phi_{2j}$ be blocks $\Phi_{11}$ and $\Phi_{22}$ in $\Phi_j$, as defined in Assumption (2.1a).

As far as the aggregate model (11)-(12) is concerned, the OLS estimators for $b$ and $\Gamma$ are

$$\hat{b} = \left[ \sum_{t=1}^{T} z_t z_t' \right]^{-1} \left[ \sum_{t=1}^{T} z_t \bar{y}_t \right]$$

$$\hat{\Gamma} = \left[\sum_{t=1}^{T} \bar{x}_t z_t'\right]\left[\sum_{t=1}^{T} z_t z_t'\right]^{-1}$$

and the following Lemma characterizes their limit distributions:

**Lemma 2.4** *The OLS estimators $\hat{b}$ and $\hat{\Gamma}$ have the following limit distribution for large $T$*

$$T\left(\hat{b} - b\right) \Rightarrow \sum_{i=1}^{n} \left[\int W_z W_z'\right]^{-1} \left[\int W_z dW_1^{\theta *} \Phi_{11,i}' + \int W_z dW_2^{\theta *} \Phi_{22,i}' \beta_i'\right] \quad (15)$$

$$T\left(\hat{\Gamma} - \Gamma\right) \Rightarrow \sum_{i=1}^{n} \left[\Phi_{22,i} \int dW_2^{*} W_z'\right] \left[\int W_z W_z'\right]^{-1} \quad (16)$$

Having derived $\hat{b}$ and $\hat{\Gamma}$, we now turn to the disaggregate model estimation. We will estimate the $\beta_i$s in equation (7) and the $A_h$s in equation (5). The OLS estimates have the following representation[4]

$$\hat{\beta}_i = \left[\sum_{t=1}^{T} x_{it} x_{it}'\right]^{-1} \left[\sum_{t=1}^{T} x_{it} y_{it}\right], \ i = 1, ...n$$

$$\hat{A}_h = \left[\sum_{t=1}^{T} x_{ht} z_t'\right] \left[\sum_{t=1}^{T} z_t z_t'\right]^{-1}, \ h = 1, ..., p$$

The following Lemma characterizes their limit distribution:

**Lemma 2.5** *Let assumption (2.2a) be valid. Then, for any $i = 1, ..., n$ and $h = 1, ..., p$ and for large $T$*

$$T\left(\hat{\beta}_i - \beta_i\right) \Rightarrow \left[\Gamma_i \int W_z W_z' \Gamma_i'\right]^{-1} \left[\Gamma_i \int W_z dW_1^{*\prime} \Phi_{11,i}' + \sum_{j=0}^{\infty} \Phi_{2j,i} \Phi_{1j,i}'\right] \quad (17)$$

$$T\left(\hat{A}_h - A_h\right) \Rightarrow \left[\Phi_{22} \int dW_2^{*} W_z'\right] \left[\int W_z W_z'\right]^{-1} \quad (18)$$

[4] We have been assuming that the $z_t$s are observable. Anyway, in order for the estimation of the $A_h$s to be feasible, this is not necessary. $\hat{A}_h$ could be the principal component estimator, for instance, instead of the OLS one, even though the probabilistic structure would possibly change - we refer among others to Phillips and Ouliaris (1988), Harris (1997) and Snell (1999) for details. Similar arguments apply to $b$ and $\Gamma$ estimation.

Proofs of both Lemmas are straightforward, and we refer to Lemma 2.1. Assumption (2.2a) ensures that $\left[\Gamma_i \int W_z W_z' \Gamma_i'\right]$ is a non degenerate Brownian motion, and therefore $\left[\Gamma_i \int W_z W_z' \Gamma_i'\right]^{-1}$ exists almost surely. Some of the limit distributions contain nuisance parameters. This problem can be overcome by replacing the parameters with their estimates, such approximation being valid for large $T$.

Having set the model and the estimation theory, we will now consider the issue of maintaining cointegration after aggregation.

# 3 PERFECT COINTEGRATION VERSUS SPURIOUS REGRESSION

In this section, we discuss the conditions under which cointegration holds in the aggregate relationship (13). According to equation (14), $\widehat{\beta} \Rightarrow S$. In order to have perfect cointegration $S$ must be a vector of constants rather than a vector of random variables. Given that $b \neq 0$ by assumption, this means having

$$\Gamma' c = b \tag{19}$$

for some nonzero $c \in R^p$. In this case,

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

and therefore cointegration holds[5]. Equation (19) states that $b$ must be a linear combination of $\Gamma$ rows, and therefore cointegration in the aggregate relationship means having a nontrivial solution for the linear system (19). According to Rouche'-Capelli's Theorem, system (19) will have a solution if and only if $rank\,(\Gamma') = rank\,(\Gamma' \mid b) = p$. Then

**Proposition 3** *Cointegration in the aggregate relationship (13) always holds if and only if* $rank\,(\Gamma' \mid b) = p$.

[5] Incidentally, note that superconsistency ensures that $c$ is the true value of the aggregate relationship parameters. The linear system (19) has therefore one and only solution, if it admits any solutions.

Proposition 3 implies the following Lemma, which is another formulation of Theorem 1 in Gonzalo (1993) when the common trends in the disaggregate system are the same across all $i$[6]:

**Lemma 3.1** *If the number of regressors in the cointegration equations (1) equals the number of stochastic trends (i.e. if $p = k$), then cointegration in the aggregate relationship (13) will always hold.*

**Proof** See Appendix I.

Having laid out the formal conditions for cointegration to hold in the aggregates, in the next section we will derive a measure for the departure from cointegration when Proposition 3 doesn't hold, and therefore, strictly speaking, equation (13) represents a spurious relationship.

# 4 A MEASURE OF NON COINTEGRATION

This section is aimed at assessing the degree of non cointegration for system (4)-(6) when Proposition 3 doesn't hold, and the aggregate relationship (13) is a spurious regression equation. Here will consider the case $p = 1$[7]. If $p = 1$, the aggregate cointegration relationship for model (4)-(6) becomes

$$\bar{y}_t = \beta \bar{x}_t + \bar{u}_t = b' z_t + \beta \bar{v}_t + \bar{u}_t$$

and the OLS estimate of $\beta$ is

$$\hat{\beta} = \frac{\sum_{t=1}^{T} \bar{y}_t \bar{x}_t}{\sum_{t=1}^{T} \bar{x}_t^2}$$

Under the case of a single common trend ($k = 1$), $S = b_1/a_1 \equiv S_1$, which is a constant. Therefore the aggregate relationship always cointegrates. Let $i_n$ be an $n$-dimensional vector of ones. Provided the $a_j = 0$ condition is ruled out for at least one $j$, when $k > 1$ $S$ is a constant and (13) is not a spurious relationship, if and only if there exists a $d \in R$ such that

$$\begin{cases} \sum_{i=1}^{n} \alpha_{1i} (\beta_i - d) = 0 \\ \quad\quad\quad . \\ \quad\quad\quad . \\ \sum_{i=1}^{n} \alpha_{ki} (\beta_i - d) = 0 \end{cases}$$

[6] Notice that the condition given by Proposition 3 is necessary and sufficient, and therefore in principle one could employ it to test whether formally cointegration can hold or not in the aggregate relationship.

[7] The case of $p > 1$ is reported in Appendix II due to its algebraic complexity.

i.e. if, in vector form, there exists a vector $\gamma = di_n$ such that, for $\tau \equiv [\beta_1, ..., \beta_n]'$

$$A'(\tau - \gamma) = 0.$$

Let: $S_j \equiv b_j/a_j$,

$$\omega \equiv \frac{1}{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}},$$

and

$$\tilde{W} \equiv \begin{bmatrix} W_{12} & . & . & W_{1k-1} \\ W_{22} & & & . \\ . & & & . \\ . & & & . \\ W_{k2} & & & W_{kk-1} \end{bmatrix}.$$

Then a convenient way to express the deviation from the case of perfect cointegration (i.e. $S$ collapsing to a constant) is the following decomposition (its derivation is reported in Appendix):

$$S - S_1 = \frac{1}{\omega}\sum_{j=2}^{k}\left[f_j^*(S_j - S_1)\right] \quad (20)$$

where $f_j^*$ is the $j$-th element in $f^* \equiv [a_2 \sum a_i W_{i2}, a_3 \sum a_i W_{i3} ... a_k \sum a_i W_{ik}]'$[8]. Also, let $\alpha \equiv diag\{a_2, a_3, ...a_k\}$ and $s \equiv [(S_2 - S_1), ..., (S_k - S_1)]$. Then equation (20) can be expressed as

$$\begin{aligned} S - S_1 &= \omega\left\{\alpha\left[I_{k-1} \otimes (i_n' A)\right] vec\left(\tilde{W}\right)\right\}' s = \\ &= \omega\left[vec\left(\tilde{W}\right)\right]'\left[I_{k-1} \otimes (i_n' A)\right]' \alpha s = \omega f' \alpha s \end{aligned}$$

where $f \equiv [I_{k-1} \otimes (i_n' A)]\, vec\left(\tilde{W}\right)$. Now, the main problem associated with (20) is its being a weighted average of the distances $(S_j - S_1)$. The following steps will lead towards a suitable transformation of equation (5) that will be of use in order to rule out the presence of different weights in (20).

Notice first that each element of $f$, namely $f_j$, is a random variable resulting from the linear combination (with weights given by the elements on the $j$-th row of $I_{k-1} \otimes i_n' A$) of $k-1$ functionals of Brownian motions distributed as $\int W_i W_j$ and one distributed as $\int W_i^2$. The $f_j$s will have the same distribution across $j$ if, for some non trivial constant $r$, $i_n' A = r i_k'$. In scalar form this means

$$\sum_{i=1}^{k} \alpha_{il} = \sum_{i=1}^{k} \alpha_{iq}$$

[8] Sums go from $i = 1$ to $k$.

for $l \neq q = 1, ..., k$. Consider now the $k$-dimensional rotation matrix $M = M(\vartheta_1, ...\vartheta_{k-1})$, $\vartheta_i$ being a rotation angle. By orthogonality

$$x_t = Az_t + v_t = AMM'z_t + v_t = \tilde{A}\tilde{z}_t + v_t,$$

and the covariance matrix of $\tilde{z}_t$ is still equal to $I_k$. The $k-1$ angles we need are those that satisfy $i_n' AM = ri_k'$ or, defining $M$ $j$-th column as $m_j$, the system of $k-1$ (nonlinear) equations

$$\begin{cases} i_n' Am_1 = i_n' Am_2 \\ i_n' Am_2 = i_n' Am_3 \\ \quad\quad . \\ \quad\quad . \\ i_n' Am_{k-1} = i_n' Am_k \end{cases}$$

in $k-1$ unknowns, whose solution will be referred to as $M^*$. If this holds, plain algebra shows that $\alpha = rI_{k-1}$. Also $\omega = \left(r^2 \sum_{i=1}^k \sum_{j=1}^k W_{ij}\right)^{-1} = \omega^*/r^2$, so that

$$\begin{aligned} S - S_1 &= \frac{1}{r^2}\omega^* \left[vec(W)\right]' \left[I_{k-1} \otimes ri_k\right] rI_{k-1}s = \\ &= \omega^* \left[vec(W)\right]' \left[I_{k-1} \otimes i_k\right] s = \sum_{j=2}^{k} \varphi_j \left(S_j - S_1\right) \end{aligned}$$

where $\varphi_i \equiv \omega^* \sum_{j=1}^k W_{ij}$ and they are identically distributed but non independent random variables. Last

$$S = S_1 + \sum_{j=2}^{k} \varphi_j \left(S_j - S_1\right). \tag{21}$$

Following the presentation of the theoretical framework, we now present a measure of non cointegration and a testing strategy. To get equation (21), we need first to find $M^*$, as defined above. Then transform equation (5) as follows

$$x_t = AM^*M^{*\prime}z_t + v_t = A^*z_t^* + v_t$$

with $A^* \equiv AM^*$. After this transformation, the system parameters can be estimated and equation (21) computed.

Such decomposition allows one to obtain a statistical framework to test the null hypothesis of cointegration. Define the vector $\tilde{h}_i$ (for $i = 2, ..., k$) as

$$\tilde{h}_{ij} \equiv \frac{\tilde{\alpha}_{ij}}{\sum_{j=1}^n \tilde{\alpha}_{ij}} - \frac{\tilde{\alpha}_{1j}}{\sum_{j=1}^n \tilde{\alpha}_{1j}}.$$

Then

$$
\begin{aligned}
Var(S) &= \sigma_\varphi^2 \left\{ \sum_{i=2}^{k} (S_i - S_1)^2 + \frac{1}{2}\rho_\varphi \sum_{j=2}^{k} \sum_{i \neq j} \left[ (S_i - S_1)(S_j - S_1) \right] \right\} = \\
&= \sigma_\varphi^2 \|\tau\|^2 \left\{ \sum_{i=2}^{k} \left\| \tilde{h}_i \right\|^2 \cos^2\left(\tau, \tilde{h}_i\right) + \right. \\
&\quad \left. + \frac{1}{2}\rho_\varphi \sum_{j=2}^{k} \sum_{i \neq j} \left\| \tilde{h}_i \right\| \left\| \tilde{h}_j \right\| \cos\left(\tau, \tilde{h}_i\right) \cos\left(\tau, \tilde{h}_j\right) \right\}
\end{aligned}
$$

where $\sigma_\varphi^2 \equiv Var(\varphi_i)$ and $\rho_\varphi = Corr(\varphi_i, \varphi_j)$ for any $i \neq j$. These are both scalars, and can be evaluated via simulation; $\rho_\varphi$ depends on $k$ only, so that $\rho_\varphi = \rho_\varphi(k)$. Table 2 reports some simulation results for $\rho_\varphi$ for different values of $k$[9]:

| $k$ | $\rho_\varphi$ | $k$ | $\rho_\varphi$ | $k$ | $\rho_\varphi$ | $k$ | $\rho_\varphi$ |
|---|---|---|---|---|---|---|---|
| 3 | -0.49610 | 9 | -0.10996 | 15 | -0.07888 | 50 | -0.01651 |
| 4 | -0.32662 | 10 | -0.12686 | 20 | -0.06221 | 60 | -0.03107 |
| 5 | -0.25423 | 11 | -0.07933 | 25 | -0.04761 | 70 | 0.02490 |
| 6 | -0.18790 | 12 | -0.08673 | 30 | -0.03339 | 80 | -0.01002 |
| 7 | -0.17044 | 13 | -0.08194 | 35 | -0.0269 | 90 | -0.00890 |
| 8 | -0.16912 | 14 | -0.07101 | 40 | -0.01813 | 100 | -0.01319 |

Table 2: Values of $\rho_\varphi$ for different number of trends.

From $Var(S)$, which is a measure of the spread of $S$ distribution, we can derive the following descriptive measure

$$
\begin{aligned}
d(S) &= \|\tau\|^2 \left\{ \sum_{i=2}^{k} \left\| \tilde{h}_i \right\|^2 \cos^2\left(\tau, \tilde{h}_i\right) + \right. \\
&\quad \left. + \frac{1}{2}\rho_\varphi \sum_{j=2}^{k} \sum_{i \neq j} \left\| \tilde{h}_i \right\| \left\| \tilde{h}_j \right\| \cos\left(\tau, \tilde{h}_i\right) \cos\left(\tau, \tilde{h}_j\right) \right\}
\end{aligned}
$$

or, to make it invariant to $\|\tau\|$, we rewrite it as

$$
D(S) = \frac{1}{\|\tau\|^2} \left\{ \sum_{i=2}^{k} \left(\tau' \tilde{h}_i\right)^2 + \frac{1}{2}\rho_\varphi \sum_{j=2}^{k} \sum_{i \neq j} \left(\tau' \tilde{h}_i\right)\left(\tau' \tilde{h}_j\right) \right\}. \tag{22}
$$

[9] The simulations were performed with GAUSS and the routine is available upon request. The number of replications we employed for each experiment was 10000, given that higher values did not bring any significant changes in the results.

This statistics assumes values in the interval $[0, +\infty)$; the larger, the less the system is close to the case of perfect cointegration, which holds for $D(S) = 0$.

In the light that we are interested in testing for the null hypothesis of cointegration versus the alternative of non cointegration, in what follows we derive the distributional counterpart for $D(S)$. Let $\theta \equiv [\tau', vec'(A)]'$ and $\hat{\theta}$ its OLS estimator, whose limit distribution for large $T$ is derived in section 2.2. The main result for testing is in the following Theorem, whose proof is omitted:

**Proposition 4** *Let $\widehat{D(S)}$ be the OLS counterpart for $D(S)$. Then, under the null hypothesis of cointegration $D(S) = 0$ and*

$$T\left[\widehat{D(S)} - D(S)\right] = T\widehat{D(S)} \Rightarrow [J(\theta)]' L_\theta, \tag{23}$$

*where the $n(k+1) \times 1$ vector $J(\theta)$ is defined as*

$$J(\theta) \equiv \frac{\partial D(S)}{\partial \theta}.$$

**Remarks**

(a) Equation (23) is an application of the Delta method for nonlinear transformation of statistics - see Greene (1993).

(b) A closed form expression for the Jacobian $J(\theta)$ is not available, as this depends upon unknown functions of the $\beta_i$s and $\alpha_{ij}$s. These functions are the $k-1$ rotation angles $(\vartheta_1, ...\vartheta_{k-1})$ in $M^* = M^*(\vartheta_1, ...\vartheta_{k-1})$.

(c) The limit distribution of $T\widehat{D(S)}$ depends upon unknown quantities, i.e. the elements of vector $J(\theta)$, since these are functions of the true values of the system parameters. This problem is solved employing $J\left(\hat{\theta}\right)$; also, $J\left(\hat{\theta}\right) \rightarrow_p J(\theta)$.

(d) Notice that Proposition 5 holds for finite $n$. For large $n$, sequential limit theory as developed in Proposition 2 holds for the aggregate relationship estimator, and therefore the testing framework developed within this section is no longer valid.

The hypothesis testing framework is therefore:

$$\begin{cases} H_0 : D(S) = 0 \\ H_1 : D(S) > 0 \end{cases} \tag{24}$$

and the statistics we will employ is $T\widehat{D(S)}$.

## 4.1 A numerical example

The practical use of the measure of departure from cointegration $D(S)$ and its testing framework as in (23) can be illustrated with the following example. Consider a panel of data with five units ($n = 5$), where the number of trends $k$ has been found to be 3 and the three largest principal components among the independent variables are estimated. Also, for the sake of simplicity, we assume $\varepsilon_t \sim N(0, I_q)$, being $q = 13$. Suppose further that appropriate tests lead to the conclusion that all the series contain a unit root and that individual units cointegrate with coefficients

$$\widehat{\tau} = \begin{pmatrix} 0.046 & 0.037 & 0.254 & 0.53 & 0.807 \end{pmatrix}'$$

In the next step, regression of independent variables on the principal components is performed. Let the estimated coefficients be

$$\widehat{A} = \begin{pmatrix} 0.98 & 0.76 & 0.60 & 0.61 & 0.60 \\ 0.49 & 0.38 & 0.30 & 0.43 & 0.38 \\ 0.45 & 0.39 & 0.35 & 0.53 & 0.44 \end{pmatrix}'.$$

From the coefficient matrices $\widehat{\tau}$ and $\widehat{A}$, we have $\|\hat{\tau}\| = 1$, $\widehat{h}_2 = \begin{pmatrix} -0.0285 & -0.0221 & -0.0175 & 0.0453 & 0.0229 \end{pmatrix}'$ and $\widehat{h}_3 = \begin{pmatrix} -0.0677 & -0.0335 & -0.0397 & 0.1126 & 0.0378 \end{pmatrix}'$. Matrix $M$ can be defined as

$$\begin{aligned} M &= \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} = \\ &= \begin{bmatrix} \cos\phi\cos\theta & -\sin\phi & \cos\phi\sin\theta \\ \sin\phi\cos\theta & \cos\phi & \sin\phi\sin\theta \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \end{aligned}$$

The rotation angles we need are, in radiants, $\phi^* = 2.573$ and $\theta^* = 1.711$, so that

$$M^* = \begin{pmatrix} 0.118 & -0.538 & -0.834 \\ -0.0752 & -0.843 & 0.533 \\ -0.990 & 0 & -0.140 \end{pmatrix}$$

and

$$\widehat{A}^* = M^*\widehat{A} = \begin{pmatrix} -0.367 & -0.325 & -0.298 & -0.485 & -0.393 \\ -0.941 & -0.729 & -0.576 & -0.691 & -0.643 \\ -0.619 & -0.486 & -0.390 & -0.354 & -0.360 \end{pmatrix}'.$$

Using the last outcome, $\widehat{h}_2 = \begin{pmatrix} 0.0664 & 0.0298 & 0.0012 & -0.0666 & -0.0308 \end{pmatrix}'$ and $\widehat{h}_3 = \begin{pmatrix} 0.0842 & 0.0461 & 0.0168 & -0.0994 & -0.0477 \end{pmatrix}'$. Gathering the results, we see that the departure from cointegration is

$$\widehat{D(S)} = 0.00744.$$

Here too we ran two simulations, for $T = 30$ and $T = 100$. The simulations gave the following results[10]:

| $T$ | *p-value* |
|---|---|
| 30 | 0.7954 |
| 100 | 0.4174 |

Table 3: P-values for the null hypothesis of cointegration.

Such p-values would both lead to accept the null hypothesis (presence of cointegration).

# 5 SMALL SAMPLE PROPERTIES

The results about the distribution of $\hat{\beta}$ were obtained under the large sample hypothesis in section 2.1, and were therefore valid only asymptotically. In this section we would like to see how well the asymptotic distribution characterizes the real small sample distribution.

In order to examine the small sample properties of the OLS estimator $\widehat{\beta}$ we evaluate data generated by the system described by equations (1)-(3), setting $p = 1$ as having a higher number of covariates in the regression equation didn't change the results very much. We consider sample sizes of $T = 30$ and $T = 100$ for their being representative of the range of annual and quarterly data in empirical applications. In our experiments we choose the values of $\tau$ and $A$ such that $M = I_{k-1}$, and $\tau' h_i = 1$ for any $i$, and we then generate the parameters $A$ and $\tau$ randomly subject to these constraints. Let $\mu_\varphi \equiv E(\varphi)$; then

$$E(S - S_1) = (k-1)\,\mu_\varphi$$

$$Var(S) = \sigma_\varphi^2 (k-1)\left[1 + \frac{1}{2}\rho_\varphi (k-2)\right]$$

We considered $n = 5$ and $k = 3$, obtaining

$$E(S - S_1) = 0.659$$

[10] The simulations were performed with GAUSS and the routine is available upon request. The number of replications we chose for each experiment was 50000.

$$Var\,(S) = 0.301$$

Different, finite $n$ didn't change much the results, and also simulations for higher values of $k$ didn't result in great changes[11]. The number of replications in all experiments is 50000. All innovations in each experiment of the simulation are set to follow processes of the ARMA (1,1) form where the noise terms are independent standard normal. The set of values for both autoregressive parameter $\rho$ and moving average parameter $\vartheta$ is $\{-0.9, -0.3, 0, 0.3, 0.9\}$ where the values $\pm 0.3$ represent a moderate departure from non autocorrelation and $\pm 0.9$ represent a nearly nonstationary or nearly non-invertible processes. In correspondence with the underlying model, we generate the stochastic trends $z$ by summing the ARMA errors and then we scale them by the square root of their long-run variance $\lambda = (1+\vartheta)^2/(1-\rho)^2$. To keep the variance of the innovations comparable across the experiments with different ARMA parameters, we normalize the stationary errors in the equations generating $x_t$ and $y_t$ by the square root of their variance $\sigma^2 = (\vartheta^2 + 2\vartheta\rho + 1)/(1-\rho^2)$. To identify the effect of the serial correlation in different parts of the system, we distinguish four cases.

First, the trends are generated as ARMA processes and no noise is assumed in the processes generating $x_t$ and $y_t$. The results for the case $T = 30$ are reported in Table 3a. We notice that the mean of the asymptotic distribution is a good guide even in sample of this size, even if it results to be slightly upwardly biased for any couple $(\rho, \vartheta)$. With very few exceptions, the variance increases monotonically with respect to both autoregressive and moving average parameters. When both ARMA parameters are negative and large, the small sample variance found to be about 5 times smaller than the asymptotic value; when instead the MA process approaches a non invertible one, the observed variance is nearly twice the true one. Such discrepancy means that the asymptotic value is not a precise guide. Nonetheless, in the case of negative parameters the spread of the distribution is actually much better than what we would conclude from the limiting distribution. On the other hand, for the samples of this size the variance will be at most three times bigger than the asymptotic variance.

[11] Simulation for $k > 3$ and any $n$, and the GAUSS code to perform them, are available upon request.

| $\vartheta$ | -0.9 | | -0.3 | | 0 | | 0.3 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | mean | var | mean | var | mean | var | mean | var | mean | var |
| -0.9 | 0.666 | 0.054 | 0.667 | 0.057 | 0.668 | 0.091 | 0.669 | 0.137 | 0.669 | 0.197 |
| -0.3 | 0.666 | 0.028 | 0.669 | 0.127 | 0.669 | 0.174 | 0.670 | 0.200 | 0.670 | 0.214 |
| 0 | 0.666 | 0.047 | 0.669 | 0.161 | 0.669 | 0.200 | 0.670 | 0.219 | 0.670 | 0.228 |
| 0.3 | 0.666 | 0.083 | 0.669 | 0.202 | 0.670 | 0.232 | 0.670 | 0.246 | 0.670 | 0.252 |
| 0.9 | 0.665 | 0.546 | 0.668 | 0.568 | 0.670 | 0.598 | 0.671 | 0.620 | 0.671 | 0.626 |

Table 4a: Mean and variance of the simulated distribution of $\widehat{\beta}$, $T = 30$.

The experiments for $T = 100$ are given in Table 3b. The results are qualitatively almost the same as in the previous case, even though the over-prediction for variance is now more shrunk towards the true value.

| $\vartheta$ | -0.9 | | -0.3 | | 0 | | 0.3 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | mean | var | mean | var | mean | var | mean | var | mean | var |
| -0.9 | 0.667 | 0.019 | 0.669 | 0.072 | 0.670 | 0.124 | 0.670 | 0.166 | 0.670 | 0.197 |
| -0.3 | 0.667 | 0.026 | 0.670 | 0.164 | 0.670 | 0.188 | 0.670 | 0.198 | 0.670 | 0.201 |
| 0 | 0.667 | 0.046 | 0.670 | 0.182 | 0.670 | 0.198 | 0.670 | 0.203 | 0.670 | 0.205 |
| 0.3 | 0.667 | 0.078 | 0.670 | 0.198 | 0.670 | 0.208 | 0.670 | 0.211 | 0.670 | 0.212 |
| 0.9 | 0.668 | 0.349 | 0.672 | 0.347 | 0.672 | 0.350 | 0.671 | 0.349 | 0.671 | 0.349 |

Table 4b: Mean and variance of the simulated distribution of $\widehat{\beta}$, $T = 100$.

The second set of experiments is carried out within same setting with addition of white noise errors into the equations generating $x_t$ and $y_t$. Table 4a reports the experiment for the case of $T = 30$. In this case, the mean of the short sample distribution tends, with few exceptions, to underestimate its asymptotic counterpart. The bias decreases with both $\rho$ and $\vartheta$. The variance of the sample distribution is now on average closer to the asymptotic value than in the previous case. Moreover, the small sample variance is now always smaller than the asymptotic value except for values of $\rho$ close to 1. That means that the real variance will be actually more favorable than its asymptotic prediction. In the case of $T = 100$ (see Table 4b), the pattern of the sample variances is preserved. The small sample values are now closer to the limiting values though the speed of convergence is perhaps not as fast as would be expected.

| $\vartheta$ | -0.9 | | -0.3 | | 0 | | 0.3 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | mean | var | mean | var | mean | var | mean | var | mean | var |
| -0.9 | 0.662 | 0.054 | 0.639 | 0.056 | 0.633 | 0.086 | 0.630 | 0.126 | 0.635 | 0.177 |
| -0.3 | 0.653 | 0.028 | 0.625 | 0.117 | 0.629 | 0.157 | 0.634 | 0.180 | 0.643 | 0.196 |
| 0 | 0.0650 | 0.047 | 0.628 | 0.146 | 0.634 | 0.180 | 0.640 | 0.197 | 0.648 | 0.210 |
| 0.3 | 0.648 | 0.081 | 0.634 | 0.181 | 0.641 | 0.209 | 0.646 | 0.223 | 0.652 | 0.234 |
| 0.9 | 0.646 | 0.493 | 0.644 | 0.474 | 0.650 | 0.508 | 0.654 | 0.532 | 0.659 | 0.558 |

Table 5a: Mean and variance of the simulated distribution of $\widehat{\beta}$, $T = 30$.

| $\vartheta$ | -0.9 | | -0.3 | | 0 | | 0.3 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | mean | var | mean | var | mean | var | mean | var | mean | var |
| -0.9 | 0.663 | 0.019 | 0.654 | 0.070 | 0.654 | 0.119 | 0.655 | 0.158 | 0.658 | 0.188 |
| -0.3 | 0.655 | 0.025 | 0.651 | 0.154 | 0.655 | 0.178 | 0.658 | 0.188 | 0.662 | 0.195 |
| 0 | 0.654 | 0.045 | 0.654 | 0.172 | 0.658 | 0.189 | 0.661 | 0.195 | 0.663 | 0.199 |
| 0.3 | 0.653 | 0.076 | 0.658 | 0.189 | 0.661 | 0.201 | 0.663 | 0.205 | 0.665 | 0.208 |
| 0.9 | 0.660 | 0.336 | 0.668 | 0.339 | 0.669 | 0.343 | 0.669 | 0.344 | 0.669 | 0.345 |

Table 5b: Mean and variance of the simulated distribution of $\widehat{\beta}$, $T = 100$.

In the third set of experiments, the innovations generating the trends $z$ are white noise while we now allow the errors in $x$ and $y$ to follow ARMA processes. The values of parameters $\rho$ and $\vartheta$ in Tables 5a and 5b now refer to the noise in the variables instead in the trends. In this setting, the asymptotic variance predicts the small sample variance remarkably better, even though the prediction is always downwardly biased. The underprediction however doesn't exceed 30 percent for the case of $T = 30$; this performance is only slightly improved for the case of $T = 100$. The mean of the sample distribution is underpredicted for $T = 30$ by up to 5 percent, but this underprediction vanishes quickly as the sample size increases.

| $\vartheta$ | -0.9 | | -0.3 | | 0 | | 0.3 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | mean | var | mean | var | mean | var | mean | var | mean | var |
| -0.9 | 0.640 | 0.182 | 0.640 | 0.181 | 0.639 | 0.181 | 0.638 | 0.181 | 0.630 | 0.178 |
| -0.3 | 0.634 | 0.176 | 0.634 | 0.177 | 0.635 | 0.179 | 0.635 | 0.180 | 0.635 | 0.183 |
| 0 | 0.634 | 0.176 | 0.634 | 0.178 | 0.635 | 0.181 | 0.636 | 0.183 | 0.636 | 0.185 |
| 0.3 | 0.634 | 0.177 | 0.635 | 0.180 | 0.636 | 0.184 | 0.637 | 0.187 | 0.638 | 0.189 |
| 0.9 | 0.631 | 0.181 | 0.648 | 0.207 | 0.651 | 0.211 | 0.651 | 0.212 | 0.652 | 0.213 |

Table 6a: Mean and variance of the simulated distribution of $\widehat{\beta}$, $T = 30$.

| $\vartheta$ | -0.9 | | -0.3 | | 0 | | 0.3 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | mean | var | mean | var | mean | var | mean | var | mean | var |
| -0.9 | 0.659 | 0.189 | 0.659 | 0.189 | 0.659 | 0.189 | 0.659 | 0.189 | 0.658 | 0.188 |
| -0.3 | 0.658 | 0.188 | 0.658 | 0.188 | 0.658 | 0.189 | 0.658 | 0.189 | 0.658 | 0.189 |
| 0 | 0.658 | 0.188 | 0.658 | 0.189 | 0.658 | 0.189 | 0.659 | 0.189 | 0.659 | 0.189 |
| 0.3 | 0.658 | 0.188 | 0.658 | 0.189 | 0.659 | 0.189 | 0.659 | 0.190 | 0.659 | 0.190 |
| 0.9 | 0.658 | 0.189 | 0.661 | 0.194 | 0.661 | 0.195 | 0.662 | 0.196 | 0.662 | 0.196 |

Table 6b: Mean and variance of the simulated distribution of $\widehat{\beta}$, $T = 100$.

Finally, in the last set of experiments we let all the innovations in the system to follow an ARMA process with identical parameter values. The mean of the small sample distribution behaves in similar way to the case in which the innovations in $x_t$ and $y_t$ variables follow only a white noise process. The mean is again underpredicted for the smaller sample sizes but the value of the mean becomes closer to the asymptotic value in larger samples, with the worst underprediction being observed for $\rho = -0.9$. The variance, on the other hand, follows the pattern of the case where there is no noise in the variables $x_t$ and $y_t$. The variance again slowly converges to the asymptotic values.

| $\vartheta$ | -0.9 | | -0.3 | | 0 | | 0.3 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | mean | var | mean | var | mean | var | mean | var | mean | var |
| -0.9 | 0.345 | 0.251 | 0.434 | 0.154 | 0.502 | 0.124 | 0.565 | 0.127 | 0.631 | 0.176 |
| -0.3 | 0.440 | 0.053 | 0.575 | 0.107 | 0.616 | 0.150 | 0.635 | 0.180 | 0.645 | 0.199 |
| 0 | 0.491 | 0.054 | 0.610 | 0.139 | 0.635 | 0.181 | 0.646 | 0.203 | 0.650 | 0.215 |
| 0.3 | 0.544 | 0.074 | 0.635 | 0.182 | 0.648 | 0.217 | 0.653 | 0.233 | 0.655 | 0.241 |
| 0.9 | 0.646 | 0.492 | 0.662 | 0.555 | 0.664 | 0.588 | 0.665 | 0.605 | 0.667 | 0.616 |

Table 7a: Mean and variance of the simulated distribution of $\widehat{\beta}$, $T = 30$.

| $\vartheta$ | -0.9 | | -0.3 | | 0 | | 0.3 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | mean | var | mean | var | mean | var | mean | var | mean | var |
| -0.9 | 0.338 | 0.090 | 0.487 | 0.078 | 0.564 | 0.106 | 0.617 | 0.143 | 0.658 | 0.188 |
| -0.3 | 0.450 | 0.032 | 0.626 | 0.143 | 0.650 | 0.175 | 0.658 | 0.189 | 0.662 | 0.195 |
| 0 | 0.507 | 0.042 | 0.646 | 0.167 | 0.658 | 0.189 | 0.662 | 0.197 | 0.664 | 0.200 |
| 0.3 | 0.564 | 0.065 | 0.658 | 0.189 | 0.664 | 0.203 | 0.665 | 0.207 | 0.666 | 0.208 |
| 0.9 | 0.660 | 0.335 | 0.670 | 0.345 | 0.670 | 0.348 | 0.670 | 0.348 | 0.670 | 0.347 |

Table 7b: Mean and variance of the simulated distribution of $\widehat{\beta}$, $T = 100$.

According to the four experiments, the small sample mean seems to be affected by the presence of noise in the processes generating $x_t$ and $y_t$. The degree of misprediction does not depend very much on the structure of the noise, which suggests that the presence of autocorrelation does not make the results worst. The variance of the small sample distribution, on the contrary, is mainly influenced by the presence of autocorrelation and moving average components in the innovation generating $z_t$. Finally, both mean and variance of the small sample distribution approach the asymptotic values as $T$ increases.

The main issue that we wanted to pursue with these experiments was whether the asymptotic distribution of $\widehat{\beta}$ was a valuable guide for the small samples case. The conclusion from the experiments is that at worst the variance in the small sample is five times smaller even for relatively large positive values of both autoregressive and moving average parameters. Furthermore, if the degree of autocorrelation is only moderate, the small sample variance is actually lower than the asymptotic value. This leads us to the conclusion that the knowledge of the limiting distribution of $\widehat{\beta}$ is of use in order to estimate the upper bound of the degree of non-cointegration in real data.

# 6 CONCLUSIONS

In nonstationary heterogeneous panels where each unit cointegrates, the aggregate relationship in general does not cointegrate unless a large number of conditions are satisfied. To satisfy aggregation conditions, literature has shown that a necessary and sufficient condition is that the micro regressors share a single common stochastic trend in a single covariate regression equation framework. Alternatively, a sufficient condition that applies to cointegration relationships with more than one covariate is that the amount of cointegration in each micro relationship is enough to preserve the existence of a long-run relationship in the aggregate equation as well.

The main contribution of this paper is twofold. First, we derive necessary and sufficient conditions to ensure the validity of the aggregate cointegration equation, by evaluating the limit distribution of the aggregate estimate, proving that this tends to a constant, thus implying the validity of the superconsistency property of the estimator.

Second, we derive the asymptotic measure of the distance, $D(S)$, between the case of perfect cointegration ($D(S) = 0$) and that with very heterogeneous response of the system ($|D(S)| > 0$). This statistics is derived from the variance of the limit distribution of the aggregate estimate when cointegration does not hold. In addition, we derive an estimation theory and a testing strategy to assess the degree of departure from perfect cointegration. This stage is aimed at verifying whether the violation of the necessary and sufficient condition is mild enough to preserve cointegration.

Finally, we conducted a series of Monte Carlo simulations in order to evaluate whether the asymptotic distribution of the $\widehat{\beta}$ estimator was a valuable guide for the small samples case. Both mean and variances show values different from the limiting distribution counterparts. However, the mean and variance of the limiting distribution are good approximations of the small sample case. Moreover, the small sample mean seems to be affected by the presence of noise in the processes generating the variables $x_t$ and $y_t$, while the degree of misprediction does not depend very much on the structure of the noise. The variance of the small sample distribution, on contrary, is mainly influenced by the presence of autocorrelation and moving average components in the innovation generating $z_t$.

Our paper is a further and formal support to the view that even if the aggregation conditions are 'slightly' violated, the aggregate regression is still useful in characterizing the macro relationship.

# References


[1] Ghose, D. (1995), "Linear Aggregation in Cointegrated Systems", *Journal of Economic Dynamic and Control,* 19, 1011-1032.

[2] Gonzalo, J. (1993), "Cointegration and Aggregation", *Ricerche Economiche,* 47, 281-291.

[3] 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.

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

[5] Greene, W. (1993), "Econometric Analysis", *Prentice Hall*, 3rd ed.

[6] Hall, S., Lazarova, S. and G. Urga (1999), "A Principal Components Analysis of Common Stochastic Trends in Heterogeneous Panel Data: Some Monte Carlo Evidence", *Oxford Bulletin of Economics and Statistics*, 61, 749-767.

[7] Harris, D. (1997), "Principal Components Analysis for Cointegrated Time Series", *Econometric Theory*, 13, 529-577.

[8] Lazarova, S., Trapani, L., Urga, G. (2003), "Common Stochastic Trends and Aggregation in Heterogeneous Panels", mimeo, Cass Business School.

[9] Park, J.Y. and Phillips, P.C.B. (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1", *Econometric Theory*, 4, 468-497.

[10] Park, J.Y. and Phillips, P.C.B. (1989), "Statistical Inference in Regressions with Integrated Processes: Part 2", *Econometric Theory*, 5, 95-131.

[11] Pesaran, M.H. and Smith R. (1995), "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels", *Journal of Econometrics*, Vol. 68, 79-113.

[12] Phillips, P.C.B. (1986), "Understanding Spurious Regressions in Econometrics", *Journal of Econometrics,* 33, 311-340.

[13] Phillips, P.C.B. and Moon, H.R. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica*, 67, 1057-1111.

[14] Phillips, P.C.B. and Ouliaris, S. (1988), ”Testing for Cointegration Using Principal Component Methods”, *Journal of Economic Dynamic and Control,* 12, 205-230.

[15] Pollock, D. (1979), ”The Algebra of Econometrics”, John Wiley & Sons.

[16] Snell, A. (1999), ”Testing for *r* versus *r-1* Cointegrating Vectors”, *Journal of Econometrics*, 88, 151-191.

# Appendix I: proofs and derivations

## Proof of Lemma 2.2

Proof is immediate. Consider equation (8)

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

The $x_{it}$ cointegrate among themselves if and only if there exists a vector $\zeta_i \in R^p \backslash \{0\}$ s.t.

$$\zeta_i' x_{it} = \zeta_i' \Gamma_i z_t + \zeta_i' v_{it} = \zeta_i' v_{it} \sim I(0),$$

i.e. if and only if $\zeta_i' \Gamma_i = 0$. This means that in order for a non trivial $\zeta_i$ to exist, the space spanned by the columns of $\Gamma_i$ must have smaller dimension than $p$. Requiring $\Gamma_i$ to have full rank rules out this possibility.

## Proof of Proposition 2

The proof follows from Phillips and Moon (1999):

$$\begin{aligned} \hat{\beta}_{n,T} &\Rightarrow S \equiv \left[\Gamma \int W_z W_z' \Gamma'\right]^{-1} \left[\Gamma \int W_z W_z' b\right] = \\ &= \left[\sum \Gamma_i \int W_z W_z' \sum \Gamma_i'\right]^{-1} \left[\sum \Gamma_i \int W_z W_z' \sum \Gamma_i' \beta_i\right] \end{aligned}$$

where the sums are for $i$ from 1 to $n$. According to assumption 2.3 and Lemma 2.2:

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

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

Also, according to Lemma 5(a) on pg. 11 of Phillips and Moon (1999), under assumption 2.4 the joint and the sequential limit are equivalent.

## Proof of Lemma 3.1

The proof follows from Proposition 3. If $\Gamma$ is a $k \times k$ matrix $rank(\Gamma) = k$ under full rank assumption, and $rank(\Gamma' \mid b) = k$ as well. This is sufficient for a non trivial $c$ to exist. Another, different proof can be found in Gonzalo (1993).

## Derivation of (20)

The derivation of equation (20) is a simple algebraic manipulation of

$$S - S_1 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i b_j W_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}} - \frac{b_1}{a_1}.$$

The following passages lead to equation (20):

$$\begin{aligned}
\frac{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i b_j W_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}} - \frac{b_1}{a_1} &= \frac{a_1\sum_{i=1}^{k}\sum_{j=2}^{k} a_i b_j W_{ij} - b_1\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}}{a_1\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}} = \\
&= \frac{\sum_{i=1}^{k}\sum_{j=2}^{k}\left[a_1 a_i b_j - b_1 a_i a_j\right] W_{ij}}{a_1\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}} = \\
&= \frac{\sum_{i=1}^{k}\sum_{j=2}^{k}\left[a_i b_j - \frac{a_i a_j}{a_1} b_1\right] W_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}} = \\
&= \frac{\sum_{i=1}^{k}\sum_{j=2}^{k} a_i a_j\left[\frac{b_j}{a_j} - \frac{b_1}{a_1}\right] W_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}}
\end{aligned}$$

and, recalling the definition of $S_j$

$$S - S_1 = \sum_{l=2}^{k}\left(\frac{\sum_{i=1}^{k} a_i W_{il} a_l}{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}}\right)(S_l - S_1)$$

which is equation (20) for

$$\left(\frac{\sum_{i=1}^{k} a_i W_{il}}{\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j W_{ij}}\right) a_l \equiv f_l$$

Also, it is immediate to derive the useful relationship $(S_i - S_1) = \tau' h_i$:

$$\begin{aligned}
(S_i - S_1) &= \frac{b_i}{a_i} - \frac{b_1}{a_1} = \frac{\sum_{j=1}^{n}\beta_i\alpha_{ij}}{\sum_{j=1}^{n}\alpha_{ij}} - \frac{\sum_{j=1}^{n}\beta_1\alpha_{1j}}{\sum_{j=1}^{n}\alpha_{1j}} = \\
&= \frac{\tau' A_{.i}}{\sum_{j=1}^{n}\alpha_{ij}} - \frac{\tau' A_{.1}}{\sum_{j=1}^{n}\alpha_{1j}} = \tau' h_i.
\end{aligned}$$

where $A_{.i}$ denotes $A$'s $i$-th column. QED.

# Appendix II: the measure *D(S)* for *p>1*

This Appendix considers the generalization of the measure $D(S)$ to the case $p > 1$. The framework we consider is similar to that which leads up to the measure $D(S)$ when the numer of covariates $p$ is equal to 1.

First, we recall that the difference between the finite $n$, large $T$ OLS estimator $\hat{\beta}$ and its true value $\beta$ is

$$\hat{\beta} - \beta \rightarrow [\Gamma B \Gamma']^{-1} [\Gamma B] [b - \Gamma' \beta]$$

where, for notational simplicity, $B \equiv \int WW'$. According to the theory developed within our framework, such difference is a random variable, and it collapses to a vector of numbers if and only if there is cointegration. Hence, the covariance matrix of the $p$-dimensional vector $\hat{\beta} - \beta$ is a null matrix if and only if cointegration is maintained after aggregation. Henceforth, this $p \times p$ matrix will be referred to as $\Sigma_\beta \equiv Var\left(\hat{\beta} - \beta\right)$. Since a matrix is null if and only if its rank is equal to zero, and since $\Sigma_\beta$ is a semipositive definite matrix, $\Sigma_\beta$ rank will be equal to zero if and only if its trace is zero. Therefore, the testing framework, analog to (24) is

$$\begin{cases} H_0 : tr(\Sigma_\beta) = 0 \\ H_1 : tr(\Sigma_\beta) > 0 \end{cases}.$$

Now, consider the $k \times k$ rotation matrix $N$. We may write

$$\bar{x}_t = \Gamma z_t + \bar{v}_t = \Gamma N' N z_t + \bar{v}_t = \Gamma^* z_t^* + \bar{v}_t$$

$$\bar{y}_t = b' z_t + \bar{s}_t = b' N' N z_t + \bar{s}_t = b^{*\prime} z_t^* + \bar{s}_t$$

and, noticing that $Var(z_t) = Var(z_t^*) = I_k$

$$\hat{\beta} - \beta \rightarrow [\Gamma^* B \Gamma^{*\prime}]^{-1} [\Gamma^* B] [b^* - \Gamma^{*\prime} \beta]. \tag{25}$$

Now, let $N$ be such that $\Gamma^* = [I_p | O]$, where $O$ is a $p \times (k - p)$ matrix of zeroes. The random variables $\Gamma^* B \Gamma^{*\prime} \equiv B_p$ and $\Gamma^* B \equiv B_{pk}$ are both made of standard brownian motions, and their moments can be obtained once and for all via simulation for every couple $(p, k)$. Equation (25) can be rewritten as

$$\hat{\beta} - \beta \rightarrow B_p^{-1} B_{pk} (b^* - \Gamma^{*\prime} \beta) = B_{ppk} (b^* - \Gamma^{*\prime} \beta) = B_{ppk} d.$$

The following passages lead to a suitable formulation for $\Sigma_\beta$. First, since $B_{ppk} d = Vec(B_{ppk} d)$, we have $\Sigma_\beta = Var[Vec(B_{ppk} d)]$; now

$$Vec(B_{ppk} d) = (d' \otimes I_p) Vec(B_{ppk}).$$

Therefore

$$\begin{aligned}\Sigma_\beta &= Var\left[Vec\left(B_{ppk}d\right)\right] = \\ &= \left(d' \otimes I_p\right) Var\left[Vec\left(B_{ppk}\right)\right] \left(d' \otimes I_p\right)'.\end{aligned}$$

Letting $Var\left[Vec\left(B_{ppk}\right)\right] \equiv V$, and since $\left(d' \otimes I_p\right)' = d \otimes I_p$ we get

$$\begin{aligned}tr\left(\Sigma_\beta\right) &= tr\left[\left(d' \otimes I_p\right) V \left(d \otimes I_p\right)\right] = \\ &= tr\left[\left(d \otimes I_p\right)\left(d' \otimes I_p\right) V\right] = \\ &= tr\left[\left(dd' \otimes I_p\right) V\right].\end{aligned}$$

Therefore, the null hypothesis becomes

$$tr\left[\left(dd' \otimes I_p\right) V\right] = 0.$$

The first order approximation of the distribution of this statistics can be obtained by employing the Delta method, as in Proposition 5. However, this would lead to a pretty awkward formulation, and as a feasible alternative we suggest bootstrap.