Lazarova, S., Trapani, L. & Urga, G. (2007). Common stochastic trends and aggregation in heterogeneous panels. Econometric Theory, 23(1), pp. 89-105. doi: 10.1017/S0266466607070041

Common stochastic trends and aggregation in heterogeneous panels*

Štěpána Lazarová
London School of Economics

Lorenzo Trapani
Cass Business School and Universita' di Bergamo

Giovanni Urga†
Cass Business School

This version: October 19, 2006

*We wish to thank Stephen Hall for inspiring discussions and comments during the initial stage of this research. We are grateful to participants in the 9th Conference on Panel Data (Geneva, 22-23 June, 2000), the 2000 Meeting of the Society for Economic Dynamics (San José, Costa Rica, 29 June - 2 July 2000) and colleagues at University of Bergamo (Italy), London Business School (UK) and Cass Business School (UK) for useful comments. We have enormously benefited from comments and suggestions from two anonymous referees, the Editor, Peter C. B. Phillips and the Co-Editor, Bruce Hansen. The usual disclaimer applies. G. Urga acknowledges ESRC (Project N. R022251032) for financial support. Lorenzo Trapani acknowledges financial support from Cass Business School (RAE Development Funds Scheme) and ESRC Postdoctoral Fellowship Scheme (A.N. PTA-026-27-1107).

†Corresponding author: Cass Business School, Faculty of Finance, 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.cass.city.ac.uk/faculty/g.urga.

Common stochastic trends and aggregation in heterogeneous panels

Abstract

In nonstationary heterogeneous panels where the number of units is finite and where each unit cointegrates, a large number of conditions needs to be satisfied for cointegration to be preserved in the aggregate relationship. In reality, the conditions most likely will not hold. This paper takes a closer look at what happens when the conditions are violated. In this case, it is of particular interest the question of whether an aggregate relationship is observationally equivalent to a cointegrating equation. We derive a measure of the degree of noncointegration of the aggregate estimates and we explore its asymptotic properties.

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

1 INTRODUCTION

The issue of aggregation has been of considerable interest in the econometric literature. Many macroeconomic theories are based on the behaviour of individual agents, households or firms, but often only aggregate data are available to test the theories. The question then arises of how well the aggregate relationship approximates the properties of the individual components. Conversely, the data may be given at a disaggregated level. The micro relationships can then be summarised in many ways, the simple aggregation of the components being one of the possibilities.

When the variables in the economic system are integrated, an important observation is that the cointegration on the micro level does not automatically imply cointegration on the macro level (see e.g. Pesaran and Smith, 1995). If cointegration does not carry through the aggregation process, the macro estimates are not consistent, rendering the information provided by macro summary meaningless. It was long thought that satisfying the representative agent assumption is the only way to preserve cointegration. Gonzalo (1993) shows that this conjecture is not correct and that the agents need not be homogeneous. When the micro cointegrating vectors are heterogeneous across units, the cointegration can still be preserved if there is a sufficient degree of cointegration among the variables in the economic system, i.e. if the series are driven by sufficiently low number of common stochastic trends.

Granger (1993) considers a case in which only few common stochastic trends are shared across virtually all of the original series of the model. The remaining trends are shared by only small groups of the series. In such a case, the coefficients of the shared common trends in the aggregate regression are larger than the coefficients for the idiosyncratic common trends by an order of magnitude. Removing the large trends from the aggregate regression leaves only "small" $I(1)$ elements in the residuals that may not be found by standard tests applied to relatively small samples. In this case, the system is described adequately by a small number of dominant components and the aggregate relationship "approximately cointegrates".

In this paper we intend to bring further insights into the aggregation conditions for nonstationary heterogenous panels. Our standpoint is that in real economic systems the tight set of aggregation conditions is indeed unlikely to be satisfied. We believe, however, that the aggregate relationship does not become entirely meaningless when the conditions get "mildly violated", in that, though the panel equation might not satisfy the formal condition for cointegration, the aggregate data may only have "small" non-stationary components and this makes the (strictly speaking spurious) macro relationship observationally equivalent to a cointegration equation. We argue that when the aggregate relationship "approximately cointegrates", in the sense stated above, then it should be treated as if it were actually a cointegrating relationship. In this paper, we derive a measure of the degree of non-cointegration of the aggregate estimate and we explore its asymptotic properties.

The remainder of the paper is organised as follows. The theoretical framework is presented in Section 2, where we set up a model of a heterogenous panel. In Section 3, we discern the factors determining the behaviour of the aggregate estimate and we derive an asymptotic measure of the distance from the case of perfect cointegration. Section 4 concludes.

2 THE MODEL

Let us consider a simple system of n cointegrated micro relationships each with one explanatory variable,

$$y_{it} = \beta_i x_{it} + u_{it}, \qquad t = 1, 2, ..., T, \quad i = 1, 2, ...n, \tag{1}$$

where the explanatory variables are I(1) processes that share k common trends,

$$x_{it} = \alpha_{i1} z_{1t} + \alpha_{i2} z_{2t} + \ldots + \alpha_{ik} z_{kt} + v_{it}, \tag{2}$$

with

$$z_{jt} = z_{jt-1} + w_{jt}, \qquad j = 1, 2, .., k,$$

where $z_{j0} = 0$. In matrix form the system can be rewritten as

$$\begin{aligned} y_t &= Bx_t + u_t, & (3)\\ x_t &= Az_t + v_t, & (4)\\ z_t &= z_{t-1} + w_t, & (5) \end{aligned}$$

where $y_t = (y_{1t}, ..., y_{nt})'$ with x_t, z_t defined similarly, $B = \text{diag}\,(\beta_1, ...\beta_n)$, $A = (\alpha_1', ...\alpha_n')'$ and $\alpha_i = (\alpha_{i1}, ..., \alpha_{ik})$, and where u_t, v_t, w_t are vectors of unobservable disturbances. The matrix A has rank k so that x_t is driven by no less than k stochastic trends. We define the vector of disturbances $\varepsilon_t = (u_t', v_t', w_t')'$ and the vector of partial sums $S_t = \sum_{i=1}^{t} \varepsilon_i$ with $S_0 = 0$.

We assume that ε_t is a stationary invertible process satisfying the following assumptions.

Assumptions The conditions (a)-(e) below hold:

(a) $E\varepsilon_t = 0 \; \forall t$,

(b) $\sup_{j,t} E \left|\varepsilon_{jt}\right|^{\zeta+\gamma} < \infty$ for some $\zeta > 2$ and $\gamma > 0$,

(c) $\Sigma = \lim_{T\to\infty} T^{-1}E(S_T S_T')$ exists,

(d) the sequence $\{\varepsilon_t\}_{t=1}^{\infty}$ is strong mixing with mixing numbers α_m satisfying $\sum_{m=1}^{\infty} \alpha_m^{1-2/\zeta} < \infty$,

(e) the components of w_t are independent of each other and of $(u_t', v_t')'$, and the trends z_t have unit long-run variance so that the lower diagonal $k \times k$ block of the matrix Σ is an identity matrix, $L_{RV}(z_t) = \lim_{T\to\infty} T^{-1}E(z_t z_t') = I_k$.

Conditions (a)-(d) are necessary for the validity of the functional central limit theorem. The assumption of orthonormality (e) ensures that the behavior of the system is fully described by the coefficients β_i and α_i, so that the trends z_{it} are neutral.

When we aggregate the regressors across the units, we obtain

$$\overline{x}_t = a_1 z_{1t} + a_2 z_{2t} + \ldots + a_k z_{kt} + \overline{v}_t = a' z_t + \overline{v}_t, \tag{6}$$

where $\bar{x}_t = \sum_{i=1}^n x_{it}$, $a = (a_1, ..., a_k)'$, $a_j = \sum_{i=1}^n \alpha_{ij}$ and $\overline{v}_t = \sum_{i=1}^n v_{it}$. We assume there is at least one j for which $a_j \neq 0$, so that $\overline{x}_t$ is $I(1)$. For the dependent variable we have

$$y_{it} = \beta_i \alpha_{i1} z_{1t} + \beta_i \alpha_{i2} z_{2t} + \ldots + \beta_i \alpha_{ik} z_{kt} + \beta_i v_{it} + u_{it},$$

so the aggregate variable is of the following form:

$$\overline{y}_t = b_1 z_{1t} + b_2 z_{2t} + \ldots + b_k z_{kt} + \overline{u}_t = b' z_t + \overline{u}_t \tag{7}$$

where $\overline{y}_t = \sum_{i=1}^n y_{it}$, $b = (b_1, ..., b_k)'$, $b_j = \sum_{i=1}^n \beta_i \alpha_{ij}$ and $\bar{u}_t = \sum_{i=1}^n \beta_i v_{it} + \sum_{i=1}^n u_{it}$. We assume there is at least one j for which $b_j \neq 0$, so that $\overline{y}_t$ contains a unit root.

In this model each unit cointegrates. Our main interest is to examine conditions under which the aggregate relationship cointegrates as well. For this purpose, let us consider the linear regression $\overline{y}_t = \beta \overline{x}_t + e_t$. This equation can either be a cointegration relationship or a spurious regression. We consider the restriction

$$L_{RCV}[\Delta \overline{x}_t, \Delta e_t] = 0,$$

where $L_{RCV}[\Delta \overline{x}_t, \Delta e_t] = \lim_{T\to\infty} T^{-1} \sum_{t=1}^T \sum_{s=1}^T E(\Delta \bar{x}_t \Delta e_s)$. Let now $L_{RV}(\Delta \bar{y}_t) = \lim_{T\to\infty} \text{var}\left[T^{-1/2} \sum_{t=1}^T \Delta \bar{y}_t\right]$ be the long run variance of $\Delta \bar{y}_t$, and indicate with $L_{RV}(\Delta \bar{x}_t)$ and $L_{RV}(\Delta e_t)$ the long run variances of $\Delta \bar{x}_t$ and Δe_t respectively. Then we have

$$L_{RV}(\Delta \bar{y}_t) = \beta^2 L_{RV}(\Delta \bar{x}_t) + L_{RV}(\Delta e_t).$$

Let us examine the behaviour of the least-squares estimator $\hat{\beta}$ of β. The following theorem characterises the limiting behaviour of $\hat{\beta}$ when n is finite.

Theorem 1 *If y_t and x_t are generated by (3)-(5) where the innovation sequence $\{\varepsilon_t\}_{t=1}^{\infty}$ satisfies Assumptions (a)-(e), then in the OLS regression of $\overline{y}_t$ on $\overline{x}_t$, as $T \to \infty$*

$$\widehat{\beta} \xrightarrow{d} S = \frac{a'\left(\int_0^1 W(r) W'(r)\, dr\right) b}{a'\left(\int_0^1 W(r) W'(r)\, dr\right) a} = \\ = \frac{\sum_{i=1}^k \sum_{j=1}^k b_i a_j W_{ij}}{\sum_{i=1}^k \sum_{j=1}^k a_i a_j W_{ij}}$$

where $W = (W_1, ...W_k)'$ is a k-dimensional vector of independent standard Brownian processes, where $W_{ij} = \int_0^1 W_i(r) W_j(r) dr$ and where "$\xrightarrow{d}$" denotes convergence in distribution of the associated probability measures as $T \to \infty$.

Proof. See Park and Phillips (1988, 1989). ■

Theorem 1 states that when the number of cross sectional units n is finite, the aggregate relationship may not cointegrate. When $n \to \infty$, this is no longer the case. Under cross sectional independence, the aggregate relationship cointegrates. Moreover, the asymptotic behavior of the OLS estimator considered here, $\hat{\beta}$, has the same probability limit as the pooled OLS estimator $\hat{\beta}^{POLS}$ in Phillips and Moon (1999), defined as

$$\hat{\beta}^{POLS} = \frac{\sum_{i=1}^n \sum_{t=1}^T y_{it} x_{it}}{\sum_{i=1}^n \sum_{t=1}^T x_{it}^2}.$$

This is stated in the following proposition.

Proposition 1 *Assume that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \beta_i \alpha_i = \left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \beta_i\right) \left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \alpha_i\right),$$

and let $\bar{\beta} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \beta_i$. Then, under cross-sectional independence, when $T \to \infty$ and $n \to \infty$,

$$\hat{\beta}^{POLS} \xrightarrow{p} \bar{\beta}$$

and

$$\hat{\beta} \xrightarrow{p} \bar{\beta}.$$

Proof. See Appendix. ■

Note that, following the conlusions of Phillips and Moon (1999), and Sun (2004) under milder conditions, when $(n, T) \to \infty$ sequentially or jointly, we have that $\hat{\beta} \xrightarrow{p} \bar{\beta}$ at the rate n if cointegration between $\bar{y}_t$ and $\bar{x}_t$ holds and at the rate $\sqrt{n}$ if cointegration between $\bar{y}_t$ and $\bar{x}_t$ fails to hold. Thus, $\hat{\beta}$ converges to the average slope coefficient of the individual relationships, irrespectively of whether $\bar{y}_t$ and $\bar{x}_t$ cointegrate.

3 PERFECT COINTEGRATION VERSUS SPURIOUS REGRESSION

Theorem 1 describes the properties of the estimator $\widehat{\beta}$ in the presence of one or more common stochastic trends. If the x_{it}s are driven by a single stochastic trend, $k = 1$, the limiting distribution of $\widehat{\beta}$ is

$$S = \frac{b_1 a_1 W_{11}}{a_1 a_1 W_{11}} = \frac{b_1}{a_1} = \frac{\sum_{i=1}^{n} \beta_i \alpha_{i1}}{\sum_{i=1}^{n} \alpha_{i1}}. \tag{8}$$

In the presence of a single trend, therefore, the aggregate relationship cointegrates. The OLS estimator $\widehat{\beta}$ converges to a weighted average of β_i coefficients where weights are given by the α_{i1} coefficients. This finding is consistent with the analysis of Gonzalo (1993). Hall, Lazarová 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 all individual units do cointegrate.

When there is more than one common stochastic trend among the right-hand side variables ($k > 1$), the aggregate regression is generally spurious and the estimator $\widehat{\beta}$ converges in distribution to a non-degenerate random variable. The form of the limiting random variable S in Theorem 1 implies that cointegration occurs only if there exists a constant c such that $a_j(b_i - ca_i) = 0$ for every i and j. This condition is satisfied if and only if either

$$a_j = \sum_{i=1}^{n} \alpha_{ij} = 0 \quad \text{for every } j \tag{9}$$

or

$$b_i - ca_i = \sum_{j=1}^{n} (\beta_j - c)\alpha_{ji} = 0 \quad \text{for every } i. \tag{10}$$

If (9) holds, the aggregate series $\overline{x}_t$ does not have a unit root. We exclude this situation by assumption. If (10) holds, either β_i's must be homogeneous,

$\beta_j = c$ for all j, or, letting $\iota = (1, \ldots, 1)$ be an n-dimensional vector of ones, the vector $\tau - c\iota$, where $\tau = (\beta_1, \beta_2, \ldots, \beta_n)'$, needs to be orthogonal to each of the columns of the matrix A. Formally, this could be represented as

$$A' (\tau - c\iota) = 0. \tag{11}$$

This can occur only if the vector $\tau - c\iota$ lies in the space spanned by the columns of matrix A. While this is true when $k = 1$, it does not necessarily hold when $k > 1$.

Condition (11), characterizing the cases under which the aggregate system cointegrates, has been derived from the limiting distribution of S in Theorem 1. One could also arrive at this condition from population regression. This alternative derivation of equation (11) is reported in Appendix.

If there is more than one trend driving the regressors and neither of conditions (9) and (10) is satisfied, the aggregate relationship does not cointegrate. This qualifies the regression as spurious. However, the dispersion of the limiting variable depends on the parameters of the underlying processes, therefore for some values of parameters the distribution can be nearly degenerate. In such a case, the estimator $\widehat{\beta}$, though inconsistent, may not be entirely worthless. It is therefore of some interest to analyse factors on which the variance of the limiting distribution depends.

Theorem 1 indicates that when $k > 1$, the limiting distribution of $\hat{\beta}$ is $S = \left(\int W_a W_b\right) / \left(\int W_a^2\right)$, where the scalars W_a and W_b are Brownian processes defined as $a'W$ and $b'W$, where W is a standard Brownian process. The process W_b can be decomposed as

$$W_b = \frac{a'b}{a'a} W_a + W_c,$$

where W_c is a Brownian process independent of W_a. The variance of W_c is $\operatorname{var} W_c(r) = (b'b - (a'b)^2 / a'a) r$. It follows that

$$\begin{aligned} S &= \frac{a'b}{a'a} + \frac{\int_0^1 W_a(r) W_c(r) \, dr}{\int_0^1 W_a^2(r) \, dr} = \\ &\overset{d}{=} \frac{a'b}{a'a} + \sqrt{\frac{(a'a)(b'b) - (a'b)^2}{(a'a)^2}} S_0, \end{aligned} \tag{12}$$

where "$\overset{d}{=}$" stands for equality in distribution,

$$S_0 = \frac{\int_0^1 B_1(r) B_2(r) \, dr}{\int_0^1 B_1^2(r) \, dr},$$

and B_1, B_2 are standard independent Brownian processes. The first two moments of S are

$$ES = \frac{a'b}{a'a}$$

and

$$\operatorname{var} S = \frac{(a'a)(b'b) - (a'b)^2}{(a'a)^2} \operatorname{var} S_0. \tag{13}$$

If cointegration occurs, S degenerates to a constant value of $a'b/a'a$. When cointegration fails to occur, the dispersion of S around $a'b/a'a$ is nonzero. The variance of S can thus be used as a measure of the distance from the case of aggregate cointegration.[1]

Let now

$$\lambda_0 = \frac{L_{RV}(\Delta \bar{y}_t)}{L_{RV}(\Delta \bar{x}_t)}.$$

Given that equations (6) and (7) are cointegration relationships, it is straightforward to see that $L_{RV}(\Delta \bar{y}_t) = b'b$ and $L_{RV}(\Delta \bar{x}_t) = a'a$. Therefore, from (13) we obtain

$$\begin{aligned} \operatorname{var} S &= \frac{b'b}{a'a}\left(1 - \frac{(a'b)^2}{(a'a)(b'b)}\right) \operatorname{var} S_0 \\ &= \lambda_0 \sin^2(a,b) \operatorname{var} S_0, \end{aligned} \tag{14}$$

where $\sin(a,b)$ denotes the sine of the angle between vectors a and b. Decomposition (14) motivates the construction of a measure of the departure from cointegration. To obtain a scale invariant statistic, we normalise $\operatorname{var} S$ and define

$$d = \frac{\operatorname{var} S}{(ES)^2 \operatorname{var} S_0} = \left(\frac{a'a}{a'b}\right)^2 \lambda_0 \sin^2(a,b).$$

Further,

$$d = \frac{(a'a)^2}{(a'b)^2}\frac{(b'b)}{(a'a)} \sin^2(a,b) = \frac{\sin^2(a,b)}{\cos^2(a,b)} = \tan^2(a,b) \tag{15}$$

where $\cos(a,b)$ and $\tan(a,b)$ denote the cosine and tangent of the angle between vectors a and b, respectively.

It is evident that d only depends on the angle between the two vectors a and b. The smaller the angle between the two vectors, the closer we are to the case of aggregate cointegration, which occurs when $d = 0$.

[1]We wish to thank an anonymous referee for suggesting this decomposition of the limiting distribution of S.

The definition (15) is based on the variance of the random variable S and the notion of superconsistency when cointegration is present. Alternatively, d may be derived as a measure of the degree of noncointegration directly from equation (11). This latter derivation, suggested by a referee, is reported in Appendix.

In general, the parameters a, b are unknown. A feasible version $\hat{d}$ of the statistic d can be computed using estimates $\hat{a}$, $\hat{b}$ of a, b as

$$\hat{d} = \tan^2(\hat{a}, \hat{b}).$$

We now explore the asymptotic properties of the estimator $\hat{d}$. When the aggregate relationship cointegrates, $\operatorname{var} S = 0$ and equation (13) implies that $(a'b)^2 = (a'a)\,(b'b)$. By the Schwarz inequality, the last equality holds if and only if there exist a constant C such that $b = Ca$. Since $a'b/a'a = C$, the aggregate relationship cointegrates if and only if $b = \frac{a'b}{a'a}a$. To model the system under no aggregate cointegration, we consider local departures from cointegration of the form

$$b = \frac{a'b}{a'a}a + \delta_T, \tag{16}$$

where $\delta_T' a = 0$ and $T\delta_T \to \delta$ as $T \to \infty$. In this formulation, when $\delta = 0$, the aggregate regression cointegrates while when $\delta \neq 0$, aggregate cointegration fails to occur. The following theorem characterizes the rate of convergence of $\widehat{d}$.

Theorem 2 *Let $\hat{a}$ and $\hat{b}$ be superconsistent estimators of a and b, that is $\hat{a} - a = O_p(T^{-1})$ and $\hat{b} - b = O_p(T^{-1})$. Under the local departure from cointegration (16) with $\delta = 0$ (cointegration) or $\delta \neq 0$ (no cointegration),*

$$T^2\left(\widehat{d} - d\right) = O_p(1)\,. \tag{17}$$

Proof. See Appendix. ■

The rate of convergence of $\widehat{d}$ is of order T^{-2}. When the stochastic trends z_t's are observable, consistent estimators of a and b can be obtained via OLS by regressing $\bar{x}_t$ and $\bar{y}_t$ on z_t, respectively. Such estimates are superconsistent because, irrespectively of the existence of a cointegration relationship between $\bar{x}_t$ and $\bar{y}_t$, $\bar{x}_t$ and $\bar{y}_t$ always cointegrate with z_t by assumption. When z_t's are unobservable, superconsistent estimators of a and b can be obtained by estimating OLS regressions of $\bar{x}_t$ and $\bar{y}_t$ on estimated trends $\hat{z}_t$, as long as $\hat{z}_t$'s estimate z_t's consistently. A procedure for estimating z_t has been proposed by Kao, Trapani and Urga (2005).

The following theorem gives the limiting distribution of $\widehat{d}$ when the z_ts are observable:

Theorem 3 *Let*

$$D = \frac{\|a\|^2}{(a'b)^2} \int dBW' \left(\int WW' \right)^{-1} \left(I - \frac{aa'}{\|a\|^2} \right) \left(\int WW' \right)^{-1} \int WdB$$

where $dB(r) = \frac{a'b}{a'a} dW_{\bar{u}}(r) - dW_{\bar{v}}(r)$, $W_{\bar{u}}$ *and* $W_{\bar{v}}$ *are Brownian processes associated with the partial sums of processes* $\bar{u}_t$ *and* $\bar{v}_t$ *in (6) and (7) respectively. Under the local departure from cointegration (16) with* $\delta = 0$ *(cointegration) or* $\delta \neq 0$ *(no cointegration),*

$$T^2 \widehat{d} \xrightarrow{d} \frac{\|a\|^2 \|\delta\|^2}{(a'b)^2} + D - 2\frac{\|a\|^2}{(a'b)^2} \delta' \left(\int WW' \right)^{-1} \int WdB. \qquad (18)$$

Proof. See Appendix. ■

Equation (18) shows that the limiting distribution of $T^2\widehat{d}$ is the sum of three terms: the positive constant $\|a\|^2 \|\delta\|^2 (a'b)^{-2}$, the nonzero mean random variable D and the zero-mean random variable $2\|a\|^2 (a'b)^{-2} \delta' \left(\int WW' \right)^{-1} \int WdB$. Under no cointegration, when $\delta = 0$, the statistic $\hat{d}$ has a different asymptotic distribution than under cointegration. This indicates that the statistic $\hat{d}$ would be a suitable basis to test for the null of aggregate cointegration against local alternatives of aggregate noncointegration.

4 CONCLUSIONS

In nonstationary heterogeneous panels where each unit cointegrates, the aggregate relationship does not cointegrate unless the coefficients describing micro relationships satisfy a set of conditions. In reality these conditions are in general not satisfied and aggregate cointegration does not hold. Our paper takes a closer look at the conditions for cointegration at macro level to hold when the conditions for perfect aggregation are violated. The question then is whether the macro relationship is observationally equivalent to a cointegrating equation. We propose a framework for the case of finite n and large T. Our results can be viewed as complementary to Phillips and Moon's (1999) where the analysis is carried out for the case $n \to \infty$.

We derive an asymptotic measure, d, of the distance from the case of aggregate cointegration. We also propose an estimator $\hat{d}$ which converges to d at a rate of $O_p(T^{-2})$. We prove that the departure from cointegration depends on the angle between vectors a and b which describe the heterogeneity of the response of $\bar{y}_t$ to $\bar{x}_t$.

Our paper gives support to the view that even if the conditions for perfect aggregation are violated, the aggregate regression is still useful in characterising the macro relationship.

APPENDIX

Derivation of Eq. (11) using population regression

An alternative way to derive condition (11) is from the population equations. Aggregating at the micro level the DGP for each y_{it}, namely

$$y_{it} = \beta_i x_{it} + u_{it},$$

one gets

$$\begin{aligned} \bar{y}_t &= \sum_{i=1}^{n} \beta_i x_{it} + \bar{u}_t = \sum_{i=1}^{n} c x_{it} + \sum_{i=1}^{n} (\beta_i - c)\, x_{it} + \bar{u}_t = \\ &= c\bar{x}_t + \sum_{j=1}^{k} \sum_{i=1}^{n} (\beta_i - c)\, \alpha_{ij} z_{jt} + \sum_{i=1}^{n} (\beta_i - c)\, v_{it} + \bar{u}_t, \end{aligned} \tag{19}$$

for some constant c. Since $\sum_{i=1}^{n} (\beta_i - c)\, v_{it} + \bar{u}_t$ is weakly dependent, $\bar{y}_t$ and $\bar{x}_t$ are cointegrated if we assume that the remainder of the error term in equation (19), $\sum_{j=1}^{k} \sum_{i=1}^{n} (\beta_i - c)\, \alpha_{ij} z_{jt}$, cancels out. This will occur if and only if

$$\sum_{i=1}^{n} (\beta_i - c)\, \alpha_{ij} = 0$$

for all js. This is the scalar form of equation (11).

Alternative measure of departure from cointegration

Condition (11) provides an alternative way to understand when perfect cointegration holds. Given that this occurs if and only if

$$A' (\tau - c\iota) = 0,$$

one possible measure of the degree of noncointegration could be obtained by considering the following quantity

$$\min_c \frac{\| A' (\tau - c\iota) \|}{\| A'\tau \|}.$$

The solution to this minimisation problem is given by

$$c = \frac{a'b}{a'a}$$

and the minimum value attained by the objective function is equal to

$$1-\frac{(a'b)^2}{(a'a)(b'b)}=\sin^2(a,b).$$

Hence, a possible measure of the departure from aggregate cointegration is given by $\sin^2(a,b)$. This measure is equivalent to definition (15) and has an appealing interpretation. When the β_is are homogeneous across all is, it is equal to zero by definition, i.e. aggregate cointegration always holds. This holds also under more general conditions, namely when the vectors a and b are parallel.

Proof of Proposition 1

Consider the OLS estimator:

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{t=1}^{T}\left(\sum_{i=1}^{n} y_{it}\right)\left(\sum_{i=1}^{n} x_{it}\right)}{\sum_{t=1}^{T}\left(\sum_{i=1}^{n} x_{it}\right)^2}\\ &= \frac{\sum_{t=1}^{T}\left(\sum_{i=1}^{n} \beta_i x_{it}+\sum_{i=1}^{n} u_{it}\right)\left(\sum_{i=1}^{n} x_{it}\right)}{\sum_{t=1}^{T}\left(\sum_{i=1}^{n} x_{it}\right)^2}.\end{aligned}$$

As $T\to\infty$ and $n\to\infty$ we have

$$\begin{aligned}\hat{\beta} &= \frac{\lim\limits_{T\to\infty}\sum_{t=1}^{T}\left[\lim\limits_{n\to\infty}\sum_{i=1}^{n}\beta_i\alpha_i' z_t+o_p(1)\right]\left[\lim\limits_{n\to\infty}\sum_{i=1}^{n}\alpha_i' z_t+o_p(1)\right]}{\lim\limits_{T\to\infty}\sum_{t=1}^{T}\left[\lim\limits_{n\to\infty}\sum_{i=1}^{n}\alpha_i' z_t+o_p(1)\right]^2},\\ &= \frac{\lim\limits_{T\to\infty}\sum_{t=1}^{T}\left(\lim\limits_{n\to\infty} n^{-1}\sum_{i=1}^{n}\beta_i\right)\left[\left(\lim\limits_{n\to\infty} n^{-1}\sum_{i=1}^{n}\alpha_i'\right) z_t\right]\left[\left(\lim\limits_{n\to\infty} n^{-1}\sum_{i=1}^{n}\alpha_i'\right) z_t\right]}{\lim\limits_{T\to\infty}\lim\limits_{n\to\infty} n^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\alpha_i' z_t\right)^2}+o_p(1)\\ &= \bar{\beta}\frac{\lim\limits_{T\to\infty}\sum_{t=1}^{T}\left(\bar{\alpha}' z_t\right)^2}{\lim\limits_{T\to\infty}\sum_{t=1}^{T}\left(\bar{\alpha}' z_t\right)^2},\end{aligned}$$

proving that $\hat{\beta}\xrightarrow{p}\bar{\beta}$.

As far as the pooled OLS estimator, $\hat{\beta}^{POLS}$, is concerned, we have

$$\begin{aligned}\hat{\beta}^{POLS} &= \frac{\sum_{t=1}^{T}\sum_{i=1}^{n} y_{it}x_{it}}{\sum_{t=1}^{T}\sum_{i=1}^{n} x_{it}^2}\\ &= \frac{\sum_{t=1}^{T}\sum_{i=1}^{n} x_{it}\left(\beta_i x_{it}+u_{it}\right)}{\sum_{t=1}^{T}\sum_{i=1}^{n} x_{it}^2},\end{aligned}$$

and as $T \to \infty$ and $n \to \infty$ it holds

$$\begin{aligned}
\hat{\beta}^{POLS} &= \frac{\lim\limits_{T\to\infty}\lim\limits_{n\to\infty} \sum_{t=1}^{T}\sum_{i=1}^{n} \beta_i \left(\alpha_i' z_t\right)^2}{\lim\limits_{T\to\infty}\lim\limits_{n\to\infty} \sum_{t=1}^{T}\sum_{i=1}^{n} \left(\alpha_i' z_t\right)^2} + o_p(1) \\
&= \frac{\lim\limits_{T\to\infty} \sum_{t=1}^{T} \left(\lim\limits_{n\to\infty} n^{-1} \sum_{i=1}^{n} \beta_i\right) \left[\lim\limits_{n\to\infty} n^{-1} \sum_{i=1}^{n} \left(\alpha_i' z_t\right)^2\right]}{\lim\limits_{T\to\infty} \sum_{t=1}^{T} \left[\lim\limits_{n\to\infty} n^{-1} \sum_{i=1}^{n} \left(\alpha_i' z_t\right)^2\right]} + o_p(1) \\
&= \bar{\beta} \frac{\lim\limits_{T\to\infty} \sum_{t=1}^{T} \left(\bar{\alpha}' z_t\right)^2}{\lim\limits_{T\to\infty} \sum_{t=1}^{T} \left(\bar{\alpha}' z_t\right)^2},
\end{aligned}$$

proving $\hat{\beta}^{POLS} \xrightarrow{p} \bar{\beta}$. This proves the equivalence of the probability limits of $\hat{\beta}$ and $\hat{\beta}_{POLS}$.

Proof of Theorem 2

Recall that $C = \frac{a'b}{a'a}$ and thus the orthogonal projection of vector b on vector a is $\frac{a'b}{a'a} a = Ca$. In this notation, $\delta_T = b - Ca$, $a'b = C \|a\|^2$ and $\|b\|^2 = C^2 \|a\|^2 + \|\delta_T\|^2$ so that $\|b\| = |C| \|a\| + O(T^{-1})$.

Lemma 1 *Under the local departures from aggregate cointegration,*

$$\cos(a, b) - \cos\left(\hat{a}, \hat{b}\right) = \operatorname{sgn}(C) \left(\frac{1}{2} Q_T - R_T\right) + o_p\left(T^{-2}\right)$$

where $\operatorname{sgn}(C)$ *is the sign of* C *and*

$$\begin{aligned}
Q_T &= \frac{1}{C^2 \|a\|^2} \left[C(\hat{a} - a) - \left(\hat{b} - b\right)\right]' \left(I - \frac{aa'}{\|a\|^2}\right) \left[C(\hat{a} - a) - \left(\hat{b} - b\right)\right] \\
R_T &= \frac{1}{C^2 \|a\|^2} \delta_T' \left[C(\hat{a} - a) - \left(\hat{b} - b\right)\right].
\end{aligned}$$

Proof. We have

$$\begin{aligned}
\cos(a, b) - \cos\left(\hat{a}, \hat{b}\right) &= \frac{a'b}{\|a\| \|b\|} - \frac{\hat{a}'\hat{b}}{\|\hat{a}\| \left\|\hat{b}\right\|} \\
&= \frac{\|\hat{a}\| \left\|\hat{b}\right\| a'b - \|a\| \|b\| \hat{a}'\hat{b}}{\|a\| \|b\| \|\hat{a}\| \left\|\hat{b}\right\|}. \qquad (20)
\end{aligned}$$

Superconsistency of $\hat{a}$ and $\hat{b}$ implies that $\hat{a} = a + O_p(T^{-1})$. The Taylor expansion gives us

$$\begin{aligned} \|\hat{a}\| &= \|a\| + \frac{1}{2}\frac{\|\hat{a}\|^2 - \|a\|^2}{\|a\|} - \frac{1}{8}\frac{\left(\|\hat{a}\|^2 - \|a\|^2\right)^2}{\|a\|^3} + O_p\left(T^{-3}\right) \\ &= \|a\| + \frac{a'\varepsilon_a}{\|a\|} + \frac{\varepsilon_a'\varepsilon_a}{2\|a\|} - \frac{(a'\varepsilon_a)^2}{2\|a\|^3} + O_p\left(T^{-3}\right) \end{aligned}$$

where $\varepsilon_a = \hat{a} - a$. Similarly for $\hat{b}$ we have $\hat{b} = b + O_p(T^{-1})$ and

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

where $\varepsilon_b = \hat{b} - b$. Therefore the denominator of (20) is equal to $\|a\|^2\|b\|^2 + O_p(T^{-2}) = C^2\|a\|^4 + O_p(T^{-2})$. As far as the numerator of (20) is concerned, simple algebra yields

$$\begin{aligned} & \|\hat{a}\|\left\|\hat{b}\right\| a'b - \|a\|\|b\|\hat{a}'\hat{b} \\ = & \frac{|C|}{C}\frac{\|a\|^2}{2}\left((C\varepsilon_a - \varepsilon_b)'\left(I - \frac{aa'}{\|a\|^2}\right)(C\varepsilon_a - \varepsilon_b) + O_p\left(T^{-3}\right)\right)(1 + o_p(1)) \\ & -\frac{C}{2|C|}\|a\|^2\left(2\delta_T'(C\varepsilon_a - \varepsilon_b) + O_p\left(T^{-3}\right)\right)(1 + o_p(1)). \end{aligned}$$

Since ε_a, ε_b and δ_T are $O_p(T^{-1})$, the last display is equal to

$$\operatorname{sgn}(C)\|a\|^2\left(\frac{1}{2}(C\varepsilon_a - \varepsilon_b)'\left(I - \frac{aa'}{\|a\|^2}\right)(C\varepsilon_a - \varepsilon_b) - \delta_T'(C\varepsilon_a - \varepsilon_b)\right) + o_p\left(T^{-2}\right).$$

This implies that

$$\begin{aligned} & \cos(a,b) - \cos\left(\hat{a},\hat{b}\right) \\ = & \frac{\operatorname{sgn}(C)}{2C^2\|a\|^2}\left((C\varepsilon_a - \varepsilon_b)'(I - aa')(C\varepsilon_a - \varepsilon_b) - 2\delta_T'(C\varepsilon_a - \varepsilon_b) + o_p\left(T^{-2}\right)\right) \\ & \times\left(1 + O_p\left(T^{-2}\right)\right) \\ = & \operatorname{sgn}(C)\left(\frac{1}{2}Q_T - R_T\right) + o_p\left(T^{-2}\right). \end{aligned}$$

■

Lemma 1 implies that, under the local departures from aggregate cointegration, $\cos\left(\hat{a},\hat{b}\right) = \cos(a,b)\left(1+O_p\left(T^{-2}\right)\right)$. Since

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

and $\cos(a,b) = \operatorname{sgn}(C)\left(1+O_p\left(T^{-2}\right)\right)$, we have

$$\hat{d} = \frac{\sin^2\left(\hat{a},\hat{b}\right)}{\cos^2\left(\hat{a},\hat{b}\right)} = \frac{\sin^2(a,b)+2\left(1+O_p\left(T^{-2}\right)\right)\operatorname{sgn}^2(C)\left[\frac{1}{2}Q_T-R_T+o_p\left(T^{-2}\right)\right]}{\cos^2(a,b)\left(1+O_p\left(T^{-2}\right)\right)}.$$

This implies that

$$\begin{aligned}\hat{d}-d &= \frac{\sin^2(a,b)}{\cos^2(a,b)}O_p\left(T^{-2}\right)+\left(Q_T-2R_T+o_p\left(T^{-2}\right)\right)\left(1+O_p\left(T^{-2}\right)\right)\\ &= Q_T-2R_T+o_p\left(T^{-2}\right)=O_p\left(T^{-2}\right) \qquad (21)\end{aligned}$$

because $\tan^2(a,b) = O_p\left(T^{-2}\right)$ and $Q_T-2R_T = O_p\left(T^{-2}\right)$.

Proof of Theorem 3

Recall that, by definition, the OLS estimators for a and b in equations (6) and (7) are given by

$$\hat{a} = \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1}\left(\sum_{t=1}^{T} z_t \bar{x}_t\right),$$

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

Therefore

$$\varepsilon_a = \hat{a}-a = \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1}\left(\sum_{t=1}^{T} z_t \bar{v}_t\right),$$

$$\varepsilon_b = \hat{b}-b = \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1}\left(\sum_{t=1}^{T} z_t \bar{u}_t\right).$$

Since both equations (6) and (7) are cointegration relationship, the OLS estimators $\hat{a}$ and $\hat{b}$ are superconsistent, i.e. it holds that $\varepsilon_a = O_p\left(T^{-1}\right)$ and

$\varepsilon_b = O_p(T^{-1})$. Let $W_{\bar{v}}(r)$ and $W_{\bar{u}}(r)$ be the Brownian motions associated with the partial sums of $\bar{v}_t$ and $\bar{u}_t$ respectively. Then the central limit theorem for functional spaces, implied by Assumptions (a)-(e), ensures that

$$T\varepsilon_a \xrightarrow{d} \left(\int WW'\right)^{-1} \int W dW_{\bar{v}}, \tag{22}$$

$$T\varepsilon_b \xrightarrow{d} \left(\int WW'\right)^{-1} \int W dW_{\bar{u}}. \tag{23}$$

Under the local departure from cointegration (16) with $\delta = 0$ (cointegration) or $\delta \neq 0$ (no cointegration) we have that, from equation (21)

$$\widehat{d} = d + Q_T - 2R_T + o_p\left(T^{-2}\right),$$

and since

$$d = \frac{\sin^2(a,b)}{\cos^2(a,b)} = \left[1 - \frac{(a'b)^2}{\|a\|^2\|b\|^2}\right]\left[\frac{(a'b)^2}{\|a\|^2\|b\|^2}\right]^{-1} = \left[\frac{\|\delta_T\|^2}{C^2\|a\|^2}\right]\left[1 - \frac{\|\delta_T\|^2}{C^2\|a\|^2}\right]^{-1}, \tag{24}$$

we have that

$$\lim_{T\to\infty} T^2 d = \frac{\|a\|^2\|\delta\|^2}{(a'b)^2}. \tag{25}$$

Also, as $T \to \infty$

$$\begin{aligned} T^2 Q_T &= \frac{1}{C^2\|a\|^2}\left(CT\varepsilon_a - T\varepsilon_b\right)'\left(I - \frac{aa'}{\|a\|^2}\right)\left(CT\varepsilon_a - T\varepsilon_b\right) \\ &\Rightarrow \frac{1}{C^2\|a\|^2}\left[\left(\int WW'\right)^{-1}\int W\left(CdW_{\bar{v}} - dW_{\bar{u}}\right)\right]' \\ &\quad \times\left(I - \frac{aa'}{\|a\|^2}\right)\left(\int WW'\right)^{-1}\int W\left(CdW_{\bar{v}} - dW_{\bar{u}}\right) \\ &= \frac{1}{C^2\|a\|^2}\int dBW'\left(\int WW'\right)^{-1} \\ &\quad \times\left(I - \frac{aa'}{\|a\|^2}\right)\left(\int WW'\right)^{-1}\int WdB. \end{aligned} \tag{26}$$

Last, we have that

$$\begin{aligned} T^2 R_T &= \frac{1}{C^2 \|a\|^2} (T\delta_T)' (CT\varepsilon_a - T\varepsilon_b) \\ &\Rightarrow \frac{1}{C^2 \|a\|^2} \delta' \left(\int WW' \right)^{-1} \int W (CdW_{\bar{v}} - dW_{\bar{u}}) \\ &= \frac{1}{C^2 \|a\|^2} \delta' \left(\int WW' \right)^{-1} \int WdB. \end{aligned} \tag{27}$$

Combining (25), (26) and (27) we finally get equation (18).

REFERENCES

Gonzalo, J. (1993) Cointegration and aggregation. *Ricerche Economiche* 47, 281-291.

Granger, C. W. J. (1993) Implications of seeing economic variables through an aggregation window. *Ricerche Economiche* 47, 269-279.

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.

Kao, C., Trapani, L. and Urga, G. (2005) The asymptotics for panel models with common shocks. Mimeo, Cass Business School.

Park, J.Y. and Phillips, P.C.B. (1988) Statistical inference in regressions with integrated orocesses: Part 1. *Econometric Theory* 4, 468-497.

Park, J.Y. and Phillips, P.C.B. (1989) Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* 5, 95-131.

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. and Moon, H.R. (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057-1111.

Sun, Y. (2004) Estimation of the long run relationship in nonstationary panel time series. *Econometric Theory* 20, 1227-1260.