

# Spurious Regression in Nonstationary Panels time Series with Cross-Member Cointegration

## Citation for published version (APA):

Urbain, J. R. Y. J., & Westerlund, J. (2008). *Spurious Regression in Nonstationary Panels time Series with Cross-Member Cointegration*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 044 <https://doi.org/10.26481/umamet.2008044>

## Document status and date:

Published: 01/01/2008

## DOI:

[10.26481/umamet.2008044](https://doi.org/10.26481/umamet.2008044)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Jean-Pierre Urbain, Joakim Westerlund

Spurious Regression in Non-Stationary Panel  
Time Series with Cross-Unit Cointegration

RM/08/044

JEL code: C13, C33



Maastricht research school of **E**conomics  
of **T**Echnology and **O**Rganizations

Universiteit Maastricht  
Faculty of Economics and Business Administration  
P.O. Box 616  
NL - 6200 MD Maastricht

phone : ++31 43 388 3830  
fax : ++31 43 388 4873

# SPURIOUS REGRESSION IN NON-STATIONARY PANEL TIME SERIES WITH CROSS-UNIT COINTEGRATION\*

Jean-Pierre Urbain<sup>†</sup>  
Universiteit Maastricht  
The Netherlands

Joakim Westerlund  
Lund University  
Sweden

October 9, 2008

## Abstract

This paper illustrates analytically the effects of cross-unit cointegration using as an example the conventional pooled least squares estimate in the spurious panel time series regression case. The results suggest that the usual result of asymptotic normality depends critically on the absence of cross-unit cointegration.

**JEL Classification:** C13; C33.

**Keywords:** Panel cointegration; Spurious regression; Cross-unit cointegration; Cross-section dependence.

---

\*A preliminary version of the paper was presented at the Conference on Common Feature in Maastricht, at Ente “Luigi Einaudi” in Rome, at ESEM2004 in Madrid and at the Econometric Seminar in Rotterdam. The authors would like to thank Christian Gengenbach, Stephane Gregoir, Richard Paap, Franco Peracchi and Fabio Buseti for helpful comments and suggestions. The second author would also like to thank the Maastricht Research School of Economics of Technology and Organizations for its hospitality during a visit at the Department of Quantitative Economics at the University of Maastricht, where a part of this paper was written. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number W2006-0068:1. The usual disclaimer applies.

<sup>†</sup>Corresponding author: Department of Quantitative Economics, Universiteit Maastricht, P. O. Box 616, 6200 MD Maastricht, The Netherlands. Telephone: +31 43 388 3660, Fax: +31 43 388 48 74, E-mail address: [j.urbain@ke.unimaas.nl](mailto:j.urbain@ke.unimaas.nl).

# 1 Introduction

The use of non-stationary time series techniques to infer both spurious and cointegrated panel data regressions has literally exploded during the last couple of years. The main reason for this being the more powerful and straightforward asymptotic results that can be obtained by pooling across a large cross-section of independent time series, henceforth denoted  $N$ . In fact, as Phillips and Moon (1999) show, since this independence works by effectively smoothing out the usual unit root dependency for each unit, valid inference is usually possible, even in the so difficult spurious regression case. In particular, they show how the pooled least squares estimator is asymptotically normal and  $\sqrt{N}$ -consistent for the so-called long-run average regression coefficient.

However, as argued by Banerjee *et al.* (2006), assuming independence between cross-sectional units in this way is usually not warranted, especially when considering macroeconomic or financial time series data with strong, possibly dynamic, linkages across the units, typically countries or regions. In particular, by using simulation methods, they show that the presence of cross-unit cointegration via common stochastic trends, a reasonable assumption in most applications involving this type of data, can have quite drastic effects.

The problem is that Banerjee *et al.* (2004) only consider the problem of testing for cointegration, and it is unclear to what extent their results have implications for estimation and inference. As a response to this, a number of authors, including Coakley *et al.* (2006) and Fuertes (2008), have recently begun to consider the small-sample behavior of various pooled estimators in cases when cross-sectional independence fails. One of the main findings is that cross-unit cointegration is likely to lead to substantial bias with deceptive inference as a result.

In this paper, we complement analytically these simulation studies, taking as an example the pooled least squares estimator in the simple bivariate case. Results complementary to those obtained in this paper are reported in Gengenbach *et al.* (2006), where the main focus is residual-based panel cointegration tests. Here we consider two possibilities, one is when the cointegration occurs across the whole panel, the other is when the cointegration occurs within groups that are otherwise independent of each other, which seems like a reasonable scenario when studying for example club convergence. In both cases, it is shown that the presence of cross-unit cointegration leads to statistics that diverge with the size of the cross-section,  $N$ .

On the other hand, if there is no cointegration across units, then normality is again possible, with the center of the distribution located at the long-run average regression coefficient. For the case with cointegration within groups, this naturally leads to the definition of a long-run average group coefficient.

A word on notation. The symbols  $\Rightarrow$ ,  $\rightarrow_p$  and  $:=$  will be used to signify weak convergence, convergence in probability and definitional equality, respectively. As usual,  $X_T = O_p(T^r)$  will be used to signify that  $X_T$  is at most order  $T^r$  in probability, while  $X_T = o_p(T^r)$  will be used in case  $X_T$  is of smaller order in probability than  $T^r$ . In the case of a double indexed sequence  $X_{NT}$ ,  $T, N \rightarrow \infty$  will be used to signify that the limit has been taken while passing both indices to infinity jointly. For a square matrix  $A$ ,  $\|A\|$  will denote its Euclidian norm. For simplicity, the Brownian motion  $B(s)$  defined on the interval  $s \in [0, 1]$  will be written  $B$ , with the measure of integration omitted. Integral such as  $\int_0^1 B(s)ds$  and  $\int_0^1 B(s)B(s)'ds$  will be written as  $\int B$  and  $\int BB'$ , respectively. Finally,  $[x]$  will be used to denote the integer part of  $x$ .

## 2 Model and assumptions

We begin by considering the case when the stochastic trends are common to the whole panel. The case with group specific stochastic trends is considered last. For simplicity and ease of exposition, the data generating process employed for this purpose consists of two scalar variables  $X_{it}$  and  $Y_{it}$  such that

$$\begin{pmatrix} X_{it} \\ Y_{it} \end{pmatrix} = \lambda_i F_t + \gamma_i U_{it} + E_{it}, \quad (1)$$

where  $t = 1, \dots, T$  and  $i = 1, \dots, N$  indexes the time series and cross-sectional units, respectively. Note that while  $F_t = (F_{xt}, F_{xt})'$  is common across  $i$ ,  $U_{it}$  and  $E_{it}$ , having a similar partition, are not and can be viewed as idiosyncratic components. The parameter matrix  $\lambda_i$  is written as

$$\lambda_i = (\lambda_{xi} \quad \lambda_{yi})' = \begin{pmatrix} \lambda_{1xi} & \lambda_{2xi} \\ \lambda_{1yi} & \lambda_{2yi} \end{pmatrix}$$

with an analogous partition of  $\gamma_i$ . These matrices are assumed to be independent of each other, and randomly distributed across  $i$ . This is made explicit in Assumption 1, where  $M < \infty$  denotes a generic positive real number.

**Assumption 1.** (i)  $\lambda_i$  and  $\gamma_i$  are *i.i.d.* across  $i$  with expected values  $\bar{\lambda}$  and  $\bar{\gamma}$ , respectively, (ii)  $E(\|\lambda_i\|^4) < M$  and  $E(\|\gamma_i\|^4) < M$ .

We further assume that the vector  $w_{it} := (\Delta F'_t, \Delta U'_{it}, E'_{it})'$  is a stationary linear process (see Phillips and Solo, 1992) such that

$$\begin{aligned} w_{it} &= \begin{pmatrix} C_{11}(L) & 0 & C_{13i}(L) \\ 0 & C_{22i}(L) & C_{23i}(L) \\ C_{31i}(L) & C_{32i}(L) & C_{33i}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2it} \\ \varepsilon_{3it} \end{pmatrix} \\ &= C_i(L)\varepsilon_{it}, \end{aligned}$$

where  $C_i(L) := I_6 - \sum_{j=1}^{\infty} C_{ij}L^j$  and  $L$  is the lag operator, with coefficient matrices

$$C_{ij} = \begin{pmatrix} C_{11j} & 0 & C_{13ij} \\ 0 & C_{22ij} & C_{23ij} \\ C_{31ij} & C_{32ij} & C_{33ij} \end{pmatrix}.$$

The rest of the assumptions can be summarized in the following way.

**Assumption 2.** (i)  $\varepsilon_{it} \sim i.i.d.(0, I_6)$  with  $E(\|\varepsilon_{it}\|^8) < M$ , (ii)  $\varepsilon_{it}$  is independent across both  $i$  and  $t$ , (iii)  $F_0, U_{i0}$  and  $\varepsilon_{i0}$  are  $O_p(1)$ , (iv)  $E(\Delta U_{it}\Delta U'_{it}) =: \Phi_i < M$  for all  $t$ .

**Assumption 3.**  $C_{ij}$  is random across  $i$  and satisfy Assumptions 1 and 2 of Phillips and Moon (1999).

**Assumption 4.**  $\lambda_i, \gamma_i$  and  $\varepsilon_{it}$  are mutually independent.

Assumptions 2 and 3 ensure that for all  $i$ , a functional central limit theorem holds for  $w_{it}$ . In other words, it holds that as  $T \rightarrow \infty$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor sT \rfloor} w_{it} \Rightarrow B_i,$$

where  $B_i = (B'_1, B'_{2i}, B'_{3i})'$  is a vector Brownian motion that is partitioned conformably with  $w_{it}$ . The covariance matrix of  $B_i$  is given by

$$\Omega_i := C_i(1)C_i(1)' = \begin{pmatrix} \Omega_{11} & 0 & \Omega_{13i} \\ 0 & \Omega_{22i} & \Omega_{23i} \\ \Omega'_{13i} & \Omega'_{23i} & \Omega_{33i} \end{pmatrix}.$$

Although very simple, this data generating process yields significant insight. In particular, note how (1) nests a variety of spurious and cointegration cases, both with and without cross-unit cointegration, depending on the values taken by  $\lambda_i$  and  $\gamma_i$ .<sup>1</sup> Here are the four leading examples that we will consider.

<sup>1</sup>In fact, it is not difficult to see that, under certain conditions, (1) is actually a system of two Bai and Ng (2004) factor models, one for  $X_{it}$  and one for  $Y_{it}$ .

- (a) If  $\gamma_i = 0$ , there is only the common unit roots in  $F_t$ , suggesting that both  $X_{it}$  and  $Y_{it}$  are cross-cointegrated. If in addition one of the columns of  $\lambda_i$  is zero,  $X_{it}$  and  $Y_{it}$  are not only cross-cointegrated but also cointegrated with each other. For example, suppose that  $\lambda_{2yi} = \lambda_{2xi} = 0$  so that the second column is zero, letting  $\beta_i = \frac{\lambda_{1yi}}{\lambda_{1xi}}$ , the cointegrating relationship for unit  $i$  is given by

$$\begin{pmatrix} -\beta_i & 1 \end{pmatrix} \begin{pmatrix} X_{it} \\ Y_{it} \end{pmatrix} = \begin{pmatrix} -\beta_i & 1 \end{pmatrix} E_{it},$$

while  $X_{it}$  and  $X_{jt}$  are cointegrated for all  $i \neq j$ , with cointegrating vector

$$\begin{pmatrix} -\frac{\lambda_{1xj}}{\lambda_{1xi}} & 1 \end{pmatrix}$$

with a similar results applying to the pairs  $Y_{it}$  and  $Y_{jt}$ .

- (b) On the other hand, if  $\gamma_i = 0$  but  $\lambda_i$  is diagonal, although still cross-cointegrated,  $X_{it}$  and  $Y_{it}$  are no longer cointegrated with each other for each  $i$ .
- (c) Another situation worth mentioning is when  $\lambda_i = 0$  and  $\gamma_i$  is diagonal, which is the conventional spurious regression case without cross-unit cointegration.
- (d) If  $\lambda_i = 0$  and one of the columns of  $\gamma_i$  is equal to zero, then  $X_{it}$  and  $Y_{it}$  are cointegrated with each other, but not across the different units of the panel.

Since the focus here is on spurious regression, cases (b) and (c) are of special interest. Cases (a) and (d) are included for comparison.

### 3 Results

Define

$$\hat{\beta} := \left( \sum_{i=1}^N \sum_{t=1}^T X_{it}^2 \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T X_{it} Y_{it}.$$

The objective of this paper is to show how  $\hat{\beta}$ , the conventional pooled least squares slope estimator in a regression of  $Y_{it}$  on  $X_{it}$ , behaves in this context. But before we get to our main result, which is summarized in Theorem 1, we need to introduce some notation. We begin by defining

$$M_i := \int B_i B_i' = \begin{pmatrix} M_{11} & M_{12i} & M_{13i} \\ M'_{12i} & M_{22i} & M_{23i} \\ M'_{13i} & M'_{23i} & M_{33i} \end{pmatrix},$$

whose expectation conditional upon  $\mathcal{F} := \sigma(F_0, \dots, F_T)$ , the sigma field generated by  $F_t$ , is given by  $\bar{M} := E(M_i|\mathcal{F})$ . In this notation, we also have  $\sigma^2 := \text{var}(\gamma'_{xi}M_{22i}\gamma_{yi})$  and

$$\bar{U}_{xy} := E\left(\lambda'_{xi}M_{11}\lambda_{yi} + \gamma'_{xi}\bar{M}_{22}\gamma_{yi}|\mathcal{F}\right)$$

with  $\bar{U}_{xx}$  similarly defined, but with  $\lambda_{xi}$  and  $\gamma_{xi}$  in place of  $\lambda_{yi}$  and  $\gamma_{yi}$ , respectively. Note that unless  $\lambda_{xi} = 0$  or  $\lambda_{yi} = 0$ , the presence of  $M_{11}$  makes  $\bar{U}_{xy}$  random. Moreover,  $B_2$  will be used to denote a Brownian motion with covariance matrix

$$\Gamma := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E(\lambda_{xi}\gamma'_{yi}\Phi_i\gamma_{yi}\lambda'_{xi}).$$

Using this notation, we may state the main results of this paper.

**Theorem 1.** *Under Assumptions 1 to 4, as  $N, T \rightarrow \infty$*

- (a)  $\hat{\beta} \rightarrow_p \frac{\bar{U}_{xy}}{\bar{U}_{xx}},$
- (b)  $\sqrt{N}\hat{\beta} - \sqrt{N}\frac{\bar{U}_{xy}}{\bar{U}_{xx}} \Rightarrow \frac{2}{\bar{U}_{xx}} \int B'_1 B_2 + \frac{\sigma}{\bar{U}_{xx}} \cdot N(0, 1).$

It is interesting to note how the distribution of  $\sqrt{N}\hat{\beta}$  depends critically on what is being assumed regarding the common and idiosyncratic components of  $X_{it}$  and  $X_{it}$ .

- (a) With cointegration both within and across units,

$$\sqrt{N}\hat{\beta} - \sqrt{N}\frac{\bar{\lambda}_{1x}\bar{\lambda}_{1y}}{E(\lambda_{1xi}^2)} \rightarrow 0,$$

where  $\bar{\lambda}_{1x}$  and  $\bar{\lambda}_{1y}$  are the elements of  $\bar{\lambda}$ . This shows that the asymptotic distribution of  $\sqrt{N}\hat{\beta}$  is degenerate, and that  $\hat{\beta}$  is inconsistent for  $\beta_i := \frac{\lambda_{1yi}}{\lambda_{1xi}}$ . One exception is when  $\lambda_{1yi}$  and  $\lambda_{1xi}$  are equal across  $i$ , in which case

$$\frac{\bar{\lambda}_{1x}\bar{\lambda}_{1y}}{E(\lambda_{1xi}^2)} = \frac{\lambda_{1y}}{\lambda_{1x}} = \beta.$$

In other words, as expected, consistency of  $\hat{\beta}$  requires homogeneity of the parameters across the cross-section. Moreover, since the remainder is  $O_p(T^{-1})$ , this suggests that  $\sqrt{NT}(\hat{\beta} - \beta)$  is  $O_p(1)$ , which is the conventional panel superconsistency result, see for example Phillips and Moon (1999).



- (b) If case (b) holds, so that the pooled panel regression is spurious but, with cointegration running across units, then Theorem 1 implies that

$$\sqrt{N}\widehat{\beta} - \sqrt{N} \frac{\bar{\lambda}_{1x}\bar{\lambda}_{1y}M_{1x1y}}{E(\lambda_{1xi}^2)M_{1x1x}} \rightarrow 0,$$

where  $M_{1x1y}$  and  $M_{1x1x}$  are the elements of  $M_{11}$ . Thus, since  $M_{11}$  is a random matrix, this means that the mean of  $\widehat{\beta}$  is no longer constant, as is usually the case. This case leads to a situation that is similar to the usual standard time series spurious regression case discussed in Phillips (1986).

- (c) By contrast, if case (c) holds, we are facing the standard panel spurious regression case. Theorem 1 here implies that

$$\sqrt{N}\widehat{\beta} - \sqrt{N} \frac{\bar{\gamma}_{1x}\bar{\gamma}_{2y}\bar{M}_{2x2y}}{E(\gamma_{1xi}^2)\bar{M}_{2x2x}} \Rightarrow \frac{\sigma}{E(\gamma_{1xi}^2)\bar{M}_{2x2x}} \cdot N(0, 1),$$

with  $\sigma^2 = \text{var}(\gamma_{1xi}\gamma_{2yi}M_{2x2yi})$ , where  $M_{2x2yi}$  is from  $M_{22i}$ , while  $\gamma_{1xi}$  and  $\gamma_{2yi}$  are from  $\gamma_i$ . It follows that while  $\widehat{\beta}$  converges to a non-standard random variable in (b), this is not the case here. However,  $\sqrt{N}\widehat{\beta}$  is still non-degenerate, and in fact converges to a normal variate.

- (d) The results for case (d) with cointegration within but not across units are similar to the ones obtained for (a), but now the distribution of  $\sqrt{N}\widehat{\beta}$  is no longer degenerate. In fact, simple calculations reveal that if we assume that  $\gamma_{2yi} = \gamma_{2xi} = 0$  so that the second column of  $\gamma_i$  is zero,

$$\sqrt{N}\widehat{\beta} - \sqrt{N} \frac{\bar{\gamma}_{1x}\bar{\gamma}_{1y}}{E(\gamma_{1xi}^2)} \Rightarrow \frac{\sigma}{E(\gamma_{1xi}^2)\bar{M}_{2x2x}} \cdot N(0, 1),$$

where  $\sigma^2 = \text{var}(\gamma_{1xi}\gamma_{1yi}M_{2x2xi})$ . This result is very interesting because it suggests that  $\sqrt{NT}\widehat{\beta}$  is divergent, even after centering, which stands in sharp contrast to the conventional normality result in independent panels with homogenous cointegration, see Phillips and Moon (1999). This difference originates with Assumption 1, which has an order effect on the randomness of  $\widehat{\beta}$ . The impact of assuming homogeneity of  $\gamma_{1yi}$  and  $\gamma_{1xi}$ , suggesting the presence of a homogenous cointegrating slope equal to  $\beta = \frac{\gamma_{1y}}{\gamma_{1x}}$ , is

most easily seen by noting that

$$\begin{aligned}
\frac{\sigma}{\bar{U}_{xx}} \cdot N(0, 1) &= \lim_{N \rightarrow \infty} \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^N (\gamma_{1xi} \gamma_{1yi} M_{2x2xi} - \bar{\gamma}_{1x} \bar{\gamma}_{1y} \bar{M}_{2x2x})}{\frac{1}{N} \sum_{i=1}^N \gamma_{1xi}^2 M_{2x2xi}} \\
&= \frac{\gamma_{1y}}{\gamma_{1x}} \cdot \lim_{N \rightarrow \infty} \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^N (M_{2x2xi} - \bar{M}_{2x2x})}{\frac{1}{N} \sum_{i=1}^N M_{2x2xi}} \\
&= \frac{\gamma_{1y}}{\gamma_{1x}} \cdot \lim_{N \rightarrow \infty} \sqrt{N} \left( 1 - \frac{\bar{M}_{2x2x}}{\frac{1}{N} \sum_{i=1}^N M_{2x2xi}} \right) = 0.
\end{aligned}$$

We therefore get

$$\sqrt{N} \hat{\beta} - \sqrt{N} \frac{\bar{\gamma}_{1x} \bar{\gamma}_{1y}}{E(\gamma_{1xi}^2)} = \sqrt{N} \hat{\beta} - \sqrt{N} \beta \rightarrow 0,$$

where the remainder is again  $O_p(T^{-1})$ , thus corroborating the superconsistency result of Phillips and Moon (1999).

A few more remarks are in order.

**Remark 1.** It is interesting to compare Theorem 1 with the results obtained by Phillips and Moon (1999) in the cross-sectionally independent case. In particular, note that if we are in case (c) with no common component, then

$$\hat{\beta} \rightarrow_p \frac{\bar{\gamma}_{1x} \bar{\gamma}_{2y} \bar{M}_{2x2y}}{E(\gamma_{1xi}^2) \bar{M}_{2x2x}},$$

which is the long-run average regression coefficient discussed in Phillips and Moon (1999). On the other hand, if  $\lambda_i \neq 0$ , then  $\hat{\beta}$  is no longer centered at the long-run average regression coefficient. We also see that normality breaks down unless  $\lambda_{xi} = 0$  or  $\lambda_{yi} = \gamma_{yi} = 0$ , or both, so that the influence of  $M_{11}$  and  $\int B_1' B_2$  is wiped out. In the former situation  $X_{it}$  has no common component, while in the latter  $Y_{it}$  is completely stationary.

**Remark 2.** Consider  $\frac{\sigma}{\bar{U}_{xx}} \cdot N(0, 1)$ , the second of the two terms in the asymptotic distribution of  $\sqrt{N} \hat{\beta}$ . If there is no common component, then  $\bar{U}_{xx}$  reduces to  $E(\gamma_{xi}' \bar{M}_{22} \gamma_{xi})$ , which is smaller than for the case when there is a common component. In other words, the presence of the common component will tend to reduce the variance of  $\sqrt{N} \hat{\beta}$ . But there is also  $\int B_1' B_2$ , which has a compensating, increasing, effect. The net outcome is therefore unclear. One exception is when it is only  $X_{it}$  that has a common component, in which case  $\int B_1' B_2$  remains but  $\bar{U}_{xx}$  is still equal to  $E(\gamma_{xi}' \bar{M}_{22} \gamma_{xi})$ , with an increase in variance as a result.

**Remark 3.** Although written in terms of standard Brownian motions, Theorem 1 holds even if  $B_i$  is demeaned or detrended, or both. Put in another way, the assumption that there are no deterministic components in the estimated regression is not restrictive. However, knowledge of the exact nature of  $B_i$  means that the results can be made even more precise. For example, if  $B_i$  is indeed a standard Brownian motion, then

$$\overline{M}_{22} = E(M_{22i}|\mathcal{F}) = \int E(B_{2i}B'_{2i}) = \frac{1}{2}E(\Omega_{22i}),$$

while if  $B_i$  is demeaned,  $\overline{M}_{22} = \frac{1}{6}E(\Omega_{22i})$ .

## 4 Group specific common stochastic trends

In the remainder of this section we make a brief digression to the case with group specific common unit roots. Specifically, suppose that there are  $n \subset N$  groups with  $K$  units in each, and let  $F_{jt}$  be the common unit root component of group  $j = 1, \dots, n$  containing the units  $i = 1 + (j - 1)K, \dots, jK$ . Let  $B_{1j}$  be the Brownian motion associated with  $F_{jt}$ , and assume that  $B_{1j}$  is independent across  $j$ . For ease of notation, we also assume that  $\lambda_i$  and  $\gamma_i$  are diagonal, and that their expectations do not differ across groups.

In the appendix we show that if  $n$  is fixed, as  $N, T \rightarrow \infty$

$$\hat{\beta} \rightarrow_p \frac{\overline{Q}_{xy}}{\overline{Q}_{xx}},$$

where

$$\overline{Q}_{xy} := \overline{\lambda}_{1x}\overline{\lambda}_{2y} \left( \frac{1}{n} \sum_{j=1}^n M_{1x1yj} \right) + \overline{\gamma}_{1x}\overline{\gamma}_{2y}\overline{M}_{2x2y}$$

with  $\overline{Q}_{xx}$  similarly defined, and where  $M_{1x1yj}$  is the upper right element of  $M_{11j} = \int B_{1j}B'_{1j}$ . If we further let  $\overline{M}_{1x1y}$  be the corresponding element of  $\overline{M}_{11} = E(M_{11j})$ , and define

$$W := \frac{E(\lambda_{1xi}^2)\overline{M}_{1x1x}}{E(\lambda_{1xi}^2)\overline{M}_{1x1x} + E(\gamma_{1xi}^2)\overline{M}_{2x2x}},$$

then we have the following result.

**Corollary 1.** *Under the assumptions laid out in the above, as  $K, n, T \rightarrow \infty$*

$$\hat{\beta} \rightarrow_p W \cdot \frac{\overline{\lambda}_{1x}\overline{\lambda}_{2y}\overline{M}_{1x1y}}{E(\lambda_{1xi}^2)\overline{M}_{1x1x}} + (1 - W) \cdot \frac{\overline{\gamma}_{1x}\overline{\gamma}_{2y}\overline{M}_{2x2y}}{E(\gamma_{1xi}^2)\overline{M}_{2x2x}}.$$

The above result is very interesting. Specifically, suppose that  $\lambda_{xi} = 0$ , so that there are no common unit roots in  $X_{it}$ , then  $W = 0$  and so  $\hat{\beta}$  again converges to the long-run average panel regression coefficient

$$\frac{\bar{\gamma}_{1x}\bar{\gamma}_{2y}\overline{M}_{2x2y}}{E(\gamma_{1xi}^2)\overline{M}_{2x2x}},$$

whereas if  $\gamma_{xi} = 0$ , so that  $W = 1$ , then

$$\hat{\beta} \rightarrow_p \frac{\bar{\lambda}_{1x}\bar{\lambda}_{2y}\overline{M}_{1x1y}}{E(\lambda_{1xi}^2)\overline{M}_{1x1x}},$$

which might be thought of as a long-run average group regression coefficient. Thus, if  $\gamma_{xi} = 0$ , as  $n \rightarrow \infty$ , we end up with something that is very similar to what we had earlier when there were no groups and  $\lambda_i = 0$  and in this case the groups essentially take the role of the individual units. The constant  $W \in [0, 1]$  may be interpreted as a measure of the relative importance of these groups for the variation in  $X_{it}$ . Note especially how  $W \rightarrow 1$  as  $\frac{E(\gamma_{1xi}^2)\overline{M}_{2x2x}}{E(\lambda_{1xi}^2)\overline{M}_{1x1x}} \rightarrow 0$ .

## 5 Concluding remarks

It is now well-known that the conventional assumption of independent cross-sections is very hard to maintain when conducting inference in non-stationary panels.

In this paper, we study the effects of cross-unit cointegration taking as an example the pooled least squares estimator in the spurious regression case. We consider both the case when the cointegration takes place across the panel as a whole, and when it takes place within groups. Our findings show that the presence of cross-unit cointegration can have dramatic effects for the usual asymptotic results based on cross-section independence, and that the consistency and asymptotic normality of the pooled least squares estimator may even be lost. In particular, it is shown that the presence of cross-unit cointegration leads to statistics that diverge with  $N$ . On the other hand, if the cross-cointegration is absent, then normality is again possible, with the center of the distribution located at the long-run average regression coefficient.

These results have important empirical implications. In particular, it is not difficult to see how most empirical work based on the pooled least squares estimator in non-stationary panel data with a possible common factor structure needs to be reevaluated.

Take as an example purchasing power parity (PPP), which is usually tested by checking whether the long-run elasticity of the nominal exchange rate with respect to the relative price

level is equal to one, see Coakley *et al.* (2005) and Jacobson *et al.* (2008) for two recent examples.

The problem is that since all variables are typically measured relative to the United States, this means that the common factors are there by construction. Moreover, since both exchange rates and relative prices are generally believed to possess unit roots, one cannot rule out the possibility that these common factors are non-stationary. Of course, this does not mean that all studies of PPP are wrong in their conclusions. However, it does mean that one should take care in interpreting empirical results of this kind, as the standard asymptotic results may not hold. Of course, the critique same applies to most, if not all, empirical macro economic and financial problems.

But the implications of our results are not confined to applications, but extend also to theoretical work. For example, it is not difficult to see how tests for the absence of cointegration in presumably independent panels based on  $\hat{\beta}$ , such as those of Kao (1999), become invalid if the units are in fact cointegrated with each other, see for example Gengenbach *et al.* (2006). It is also not difficult to see how bootstrap tests of the long-run average regression coefficient, such as the one proposed by Fuertes (2008), can become misleading if the cross-sectional dependence is in the form of non-stationary common factors.

## References

- Bai, J., and S. Ng (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica* **72**, 1127–1177.
- Banerjee, A., M. Marcellino and C. Osbat (2004). Some Cautions on the use of Panel Methods for Integrated Series of Macroeconomic Data. *Econometrics Journal* **7**, 322–340.
- Billingsley, P. (1986). *Convergence of Probability Measures*. Wiley, New York.
- Coakley, J., R. P. Flood, A.-M. Fuertes and M. P. Taylor (2005). Purchasing Power Parity and the Theory of General Relativity: The First Tests. *Journal of International Money and Finance* **24**, 293–316.
- Coakley, J., A.-M. Fuertes and R. Smith (2006). Unobserved Heterogeneity in Panel Time Series Models. *Computational Statistics and Data Analysis* **50**, 2361–2380.
- Fuertes, A.-M. (2008). Sieve Bootstrap  $t$ -Tests on Long-Run Average Parameters. *Computational Statistics and Data Analysis* **52**, 3354–3370.
- Gengenbach, C., F. C. Palm and J.-P. Urbain (2006). Panel Cointegration Testing in the Presence of Common Factors. *Oxford Bulletin of Economics and Statistics* **68**, 683–719.
- Jacobson, T., J. Lyhagen, R. Larsson and M. Nessén (2008). Inflation, Exchange Rates and PPP in a Multivariate Panel Cointegration Model. *Econometrics Journal* **11**, 58–79
- Kao, C. (1999). Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics* **90**, 1–44.
- Phillips, P. C. B. (1986). Understanding Spurious Regressions in Econometrics. *Journal of Econometrics* **33**, 311–340.
- Phillips, P., and V. Solo (1992). Asymptotics for Linear Processes. *Annals of Statistics* **20**, 971–1001.
- Phillips, P. C. B., and H. R. Moon (1999). Linear Regression Limit Theory of Nonstationary Panel Data. *Econometrica* **67**, 1057–1111.

## Mathematical appendix

In this appendix, Theorem 1 and Corollary 1 are proven. Only necessary details are provided. Unreported results can be obtained from the corresponding author upon request.

### Proof of Theorem 1.

Consider (a). The least squares estimator of  $\beta$  can be written as

$$\sqrt{N}\hat{\beta} = \left( \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T X_{it}^2 \right)^{-1} \frac{1}{\sqrt{NT^2}} \sum_{i=1}^N \sum_{t=1}^T X_{it} Y_{it}. \quad (\text{A1})$$

Consider the denominator of  $\sqrt{N}\hat{\beta}$ . By first using the Beveridge–Nelson decomposition of  $C_{11}(L)$  as  $C_{11}(L) = C_{11}(1) + C_{11}^*(L)(1 - L)$ , and then taking the limit as  $T \rightarrow \infty$ , we get

$$\begin{aligned} \frac{1}{\sqrt{T}} F_t &= C_{11}(1) \frac{1}{\sqrt{T}} \sum_{s=1}^t \varepsilon_s + \frac{1}{\sqrt{T}} (C_{11}^*(L)(\varepsilon_t - \varepsilon_0) + F_0) \\ &= C_{11}(1) \frac{1}{\sqrt{T}} \sum_{s=1}^t \varepsilon_s + O_p(T^{-1/2}) \Rightarrow B_1, \end{aligned}$$

and similarly,  $\frac{1}{\sqrt{T}} U_{it} \Rightarrow B_{2i}$ . Thus, since  $E_{xit}$  is  $O_p(1)$  by Assumption 2 it follows that as  $T \rightarrow \infty$

$$\frac{1}{\sqrt{T}} X_{it} = \lambda'_{xi} \left( \frac{1}{\sqrt{T}} F_t \right) + \gamma'_{xi} \left( \frac{1}{\sqrt{T}} U_{it} \right) + O_p(T^{-1/2}) \Rightarrow \lambda'_{xi} B_1 + \gamma'_{xi} B_{2i} =: X_i,$$

which, together with the continuous mapping theorem, gives

$$\frac{1}{T^2} \sum_{t=1}^T X_{it}^2 \Rightarrow \int X_i^2.$$

Now, conditional on  $\mathcal{F}$ , the integral  $\int X_i^2$  is independent across  $i$ , which gives us the following sequential limit as  $T \rightarrow \infty$  and then  $N \rightarrow \infty$

$$\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T X_{it}^2 \rightarrow_p E \left( \int X_i^2 | \mathcal{F} \right). \quad (\text{A2})$$

According to Corollary 1 of Phillips and Moon (1999), since the scaling of  $\frac{1}{T^2} \sum_{t=1}^T X_{it}^2$  is just unity, if we can show that this quantity is uniformly integrable in  $T$ , then (A2) is not only a sequential but also a joint limit as  $N, T \rightarrow \infty$ . But since  $\frac{1}{T^2} \sum_{t=1}^T X_{it}^2$  converges to  $\int X_i^2$ , we have from Theorem 5.4 of Billingsley (1968) that uniform integrability of  $\frac{1}{T^2} \sum_{t=1}^T X_{it}^2$  is equivalent to requiring that

$$E \left( \frac{1}{T^2} \sum_{t=1}^T X_{it}^2 | \mathcal{F} \right) \rightarrow E \left( \int X_i^2 | \mathcal{F} \right),$$

which holds since  $X_{it}$  is a scalar so (A2) is indeed a joint limit as  $N, T \rightarrow \infty$ , see Appendix C of Phillips and Moon (1999).

By definition,

$$M_i = \int \begin{pmatrix} B_1 B'_1 & B_1 B'_{2i} & B_1 B'_{3i} \\ B_{2i} B'_1 & B_{2i} B'_{2i} & B_{2i} B'_{3i} \\ B_{3i} B'_1 & B_{3i} B'_{2i} & B_{3i} B'_{3i} \end{pmatrix}.$$

Consider  $M_{12i}$ . Since  $B_{2i}$  is mean zero, we get  $E(M_{12i}|\mathcal{F}) = \int B_1 E(B'_{2i}|\mathcal{F}) = 0$ . But the same goes for all off diagonal elements of  $M_i$  involving  $B_1$ . Note also that  $E(M_{11}|\mathcal{F}) = M_{11}$ . It follows that

$$\bar{M} = \begin{pmatrix} M_{11} & 0 & 0 \\ 0 & \bar{M}_{22} & \bar{M}_{23} \\ 0 & \bar{M}_{32} & \bar{M}_{33} \end{pmatrix}.$$

Now, consider

$$\begin{aligned} E\left(\int X_i^2|\mathcal{F}\right) &= E\left(\lambda'_{xi} M_{11} \lambda_{xi} + \lambda'_{xi} M_{12i} \gamma_{xi} + \gamma'_{xi} M'_{12i} \lambda_{xi} + \gamma'_{xi} M_{22i} \gamma_{xi}|\mathcal{F}\right) \\ &= E\left(\lambda'_{xi} M_{11} \lambda_{xi} + \gamma'_{xi} \bar{M}_{22} \gamma_{xi}|\mathcal{F}\right) = \bar{U}_{xx}, \end{aligned} \quad (\text{A3})$$

where the second equality follows from the fact that  $\bar{M}_{12} = \bar{M}_{21} = 0$ , while

$$E(\gamma'_{xi} M_{22i} \gamma_{xi}|\mathcal{F}) = E\left[\gamma'_{xi} E\left(\int B_{2i} B'_{2i}|\mathcal{F}\right) \gamma_{xi}|\mathcal{F}\right] = E(\gamma'_{xi} \bar{M}_{22} \gamma_{xi}|\mathcal{F}).$$

By combining (A2) and (A3) it follows that

$$\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T X_{it}^2 \rightarrow_p \bar{U}_{xx}. \quad (\text{A4})$$

The numerator can be written as

$$\begin{aligned} \frac{1}{\sqrt{NT^2}} \sum_{i=1}^N \sum_{t=1}^T X_{it} Y_{it} &= \frac{1}{\sqrt{NT^2}} \sum_{i=1}^N \sum_{t=1}^T \left( \lambda'_{xi} F_t F'_t \lambda_{yi} + \lambda_{xi} F_t U'_{it} \gamma_{yi} + \gamma'_{xi} U_{it} F'_t \lambda_{yi} \right. \\ &\quad \left. + \gamma'_{xi} U_{it} U'_{it} \gamma_{yi} \right) + O_p(T^{-1}) \\ &= I + II + III + IV + O_p(T^{-1}). \end{aligned} \quad (\text{A5})$$

Part *I* is the easiest. In fact, from the above we have that as  $N, T \rightarrow \infty$

$$\frac{1}{\sqrt{N}} I = \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \lambda'_{xi} F_t F'_t \lambda_{yi} \rightarrow_p E(\lambda'_{xi} M_{11} \lambda_{yi}|\mathcal{F}). \quad (\text{A6})$$



Part *II* can be written as

$$\begin{aligned} II &= \frac{1}{\sqrt{NT^2}} \sum_{i=1}^N \sum_{t=1}^T \lambda'_{xi} F_t U'_{it} \gamma_{yi} = \frac{1}{\sqrt{NT^2}} \sum_{t=1}^T F'_t \sum_{s=1}^t \sum_{i=1}^N \lambda_{xi} \Delta U'_{is} \gamma_{yi} \\ &= \frac{1}{T^2} \sum_{t=1}^T F'_t \sum_{s=1}^t \phi_s, \end{aligned}$$

where

$$\phi_s := \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_{xi} \Delta U'_{is} \gamma_{yi} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_{is}.$$

The random variable  $\phi_{is}$  has mean zero and is independently distributed across  $i$ . Hence, since  $\frac{1}{N} \sum_{i=1}^N \phi_{is}^2 = O_p(1)$ , according to the Lindeberg central limit theorem given in Billingsley (1986, Theorem 7.2), if we can show that for all  $\varepsilon > 0$ ,

$$\sum_{i=1}^N E\left((N^{-1/2} \phi_{is})^2 1(|N^{-1/2} \phi_{is}| > \varepsilon)\right) \rightarrow_p 0$$

as  $N \rightarrow \infty$ , then  $\phi_s \Rightarrow N(0, \Gamma)$  for each  $s$ , where

$$\Gamma = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \text{var}(\phi_{is}) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E(\lambda_{xi} \gamma'_{yi} \Phi_i \gamma_{yi} \lambda'_{xi}).$$

But this condition holds by the same arguments used for verifying (A9) below. Moreover, by letting  $T \rightarrow \infty$ , we get  $\frac{1}{\sqrt{T}} \sum_{s=1}^t \phi_s \Rightarrow B_2$ , where  $B_2$  is a Brownian motion with covariance matrix  $\Gamma$ . It follows that as  $N, T \rightarrow \infty$

$$II = \frac{1}{T^2} \sum_{t=1}^T F'_t \sum_{s=1}^t \phi_s \Rightarrow \int B'_1 B_2, \quad (\text{A7})$$

which is also true for *III*.

Next, consider *IV*. We know from before that as  $N, T \rightarrow \infty$

$$\frac{1}{\sqrt{N}} IV = \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \gamma'_{xi} U_{it} U'_{it} \gamma_{yi} \rightarrow_p E(\gamma'_{xi} \bar{M}_{22} \gamma_{yi}). \quad (\text{A8})$$

By inserting (A6) to (A8) into (A5),

$$\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T X_{it} Y_{it} \rightarrow_p E\left(\lambda'_{xi} M_{11} \lambda_{yi} + \gamma'_{xi} \bar{M}_{22} \gamma_{yi} | \mathcal{F}\right) = \bar{U}_{xy},$$

which in turn can be inserted together with (A4) into (A1) to obtain

$$\hat{\beta} \rightarrow_p \frac{\bar{U}_{xy}}{\bar{U}_{xx}}.$$

This establishes (a).

Next, consider (b). Given the result in (A8) and the fact that  $\gamma'_{xi}U_{it}U'_{it}\gamma_{yi}$  is independent across  $i$ , we would expect a central limit theorem to apply to  $IV$  after centering by  $\sqrt{N}E(\gamma'_{xi}\bar{M}_{22}\gamma_{yi})$ . We now show that this is indeed the case. Let us therefore consider

$$G_i := \gamma'_{xi}M_{22i}\gamma_{yi} - E(\gamma'_{xi}\bar{M}_{22}\gamma_{yi}).$$

The above Lindeberg condition can be written in terms of  $G_i$  as

$$\sum_{i=1}^N E\left((N^{-1/2}G_i)^2 1(|N^{-1/2}G_i| > \varepsilon)\right) = \frac{1}{N} \sum_{i=1}^N E\left(G_i^2 1(|G_i| > \sqrt{N}\varepsilon)\right) \rightarrow_p 0. \quad (\text{A9})$$

In order to verify this we make use of the Cauchy–Schwarz inequality, which yields

$$E\left(G_i^2 1(|G_i| > \sqrt{N}\varepsilon)\right) \leq \sqrt{E(G_i^4)E(1(|G_i| > \sqrt{N}\varepsilon))},$$

and by further application of the Markov inequality,

$$E(1(|G_i| > \sqrt{N}\varepsilon)) \leq \frac{1}{N\varepsilon^2}E(G_i^2).$$

Thus,

$$\frac{1}{N} \sum_{i=1}^N E\left(G_i^2 1(|G_i| > \sqrt{N}\varepsilon)\right) \leq \frac{1}{\sqrt{N}\varepsilon} \frac{1}{N} \sum_{i=1}^N \sqrt{E(G_i^4)E(G_i^2)},$$

which is  $O_p(N^{-1/2})$ , provided that  $E(G_i^4)$  and  $E(G_i^2)$  exist. Thus, since (A9) holds, we have that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N G_i \Rightarrow N(0, \sigma^2)$$

as  $N \rightarrow \infty$ , where  $\sigma^2 = \text{var}(G_i)$ . But we also have that as  $T \rightarrow \infty$

$$G_{iT} = \frac{1}{T^2} \sum_{t=1}^T \gamma'_{xi}U_{it}U'_{it}\gamma_{yi} - E(\gamma'_{xi}\bar{M}_{22}\gamma_{yi}) \Rightarrow G_i,$$

suggesting the following sequential limit as  $N \rightarrow \infty$  and then  $T \rightarrow \infty$ ,

$$IV - \sqrt{N}E(\gamma'_{xi}\bar{M}_{22}\gamma_{yi}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N G_{iT} \Rightarrow N(0, \sigma^2). \quad (\text{A10})$$

We now show that this result continues to hold in the joint limit as  $N, T \rightarrow \infty$ , which is accomplished by verifying conditions (i) to (iv) in Theorem 3 of Phillips and Moon (1999).

Conditions (i), (ii) and (iv) are obviously satisfied. Thus, for this theorem to apply we only need to verify (iii), which requires that  $G_{iT}^2$  is uniformly integrable in  $T$ . Towards this end, note that by the continuous mapping theorem,  $G_{iT}^2 \Rightarrow G_i^2$  as  $T \rightarrow \infty$ , which together with  $E(G_{iT}^2) \rightarrow E(G_i^2)$  shows that  $G_{iT}^2$  is uniformly integrable in  $T$ .

By adding these results, and then taking the limit as  $N, T \rightarrow \infty$ , we obtain

$$\frac{1}{\sqrt{NT^2}} \sum_{i=1}^N \sum_{t=1}^T X_{it} Y_{it} - \sqrt{N} \bar{U}_{xy} \Rightarrow 2 \int B_1' B_2 + N(0, \sigma^2),$$

which, together with (A1) and (A4), in turn implies that

$$\sqrt{N} \hat{\beta} - \sqrt{N} \frac{\bar{U}_{xy}}{\bar{U}_{xx}} \Rightarrow \frac{2}{\bar{U}_{xx}} \int B_1' B_2 + \frac{\sigma}{\bar{U}_{xx}} \cdot N(0, 1).$$

This completes the proof of (b). ■

### Proof of Corollary 1.

Corollary 1 is easily established by a simple manipulation of the proof of Theorem 1 (a). In so doing it is useful to define  $\mathcal{F}_j := \sigma(F_{j0}, \dots, F_{jT})$ . By the same arguments used in the proof of Theorem 1 (a),

$$\frac{1}{\sqrt{T}} X_{it} = \lambda_{1xi} \left( \frac{1}{\sqrt{T}} F_{xjt} \right) + \gamma_{1xi} \left( \frac{1}{\sqrt{T}} U_{xit} \right) + O_p(T^{-1/2}) \Rightarrow \lambda_{1xi} B_{1xj} + \gamma_{1xi} B_{2xi} = X_i,$$

as  $T \rightarrow \infty$ , where the index  $j$  has been suppressed in  $X_i$ , as  $i$  runs over  $j$ . This implies that

$$\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T X_{it}^2 \Rightarrow \frac{1}{N} \sum_{i=1}^N \int X_i^2 = \frac{1}{nK} \sum_{j=1}^n \sum_{i=1+(j-1)K}^{jK} \int X_i^2. \quad (\text{A11})$$

But we know from before that if  $K, T \rightarrow \infty$

$$\begin{aligned} \frac{1}{K} \sum_{i=1+(j-1)K}^{jK} \int X_i^2 &\rightarrow_p E \left( \int X_i^2 | \mathcal{F}_j \right) = E \left( \lambda_{1xi}^2 M_{1x1xj} + \gamma_{1xi}^2 \bar{M}_{2x2x} | \mathcal{F}_j \right) \\ &= E(\lambda_{1xi}^2) M_{1x1xj} + E(\gamma_{1xi}^2) \bar{M}_{2x2x} = \bar{U}_{xxj} \end{aligned}$$

for each  $j$ , and where the  $i$  index in  $\lambda_{xi}$  and  $\gamma_{xi}$  can be ignored since that variation is eliminated when taking expectations. Thus, taking the limit as  $n \rightarrow \infty$ ,

$$\begin{aligned} \bar{Q}_{xx} &= \frac{1}{n} \sum_{j=1}^n \bar{U}_{xxj} = E(\lambda_{1xi}^2) \frac{1}{n} \sum_{j=1}^n M_{1x1xj} + E(\gamma_{1xi}^2) \bar{M}_{2x2x} \\ &\rightarrow_p E(\lambda_{1xi}^2) \bar{M}_{1x1x} + E(\gamma_{1xi}^2) \bar{M}_{2x2x} =: \bar{P}_{xx}, \end{aligned} \quad (\text{A12})$$

and by similar arguments,

$$\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T X_{it} Y_{it} \rightarrow_p \bar{\lambda}_{1x} \bar{\lambda}_{2y} \bar{M}_{1x1y} + \bar{\gamma}_{1x} \bar{\gamma}_{2y} \bar{M}_{2x2y} =: \bar{P}_{xy}. \quad (\text{A13})$$

Adding (A11) to (A12) yields

$$\begin{aligned} \hat{\beta} &\rightarrow_p \frac{\bar{P}_{xy}}{\bar{P}_{xx}} = \frac{E(\lambda_{1xi}^2) \bar{M}_{1x1x}}{\bar{P}_{xx}} \cdot \frac{\bar{\lambda}_{1x} \bar{\lambda}_{2y} \bar{M}_{1x1y}}{E(\lambda_{1xi}^2) \bar{M}_{1x1x}} + \frac{E(\gamma_{1xi}^2) \bar{M}_{2x2x}}{\bar{P}_{xx}} \cdot \frac{\bar{\gamma}_{1x} \bar{\gamma}_{2y} \bar{M}_{2x2y}}{E(\gamma_{1xi}^2) \bar{M}_{2x2x}} \\ &= W \cdot \frac{\bar{\lambda}_{1x} \bar{\lambda}_{2y} \bar{M}_{1x1y}}{E(\lambda_{1xi}^2) \bar{M}_{1x1x}} + (1 - W) \cdot \frac{\bar{\gamma}_{1x} \bar{\gamma}_{2y} \bar{M}_{2x2y}}{E(\gamma_{1xi}^2) \bar{M}_{2x2x}}, \end{aligned}$$

which establishes the proof. ■