Representation and statistical analysis of weakly linearly exchangeable dynamic panels

Stéphane Gregoir
CREST - INSEE *

March 8, 2005

Abstract

In the recent past, a lot of attention has been devoted to the analysis of integratedness in a large set of univariate time series. Bai and Ng (2001), Chang and Song (2002) or Moon and Perron (2001) among others have recently improved this approach by considering the situation in which there exist sources of cross-correlation between the individual times series captured by common factors. If these strategies are well adapted to the description of a large set a heterogeneous time series one wants to summarize through some key components, it is difficult to draw on them to deal with the modeling of a large set of multivariate time series. We develop a simple framework to deal with a large number of multivariate homogeneous time series with a possible cross-section dependence. We propose to rely on the idea that there does not exist any relevant ordering of the (multivariate) time series in a panel. We limit our attention to weak linear representation and make the most of this property in this set-up. This correspond to a restricted form of the property of exchangeability introduced by de Finetti (1937). We derive the Wold-type representation of such a panel and from this Wold representation, derive the forms of autoregressive models it induces. Focusing on a particular approximation, we then turn to the problem of testing for integratedness or stationarity and then to that of the estimation and analysis of cointegrated variables for such panels.

1 Introduction

Testing for time series stationarity or integratedness is an important step of their modeling. In the recent past, a lot of attention has been devoted to the analysis of this question when dealing with a large set of univariate time series. When one works with a large number of homogeneous independent time series, this may naturally improve the quality of the decision the statistician takes. When a large number of heterogeneous times series is under study, the simultaneous use of a large number of time series can be motivated by the fact that using additional information in the usual unit root test procedure can improve their properties in term of power that are known to be weak. This is relevant when one assumes that there exist some links between the variables. This was for instance the approach proposed by Hansen (1994) who propose to introduce covariates in the unit root test regressions. Notwithstanding, in the initial works on heterogeneous panels, attention was essentially focused on sets of independent times series. But more recently, works by

*First version April 15,2003. Address for correspondance: CREST Timbre J301, 15 Boulevard Gabriel Péri 92245 Malakoff Cedex France, gregoir@ensae.fr. I thank Jan Magnus, Jean-Pierre Urbain and participants to the conference “Common Features in Maastricht” (December 2003) and ESEM2004 for helpful comments.

Bai and Ng (2001), Chang and Song (2002) or Moon and Perron (2002) among others have considered the situation in which there exist sources of cross-correlation between the individual times series eventually heterogeneous. Bai and Ng (2001) and Moon and Perron (2002) model the cross-sectional dependence by common factors that can be integrated or covariance stationary processes and whose influence on each time series is idiosyncratic. This seems an appealing way to describe a large set of heterogeneous data. This is needed for instance when one wants to analyze the short-term changes in a large set of various economic indexes. This way of modeling the dependence between numerous time series corresponds to the introduction of a large number of cointegration relationships between the components of the vector obtained in stacking up all the univariate time series. This alternative modeling strategy is followed by Chang and Song (2002). In both approaches, the panel structure is such that covariance between a particular couple of time series is idiosyncratic. In practice, interest in these relationships is not emphasized and their analysis may be eventually difficult. The panel is not really structured and the framework is more related to that of a very large multivariate time series.

If these works with factors are well adapted to the description of a large set a heterogenous time series we want to summarize through some key components, it is difficult to draw on them to model a large set of multivariate time series. For instance, it is difficult to construct a VECM representation in this framework if there exist some cointegration relationship between different components of the multivariate process under study. This is nevertheless what applied economists would like to have at their disposal when they want to analyze with a panel some particular issues such as sources of growth or employer-employee dynamic relationships. They can resort to the classic dynamic panel procedures but at the cost of the absence of a cross-sectional autocorrelated dependence. It seems however of interest to be able to capture such a spatial link. Banerjee, Marcellino and Osbat (2001) question the appropriateness of panel approaches without this feature.

We are interested in developing such a framework, i.e. that enables us to deal with a large number of multivariate time series with a possible cross-section dependence. We limit our attention to weak linear representations. This is possible in considering simple and natural constraints on the covariance structure of the time series in the panel. We propose to rely on the idea that there does not exist any relevant ordering of the (multivariate) time series of a panel and make the most of this property in a linear framework. This corresponds to the fact that covariance structure between two different time series must not depend on the selected couple. This amounts to a simple form of the exchangeability property introduced by de Finetti (1937). This idea was already implicitly used in the classical panel literature (Maddala (1992)) but mainly in the spectral analysis of the error terms. In an Hilbertian framework, we derive the Wold-type representation of a such a panel. Starting from this Wold representation, we can derive the general form of the VAR models/VECM that are to be satisfied by the panel. We then focus on the situation when N is not very large and derive in these circumstances the VECM type approximation that is satisfied by the data. We next turn to the problem of testing for integratedness or stationarity for such a panel and then to that of the estimation and analysis of cointegrated variables.

We limit our attention to homogeneous time series. Keeping our weak exchangeability property, it is possible to consider heterogenous panels with stochastic coefficients whose probability structure is constrained to ensure the invariance of the covariance matrices for any couple. This more elaborate parameterization will be the subject of further researches.

2 Definition and properties

Exchangeability (or interchangeability) of a sequence of random variables was introduced by de Finetti (1937) to weaken the assumption that random variables are independent and identically distributed that is in many situations unreasonable. A sequence of random variables is said exchangeable if the joint distribution of every finite subset of k random variables depends only upon the integer k and not upon the components of the particular subset. Numerous results about their representation and their convergence properties have been proved as well for sequences as for arrays of exchangeable variables (see *inter alios* Aldous (1980), Chow and Teicher (1997), Kallenberg (1991)). In particular Aldous (1980) introduced powerful representations in terms of probability of row and/or column exchangeable arrays. We do not follow explicitly this way as we propose to focus only on weak linear representation of exchangeable time series. Our result is therefore related to Aldous (1980)'s ones but derived in simpler framework.

We consider a set of m-dimensional stochastic processes $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ doubly indexed where i indexes units (that can be individuals, firms, regions,...) and t time periods. We define a weak linear property of exchangeability by the following characterization.

Definition 1 *A panel of m-dimensional stochastic processes $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ is said weakly linearly exchangeable (WLE) if*

> (i) $\forall i \in \mathbb{N}, (y_{i,t})_{t\in\mathbb{Z}}$ *is a purely non deterministic (regular) covariance stationary process,*
> (ii) $\forall (i,j) \in \mathbb{N}^2, \forall h \in \mathbb{Z}, cov\,(y_{i,t}, y_{j,t+h}) = \Gamma(h)\,\delta_{i,j} + \Xi(h)\,(1-\delta_{i,j})$

(*where $\delta_{i,j}$=1 if i=j and $\delta_{i,j}$=0 if $i \neq j$*) .

This means that the covariance structure is invariant by translation with respect to the time index and invariant by permutation with respect to the unit index (the same covariance function for any couple (i,j) in $\mathbb{N}^2$ with $i \neq j$). Two individual time series cannot be distinguished from their autocovariance function and their cross-covariance. This can be read as an anonymity properties from a stochastic point of view, which may appear very strict but is natural as soon as a modeler thinks that once taken into account some information characteristic to each individual, (s)he is dealing with a large number of homogeneous data. Indeed, on the one hand, the above definition can be extended by allowing for the presence of heterogenous deterministic terms $\mu_i(t)$ such that $(y_{i,t} - \mu_i(t))_{i\in\mathbb{N},t\in\mathbb{Z}}$ is a WLE panel. On the other hand, we can consider panel data whose data generating process would involve exogenous $(x_{i,t})_{i,t}$ variables with heterogeneous parameters, but such that (weakly) conditionally on these variables the remainder satisfies a weak linear exchangeability property. In short, this exchangeability property is adapted for panel of homogeneous multivariate variables that cannot be ordered and whose second order stochastic properties cannot be distinguished once taken into account some exogenous features that are characteristic of each unit. When the ordering of the units matters and some time interaction between the units is expected, modelers may introduce the sources of these interactions on the basis of descriptive (distance between the units may be introduced for instance) or structural information.

We are interested in deriving the basic properties of a WLE panel in relying only on the second order properties. We propose first to establish a representation very similar to the Wold representation of a stationary time series. It will be the keystone of almost all the results. We first introduce some notations. We denote H_y the Hilbert space generated by all the random variables that compose the panel and H_{y_i} the Hilbert space generated the stochastic process $(y_{i,t})_{t\in\mathbb{Z}}$. Similarly, we denote $H_{y,t}$ and $H_{y_i,t}$ the subspaces generated by the random variables that compose respectively the panel and the stochastic process $(y_{i,t})_{t\in\mathbb{Z}}$ up to the date t. At last, p_G stands for the orthogonal projector on the Hilbert space G.

Lemma 2 *If $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ is WLE, there exists $(z_t)_{t\in\mathbb{Z}}$ a covariance stationary process in H_y such that $\forall t\in\mathbb{Z}$*

$$\begin{cases} (i) & \forall h\in\mathbb{Z},\quad cov\,(z_t, z_{t+h}) = \Xi\,(h)\,, \\ (ii) & \forall\,(h,i)\in\mathbb{Z}\times\mathbb{N},\quad cov\,(z_t, y_{i,t+h}) = \Xi\,(h)\,, \\ (iii) & \forall\,(h,i,j)\in\mathbb{Z}\times\mathbb{N}^2,\quad cov\,(y_{i,t}-z_t, y_{j,t+h}-z_{t+h}) = \delta_{i,j}\,(\Gamma\,(h)-\Xi\,(h)) \end{cases}$$

From Lemma 2 (i) and (ii), we deduce that for every unit i

$$\forall\,(t,h)\in\mathbb{Z}^2,\qquad cov\,(z_t, y_{i,t+h}-z_{t+h}) = 0$$

which means that $H_z\perp H_{y_i-z}$. Moreover, we have

$$y_{i,t} = y_{i,t} - z_t + z_t$$

i.e. $H_{y_i} = H_z \oplus H_{y_i-z}$. It follows that

$$\begin{aligned} H_y &= \oplus_{i\in\mathbb{N}} H_{y_i} \\ &= H_z \oplus (\oplus_{i\in\mathbb{N}} H_{y_i-z}) \end{aligned}$$

in which all the components are orthogonal from (iii) in Lemma 2 and the above consequence of the same Lemma ($H_z\perp H_{y_i-z}$). This holds also for the subspaces constructed with the random variables up to date t, so that

$$\begin{aligned} p_{H_{y,t-1}} y_{i,t} &= p_{H_{y,t-1}}\,(y_{i,t}-z_t+z_t) \\ &= p_{H_{y_i-z,t-1}}\,(y_{i,t}-z_t) + p_{H_{z,t}}\,(z_t) \end{aligned}$$

and $\left(y_{i,t}-z_t-p_{H_{y_i,t-1-z,t-1}}\,(y_{i,t}-z_t)\right)_{t\in\mathbb{Z}}$ is the innovation process of $(y_{i,t}-z_t)_{t\in\mathbb{Z}}$ and $\left(z_t - p_{H_{z,t-1}}\,(z_t)\right)_{t\in\mathbb{Z}}$ the innovation process of $(z_t)_{t\in\mathbb{Z}}$. By assumption for every unit i,

$$\begin{aligned} \cap_t H_{y_i,t} &= \{0\} \\ \cap_t H_{z,t}\oplus H_{y_i-z,t} &= \{0\} \\ \cap_t H_{z,t}\oplus \cap_t H_{y_i-z,t} &= \{0\} \end{aligned}$$

where the last equation comes from the orthogonality property and therefore $\cap_t H_{z,t} = \{0\}$ and $\cap_t H_{y_i-z,t} = \{0\}$. At last, the covariance matrix function of any time series $(y_{i,t}-z_t)_{t\in\mathbb{Z}}$ is the same one: $(\Gamma\,(h)-\Xi\,(h))_{h\in\mathbb{Z}}$, a Wold representation of $(y_{i,t})_{t\in\mathbb{Z}}$ is therefore given by :

$$y_{i,t} = \sum_{k=0}^{+\infty} C_k\varepsilon_{t-k} + \sum_{k=0}^{+\infty} D_k\eta_{i,t-k}$$

where

$$\begin{array}{ll} C_0 = I_p & D_0 = I_p \\ \sum_{k=0}^{+\infty}\sqrt[2]{Tr\left(C_kC_k'\right)} < +\infty & \sum_{k=0}^{+\infty}\sqrt[2]{Tr\left(D_kD_k'\right)} < +\infty \\ \varepsilon_t \in H_{z,t}\cap\left(H_{z,t-1}\right)_\perp & \eta_{i,t}\in H_{y_i-z,t}\cap\left(H_{y_i-z,t-1}\right)_\perp \end{array}$$

$(\varepsilon_t)_{t\in\mathbb{Z}}$ and $(\eta_{i,t})_{t\in\mathbb{Z}}$ are respectively the innovation processes of $(z_t)_{t\in\mathbb{Z}}$ and $(y_{i,t}-z_t)_{t\in\mathbb{Z}}$ of variance-covariance matrices Ω_ε and Ω_η.

Remark 3 *A natural approximation of the process* $(z_t)_{t\in\mathbb{Z}}$ *that appears in Lemma 2 is the natural estimates of the common factor in the set of homogeneous time series under study, i.e.* $\frac{1}{N}\sum_{j=1}^{N} y_{j,t}$.

The basic properties of autocovariance matrix functions are satisfied by $(\Gamma(h))_{h\in\mathbb{N}}$ and $(\Xi(h))_{h\in\mathbb{N}}$, but there exists an additional link between both functions. This is introduced in the following Corollary that is a by-product of the preceding Lemma.

Corollary 4 $(\Gamma(h))_{h\in\mathbb{N}}$ *and* $(\Xi(h))_{h\in\mathbb{N}}$ *are non-negative matrix functions such that* $(\Gamma(h)-\Xi(h))_{h\in\mathbb{N}}$ *is also a non-negative matrix function.*

The above form of exchangeability may be weakened if we want to introduce some scale heterogeneity and then limit our attention to the autocorrelation structure, allowing for individual heteroskedasticity. In this case, its definition can be phrased as follows:

Definition 5 *A panel of* m*-dimensional stochastic processes* $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ *is said weakly linearly correlation exchangeable (WLCE) if*

(i) $\forall i\in\mathbb{N}, (y_{i,t})_{t\in\mathbb{Z}}$ *is a purely non deterministic covariance stationary process,*
(ii) $\forall (i,j)\in\mathbb{N}^2, \forall h\in\mathbb{Z}$,

$$cov(y_{i,t},y_{j,t+h}) = dvar(y_{i,t})^{1/2}\left(\delta_{i,j}\widetilde{\Gamma}(h)+(1-\delta_{i,j})\widetilde{\Xi}(h)\right)dvar(y_{j,t})^{1/2}$$

where $dvar(y_{i,t})$ *stands for the diagonal matrix whose diagonal terms are those of* $var(y_{i,t})$.

This definition can also be extended by allowing for the presence of heterogenous deterministic terms $\mu_i(t)$ such that $(y_{i,t}-\mu_i(t))_{i\in\mathbb{N},t\in\mathbb{Z}}$ is a WLCE panel or in conditioning on some exogenous variables $(x_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$. Similarly, the basic properties of autocovariance matrix functions are satisfied by $\left(\widetilde{\Gamma}(h)\right)_{h\in\mathbb{N}}$ and $\left(\widetilde{\Xi}(h)\right)_{h\in\mathbb{N}}$.

Corollary 6 $\left(\widetilde{\Gamma}(h)\right)_{h\in\mathbb{N}}$ *and* $\left(\widetilde{\Xi}(h)\right)_{h\in\mathbb{N}}$ *are non negative matrix functions such that* $\left(\widetilde{\Gamma}(h)-\widetilde{\Xi}(h)\right)_{h\in\mathbb{N}}$ *is also a non negative matrix function.*

At last Lemma 2 takes the following form

Lemma 7 *If* $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ *is WLCE, there exists* $(\widetilde{z}_t)_{t\in\mathbb{Z}}$ *a covariance stationary process in* H_y *such that* $\forall t\in\mathbb{Z}$,

$$\left\{\begin{array}{c}
(i)\forall h\in\mathbb{N}, cov(\widetilde{z}_t,\widetilde{z}_{t+h}) = \widetilde{\Xi}(h),\\
(ii)\forall (h,i)\in\mathbb{Z}\times\mathbb{N}, cov(\widetilde{z}_t,y_{i,t+h}) = \widetilde{\Xi}(h)\,dvar(y_{i,t})^{1/2}\\
(iii)\,\forall (h,i,j)\in\mathbb{Z}\times\mathbb{N}^2,\\
cov\left(y_{i,t}-var(y_{i,t})^{1/2}\widetilde{z}_t, y_{j,t+h}-var(y_{j,t})^{1/2}\widetilde{z}_{t+h}\right) =\\
\delta_{i,j}dvar(y_{i,t})^{1/2}\left(\widetilde{\Gamma}(h)-\widetilde{\Xi}(h)\right)dvar(y_{i,t})^{1/2}
\end{array}\right.$$

A natural approximation of the process $(\widetilde{z}_t)_{t\in\mathbb{Z}}$ that appears in Lemma 7 is $\frac{1}{N}\sum_{j=1}^{N} dvar(y_{i,t})^{-1/2}y_{j,t}$ and the Wold type representation equation satisfied by $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ is given by

$$y_{i,t} = \Lambda_i\left(\sum_{k=0}^{+\infty}\widetilde{C}_k\widetilde{\varepsilon}_{t-k}+\sum_{k=0}^{+\infty}\widetilde{D}_k\widetilde{\eta}_{i,t-k}\right)$$

where Λ_i is a diagonal matrix and

$$\begin{array}{ll}
\widetilde{C}_0 = I_p & \widetilde{D}_0 = I_p \\
\sum_{k=0}^{+\infty} \sqrt[2]{Tr\left(\widetilde{C}_k \widetilde{C}_k'\right)} < +\infty & \sum_{k=0}^{+\infty} \sqrt[2]{Tr\left(\widetilde{D}_k \widetilde{D}_k'\right)} < +\infty \\
\widetilde{\varepsilon}_t \in H_{\{z_{t'}, t' \leq t\}} \cap H_{\{z_{t'}, t' \leq t-1\}, \perp} & \widetilde{\eta}_{i,t} \in H_{\left\{y_{i,t'} - z_{t'}, t' \leq t\right\}} \cap H_{\left\{y_{i,t'} - z_{t'}, t' \leq t-1\right\}, \perp} \\
V\widetilde{\varepsilon}_t = \Omega_\varepsilon & V\widetilde{\eta}_{i,t} = \Omega_\eta \\
diag\left(\sum_{k=0}^{+\infty} \widetilde{C}_k \Omega_\varepsilon \widetilde{C}_k'\right) = I_p & diag\left(\sum_{k=0}^{+\infty} \widetilde{D}_k \Omega_\eta \widetilde{D}_k'\right) = I_p
\end{array}$$

Remark 8 *It is possible to extend this framework to take into account additional exchangeability properties and derive more worked out Wold representation. For instance, if we deal with data that can be organized in subsets for which it seems reasonable to assume that there exist exchangeability properties between subsets and within subsets once some exogenous information has been taken into account, a similar Wold representation can be derived. This may be the case when statisticians have at their disposal a large set of time series associated to firms in a large number of industries. If $i\,(j)$ means that unit i is in subset j and there are N_j units in M subsets, we can consider the situation in which :*

$$\begin{aligned}
cov\left(y_{i(j),t}, y_{k(l),t+h}\right) &= \Gamma_j\,(h)\,\delta_{i(j),k(l)} + \Xi_j\,(h)\left(1 - \delta_{i(j),k(l)}\right)\delta_{j,l} \\
&\quad + \Lambda\,(h)\left(1 - \delta_{i(j),k(l)}\right)\left(1 - \delta_{j,l}\right)
\end{aligned}$$

where $\lim_{M \to +\infty} \frac{1}{M} \sum_{j=1}^{M} \Xi_j\,(h) = \Xi^*\,(h)$ for all $h \in \mathbb{Z}$ and get the following Wold representation

$$y_{i(j),t} = \sum_{k=0}^{+\infty} B_k \xi_{t-k} + \sum_{k=0}^{+\infty} C_{j,k} \varepsilon_{j,t-k} + \sum_{k=0}^{+\infty} D_{j,k} \eta_{i(j),t-k}$$

The first component can be read as the general environment component, the second one as that of the subset j and the last one as that of the unit i in the subset j.

3 ECM representation theorem for WLE panel

When there exists a cross-correlation between individual time series, the panels that are considered in recent works (Bai and Ng (2001), Moon and Perron (2001), Chang and Song (2002)) are composed of various heterogenous univariate time series in presence of common factors. They can be viewed as a very large multivariate time series. When panels of multivariate times series under an autoregressive form are considered, an independence assumption between the units is usually added (Phillips and Moon (1999), Pedroni (1999) *inter alios*). Otherwise, ergodicity properties may be lost (Phillips and Moon (1999)) when this dependence is not explicitly modeled.

We are interested in dealing with a large number of multidimensional processes in presence of a cross-section dependence but under an autoregressive representation rather than the moving average one we derived in the preceding section. To be able to handle this kind of DGP, we have to introduce some constraints, but these constraints must be reasonable and well grounded. The exchangeability property we used in the preceding section allows us just to derive an interesting representation under an autoregressive form. It can take two forms related to a VAR or a VECM representation. This second model family is quite useful when some components of the multivariate time series are integrated and possibly cointegrated, which is usually considered when working with macroeconomic data. Knowing the general VAR/VECM type equation they satisfy, we can then study the problem of their estimation.

We proceed in two steps. First, we derive the general VECM type equation from the Wold representation established in the preceding section, but we restrict our attention to a particular and relevant sub-family of models. In general it will not be possible to handle easily the form obtained for this sub-family with simple statistical tools. Moreover its derivation is obtained under the assumption that N is very large and an accurate approximation of the multivariate common component is available that may not always be adapted in practice We then focus on the situation when N is not very large i.e. when we are handling a subsample of a WLE panel and derive in these circumstances the VECM type approximation that is satisfied by the data. In a second step, we select one particular and relevant approximation of this VECM form as a DGP and compute its Wold representation in order to be able to study the properties of the estimators we introduce in the following section.

3.1 From the Wold representation to the autoregressive representation

We start from a collection of covariance stationary weakly exchangeable processes. To represent the integrated processus that could be built from them, we use the integral operator S that plays the role of the inverse of the first-difference operator $\Delta = (1 - L)$. It is defined by

Definition 9 *The integral operator S associates to any sequence $\varepsilon_t = (\varepsilon_t, t = ..., \ -1, \ 0, \ 1, ...)$ of real numbers a sequence $S\varepsilon_t$ defined by : F*

$$S\varepsilon_t = \begin{cases} \sum_{\tau=1}^{t} \varepsilon_\tau & \text{for } t > 0 \\ 0 & \text{for } t = 0 \\ -\sum_{\tau=t+1}^{0} \varepsilon_\tau & \text{for } t < 0 \end{cases}$$

In practice, if we start from an integrated of order one process $(x_t)_t$, we would work in the sequel on (Δx_t) so that $S(\Delta x_t) = x_t - x_0$. All the integrated processes constructed with the help of the integral operator are equal to 0 in date $t = 0$. This is a convention. Its use implies the introduction of deterministic terms in cointegration relationships.

We consider *the family $\mathcal{P}$* of WLE panels that satisfy the following Wold representation :

$$y_{i,t} = \sum_{k=0}^{+\infty} C_k \varepsilon_{t-k} + \sum_{k=0}^{+\infty} D_k \eta_{i,t-k} \tag{1}$$

where the polynomial matrices $C(L) = \sum_{k=0}^{+\infty} C_k L^k$ and $D(L) = \sum_{k=0}^{+\infty} D_k L^k$ are such that

$$\begin{array}{ll} (i)_{\mathcal{P}} & \det(C(L)) = (1-L)^r c(L) \\ (ii)_{\mathcal{P}} & \text{rank} C(1) = m - r \\ (iii)_{\mathcal{P}} & \det(D(L)) = (1-L)^s d(L) \\ (iv)_{\mathcal{P}} & \text{rank} D(1) = m - s \end{array}$$

with $(v)_{\mathcal{P}}$ $r < m$ and $s < m$, and $(vi)_{\mathcal{P}}$ the roots of the polynomials $c(L)$ and $d(L)$ have modulus strictly larger than one and $(vii)_{\mathcal{P}}$ $\sum_{k=0}^{+\infty} kTr(C_k C_k')^{1/2} < +\infty$ and $\sum_{k=0}^{+\infty} kTr(D_k D_k')^{1/2} < +\infty$. For panel data in this family, we get the following Representation theorem

Theorem 10 *Let $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ be a WLE panel in $\mathcal{P}$, then let β be a full rank $m \times r$ matrix whose columns span the vector space $KerC(1)'$*[1] *and γ be a full rank $m \times s$ matrix whose columns span the vector space $KerD(1)'$*[2]*, there exist α and δ full rank matrices*

[1] i.e, $\beta' C(1) = 0$.

[2] i.e., $\gamma' D(1) = 0$.

of respective dimensions $m \times r$ and $m \times s$, two polynomial matrices $\psi(L)$ and $\phi(L)$, a set of constant $(s \times 1)$ vectors $(\mu_i)_{i \in \mathbb{N}}$ such that $\lim_{N \longrightarrow +\infty} \frac{1}{N} \sum_{i=1}^{N} \mu_i = 0$ and a constant $(r \times 1)$ vector μ such that

$$\begin{aligned} y_{i,t} &= \alpha\left(\beta' S z_{t-1} + \mu\right) + \delta\left(\gamma'\left(S y_{i,t-1} - S z_{t-1}\right) + \mu_i\right) \\ &\quad + \psi(L) z_{t-1} + \phi(L)\left(y_{i,t-1} - z_{t-1}\right) + \varepsilon_t + \eta_{i,t} \end{aligned}$$

where $(z_t)_{t \in \mathbb{Z}}$ is unobserved and satisfies the VECM

$$z_t = \alpha\left(\beta' S z_{t-1} + \mu\right) + \psi(L) z_{t-1} + \varepsilon_t$$

and $\psi(0) = I_p$, $\phi(0) = I_p$, $\sum_{k=0}^{+\infty} Tr\left(\Psi_k \Psi_k'\right) < +\infty$ and $\sum_{k=0}^{+\infty} Tr\left(\phi_k \phi_k'\right) < +\infty$, and $(\varepsilon_t)_{t \in \mathbb{Z}}$ and $(\eta_{i,t})_{t \in \mathbb{Z}}$ are the white noise processes that appear in the Wold representation (1) *such that $V\varepsilon_t = \Omega_\varepsilon$ and $V\eta_{i,t} = \Omega_\eta$.*

Two sets of cointegrating vectors appear in this representation. One the one hand, β corresponds to the set of cointegrating direction between the common components. On the other hand, γ spans the cointegration space of the idiosyncratic components. These cointegration spaces are not necessary identical.

Remark 11 *When T is not very large, the usual practice by panel data modelers is to introduce time dummies to capture the contribution of the common components $(z_t)_{t \in \mathbb{Z}}$. This approach may be consistent with the framework under study. When the variables are assumed to be parcimoniously described as integrated processes, it is consistent only when the cointegration relationships between the idiosyncratic components $(y_{i,t} - z_t)_{t \in \mathbb{Z}}$ is a subset (possibly void) of those of the common ones and their loading factors are equal. Indeed, even if the cointegration relationships are the same ones $(\beta = \gamma)$, it may happen that the idiosyncratic and common error correction mechanisms affect differently the changes of $y_{i,t}$ i.e. $\alpha \neq \delta$. In the other cases, in order to get a correct autoregressive approximation of the process the two sets of cointegration relationships must be included.*

Remark 12 *When $r = 0$ and $s = 0$ or the variables under study are covariance stationary, the autoregressive models we wind up with are VAR-type models that involves lagged values of the common factors $(z_t)_{t \in \mathbb{Z}}$ and the idiosyncratic components $(y_{i,t} - z_t)_{t \in \mathbb{Z}}$. A standard VAR representation of WLE panel is obtained when $\psi(L) = \phi(L)$ but with a two-way error term.*

Remark 13 *When the Wold representation of the processes under study involves deterministic terms as in*

$$y_{i,t} = \kappa_{i,1} + \kappa_{i,2} t + \sum_{k=0}^{+\infty} C_k \varepsilon_{t-k} + \sum_{k=0}^{+\infty} D_k \eta_{i,t-k}$$

deterministic trends are present in the cointegration relationships that take the form

$$\gamma'\left(S y_{i,t-1} - S z_{t-1} - \left[\kappa_{i,1} - \kappa_1 + \frac{\kappa_{2,i} - \kappa_2}{2}\right] t - \left[\frac{\kappa_{2,i} - \kappa_2}{2}\right] t^2\right) + \mu_i$$

and

$$\left(\beta'\left(S z_{t-1} - \left(\kappa_1 + \frac{\kappa_2}{2}\right) t - \frac{\kappa_2}{2} t^2\right) + \mu\right)$$

If we do not impose conditions $(ii)_{\mathcal{P}}$ and $(iv)_{\mathcal{P}}$ in the definition of the process family under study, we can get more worked out representation theorem with polynomial cointegration in the spirit of what was obtained by Gregoir and Laroque (1993). Similarly, if we consider various unit roots on the unit circle, the representation equation would be more complicated and would involve seasonal polynomial cointegration terms as in Gregoir(1999).

In practice the above representation equation cannot be easily used as $(z_t)_{t\in\mathbb{Z}}$ is generally unobserved. The use of unobserved components would necessitate the introduction of distributional assumptions for the error terms and identification constraints. It would also lead to complicated estimation procedures we do not want to consider here. Nevertheless, three cases can be distinguished.

a- In some circumstances, the process $(z_t)_{t\in\mathbb{Z}}$ might be observable. For instance, when $(y_{i,t})_t$ are firm level data that are homogeneous to variables measured in the national system of accounts (in an appropriate classification with respect to the exchangeability hypothesis). The common component $(z_t)_{t\in\mathbb{Z}}$ can be approximated by the m-dimensional vector of national account components divided by the size of this population. This is not possible if among the variables some are ratios such as for instance margin rate. Similarly, if among the variables under study, one component is an individual forecast at horizon 1 of an exchange rate on a monthly basis, the common component can be approximated in case of rational use of information by the exchange rate in the appropriate futures market. It is notwithstanding not always possible to have such variables at the modeler's disposal.

b- When N is very large, the representation can be approximated in replacing z_t by $\frac{1}{N}\sum_{i=1}^{N} y_{i,t}$. We then get

$$\begin{aligned} y_{i,t} &= \alpha\left(\beta'\frac{1}{N}\sum_{j=1}^{N} Sy_{j,t-1}+\mu\right)+\delta\left(\gamma'\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N} Sy_{j,t-1}\right)+\mu_i\right) \qquad (2)\\ &\quad +\psi(L)\left(\frac{1}{N}\sum_{j=1}^{N} y_{j,t-1}\right)+\phi(L)\left(y_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N} y_{j,t-1}\right)+\varepsilon_t+\eta_{i,t} \end{aligned}$$

and the modeler can work with such a representation.

c- When N is not very large, the above approximation can be inaccurate. Simply because the cointegration relationships between the components of Sz_t can be poorly approximated by the same linear combinations between those of $\frac{1}{N}\sum_{j=1}^{N} Sy_{j,t}$. We are interested in deriving, if it is possible, an alternative Representation Theorem for a finite subsample of exchangeable processes that may provide us with a better approximation in this case. It certainly implies a more constrained cointegration structure since part of the cointegrating relations of the process $(z_t)_{t\in\mathbb{Z}}$ might not be covariance stationary when applied to the process $\left(\frac{1}{N}\sum_{i=1}^{N} y_{i,t}\right)_{t\in\mathbb{Z}}$. A preliminary remark can clarify this point. The spectral matrix in 0 of the process $\left(\frac{1}{N}\sum_{i=1}^{N} y_{i,t}\right)_{t\in\mathbb{Z}}$ is equal to

$$\frac{1}{2\pi}\left(\frac{1}{N}D(1)\,\Omega_\eta D(1)'+C(1)\,\Omega_\varepsilon C(1)'\right)$$

Any cointegrating vector at frequency 0 of this process is in the kernel of the above matrix, but if a is in its kernel, then

$$\frac{1}{N}aD(1)\,\Omega_\eta D(1)'\,a=-aC(1)\,\Omega_\varepsilon C(1)'\,a$$

Both quantities must be positive or equal to 0 as they are quadratic form associated to positive matrices, they are necessarily equal to 0. A cointegrating vector of $\left(\frac{1}{N}\sum_{i=1}^{N} y_{i,t}\right)_{t\in\mathbb{Z}}$

is in the kernel of $C(1)\Omega_{\varepsilon}C(1)'$and in that of $D(1)\Omega_{\eta}D(1)'$ .In other words, a cointegration vector of the approximate common component is necessarily a cointegrating vector of the idiosyncratic ones.

Remark 14 *When dealing with WLEC processes, the process z_t is approximated by a weighted average of contemporaneous terms. More generaly, common components could be approximated by weighted averages when some heteroskedasticity is present in the individual time series and can be described with some well-chosen variable.*

When $KerD(1)' \cap KerC(1)' \neq \{0\}$, there exist common cointegrating relationships between the processes $\left(y_{i,t} - \frac{1}{N}\sum_{j=1}^{N} y_{j,t}\right)_{t\in\mathbb{Z}}$and the process $\left(\frac{1}{N}\sum_{j=1}^{N} y_{j,t}\right)_{t\in\mathbb{Z}}$. Necessarily, these are also cointegration relationships of the processes $(y_{i,t})_{t\in\mathbb{Z}}$. We have in this case the following result :

Proposition 15 *Let $(y_{i,t})_{i\in\{1,...,N\},t\in\mathbb{Z}}$ be a size N subsample of a WLE panel in $\mathcal{P}$ such that $KerD(1)' \cap KerC(1)' \neq \{0\}$, then let γ_1 be a full rank $m\times s_1$ matrix whose columns span the vector space $KerD(1)' \cap KerC(1)'$, let γ_2 be a full rank $m \times (s - s_1)$ matrix whose columns span the vector space orthogonal to $KerD(1)' \cap KerC(1)'$ in $KerD(1)'$, there exist α, δ_1 and δ_2 full rank matrices of respective dimensions $m \times s_1$, $m \times s_1$ and $m\times(s - s_1)$, two polynomial matrices $\psi(L)$ and $\phi(L)$,and a set of constant $(s \times 1)$ vectors $(\mu_i)_{i\in\{1,...,N\}}$ and a constant $(s_1 \times 1)$ vector μ such that*

$$\begin{aligned} y_{i,t} - z_t^N &= \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix} \left(\begin{pmatrix} \gamma_1' \\ \gamma_2' \end{pmatrix} \left(Sy_{i,t-1} - Sz_{t-1}^N\right) + \mu_i \right) \\ &\quad +\phi(L)\left(y_{i,t-1} - z_{t-1}^N\right) + \eta_{i,t} - \frac{1}{N}\sum_{j=1}^{N}\eta_{j,t} \end{aligned} \tag{3}$$

and

$$z_t^N = \alpha\left(\gamma_1' S z_{t-1}^N + \mu\right) + \psi(L) z_{t-1}^N + v_t \tag{4}$$

where $\left(z_t^N\right)_{t\in\mathbb{Z}}$ is given by $z_t^N = \frac{1}{N}\sum_{j=1}^{N} y_{j,t}$ and $\psi(0) = I_p$, $\phi(0) = I_p$, $\sum_{k=0}^{+\infty} Tr\left(\Psi_k \Psi_k'\right) < +\infty$ and $\sum_{k=0}^{+\infty} Tr\left(\phi_k \phi_k'\right) < +\infty$, $(\eta_{i,t})_{t\in\mathbb{Z}}$ is the white noise process that appears in the Wold representation (1) such that $V\eta_{i,t} = \Omega_\eta$ and $(v_t)_{t\in\mathbb{Z}}$ is a white noise process such that $v_t \in \left(H_{z^N,t-1}\right)_\perp \cap H_{z^N,t}$.

In other words, from a subsample of WLE panel, we can derive two sets of VECM representations. One for the common component and one for the idiosyncratic ones. Unfortunately, both sets of equations cannot generally be used simultaneously as the error term in the second one (4) may be correlated to past values of the idiosyncratic components. It is not always possible to add (3) and (4) to get a representation equation for $(y_{i,t})_{t\in\mathbb{Z}}$.

When $KerD(1)' \cap KerC(1)' = \{0\}$, the above Proposition remains valid with $s_1 = s$. The only cointegrating relationships we can capture are those that are associated to the polynomial matrix $D(L)$ and are associated to the processes $\left(y_{i,t} - \frac{1}{N}\sum_{j=1}^{N} y_{j,t}\right)_{t\in\mathbb{Z}}$ for which we get a VECM representation similar to (3) . The process $\left(z_t^N\right)_{t\in\mathbb{Z}}$ satisfies a VAR representation.

In the sequel, we propose to consider an approximation of the VECM equation mixing the results obtained in case b for the error term structure and in case c for the cointegration relationship structure. We therefore postulate this autoregressive approximation that may be a relevant DGP in several situations and derive its Wold-type representation to analyze the properties of estimates we can construct in this context.

3.2 A subclass of VECM of interest

In the preceding section, we saw that working with a not very large number of units (N) may lead to situations in which it is not always possible to relevantly add equation (3) and (4) to get a constrained VECM representation with respect to (2). We propose to focus on situation in which this operation is legitimate. Postulating such a DGP, we then compute its Wold representation. We introduce **a definition** : A set of panel data satisfies a **WLE-VECM**(p_1, p_2) if it satisfies the following representation equation :

$$\begin{aligned} y_{i,t} &= \delta_1 \left(\gamma_1' \left(Sy_{i,t-1} - Sz_{t-1}^N\right) + \mu_{1,i}\right) + \delta_2 \left(\gamma_2' \left(Sy_{i,t-1} - Sz_{t-1}^N\right) + \mu_{2,i}\right) \\ &\quad + \alpha \left(\gamma_1' S z_{t-1}^N + \mu\right) + \sum_{j=1}^{p_1} \psi_j z_{t-j}^N + \sum_{k=1}^{p_2} \phi_k \left(y_{i,t-k} - z_{t-k}^N\right) + \varepsilon_t + \eta_{i,t} \end{aligned} \tag{5}$$

where $z_t^N = \frac{1}{N}\sum_{j=1}^N y_{j,t}$, γ_1, δ_1 and α (resp. γ_2 and δ_2) are $m \times s_1$ (resp. $m \times s_2$) matrices of rank s_1 (resp. s_2), $\frac{1}{N}\sum_{j=1}^N \mu_{1,j} = 0$ and $\frac{1}{N}\sum_{j=1}^N \mu_{2,j} = 0$. In contrast with what we obtained in Proposition 15, we add an orthogonality assumption between the common error term $(\varepsilon_t)_{t\in\mathbb{Z}}$ and the idiosyncratic ones $(\eta_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$. Additional conditions have to be given to ensure that the error term that appear in the equation are orthogonal to the past of the processes : we assume that

M1 : the roots of the polynomials

$$\det\left((1-L)\, I_p - \alpha\gamma_1' L - \sum_{j=k}^{p_1} \psi_k (1-L)\, L^k\right) \tag{6}$$

and

$$\det\left((1-L)\, I_p - \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix} \begin{pmatrix} \gamma_1' \\ \gamma_2' \end{pmatrix} L - \sum_{j=k}^{p_2} \phi_k (1-L)\, L^k\right) \tag{7}$$

are equal to one or have a modulus strictly large than one.

We need an additional condition to ensure that equation (5) does not involve polynomial cointegration with processes integrated of order larger than one. This condition is necessary to derive the Wold representation as was shown by Johansen (1995). We get similar conditions to his :

M2 : the matrices

$$\left(\alpha_\perp' \left(I_p - \sum_{j=1}^{p_1} \psi_j\right) \gamma_{1,\perp}\right) \tag{8}$$

and

$$\left(\begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix}_\perp' \left(I_p - \sum_{j=1}^{p_2} \phi_j\right) \begin{pmatrix} \gamma_1 & \gamma_2 \end{pmatrix}_\perp\right) \tag{9}$$

must be of full rank (respectively $m - s_1$ and $m - s$).

We can now give the Wold representation of $(y_{i,t})_{i\in\{1,\dots,N\},t\in\mathbb{Z}}$, to do so, we define two matrices

$$\begin{aligned} C_1 &= \gamma_{1,\perp} \left(\alpha_\perp' \left(I_p - \sum_{j=1}^{p_1} \psi_j\right) \gamma_{1,\perp}\right)^{-1} \alpha_\perp' \\ D_1 &= \begin{pmatrix} \gamma_1 & \gamma_2 \end{pmatrix}_\perp \left(\begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix}_\perp' \left(I_p - \sum_{j=1}^{p_2} \phi_j\right) \begin{pmatrix} \gamma_1 & \gamma_2 \end{pmatrix}_\perp\right)^{-1} \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix}_\perp' \end{aligned}$$

and get

Proposition 16 *If $(y_{i,t})_{i\in\{1,...,N\},t\in\mathbb{Z}}$ is a WLE panel that satisfies equation* (5), *conditions* **M1** *and* **M2** *then there exist two polynomial matrices $C(L)$ and $D(L)$ such that $\sum_{k=0}^{+\infty}\sqrt[2]{Tr\left(C_kC_k'\right)}<+\infty$, $\sum_{k=0}^{+\infty}\sqrt[2]{Tr\left(D_kD_k'\right)}<+\infty$, $\sum_{k=0}^{+\infty}k\sqrt[2]{Tr\left(C_kC_k'\right)}<+\infty$ and $\sum_{k=0}^{+\infty}k\sqrt[2]{Tr\left(D_kD_k'\right)}<+\infty$*

$$y_{i,t}=C(L)\varepsilon_t+D(L)\eta_{i,t}+(C(L)-D(L))\frac{1}{N}\sum_{j=1}^{N}\eta_{j,t} \tag{10}$$

with $C(1)=C_1$ and $D(1)=D_1$.

Remark 17 *In case c, when N becomes larger and larger, the cointegrating relationships between the components of Sz_t may be more and more accurately detected and the properties of the multivariate time series can be described with larger accuracy. The error term $(v_t)_{t\in\mathbb{Z}}$ in* (4) *is less and less correlated with the past values of the idiosyncratic components. The above model should be changed with the number of available units N. Nevertheless if we consider asymptotic analysis as a way to approximate the distribution of functions of stochastic processes for a particular model when the size of the sample is large in N, we may keep to these approximations and analyze the asymptotic behaviors of estimates derived with such a DGP .*

4 Testing for integratedness and stationarity

Recently, a lot of tests for integration and stationarity have been introduced in the panel literature. A large number of them consider situations in which there does not exist any cross-correlation between the series (see Levin, Lin and Chu (2002) as one of the first attempt to tackle this issue and Baltaji and Kao (2001) for a survey). More recently, Bai and Ng (2001), Moon and Perron (2001) and Chang and Song (2002) have introduced unit root test procedure in panel with cross-section dependence. The procedures we introduce now for the WLE panel date are different even if formally the DGP we consider here is close to those they considered. Our DGP is more constrained. We impose by the exchangeability property the presence of one factor that affect similarly every individual time series.

In this section, the results hold under the following set of assumptions.

Assumption A: $\sum_{j=0}^{+\infty}j\left|c_j\right|<+\infty$ and $\sigma_\varepsilon^2 c(1)^2>0$, $\sum_{j=0}^{+\infty}j\left|d_j\right|<+\infty$ and $\sigma_\varepsilon^2 d(1)^2>0$

Assumption B: $\varepsilon_t, t\in\mathbb{Z}$ is a martingale difference satisfying $E\left(\varepsilon_t\left|\mathcal{F}_{t-1}\right.\right)=0$, $E\left(\varepsilon_t^2\left|\mathcal{F}_{t-1}\right.\right)=\sigma_\varepsilon^2$ and $E\left(\varepsilon_t^4\left|\mathcal{F}_{t-1}\right.\right)<+\infty$ a.s., where $\mathcal{F}_{t-1}$ is the $\sigma-field$ generated by $\{\varepsilon_{t-\tau},\tau=1,2,...\}$

Assumption C: $\forall i\in\mathbb{N}$, $\eta_{i,t}, t\in\mathbb{Z}$ is a martingale difference satisfying $E\left(\eta_{i,t}\left|\mathcal{F}_{i,t-1}\right.\right)=0$, $E\left(\eta_{i,t}^2\left|\mathcal{F}_{i,t-1}\right.\right)=\sigma_\eta^2$ and $E\left(\eta_{i,t}^4\left|\mathcal{F}_{i,t-1}\right.\right)<+\infty$ a.s., where $\mathcal{F}_{i,t-1}$ is the $\sigma-field$ generated by $\{\eta_{i,t-\tau},\tau=1,2,...\}$

Assumption D: $(\eta_{i,t})_{t\in\mathbb{Z}}$ are mutually independent and independent with $(\varepsilon_t)_{t\in\mathbb{Z}}$.

Assumption E: $(\eta_{i,t})_{t\in\mathbb{Z}}$ have the same probability distribution

These assumptions can be weakened in some directions, particularly we can consider situation in which there may remain some cross-correlation in $o_p\left(\frac{1}{N}\right)$ between the processes $(\eta_{i,t})_{i,t}$. Nuisance parameters that appear in the asymptotic distributions of the usual test statistics involves the autocovariance function and the spectral density in 0 we denote in the sequel for any time series x_t respectively $\gamma_x(0)$ and $f_x(0)$

All the limits we derive here are obtain by a sequential approach, we first let T tend toward $+\infty$, then N tends toward $+\infty$. However as shown by Phillips and Moon (1999), this approach may sometimes give misleading asymptotic results. Here, for some situations, it can be shown that the limit we get is the same one if we commute the two indexes and let first N tends toward $+\infty$ and then T .

4.1 Unit root test in a WLE panel

We follow the approach introduced by Phillips and Perron (1989) and deal explicitly with the nuisance parameters as they did. We consider various data generating processes :

$$\begin{array}{rrcl} (I) & y_{i,t} & = & u_t + v_{i,t} \\ (II) & y_{i,t} & = & \mu_i + u_t + v_{i,t} \\ (III) & y_{i,t} & = & \mu_i + \lambda_i t + u_t + v_{i,t} \end{array}$$

$y_{i,t}$ is a covariance stationary process and we consider the test that $Sy_{i,t}$ is an integrated process of order one. By use of the integral operator, integrated processes are equal to 0 at date $t = 0$. If $y_{i,t} = (1-L)\,\widetilde{y}_{i,t}$ and $\widetilde{y}_{i,t}$ is observed, the following results remain valid if $\widetilde{y}_{i,0} = O_p(1)$. We derive the test statistics under the assumption that $(u_t)_t$ and $(v_{i,t})_{i,t}$ are white noises of respective variance σ_u^2 and σ_v^2 and then in a second step study the behaviour of the these test statistics when the DGP of these processes is given by $u_t = \sum_{k=0}^{+\infty} c_k \varepsilon_{t-k}$ and $v_{i,t} = \sum_{k=0}^{+\infty} d_k \eta_{i,t-k}$ where $(\varepsilon_t)_t$ and $(\eta_{i,t})_{i,t}$ are white noises of respective variance σ_ε^2 and σ_η^2. We use various test regressions and null hypotheses, but present them in case (I). Their extension to the two other cases are straightforward. The simplest test amounts to the following null hypothesis in the linear model :

$$\begin{array}{rcl} H_0 & : & \rho = 0 \\ H_a & : & \rho < 0 \\ y_{i,t} & = & \rho S y_{i,t-1} + u_t + v_{i,t} \end{array} \tag{11}$$

Using the Representation Theorem we get in the preceding section in a univariate set-up, we are prone to consider the following test specification

$$\begin{array}{rcl} H_{0,1} & : & \rho_1 = 0 \text{ and } H_{0,2} : \rho_2 = 0 \\ H_{a,1} & : & \rho_1 < 0 \text{ and } H_{a,2} : \rho_2 < 0 \\ y_{i,t} & = & \rho_1 \left(S y_{i,t-1} - S z_{t-1}^N \right) + \rho_2 S z_{t-1}^N + u_t + v_{i,t} \end{array} \tag{12}$$

A last test procedure that seems natural and in line with the recommendation made by Im, Pesaran and Shin (1997), would be to consider

$$\begin{array}{rcl} H_{0,1}^* & : & \forall i \in \{1,...,N\} \quad \rho_{i,1} = 0 \text{ and } H_{0,2}^* : \rho_2 = 0 \\ H_{a,1}^* & : & \exists i \in \{1,...,N\} \quad \rho_{i,1} = 0 \text{ and } H_{a,2}^* : \rho_2 = 0 \\ y_{i,t} & = & \rho_{i,1} \left(S y_{i,t-1} - S z_{t-1}^N \right) + \rho_2 S z_{t-1}^N + u_t + v_{i,t} \end{array} \tag{13}$$

but we show below that the situation is particularly complicated in this case. All these estimations can be run with OLS regression omitting the cross-correlation between the components or with GLS regression taking it explicitly into account. In the approach given by regression equation (11), OLS and GLS estimates have different asymptotic behaviors. This is not the case when dealing with (12).

From equation (11) ,we get that

$$\widehat{\rho}_{OLS}=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t}Sy_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}\right)^{2}}$$

and when both processes $(u_t)_t$ and $(v_{i,t})_t$ are pure white noises equal to $(\varepsilon_t)_t$ and $(\eta_{i,t})_t$

$$\widehat{\rho}_{GLS}=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t}Sy_{i,t-1}-\frac{1}{\omega+N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}y_{i,t}\right)\left(\sum_{i=1}^{N}Sy_{i,t-1}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}\right)^{2}-\frac{1}{\omega+N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}Sy_{i,t-1}\right)^{2}}$$

where $\omega=\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$. In this case, ω can be estimated under the null by the maximum likelihood estimator under a gaussianity assumption that is equal to

$$\widehat{\omega}=\frac{N\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t}^{2}-\sum_{t=1}^{T}\left(\sum_{i=1}^{N}y_{i,t}\right)^{2}}{\sum_{t=1}^{T}\left(\sum_{i=1}^{N}y_{i,t}\right)^{2}-\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t}^{2}}$$

When the processes are not pure white noises, a natural alternative estimate would be

$$\widetilde{\omega}=\frac{f_{v_{i,t}}(0)}{f_{u_t}(0)}$$

Nevertheless, asymptotically this parameter does not play a role. It is negligeable when N becomes large, nevertheless a judicious choice can improve the properties of the test procedure when N is not large.

Proposition 18 *When the DGP of a WLE panel $(y_{i,t})_{i,t}$ is given by (A) $[A\in\{I,II,III\}]$, then we have under assumptions A-D when sequentially T then N tend to $+\infty$*

$$\frac{T\left(\widehat{\rho}_{OLS}-\frac{\frac{1}{2}\left(2\pi f_{y_i}(0)-\gamma_{y_i}(0)\right)+q_A 2\pi f_{v_i}(0)}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}^{(A)}\right)^{2}}\right)}{1-\frac{r_A 2\pi f_{v_i}(0)}{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}^{(A)}\right)^{2}}}\Longrightarrow\frac{\int_0^1 W^{(A)}dW^{(A)}}{\int_0^1\left(W^{(A)}(\tau)\right)^2 d\tau}$$

where $Sy_{i,t-1}^{(A)}$ stands for the residual of the projection of $Sy_{i,t-1}$ on the appropriate set of regressors according to the DGP we consider and $W^{(A)}$ is the associate standard Brownian motion, $q_I=0$, $q_{II}=-\frac{1}{2}$ and $q_{III}=-\frac{1}{2}$ and $r_I=\frac{1}{2}$, $r_{II}=\frac{1}{6}$ and $r_{III}=\frac{1}{15}$.Under assumptions A-E, the same result holds when sequentially N then T tend to $+\infty$.

The OLS estimate has an asymptotic distribution which is close to that of a standard unit root process. An additional correction must be incorporated to take into account the presence of a nuisance parameter in the limit of the denominator. To use this test statistic we need an estimate of the nuisance parameter

$$2\pi f_{v_i}(0)$$

It can be obtained in using the process $y_{i,t}-\frac{1}{N}\sum_{j=1}^{N}y_{j,t}=v_{i,t}-\frac{1}{N}\sum_{j=1}^{N}v_{j,t}$ that is such that under our assumptions

$$\begin{aligned}\gamma_{y_{i,t}-\frac{1}{N}\sum_{j=1}^{N}y_{j,t}}(h) &= \left(1-\frac{1}{N}\right)\gamma_{v_i}(h),\forall h\in\mathbb{N}\\ 2\pi f_{y_{i,t}-\frac{1}{N}\sum_{j=1}^{N}y_{j,t}}(0) &= \left(1-\frac{1}{N}\right)2\pi f_{v_i}(0)\end{aligned}$$

so that when N becomes large this is a correct approximation. Similar computations lead to

$$\begin{aligned}\gamma_{\frac{1}{N}\sum_{j=1}^{N} y_{j,t}}(h) &= \gamma_u(h) + \frac{1}{N}\gamma_{v_i}(h), \forall h \in \mathbb{N} \\ 2\pi f_{\frac{1}{N}\sum_{j=1}^{N} y_{j,t}}(0) &= 2\pi\left(f_u(0) + \frac{1}{N}f_{v_i}(0)\right)\end{aligned}$$

so that we have at our disposal good candidates to estimate nuisance parameters related to the process $(u_t)_{t\in\mathbb{Z}}$.

If we now turn to the t-statistic, we have

$$V\widehat{\rho}_{OLS} = \frac{\gamma_u(0) + \gamma_{v_i}(0)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}^{(A)}\right)^2}$$

whose denominator is $O_p\left(T^2N\right)$. The t test statistic has not a standard behavior, $T\sqrt{N}\sqrt{V\widehat{\rho}_{OLS}} = O_p(1)$ and $t_{\widehat{\rho}_{OLS}} = O_p\left(\sqrt{N}\right)$. Levin and Lin (1992) have also in certain circumstances a t test statistic that must be corrected by a diverging factor ($O_p\left(\sqrt{N}\right)$) to obtain a standard asymptotic normal distribution, which means that the statistic itself is diverging.

The GLS estimator deals with the presence of cross-correlation between the individual variables, we get in this case an asymptotic behavior close to that obtained in case of panel without any spatial cross-correlation.

Proposition 19 *When the DGP of a WLE panel $(y_{i,t})_{i,t}$ is given by (A) $[A \in \{I, II, III\}]$, then we have*

$$T\sqrt{N}\left(\widehat{\rho}_{GLS} - \frac{\frac{1}{2}\left(2\pi f_{v_i}(0) - \gamma_{v_i}(0)\right) + q_A 2\pi f_{v_i}(0)}{\frac{1}{NT}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}^{(A)}\right)^2 - \frac{1}{\omega+N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N} Sy_{i,t-1}^{(A)}\right)^2\right]}\right) \Longrightarrow \mathcal{N}\left(0, V_{(A)}\right)$$

where $Sy_{i,t-1}^{(A)}$ stands for the residual of the projection of $Sy_{i,t-1}$ on the appropriate set of regressors according to the DGP we consider, $V_I = 2$, $V_{II} = \frac{4}{3}$ and $V_{III} = 3$.

If we turn to the t test statistic we get that

$$V\widehat{\rho}_{GLS} = \frac{\gamma_u(0) + \gamma_{v_i}(0)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}^{(A)}\right)^2 - \frac{1}{\omega+N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N} Sy_{i,t-1}^{(A)}\right)^2}$$

whose denominator is $O_p\left(NT^2\right)$. The $t_{\widehat{\rho}_{GLS}}$ is such that

$$\sqrt{\frac{\gamma_{y_i}(0)}{r_A 2\pi f_{v_i}(0)}}t_{\widehat{\rho}_{GLS}} - \sqrt{N}\frac{\frac{1}{2}\left(2\pi f_{v_i}(0) - \gamma_{v_i}(0)\right) + q_A 2\pi f_{v_i}(0)}{\sqrt{\frac{1}{NT^2}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}^{(A)}\right)^2 - \frac{1}{\omega+N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N} Sy_{i,t-1}^{(A)}\right)^2\right]}}$$

has $\mathcal{N}\left(0, V_{(A)}\right)$ as an asymptotic distribution.

It may be useful on the other hand to consider the second kind of regression equation as it should allow us to test for a common integratedness driven by the process u_t (

$H_0 : \rho_2 = 0$) or for an individual integratedness ($H_0 : \rho_1 = 0$). The estimates of ρ_1 and ρ_2 are not correlated. In this case, the GLS and OLS estimators are equal.

$$\widehat{\rho}_1 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1} - \frac{1}{N}\sum_{j=1}^{N} Sy_{j,t-1}\right)\left(y_{i,t} - \frac{1}{N}\sum_{j=1}^{N} y_{j,t}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1} - \frac{1}{N}\sum_{j=1}^{N} Sy_{j,t-1}\right)^2}$$

and

$$\widehat{\rho}_2 = \frac{\sum_{t=1}^{T}\left(\sum_{j=1}^{N} Sy_{j,t-1}\right)\left(\sum_{j=1}^{N} y_{j,t}\right)}{\sum_{t=1}^{T}\left(\sum_{j=1}^{N} Sy_{j,t-1}\right)^2}$$

and their GLS variance is given by

$$V\widehat{\rho}_1 = \sigma_v^2\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1} - \frac{1}{N}\sum_{j=1}^{N} Sy_{j,t-1}\right)^2\right)^{-1}$$

and

$$V\widehat{\rho}_2 = \left(\sigma_u^2 + \frac{1}{N}\sigma_v^2\right)\left(\sum_{t=1}^{T}\left(\sum_{j=1}^{N} Sy_{j,t-1}\right)^2\right)^{-1}$$

We get the following weak convergence result.

Proposition 20 *When the DGP of a WLE panel $(y_{i,t})_{i,t}$ is given by (A) $[A \in \{I, II, III\}]$, then we have*

$$T\sqrt{N}\left(\widehat{\rho}_1 - \left(1 - \frac{1}{N}\right)\frac{\frac{1}{2}\left(2\pi f_{v_i}(0) - \gamma_{v_i}(0)\right) + q_A 2\pi f_{v_i}(0)}{\frac{1}{NT}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}^{(A)} - \frac{1}{N}\sum_{j=1}^{N} Sy_{j,t-1}^{(A)}\right)^2\right]}\right) \Longrightarrow \mathcal{N}\left(0, V_{(A)}\right)$$

$$T\left(\widehat{\rho}_2 - \frac{\frac{1}{2}\left(2\pi f_u(0) - \gamma_u(0)\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} Sy_{i,t-1}^{(A)}\right)^2}\right) \Longrightarrow \frac{\int_0^1 W^{(A)} dW^{(A)}}{\int_0^1 \left(W^{(A)}(\tau)\right)^2 d\tau}$$

where $Sy_{i,t-1}^{(A)}$ stands for the residual of the projection of $Sy_{i,t-1}$ on the appropriate set of regressors according to the DGP we consider, $W^{(A)}$ is the associate standard Brownian motion, $V_I = 2$, $V_{II} = \frac{4}{3}$ and $V_{III} = 3$.

If we turn to the t test statistics the result are similar to what we have obtained in the preceding approach. The following test statistic

$$\sqrt{\frac{\gamma_{y_i}(0)}{r_A 2\pi f_{v_i}(0)}}\, t_{\widehat{\rho}_1} - \left(\sqrt{N} - \frac{1}{\sqrt{N}}\right)\frac{\frac{1}{2}\left(2\pi f_{v_i}(0) - \gamma_{v_i}(0)\right) + q_A 2\pi f_{v_i}(0)}{\sqrt{\frac{1}{NT^2}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}^{(A)} - \frac{1}{N}\sum_{j=1}^{N} Sy_{j,t-1}^{(A)}\right)^2\right]}}$$

weakly converges towards $\mathcal{N}\left(0, V_{(A)}\right)$ and $t_{\widehat{\rho}_2} = O_p\left(\sqrt{N}\right)$.

We can now consider the third approach that estimates ρ_1 in each individual time series. In this case we get estimates that do not converge to a simple expression. More precisely, we get that in a first order expansion in $\frac{1}{N}$,

$$\begin{aligned}\widehat{\rho}_{i,1} &= \widetilde{\rho}_{i,1} + \frac{\omega}{\omega+N}\frac{\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)\left(\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)}{\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)^2}\widetilde{\rho}_2 \\ &-\frac{1}{\omega+N}\sum_{k=1}^{N}\frac{\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)\left(Sy_{k,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)}{\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)^2}\widetilde{\rho}_{k,1}\end{aligned}$$

and

$$\widehat{\rho}_2 = \widetilde{\rho}_2 + \frac{\omega}{\omega+N}\sum_{k=1}^{N}\frac{\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)\left(Sy_{k,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)}{\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)^2}\widetilde{\rho}_{k,1}$$

where

$$\begin{aligned}\widetilde{\rho}_{i,1} &= \frac{\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)\left(y_{i,t-1}-\frac{1}{\omega+N}\sum_{j=1}^{N}y_{j,t-1}\right)}{\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)^2} \\ \widetilde{\rho}_2 &= \frac{\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)\left(\frac{1}{N}\sum_{j=1}^{N}y_{j,t-1}\right)}{\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)^2}\end{aligned}$$

$(\widehat{\rho}_{i,1})_{i=1,\dots,N}$ are exchangeable random variables. Using Theorem 2 (Chap 7, sect 3) in Chow and Teicher (1997) we know that $\frac{1}{N}\sum_{i=1}^{N}\widehat{\rho}_{i,1}$ has a stochastic limit as soon as $\widehat{\rho}_{i,1}$ is a L^1 variable, but this limit is a complicated function of the Brownian motion W associated to the common component $(u_t)_t$.

Remark 21 *A WLE panel data can be seen as a factor model with one dynamic factor. The approach proposed by Bai and Ng (2002) that consider panel data with possibly a large number of factors can then be used. It consists of first estimating the factor and second testing for integratedness of the set of factor and of the idiosyncratic components. This would correspond here in testing for integratedness on the one hand of the process* $\left(\frac{1}{N}\sum_{j=1}^{N}y_{j,t}\right)_{t\in\mathbb{Z}}$ *and on the other hand of the processes* $\left(y_{i,t}-\frac{1}{N}\sum_{j=1}^{N}y_{j,t}\right)_{i\in\mathbb{N},t\in\mathbb{Z}}$.

Remark 22 *When the processes under study are WLEC, the above results remain valid if we replace* $y_{i,t}$ *by* $\sqrt{\gamma_{y_{i,t}}(0)}^{-1}y_{i,t}$ *and* $z_t^N=\frac{1}{N}\sum_{j=1}^{N}y_{j,t}$ *by* $z_t^N=\frac{1}{N}\sum_{j=1}^{N}\sqrt{\gamma_{y_{i,t}}(0)}^{-1}y_{i,t}$

4.2 Stationarity test in a WLE panel

Hadri (2001) develops the alternative approach that consists of testing for stationarity with the help of a LM test when there does not exist cross-correlation between the individual series. We can follow also this way and adapt it to different purposes.

We propose to test for the stationarity of the time series $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$, for that of its idiosyncratic components or that of its common one. We consider the three data generating process (I), (II) and (III) in a framework similar to that introduced by Nabeya and

Tanaka (1987). For each DGP, we consider two LM tests under a gaussianity assumption. We only present the approach for DGP (I). For the two remaining DGPs, the appropriate deterministic function is to be introduced in the DGP equation. When testing for the stationarity of the common component, we write :

$$\begin{aligned} y_{i,t} &= r_t^{(c)} + \varepsilon_t + \eta_{i,t} \\ r_t^{(c)} &= r_{t-1}^{(c)} + w_t^{(c)} \end{aligned}$$

where $\sigma^2_{w(c)} = V w_t^{(c)}$ and test for $H_{0(c)} : \sigma^2_{w(c)} = 0$. When testing for the stationarity of the idiosyncratic component, we write :

$$\begin{aligned} y_{i,t} &= r_{i,t}^{(i)} + \varepsilon_t + \eta_{i,t} \\ r_{i,t}^{(i)} &= r_{i,t-1}^{(i)} + w_{i,t}^{(i)} \end{aligned}$$

where $\sigma^2_{w(i)} = V w_{i,t}^{(i)}$ and test for $H_{0(i)} : \sigma^2_{w(i)} = 0$. The test statistics are derived under the assumption that $(\varepsilon_t)_{t\in\mathbb{Z}}$ and $(\eta_{i,t})_{t\in\mathbb{Z},i\in\mathbb{N}}$ are i.i.d gaussian white noises and modified to take into account of nuisance parameters in the general case of dependent processes. When testing for the null $H_{0(c)}$, we wind up with the following LM test statistic

$$\zeta_1^A = \frac{1}{T^2} \sum_{t=1}^{T} \frac{\left(\frac{1}{N} \sum_{j=1}^{N} S\left(y_{j,t}^{(A)} \right) \right)^2}{2\pi f_u(0)}$$

where $y_{j,t}^{(A)}$ stands for the residual of the projection of $y_{j,t}$ on the appropriate set of regressors according to the DGP we consider and whose asymptotic behavior is given by

Proposition 23 *When the DGP of a WLE panel $(y_{i,t})_{i,t}$ is given by (A) $[A \in \{I, II, III\}]$, then we have*

$$\zeta_1^A \Longrightarrow \int_0^1 \left(BW^{(A)}(\tau) \right)^2 d\tau$$

where $BW^{(A)}$ is the Brownian Bridge derived from the residuals of the OLS regression of a white noise projected on the associated deterministic function.

When testing for the null $H_{0(i)}$, we get the following LM test statistic,

$$\zeta_2^A = \frac{1}{NT^2} \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(S y_{i,t}^{(A)} - \frac{2\omega+N}{\omega+N} \sum_{t=1}^{T} \sum_{i=1}^{N} S\left(y_{i,t}^{(A)} \right) \right)^2}{2\pi f_v(0)}$$

whose asymptotic behavior is described now:

Proposition 24 *When the DGP of a WLE panel $(y_{i,t})_{i,t}$ is given by (A) $[A \in \{I, II, III\}]$, then we have*

$$\sqrt{N} \frac{\left(\zeta_2^A - r_A \right)}{s_A} \Longrightarrow \mathcal{N}(0,1)$$

where $(r_I, s_I) = \left(\frac{1}{2}, \frac{1}{3}\right)$, $(r_{II}, s_{II}) = \left(\frac{1}{6}, \frac{1}{45}\right)$ and $(r_{III}, s_{III}) = \left(\frac{1}{15}, \frac{11}{6300}\right)$.

Notice at last that a direct application of KPSS-type test statistic can be used. Taking into account the nuisance parameters, it takes the form

$$\zeta_4^A = \frac{1}{NT^2} \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(S\left(y_{i,t}^{(A)} \right) \right)^2}{2\pi f_u(0)} - r_A \frac{f_v(0)}{f_u(0)}$$

and we have

Proposition 25 *When the DGP of a WLE panel* $(y_{i,t})_{i,t}$ *is given by* (A) $[A \in \{I, II, III\}]$, *then we have*

$$\zeta_4^A \Longrightarrow \int_0^1 \left(BW^{(A)}(\tau)\right)^2 d\tau$$

where $BW^{(A)}$ *is the associate standard Brownian bridge and* $r_I = \frac{1}{2}$, $r_{II} = \frac{1}{6}$ *and* $r_{III} = \frac{1}{15}$.

4.3 Finite sample properties of the unit root test procedure

We proceed in this section to a simulation study of the properties of the different test procedures introduced in this section. We consider the situation of one unit-root in the Data Generating Process of the common and idiosyncratic components . We particularly want to investigate how well the asymptotic theory describes the small sample properties of our test procedures as we usually work with a not too large number of units and time series of dimension of 50 to 100 observations. Data generating processes considered elsewhere in the literature (Phillips and Perron (1988), Schwert (1989), DeJong, Nankervis, Savin and Whiteman (1992), *inter alios*) are used. We consider for the two kinds of components the three following models:

1. $MA(1):$ $\quad \xi_{(i,)t} = w_{(i,)t} - \theta w_{(i,)t-1}$ $\quad (\theta = .8, .5, 0, -.5, -.8)$
2. $AR(1):$ $\quad \xi_{(i,)t} = \phi w_{(i,)t-1} + w_{(i,)t}$ $\quad (\phi = .5, -.5)$
3. $GARCHMA(1):$ $\left\{ \begin{array}{ll} \xi_{(i,)t} = \zeta_{(i,)t} - \theta \zeta_{(i,)t-1}, & (\theta = .5, 0, -.5) \\ \zeta_{(i,)t} = h_{(i,)t}^{\frac{1}{2}} w_{(i,)t} & \\ h_{(i,)t} = 1 + .65 h_{(i,)t-1} + .25 \zeta_{(i,)t-1}^2, \quad h_{(i,)0} = 0 & \end{array} \right.$

and construct a large number (5000) of processes $(y_{i,t})_{i=1,\ldots N, t=1,\ldots,T+50}$ with $(N,T) \in \{(25,50), (25,100), (50,50), (50,100)\}$ generated by

$$(1-\alpha_i L)(1-\alpha L) y_{i,t} = (1-\alpha_i L)\xi_t + (1-\alpha L)\xi_{i,t} \qquad (\alpha_i, \alpha) \in \{1, .99, .95, .90, .80\}^2$$

In each simulation of these models, the initial conditions are set to 0 and the first 50 observations are dropped. The autocorrelation structure of v_t is assumed to be unknown to the econometrician, so we have to estimate the nuisance parameters it involves. The estimation of $f_y(\omega)$ relies on the properties of kernel estimator of spectral density. These non-parametric estimators entail the choice of a "bandwidth" number whose influence on the quality of the estimate can be drastic. We follow Newey and West (1994) in their way of selecting this number when using a Bartlett window.

TO BE COMPLETED

5 Cointegration analysis

In the third section we introduced the form of a VECM that a WLE panel can satisfy under mild regularity assumptions. We now turn to the problem of estimating the cointegration relationship and testing for the dimension of the cointegration space. A maximum likelihood approach involves complicated mathematical problems. We limit our attention to simple and easily implemented procedures. This is at the expense of efficiency. Regarding the error terms, in this section, the results hold under **D** (or **E**) and the two assumptions :

Assumption B': $\varepsilon_t, t \in \mathbb{Z}$ is a m-dimensional martingale difference satisfying $E\left(\varepsilon_t \left|\mathcal{F}_{t-1}\right.\right) = 0$, $E\left(\varepsilon_t \varepsilon_t' \left|\mathcal{F}_{t-1}\right.\right) = \Omega_\varepsilon$ and $E\left(\max_{j=1,\ldots,m}\left\{\varepsilon_{j,t}^4\right\}\left|\mathcal{F}_{t-1}\right.\right) < +\infty$ a.s., where $\mathcal{F}_{t-1}$ is the $\sigma - field$ generated by $\left\{\varepsilon_{t-\tau}, \tau = 1, 2, \ldots\right\}$

Assumption C': $\forall i \in \mathbb{N}$, $\eta_{i,t}, t \in \mathbb{Z}$ is a m-dimensional martingale difference satisfying $E\left(\eta_{i,t}\left|\mathcal{F}_{i,t-1}\right.\right) = 0$, $E\left(\eta_{i,t}\eta_{i,t}'\left|\mathcal{F}_{i,t-1}\right.\right) = \Omega_\eta$ and $E\left(\max_{j=1,\ldots,m}\left\{\eta_{j,i,t}^4\right\}\left|\mathcal{F}_{i,t-1}\right.\right) < +\infty$ a.s., where $\mathcal{F}_{i,t-1}$ is the $\sigma - field$ generated by $\left\{\eta_{i,t-\tau},\ \tau = 1, 2, \ldots\right\}$.

5.1 Estimation of the cointegrating relationships

In the sequel we consider that the WLE panel under study satisfies the assumptions **M1** and **M2** relative to equations (6), (7), (8) and (9) and for sake of simplicity under the additional assumption that $\mu = 0$ and $\forall i \in \{1, \ldots, N\}$ $\mu_{i,1} = \mu_{i,2} = 0$. We propose to use principal component estimates to estimates the directions spanned by the column vector of γ_1 and γ_2. We first notice that γ_1 is associated to cointegration relationships between the components of $\left(Sz_t^N = \frac{1}{N}\sum_{j=1}^N Sy_{j,t}\right)_{t\in\mathbb{Z}}$ and $\left(Sy_{i,t} - Sz_t^N\right)_{i\in\mathbb{N},t\in\mathbb{Z}}$ and therefore between those of $\left(Sy_{i,t}\right)_{i\in\mathbb{N},t\in\mathbb{Z}}$. We should find consistent estimates of γ_1 in selecting the s_1 smallest principal components (i.e. the s_1 eigenvectors associated to the s_1 smallest eigenvalues) of

$$\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T Sy_{i,t}Sy_{i,t}'$$

An estimate of $\gamma_{1,\perp}$ would correspond to the $m - s_1$ largest principal component. A consistent estimate of γ_2 would then be given in selecting the s_2 smallest principal components of

$$\frac{1}{TN}\gamma_{1,\perp}'\sum_{i=1}^N\sum_{t=1}^T\left(Sy_{i,t} - Sz_t^N\right)\left(Sy_{i,t} - Sz_t^N\right)'\gamma_{1,\perp}$$

we denote γ_{21} and in defining $\gamma_2 = \gamma_{1,\perp}\gamma_{21}$. We denote γ_{22} the $(m - s_1 - s_2)$ largest principal component in the second program. We need additional assumptions to ensure that this procedure would lead to consistent estimates of the cointegrating vector space and to identify the particular basis of the cointegration space we select. We follow the identification criterion introduced by Gregoir and Laroque (1994). We need an additional assumption

Assumption F: the eigenvalues of the matrices

$$\gamma_1' V\left(Sy_{i,t}\right)\gamma_1$$

and

$$\gamma_{21}'\gamma_{1,\perp}' V\left(Sy_{i,t} - Sz_t^N\right)\gamma_{1,\perp}\gamma_{21}$$

are all different.

Criterion 26 *The VECM representation* $(5, 6, 7, 8, 9)$ *can be selected with the particular choice of a matrix* $\gamma = \begin{pmatrix}\gamma_1 & \gamma_2 & \gamma_3\end{pmatrix}$ *such that* γ *is unitary,*

$$\gamma_1' V\left(Sy_{i,t}\right)\gamma_1$$

and

$$\gamma_{21}'\gamma_{1,\perp}' V\left(Sy_{i,t} - Sz_t^N\right)\gamma_{1,\perp}\gamma_{21}$$

are diagonal with non increasing terms.

With this identification criterion, we get the following result:

Proposition 27 *If* $(y_{i,t})_{i,t}$ *satisfies the VECM representation* $(5,6,7,8,9)$, *and* $\widehat{\gamma_1}$ *and,* $\widehat{\gamma_{1,\perp}}$ *and* $\widehat{\gamma_{21}}$ *denote the estimates of matrices associated to the principal components as introduced above, then under assumptions* ***B'***, ***C'*** *and* ***F*** *as* T *goes to infinity*

$$\widehat{\gamma_1} - \gamma_1 = \begin{pmatrix} \gamma_1 & \gamma_{1,\perp} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{T}} I_{s_1} & 0 \\ 0 & \frac{1}{T} I_{m-s_1} \end{pmatrix} O_p(1)$$

where $O_p(1)$ *represents a matrix with the appropriate dimensions which converges weakly .*

Notice that if, under assumption **F**, γ_1 and γ_2 are identified by Criterion 26, this is not the case of γ_3. Equivalently, the space spanned by the estimates $\gamma_{1,\perp}$ is superconsistently estimated, but not completely identified. The directions we obtained define a possible basis of this space, in other words, for a particular choice of $\gamma_{1,\perp}$, there exists a full rank matrix $\Theta_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}$ such that

$$\widehat{\gamma_{1,\perp}} - \gamma_{1,\perp}\Theta_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}} = O_p\left(\frac{1}{T}\right)$$

Consequently the asymptotic behavior of $\widehat{\gamma_2} = \widehat{\gamma_{1,\perp}}\widehat{\gamma_{21}}$ is affected by the behavior of both estimates. Nevertheless, by the second identification criterion γ_{21} must be replaced by $\Theta^{-1}_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}\gamma_{21}$ in order to ensure its definition[3]. The behavior of $\widehat{\gamma_2}$ is derived from :

$$\widehat{\gamma_2} - \gamma_2 = \left(\widehat{\gamma_{1,\perp}} - \gamma_{1,\perp}\Theta_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}\right)\widehat{\gamma_{21}} + \gamma_{1,\perp}\Theta_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}\left(\widehat{\gamma_{21}} - \Theta^{-1}_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}\gamma_{21}\right)$$

and the following result.

Proposition 28 *If* $(y_{i,t})_{i,t}$ *satisfies the VECM representation* $(5,6,7,8,9)$ *and* $\widehat{\gamma_{21}}$ *denotes the estimates of the second principal component problem as introduced above, then as* T *and* N *go to infinity, there exists a full rank matrix* $\Theta_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}$ *such that :*

$$\widehat{\gamma_{21}} - \Theta^{-1}_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}\gamma_{21} = \begin{pmatrix} \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{NT}} I_{s_2} & 0 \\ 0 & \frac{1}{T\sqrt{N}} I_{m-s_1-s_2} \end{pmatrix} O_p(1)$$

and $\widehat{\gamma_{1,\perp}} - \gamma_{1,\perp}\Theta_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}} = O_p\left(\frac{1}{T}\right)$ *where* $O_p(1)$ *represents a matrix with the appropriate dimensions which converges weakly .*

5.2 Test procedure

The test procedure we propose draws on the idea presented in Gregoir and Laroque (1994) and Andrade, Bruneau and Gregoir (2001). It is a two step procedure and consists of introducing in the VECM the estimates of the non-stationary terms and of testing for the non-significance of the associated coefficients. Let γ_3 be a $m \times (m - s_1 - s_2)$ matrix of full rank such that

$$\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix}$$

is a full rank $m \times m$ matrix and $\gamma_1'\gamma_3 = 0_{(s_1 \times (m-s_1-s_2))}$ and $\gamma_2'\gamma_3 = 0_{(s_2 \times (m-s_1-s_2))}$. We know that under the null hypothesis characterized by equation (5) and conditions **M1** and **M2**, $\gamma_3'\left(Sy_{i,t} - Sz_t^N\right)$ is a $(m - s_1 - s_2)$-dimensional I(1) process without cointegration

[3] $\gamma_{1,\perp}\Theta_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}\Theta^{-1}_{\widehat{\gamma_{1,\perp}},\gamma_{1,\perp}}\gamma_{21}$ is identified.

relationships and similarly $\begin{pmatrix} \gamma_2 & \gamma_3 \end{pmatrix}' S z_t^N$ is a $(m - s_1)$-dimensional I(1) process without cointegration relationships. If we introduce these terms in the regression

$$\begin{aligned} y_{i,t} &= \delta_1 \gamma_1' \left(S y_{i,t-1} - S z_{t-1}^N \right) + \delta_2 \gamma_2' \left(S y_{i,t-1} - S z_{t-1}^N \right) + \alpha \gamma_1' S z_{t-1}^N \\ &+ \sum_{j=1}^{p_1} \psi_j z_{t-j}^N + \sum_{k=1}^{p_2} \phi_k \left(y_{i,t-k} - z_{t-k}^N \right) + \delta_3 \gamma_3' \left(S y_{i,t-1} - S z_{t-1}^N \right) \\ &+ \begin{pmatrix} \alpha_2 & \alpha_3 \end{pmatrix} \begin{pmatrix} \gamma_2 & \gamma_3 \end{pmatrix}' S z_{t-1}^N + \varepsilon_t + \eta_{i,t} \end{aligned} \tag{14}$$

the set of matrices $(\delta_3, \alpha_2, \alpha_3)$ must not be significantly different from 0 and must superconverge towards 0. In practice, we do not know the matrices (γ_2, γ_3) and must replace the true value by some consistent estimate we can get from the principal component procedure. They may introduce nuisance parameters. An additional projection will be needed to get rid of them. We propose therefore to test for the rank of cointegration s_1 and s_2 by considering the following null hypotheses

$$H_{0,1}^c : \delta_3 = 0$$

and

$$H_{0,2}^c : \begin{pmatrix} \alpha_2 & \alpha_3 \end{pmatrix} = 0$$

in the above equation where $\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix}$ have been replaced by their estimates $\begin{pmatrix} \widehat{\gamma_1} & \widehat{\gamma_2} & \widehat{\gamma_3} \end{pmatrix}$ obtained in the principal component procedure. Estimation of the model can be done with GLS procedure. We suppose in a first step that Ω_ε and Ω_η are given.

Theorem 29 *Let $(y_{i,t})_{i,t}$ satisfies the VECM representation given by equation* (5) *under conditions **M1** and **M2** : $\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix}$ have been replaced by their estimates $\begin{pmatrix} \widehat{\gamma_1} & \widehat{\gamma_2} & \widehat{\gamma_3} \end{pmatrix}$ and*

$$\begin{pmatrix} \widehat{\delta_1} & \widehat{\delta_2} & \widehat{\alpha} & \left(\widehat{\psi_j}\right)_j & \left(\widehat{\phi_k}\right)_k & \widehat{\delta_3} & \widehat{\alpha_2} & \widehat{\alpha_3} \end{pmatrix}$$

be the GLS estimates in equation (14)*, let $\widehat{\alpha_\perp}$ be a $(m \times m - s_1)$ full rank matrix such that $\widehat{\alpha_\perp}'\widehat{\alpha} = 0$ and $\widehat{\delta_\perp}$ be a $(m \times m - s_1 - s_2)$ full rank matrix such that $\widehat{\delta_\perp}' \begin{pmatrix} \widehat{\delta_1} & \widehat{\delta_2} \end{pmatrix} = 0$, then when T goes to $+\infty$, the following convergence in distribution hold :*

$$\begin{aligned} & Trace \left\{ T \left(\widehat{\alpha_\perp}' \left(\Omega_\varepsilon + \frac{1}{N} \Omega_\eta \right) \widehat{\alpha_\perp} \right)^{-1} \widehat{\alpha_\perp}' \begin{pmatrix} \widehat{\alpha_2} & \widehat{\alpha_3} \end{pmatrix} S_{\widehat{\gamma_2}, \widehat{\gamma_3}} \begin{pmatrix} \widehat{\alpha_2} & \widehat{\alpha_3} \end{pmatrix}' \widehat{\alpha_\perp} \right\} \\ \Longrightarrow \; & Trace \left\{ \int dWW' \left(\int WW' \right)^{-1} \int WdW' \right\} \end{aligned}$$

where $S_{\widehat{\gamma_2}, \widehat{\gamma_3}} = \left[\frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} \widehat{\gamma_2} & \widehat{\gamma_3} \end{pmatrix}' S z_{t-1}^N S z_{t-1}^{N\prime} \begin{pmatrix} \widehat{\gamma_2} & \widehat{\gamma_3} \end{pmatrix} \right]$ and W is a standard $(m - s_1)$-dimensional Brownian motion and when sequentially T then N go to $+\infty$, the following convergence in distribution hold :

$$Trace \left\{ TN \left(\widehat{\delta_\perp}' \Omega_\eta \widehat{\delta_\perp} \right)^{-1} \widehat{\delta_\perp}' \widehat{\delta_3} S_{\widehat{\gamma_3}} \widehat{\delta_3}' \widehat{\delta_\perp} \right\} \Longrightarrow \chi_2 \left((m - s_1 - s_2)^2 \right)$$

where $S_{\widehat{\gamma_3}} = \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\gamma_3}' \left(S y_{i,t-1} - S z_{t-1}^N \right) \left(S y_{i,t-1} - S z_{t-1}^N \right)' \widehat{\gamma_3} \right]$.

In practice we do not know the matrices Ω_η and Ω_ε. First step estimates can be derived from the two following auxiliary regressions. An estimate of $\widehat{\Omega_\eta}$ can be obtained as follows

: run the ordinary least squares regression for $(i,t) \in \{1,..,N\} \times \{1,..,T\}$

$$\begin{aligned} y_{i,t} - z_t^N &= \delta_1 \gamma_1' \left(S y_{i,t-1} - S z_{t-1}^N \right) + \delta_2 \gamma_2' \left(S y_{i,t-1} - S z_{t-1}^N \right) \\ &\quad + \sum_{k=1}^{p_2} \phi_k \left(y_{i,t-k} - z_{t-k}^N \right) + w_{i,t} \end{aligned}$$

collect the residuals and compute

$$\widehat{\Omega_\eta} = \frac{N-1}{TN^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{w_{i,t}} \widehat{w_{i,t}}'$$

An estimate of $\widehat{\Omega_\varepsilon}$ can be obtained in a similar way : run the ordinary least squares regression for $t \in \{1,..,T\}$

$$z_t^N = \alpha \gamma_1' S z_{t-1}^N + \sum_{j=1}^{p_1} \psi_j z_{t-j}^N + w_t$$

and compute

$$\widehat{\Omega_\varepsilon} = \frac{1}{T} \sum_{t=1}^{T} \widehat{w_t} \widehat{w_t}'$$

In a second step, a GLS procedure can be used once we have noticed that

$$\left(I_N \otimes \Omega_\eta + J_N \otimes \Omega_\varepsilon \right)^{-1} = I_N \otimes \Omega_\eta^{-1} - J_N \otimes \left(N\Omega_\eta + \Omega_\eta \Omega_\varepsilon^{-1} \Omega_\eta \right)^{-1}$$

6 Conclusion

We introduce a family of dynamic models for panel data, structure of which is derived from the fact that there does not exist any relevant ordering between the time series of this panel data. When such an ordering exists, the modelers have to take it into account explicitly. In this set-up, we can decompose the multivariate time series into the sum of two uncorrelated processes : a common component and a idiosyncratic one. We then derive a VECM representation for these data that involves two sets of cointegration relationships for each type of components which may affect differently the changes of the multivariate time series. We then focus on a particular sub-family of interest when the number of units is not too large and the common component is approximated by the empirical average of the available units, which implies that the space of cointegration between the common component is a subspace of the cointegration vector space of the idiosyncratic ones. We then introduce test procedures to base decision on the integratedness and cointegratedness of the variables under study. The cointegration test of the idiosyncratic components is asymptotically χ_2 distributed while that of the common component has a Johansen's type asymptotic distribution.

7 Reference

Aldous (1980) Séminaire de Probabilité, Ecole d'Eté de Saint Flour, Springer Verlag

Andrade, P., C. Bruneau and S. Gregoir (2001) "Testing for the cointegration rank when some cointegrating directions are changing" forthcoming in *Journal of Econometrics*

Bai, J. and S. Ng (2002) "Determining the Number of Factors in Approximate Factor Models" *Econometrica*, 70 : 191-221

Bai, J. and S. Ng (2003) "A PANIC Attack on Unit roots and Cointegration", mimeo forthcoming in *Econometrica*

Baltagi B.H. and C. Kao (2001) "Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey", mimeo

Banerjee, A. (1999) "Panel Data Unit Roots and Cointegration: an Overview" *Oxford Bulletin of Economics and Statistics*, 61: 607-629

Banerjee, A. , M. Marcellino and C. Osbat (2001) "Testing for PPP: Should we use Panel Methods ?" mimeo, European University Institute

Boumahdi, R. and A. Thomas (1991) "Testing for Unit Roots Using Panel Data", *Economics Letters*, 37: 77-79

Chang Y. and W. Song (2002) "Panel Unit Root Test in the Presence of Cross-Sectional Dependency and Heterogeneity", mimeo, Rice University

Chow Y.S. and H. Teicher (1997) *Probability Theory*, Springer Verlag

Dickey, D.. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root" *Journal of the American Statistical Association*, 74 : 427-431

de Finetti, B. (1937) "La prévision; ses lois logiques, ses sources subjectives", *Annales de l'Institut Henri Poincaré*, 7: 1-68

Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000) "The generalized Dynamic Factor Model: Identification and Estimation", *Review of Economics and Statistics*, 82 : 540-554

Hadri, K. (2001) "Testing for Stationarity in Heterogenous Panel Data" *Econometrics Journal, 1:1-14*

Hadri, K. (1999) "Testing the null hypothesis of Stationarity Against the Alternative of a Unit Root in Panel Data with Serially Correlated Errors", mimeo

Hansen, B.(1995) "Rethinking the univariate approach of unit root testing" *Econometric Theory,* 11, 1148-1171

Im, K., M.H. Pesaran and Y. Shin (1997) "Testing for unit roots in heterogeneous panels", mimeo, Cambridge University

Johansen S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models,* Oxford University Press

Kallenberg O. (1991) Springer Verlag

Gregoir S. and **G. Laroque (1993)**, 'Multivariate time series: A polynomial error correction representation theorem', *Econometric Theory*, 9, 329-342.

Gregoir S. and **G. Laroque (1994)**, "Polynomial cointegration estimation and test ", *Journal of Econometrics*, 63, 183-214.

Gregoir S. (1999), 'Multivariate Time Series with Various Hidden Unit Roots: Part I : Integral Operator Algebra and Representation Theory', *Econometric Theory,*15, 435-468.

Levin, A , C.F. Lin and J. Chu (2002) "Unit root tests in panel data : Asymptotic and Finite Sample Properties" *Journal of Econometrics*

Maddala, G.S. (1992) *Introduction to Econometrics* MacMillan

Maddala, G.S. (1999) "On the use of panel data methods with cross country data" *Annales d'Economie et de Statistique*, 55-56 : 429-448

Magnus, J. (1982) "Multivariate Error Components Analysis of Linear and Nonlinear Regression Models by Maximum Likelihood", *Journal of Econometrics*, 19, 239-285

Moon, R. and B. Perron (2002) "Testing for a Unit Root in Panels with Dynamic Factors", mimeo, Université de Montréal

Nabeya S. and K. Tanaka (1988) "Asymptotic Theory of a Test for the Constancy of Regression Coefficients against the Random Walk Alternative", *Annals of Statistics*, 16, 218-235

Pedroni, P (2001) "Purchasing Power Parity Tests in Cointegrated Panels," *The Review of Economics and Statistics*, November, 83, 4.

Phillips, P.C.B. (1988) "Weak Convergence to the Matrix Stochastic Integral $\int BdB$." *Journal of Multivariate Analysis*, 24(2) .

Phillips, P.C.B. and S. Durlauf (1987) "Multiple Time Series Regression with Integrated Processes." *Review of Economic Studies*, 53

Phillips, P.C.B. and P. Perron (1988) "Testing for a unit root in time series regression" *Biometrika*, 75 (2), 335-346

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

Phillips, P.C.B. and H. Moon (1999) "Nonstationary Panel Data analysis: an Overview of Some Recent Developments", *Econometric Reviews*

Appendix

Proof of Lemma 2 : Let $z_t^N = \frac{1}{N}\sum_{j=1}^N y_{j,t}$, we have $\forall t \in \mathbb{Z}$, when $M > N$

$$V\left(z_t^N - z_t^M\right) = \frac{M-N}{MN}\left(\Gamma(0) - \Xi(0)\right) \ll \frac{1}{N}\left(\Gamma(0) - \Xi(0)\right)$$

so that $\left(\max_{k=1,\ldots,m}\left\|z_{k,t}^N - z_{k,t}^M\right\|_{L^2}\right)_{t\in\mathbb{Z}}$ is a Cauchy sequence in an Hilbertian space. There exists a multivariate process $(z_t)_{t\in\mathbb{Z}}$ limit of this sequence, such that

$$\lim_{N\longrightarrow+\infty} V\left(z_t^N - z_t\right) = 0 \Longleftrightarrow \lim_{N\longrightarrow+\infty} z_t^N = z_t \quad \left(L^2\right)$$

On the one hand, we have

$$\max_{k,l}\left|cov\left(z_t - \frac{1}{N}\sum_{j=1}^N y_{j,t}, z_{t+h}\right)_{k,l}\right| \leq \max_k \left\|z_{k,t} - \frac{1}{N}\sum_{j=1}^N y_{k,j,t}\right\|_{L^2}^{1/2} \max_k \left\|z_{k,t}\right\|_{L^2}^{1/2}$$

and

$$\max_{k,l}\left|cov\left(z_{t+h} - \frac{1}{N}\sum_{j=1}^N y_{j,t+h}, \frac{1}{N}\sum_{j=1}^N y_{j,t}\right)_{k,l}\right| \leq$$
$$\max_k \left\|z_{k,t+h} - \frac{1}{N}\sum_{j=1}^N y_{k,j,t+h}\right\|_{L^2}^{1/2} \max_k \left\|\frac{1}{N}\sum_{j=1}^N y_{k,j,t}\right\|_{L^2}^{1/2}$$

so that, as $\max_j \left\|z_{k,t}\right\|_{L^2}^{1/2}$ and $\max_j \left\|\frac{1}{N}\sum_{j=1}^N y_{k,j,t}\right\|_{L^2}^{1/2}$ are bounded :

$$\begin{aligned}
cov\left(z_t, z_{t+h}\right) &= \lim_{N\longrightarrow+\infty} cov\left(\frac{1}{N}\sum_{j=1}^N y_{j,t}, z_{t+h}\right) \\
&= \lim_{N\longrightarrow+\infty} cov\left(\frac{1}{N}\sum_{j=1}^N y_{j,t}, \frac{1}{N}\sum_{j=1}^N y_{j,t+h}\right) \\
&= \lim_{N\longrightarrow+\infty} \frac{1}{N}\Gamma(h) + \left(1 - \frac{1}{N}\right)\Xi(h) \\
&= \Xi(h)
\end{aligned}$$

On the other hand,

$$\max_{k,l}\left|cov\left(z_{t+h} - \frac{1}{N}\sum_{i=1}^N y_{i,t+h}, y_{j,t}\right)_{k,l}\right| \leq$$
$$\max_k \left\|z_{k,t+h} - \frac{1}{N}\sum_{j=1}^N y_{k,j,t+h}\right\|_{L^2}^{1/2} \max_k \left\|y_{k,j,t}\right\|_{L^2}^{1/2}$$

whence

$$\begin{aligned}
cov\left(z_{t+h}, y_{j,t}\right) &= \lim_{N\longrightarrow+\infty} cov\left(\frac{1}{N}\sum_{i=1}^N y_{i,t+h}, y_{j,t}\right) \\
&= \lim_{N\longrightarrow+\infty} \left(1 - \frac{1}{N}\right)\Xi(h)' + \frac{1}{N}\Gamma(h)' \\
&= \Xi(h)'
\end{aligned}$$

At last,

$$\begin{aligned} cov\left(y_{i,t}-z_t, y_{i,t+h}-z_{t+h}\right) &= \delta_{i,j}\Gamma\left(h\right)+\left(1-\delta_{i,j}\right)\Xi\left(h\right)-\Xi\left(h\right) \\ &= \delta_{i,j}\left(\Gamma\left(h\right)-\Xi\left(h\right)\right) \end{aligned}$$

Proof of Lemma 7 : The proof is similar to the preceding one in which $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ is replaced by $\left(dvar\left(y_{i,t}\right)^{-1/2}y_{i,t}\right)_{i\in\mathbb{N},t\in\mathbb{Z}}$.

Proof of Theorem 10: The result is derived from Gregoir and Laroque (1993) Representation Theorem applied to the process $(y_{i,t}-z_t)_{t\in\mathbb{Z}}$ and $(z_t)_{t\in\mathbb{Z}}$. The constraint on the constant terms in the idiosyncratic cointegration relationships comes from the fact that by construction for all date t, $\lim_{N\longrightarrow+\infty}\frac{1}{N}e_N'\otimes I_m\left(y_t-z_t\right)=0$ (where e_N is a $(N\times 1)$ vector of ones).

Proof of Proposition 15 : We stack up at date t all the unit vectors $y_{i,t}$ and innovations $\eta_{i,t}$ in a $(Nm\times 1)$ vector we denote y_t and η_t. We have

$$y_t=e_N\otimes C\left(L\right)\varepsilon_t+I_N\otimes D\left(L\right)\eta_t$$

where e_N is a $(N\times 1)$ vector of 1 such that $J_N=e_Ne_N'$. Let $\gamma=\begin{pmatrix}\gamma_1 & \gamma_2\end{pmatrix}$ be a $(m\times s)$ full rank matrix such that $\gamma' D\left(1\right)=0_{(s\times m)}$ and $\gamma_1' C\left(1\right)=0_{(s_1\times m)}$, we denote $\gamma_\perp$ the $(m\times(m-s))$ full rank matrix such that $\gamma_\perp'\gamma=0_{((m-s)\times s)}$. We start from

$$\left(I_N-\frac{1}{N}J_N\right)\otimes I_m y_t=\left(I_N-\frac{1}{N}J_N\right)\otimes D\left(L\right)\eta_t$$

where the left-hand side can be read as

$$\begin{aligned}\left(I_N-\frac{1}{N}J_N\right)\otimes I_m y_t &= y_t-e_N\otimes\frac{1}{N}\sum_{j=1}^{N}y_{j,t} \\ &= y_t-e_N\otimes z_t^N\end{aligned}$$

and the right hand one

$$\begin{aligned}\left(I_N-\frac{1}{N}J_N\right)\otimes D\left(L\right)\eta_t &= I_N\otimes D\left(L\right)\eta_t-e_N\otimes D\left(L\right)\eta_t^N \\ &= I_N\otimes D\left(L\right)\left(\eta_t-e_N\otimes\eta_t^N\right)\end{aligned}$$

with $\eta_t^N=\frac{1}{N}\sum_{j=1}^{N}\eta_{j,t}$, so that

$$I_N\otimes\begin{pmatrix}\gamma'\\ \gamma_\perp'\end{pmatrix}\left(y_t-e_N\otimes z_t^N\right)=I_N\otimes\begin{pmatrix}\gamma'\\ \gamma_\perp'\end{pmatrix}\left(D\left(1\right)+\left(1-L\right)\widetilde{D}\left(L\right)\right)\left(\eta_t-e_N\otimes\eta_t^N\right)$$

$$I_N\otimes\begin{pmatrix}\gamma'\left(1-L\right)S\\ \gamma_\perp'\end{pmatrix}\left(y_t-e_N\otimes z_t^N\right)=I_N\otimes\begin{pmatrix}\gamma'\left(1-L\right)\widetilde{D}\left(L\right)\\ \gamma_\perp' D\left(L\right)\end{pmatrix}\left(\eta_t-e_N\otimes\eta_t^N\right)$$

$$I_N\otimes\begin{pmatrix}\left(1-L\right)I_s & 0\\ 0 & I_{m-s}\end{pmatrix}\begin{pmatrix}\gamma' S\\ \gamma_\perp'\end{pmatrix}\left(y_t-e_N\otimes z_t^N\right)=$$

$$I_N\otimes\begin{pmatrix}\left(1-L\right)I_s & 0\\ 0 & I_{m-s}\end{pmatrix}\begin{pmatrix}\gamma'\widetilde{D}\left(L\right)\\ \gamma_\perp' D\left(L\right)\end{pmatrix}\left(\eta_t-e_N\otimes\eta_t^N\right)$$

and

$$I_N \otimes \gamma' \left(Sy_t - e_N \otimes Sz_t^N\right) - \left(I_N \otimes \gamma' \widetilde{D}(L)\right) \left(\eta_t - e_N \otimes \eta_t^N\right)$$

is in the kernel of $(1-L)\left(I_N - \frac{1}{N}J_N\right) \otimes I_s$ that is composed of $(Ns \times 1)$ constant processes whose sum over the units is equal to 0. Therefore

$$I_N \otimes \gamma' \left(Sy_t - e_N \otimes Sz_t^N\right) + \widetilde{\mu} = \left(I_N \otimes \gamma' \widetilde{D}(L)\right) \left(\eta_t - e_N \otimes \eta_t^N\right)$$

is a covariance stationary process. On the other hand

$$\begin{pmatrix} \gamma' \widetilde{D}(1) \\ \gamma'_\perp D(1) \end{pmatrix}$$

is a full rank matrix since

$$\begin{aligned} \det \begin{pmatrix} \gamma' \\ \gamma'_\perp \end{pmatrix} D(L) &= (1-L)^s d(L) \det \begin{pmatrix} \gamma' \\ \gamma'_\perp \end{pmatrix} \\ &= \det \begin{pmatrix} (1-L) I_s & 0 \\ 0 & I_{m-s} \end{pmatrix} \begin{pmatrix} \gamma' \widetilde{D}(L) \\ \gamma'_\perp D(L) \end{pmatrix} \\ &= (1-L)^s \det \begin{pmatrix} \gamma' \widetilde{D}(L) \\ \gamma'_\perp D(L) \end{pmatrix} \end{aligned}$$

and by assumption $d(L)$ has all its roots outside the unit circle. It follows that $\widehat{D}(L) = \begin{pmatrix} \gamma' \widetilde{D}(L) \\ \gamma'_\perp D(L) \end{pmatrix}$ is invertible. Simple algebra (Gregoir and Laroque (1993)) gives there exist a full rank $(m \times s)$ matrix $\delta = \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix}$ and a $(m \times m)$ polynomial matrix $\phi(L)$ that is related to the inverse of the above matrix such that

$$\begin{aligned} y_t - e_N \otimes z_t^N &= I_N \otimes \delta \left(\left(I_N \otimes \gamma'\right)\left(Sy_{t-1} - e_N \otimes Sz_{t-1}^N\right) + \mu\right) \\ &\quad + I_N \otimes \phi(L) \left(y_{t-1} - e_N \otimes z_{t-1}^N\right) + \eta_t - e_N \otimes \eta_t^N \end{aligned} \tag{15}$$

We turn to

$$\begin{aligned} \frac{1}{N} e'_N \otimes I_m y_t &= C(L)\varepsilon_t + \frac{1}{N} e'_N \otimes D(L) \eta_t \\ z_t^N &= C(L)\varepsilon_t + D(L)\eta_t^N \end{aligned}$$

Its spectral density matrix is under our assumptions :

$$C\left(e^{-i\omega}\right)\Omega_\varepsilon C\left(e^{i\omega}\right)' + \frac{1}{N} D\left(e^{-i\omega}\right)\Omega_\eta D\left(e^{i\omega}\right)'$$

Following the same reasoning that the one introduced subsection 3.1 (case c), we conclude that the only complex vectors v in the kernel of this matrix are those that are simultaneously in the kernel of $C\left(e^{-i\omega}\right)\Omega_\varepsilon C\left(e^{i\omega}\right)'$ and of $D\left(e^{-i\omega}\right)\Omega_\eta D\left(e^{i\omega}\right)'$ since any vector in its kernel is such that

$$0 \leq v' C\left(e^{-i\omega}\right)\Omega_\varepsilon C\left(e^{i\omega}\right)' \overline{v} = -\frac{1}{N} v' D\left(e^{-i\omega}\right)\Omega_\eta D\left(e^{i\omega}\right)' \overline{v} \leq 0$$

Under assumptions, these matrices are of full rank for all the frequencies except $\omega = 0$ and in this case, those directions are spanned by γ_1. We have

$$\begin{aligned} \begin{pmatrix} \gamma'_1 \\ \gamma'_2 \\ \gamma'_\perp \end{pmatrix} z_t^N &= \begin{pmatrix} \gamma'_1 \\ \gamma'_2 \\ \gamma'_\perp \end{pmatrix} \left(C(1) + (1-L)\widetilde{C}(L)\right)\varepsilon_t \\ &\quad + \begin{pmatrix} \gamma'_1 \\ \gamma'_2 \\ \gamma'_\perp \end{pmatrix} \left(D(1) + (1-L)\widetilde{D}(L)\right)\eta_t^N \end{aligned}$$

so that

$$\begin{pmatrix} I_{s_1}(1-L) & 0 \\ 0 & I_{m-s_1} \end{pmatrix} \begin{pmatrix} \gamma_1' S \\ \gamma_2' \\ \gamma_\perp' \end{pmatrix} z_t^N = \begin{pmatrix} I_{s_1}(1-L) & 0 \\ 0 & I_{m-s_1} \end{pmatrix} \begin{pmatrix} \gamma_1' \widetilde{C}(L) \\ \gamma_2' C(L) \\ \gamma_\perp' C(L) \end{pmatrix} \varepsilon_t + \begin{pmatrix} I_{s_1}(1-L) & 0 \\ 0 & I_{m-s_1} \end{pmatrix} \begin{pmatrix} \gamma_1' \widetilde{D}(L) \\ \gamma_2' D(L) \\ \gamma_\perp' D(L) \end{pmatrix} \eta_t^N$$

that allows us to say that

$$\gamma_1' S z_t^N + \mu = \gamma_1' \left(\widetilde{C}(L) \varepsilon_t + \widetilde{D}(L) \eta_t^N \right)$$

is a covariance stationary process. Therefore

$$\begin{pmatrix} \gamma_1' S z_t^N + \mu \\ \gamma_2' z_t^N \\ \gamma_\perp' z_t^N \end{pmatrix}$$

is a covariance stationary process whose determinant of the polynomial matrix associated to its Wold representation has no unit root. Using similar arguments to what was done in the study of the idiosyncratic components, there exists a multivariate white noise $(v_{1t})_{t\in\mathbb{Z}}$, a full rank $(m \times s_1)$ matrix α and a $(m \times m)$ polynomial matrix $\psi(L)$ such that

$$z_t^N = \alpha \left(\gamma_1' S z_{t-1}^N + \mu \right) + \psi(L) z_{t-1}^N + v_{1t} \tag{16}$$

Proof of proposition 16: We start from (5) and sum over the unit dimension the set of equations at a given date and divide by N. We get

$$z_t^N = \alpha \left(\gamma_1' S z_{t-1}^N + \mu \right) + \sum_{j=1}^{p_1} \psi_j z_{t-j}^N + \varepsilon_t + \frac{1}{N} \sum_{i=1}^{N} \eta_{i,t} \tag{17}$$

This equation is the standard equation of a VECM and from Johansen (1995) we get that there exists a polynomial matrix

$$C(L) = C_1 + (1-L)\widetilde{C}(L)$$

where $\sum_{k=0}^{+\infty} \sqrt[2]{Tr\left(C_k C_k'\right)} < +\infty$, $\sum_{k=0}^{+\infty} k \sqrt[2]{Tr\left(C_k C_k'\right)} < +\infty$ and such that

$$\begin{aligned} S z_t^N &= C_1 \sum_{\tau=1}^{t} \left(\varepsilon_\tau + \frac{1}{N} \sum_{i=1}^{N} \eta_{i,\tau} \right) + \\ &\quad \widetilde{C}(L) \left(\alpha\mu + \varepsilon_t + \frac{1}{N} \sum_{i=1}^{N} \eta_{i,t} \right) + \gamma_{1,\perp} \left(\gamma_{1,\perp}' \gamma_{1,\perp} \right)^{-1} \gamma_{1,\perp}' S z_0^N \end{aligned}$$

or

$$z_t^N = C(L) \left(\varepsilon_t + \frac{1}{N} \sum_{i=1}^{N} \eta_{i,t} \right)$$

If we now substract (17) from (5), we get that

$$\begin{aligned} y_{i,t} - z_t^N &= \delta_1 \gamma_1' \left(S y_{i,t-1} - S z_{t-1}^N \right) + \delta_2 \gamma_2' \left(S y_{i,t-1} - S z_{t-1}^N \right) \\ &\quad + \sum_{j=1}^{p_2} \phi_j \left(y_{i,t-j} - z_{t-j}^N \right) + \eta_{i,t} - \frac{1}{N} \sum_{k=1}^{N} \eta_{k,t} \end{aligned}$$

This is again a standard equation of a VECM. There exists a polynomial matrix $D(L)$

$$D(L) = D_1 + (1-L)\widetilde{D}(L)$$

where $\sum_{k=0}^{+\infty} \sqrt[2]{Tr\left(D_k D_k'\right)} < +\infty$, $\sum_{k=0}^{+\infty} k \sqrt[2]{Tr\left(D_k D_k'\right)} < +\infty$ and such that

$$\begin{aligned} Sy_{i,t} - Sz_t^N &= D_1 \sum_{\tau=1}^{t} \left(\eta_{i,\tau} - \frac{1}{N} \sum_{k=1}^{N} \eta_{k,\tau} \right) \\ &+ \widetilde{D}(L) \left(\eta_{i,t} - \frac{1}{N} \sum_{k=1}^{N} \eta_{k,t} \right) + \gamma_\perp \left(\gamma_\perp' \gamma_\perp \right)^{-1} \gamma_\perp' \left(Sy_{i,0} - Sz_0^N \right) \end{aligned}$$

or

$$y_{i,t} - z_t^N = D(L) \left(\eta_{i,t} - \frac{1}{N} \sum_{k=1}^{N} \eta_{k,t} \right)$$

We thus conclude that

$$y_{i,t} = C(L) \left(\varepsilon_t + \frac{1}{N} \sum_{i=1}^{N} \eta_{i,t} \right) + D(L) \left(\eta_{i,t} - \frac{1}{N} \sum_{k=1}^{N} \eta_{k,t} \right)$$

Proof of Proposition 18: We start from the OLS estimator when the DGP is (I)

$$\begin{aligned} \widehat{\rho}_{OLS,N} &= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t} Sy_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(Sy_{i,t-1}\right)^2} \\ \widehat{\rho}_{OLS,N} &= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{2} \left(y_{i,t} + Sy_{i,t-1}\right)^2 - \frac{1}{2} \left(Sy_{i,t-1}\right)^2 - \frac{1}{2} y_{i,t}^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(Sy_{i,t-1}\right)^2} \\ T\widehat{\rho}_{OLS,N} &= \frac{1}{2} \frac{A_{OLS}}{B_{OLS}} \end{aligned}$$

with

$$\begin{aligned} A_{OLS} &= \frac{1}{TN} \sum_{i=1}^{N} \left(\left(Sy_{i,T}\right)^2 - \left(Sy_{i,0}\right)^2 - \sum_{t=1}^{T} y_{i,t}^2 \right) \\ B_{OLS} &= \frac{1}{T^2 N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Sy_{i,t-1}\right)^2 \end{aligned}$$

where

$$\begin{aligned} y_{i,t} &= u_t + v_{i,t} \\ Sy_{i,t} &= Su_t + Sv_{i,t} \\ Sy_{i,0} &= 0 \end{aligned}$$

so that

$$\begin{aligned}
A_{OLS} &= \frac{1}{T}\left((Su_T)^2 + \frac{1}{N}\sum_{i=1}^{N}(Sv_{i,T})^2 + 2Su_T\frac{1}{N}\sum_{i=1}^{N}Sv_{i,T}\right) \\
&\quad -\frac{1}{T}\left(\sum_{t=1}^{T}u_t^2 + \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{i,t}^2 + 2\sum_{t=1}^{T}u_t\frac{1}{N}\left(\sum_{i=1}^{N}v_{i,t}\right)\right) \\
&= \left(\frac{1}{\sqrt{T}}Su_T\right)^2 + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}Sv_{i,T}\right)^2 + 2\frac{1}{\sqrt{T}}Su_T\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}Sv_{i,T} \\
&\quad -\frac{1}{T}\left(\sum_{t=1}^{T}u_t^2 + \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{i,t}^2 + 2\sum_{t=1}^{T}u_t\frac{1}{N}\left(\sum_{i=1}^{N}v_{i,t}\right)\right)
\end{aligned}$$

and

$$B_{OLS} = \frac{1}{T^2}\sum_{t=1}^{T}(Su_{t-1})^2 + \frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}(Sv_{i,t})^2 + 2\frac{1}{T^2}\sum_{t=1}^{T}\left(Su_{t-1}\frac{1}{N}\sum_{i=1}^{N}Sv_{i,t}\right)$$

When T tends toward $+\infty$, under assumptions **A-D**, we have the following joint weak convergences (see *inter alios* Phillips and Durlauf (1987)) for all i

$$\begin{aligned}
\frac{1}{\sqrt{T}}Su_T &\Longrightarrow \sigma_\varepsilon c(1)\,W(1) \\
\frac{1}{\sqrt{T}}Sv_{i,T} &\Longrightarrow \sigma_\eta d(1)\,B_i(1) \\
\frac{1}{T^2}\sum_{t=1}^{T}(Su_{t-1})^2 &\Longrightarrow \sigma_\varepsilon^2 c(1)^2\int_0^1 W(s)^2\,ds \\
\frac{1}{T^2}\sum_{t=1}^{T}(Sv_{i,t-1})^2 &\Longrightarrow \sigma_\eta^2 d(1)^2\int_0^1 B_i(s)^2\,ds \\
\frac{1}{T^2}\sum_{t=1}^{T}Sv_{i,t-1}Su_{t-1} &\Longrightarrow \sigma_\varepsilon c(1)\,\sigma_\eta d(1)\int_0^1 B_i(s)\,W(s)\,ds
\end{aligned}$$

and also $\frac{1}{T}\sum_{t=1}^{T}u_t^2 \longrightarrow_P \gamma_u(0)$, $\frac{1}{T}\sum_{t=1}^{T}v_{i,t}^2 \longrightarrow_P \gamma_v(0)$ and $\frac{1}{T}\sum_{t=1}^{T}u_tv_{i,t} \longrightarrow_P 0$. We conclude that

$$T\widehat{\rho}_{OLS,N} \Longrightarrow$$

$$\frac{1}{2}\frac{\sigma_\varepsilon^2 c(1)^2 W(1)^2 + \sigma_\eta^2 d(1)^2\frac{1}{N}\sum_{i=1}^{N}B_i(1)^2 + 2\sigma_\varepsilon c(1)\,\sigma_\eta d(1)\,W(1)\frac{1}{N}\sum_{i=1}^{N}B_i(1) - \gamma_u(0) - \gamma_v(0)}{\sigma_\varepsilon^2 c(1)^2\int_0^1 W(s)^2\,ds + \sigma_\eta^2 d(1)^2\frac{1}{N}\sum_{i=1}^{N}\int_0^1 B_i(s)^2\,ds + 2\sigma_\varepsilon c(1)\,\sigma_\eta d(1)\int_0^1 W(s)\frac{1}{N}\sum_{i=1}^{N}B_i(s)\,ds}$$

Then when N tends toward $+\infty$, using the independence and identique distribution of the random variables $\left(B_i(s)_{s\in[0,1]}\right)_{i\in\mathbb{N}}$ and $\left(W(s)\,B_i(s)_{s\in[0,1]}\right)_{i\in\mathbb{N}}$, we get that

$$\begin{aligned}
\lim_{N\longrightarrow+\infty} T\widehat{\rho}_{OLS,N} &\Longrightarrow \frac{1}{2}\frac{\sigma_\varepsilon^2 c(1)^2 W(1)^2 + \sigma_\eta^2 d(1)^2 - \gamma_u(0) - \gamma_v(0)}{\sigma_\varepsilon^2 c(1)^2\int_0^1 W(s)^2\,ds + \frac{1}{2}\sigma_\eta^2 d(1)^2} \\
&\Longrightarrow \frac{\int_0^1 W(s)\,ds + \frac{1}{2}\frac{\sigma_\eta^2 d(1)^2 - \gamma_v(0)}{\sigma_\varepsilon^2 c(1)^2} + \frac{1}{2}\left(1 - \frac{\gamma_u(0)}{\sigma_\varepsilon^2 c(1)^2}\right)}{\int_0^1 W(s)^2\,ds + \frac{\sigma_\eta^2 d(1)^2}{2\sigma_\varepsilon^2 c(1)^2}}
\end{aligned}$$

Under the additional assumption **E** that the probability distribution of the processes $(\eta_{i,t})_{t\in\mathbb{Z}}$ is the same, $\left(\frac{1}{\sqrt{T}}Sv_{i,T}\right)_{i\in\mathbb{N}}$ is a sequence of independent and identiquely distributed random variables such that $E\left(\frac{1}{\sqrt{T}}Sv_{i,T}\right)^2<+\infty$ and independent on i, applying Corollary 2 in Chow and Teicher (Chapter 5.2), we have $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}Sv_{i,T}\longrightarrow_{a.c.}0$ and $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}Sv_{i,T}\right)^2\longrightarrow_{a.c.}E\left(\frac{1}{\sqrt{T}}Sv_{i,T}\right)^2$

$$\lim_{N\longrightarrow+\infty}T\widehat{\rho}_{OLS,N}=\frac{1}{2}\frac{\left(\frac{1}{\sqrt{T}}Su_T\right)^2+E\left(\frac{1}{\sqrt{T}}Sv_{i,T}\right)^2-\frac{1}{T}\sum_{t=1}^{T}u_t^2-E\frac{1}{T}\sum_{t=1}^{T}v_{i,t}^2}{\frac{1}{T^2}\sum_{t=1}^{T}(Su_{t-1})^2+\frac{1}{T}\sum_{t=1}^{T}E\left(\frac{1}{\sqrt{T}}Sv_{i,t}\right)^2}$$

where

$$\begin{aligned}E\left(\frac{1}{\sqrt{T}}Sv_{i,t}\right)^2&=\frac{1}{T}E\left(d(1)\sum_{\tau=1}^{t}\eta_{i,\tau}+\widetilde{d}(L)\eta_{i,t}\right)^2\\&=\frac{1}{T}d(1)^2t\sigma_\eta^2+\frac{1}{T}d(1)\sum_{\tau=1}^{t}E\left(\eta_{i,\tau}\widetilde{d}(L)\eta_{i,t}\right)+\frac{1}{T}E\left(\widetilde{d}(L)\eta_{i,t}\right)^2\\&=\frac{1}{T}d(1)^2t\sigma_\eta^2+\frac{\sigma_\eta^2}{T}d(1)\sum_{\tau=1}^{t}\widetilde{d}_{t-\tau}+\frac{1}{T}E\left(\widetilde{d}(L)\eta_{i,t}\right)^2\end{aligned}$$

From direct algebra, it follows that

$$\lim_{T\longrightarrow+\infty}\lim_{N\longrightarrow+\infty}T\widehat{\rho}_{OLS,N}\Longrightarrow\frac{1}{2}\frac{\sigma_\varepsilon^2c(1)^2W(1)^2+\sigma_\eta^2d(1)^2-\gamma_u(0)-\gamma_v(0)}{\sigma_\varepsilon^2c(1)^2\int_0^1W(s)^2ds+\frac{1}{2}\sigma_\eta^2d(1)^2}$$

By introducing estimates of the nuisance parameters in the above test statistic in the numerator and denominator to get asymptotically nuisance parameter free statistic, we obtain a test statistic that follows asymptotically a Dickey-Fuller distribution. The same arguments are used when the DGPs are (II) or (III).

Proof of Proposition 19: When the DGP is (I) and $(u_t)_{t\in\mathbb{Z}}=(\varepsilon_t)_{t\in\mathbb{Z}}$ and $(v_{i,t})_{t\in\mathbb{Z}}=(\eta_{i,t})_{t\in\mathbb{Z}}$, if we denote v_t the $(N\times1)$ vector obtained in stacking up all the error terms $v_{i,t}$ at date t,we have

$$V(e_N\otimes\varepsilon_t+v_t)=\sigma_\varepsilon^2J_N+\sigma_\eta^2I_N$$

and

$$\begin{aligned}\left(\sigma_\varepsilon^2J_N+\sigma_\eta^2I_N\right)^{-1}&=I_N-\frac{\sigma_\varepsilon^2}{N\sigma_\varepsilon^2+\sigma_\eta^2}J_N\\&=I_N-\frac{1}{N+\omega}J_N\end{aligned}$$

with $\omega=\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$. The GLS estimate is given by

$$\begin{aligned}\widehat{\rho}_{GLS,N}&=\left(\sum_{t=1}^{T}Sy_{t-1}'\left(I_N-\frac{1}{N+\omega}J_N\right)Sy_{t-1}\right)^{-1}\sum_{t=1}^{T}Sy_{t-1}'\left(I_N-\frac{1}{N+\omega}J_N\right)y_t\\\widehat{\rho}_{GLS,N}&=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t}Sy_{i,t-1}-\frac{1}{\omega+N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}y_{i,t}\right)\left(\sum_{i=1}^{N}Sy_{i,t-1}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(Sy_{i,t-1})^2-\frac{1}{\omega+N}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}Sy_{i,t-1}\right)^2}\\T\widehat{\rho}_{GLS,N}&=\frac{1}{2}\frac{A_{OLS}-C_{GLS}}{B_{OLS}-D_{GLS}}\end{aligned}$$

with

$$
\begin{aligned}
C_{GLS} &= \frac{N}{N+\omega}\left(\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}Sy_{i,T}\right)^2-\left(\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}Sy_{i,0}\right)^2-\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}y_{i,t}\right)^2\right)\\
D_{GLS} &= \frac{N}{N+\omega}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}Sy_{i,t}\right)^2
\end{aligned}
$$

When T tends toward $+\infty$, under assumptions **A-D**, the following joint convergence holds

$$
\begin{aligned}
A_{OLS}-C_{GLS} \Longrightarrow\ & \frac{\omega}{\omega+N}\left(\sigma_\varepsilon^2\left(W(1)^2-1\right)+2\sigma_\varepsilon\sigma_\eta W(1)\frac{1}{N}\sum_{i=1}^{N}B_i(1)\right)\\
& +\sigma_\eta^2\left(\frac{1}{N}\sum_{i=1}^{N}\left(B_i(1)^2-1\right)-\frac{N}{\omega+N}\left(\frac{1}{N}\sum_{i=1}^{N}B_i(1)\right)^2\right)
\end{aligned}
$$

and

$$
\begin{aligned}
B_{OLS}-D_{GLS} \Longrightarrow\ & \frac{\omega}{\omega+N}\left(\sigma_\varepsilon^2\int_0^1 W(s)^2\,ds+\sigma_\varepsilon\sigma_\eta\int_0^1 W(s)\left(\frac{1}{N}\sum_{i=1}^{N}B_i(s)\right)ds\right)\\
& +\sigma_\eta^2\left(\frac{1}{N}\sum_{i=1}^{N}\int_0^1 B_i(s)^2\,ds-\frac{N}{\omega+N}\int_0^1\left(\frac{1}{N}\sum_{i=1}^{N}B_i(s)\right)^2 ds\right)
\end{aligned}
$$

When N tends toward $+\infty$, $T\widehat{\rho}_{GLS,N}$ then tends to 0 as the numerator tends to 0 and the denominator to $\frac{\sigma_\eta^2}{2}$(see for instance Levin, Lin and Chu (2002)). The first part of the two above formulas multiplied $\sqrt{N}$ is an $O_p\left(\frac{1}{\sqrt{N}}\right)$ and also tends to 0 which eliminates the common component persistence. We can use a CLT for the set of i.i.d. random variables $(B_i(1))_{i\in\mathbb{N}}$ that gives

$$
\sqrt{N}\begin{pmatrix}\frac{1}{N}\sum_{i=1}^{N}B_i(1)\\ \frac{1}{N}\sum_{i=1}^{N}B_i(1)^2-1\end{pmatrix}\Longrightarrow\mathcal{N}\left(0,\begin{pmatrix}1&0\\0&2\end{pmatrix}\right)
$$

and

$$
\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(B_i(1)^2-1\right)-\left(\frac{1}{N}\sum_{i=1}^{N}B_i(1)\right)^2\right)\Longrightarrow\mathcal{N}(0,2)
$$

We notice furthermore, that the parameter ω does not play any role in the limit. If we first let N goes to $+\infty$, with the additional assumption **E** , we obtain using Corollary 2 in Chow and Teicher (Chapter 5.2) that

$$
\begin{aligned}
\lim_{N\longrightarrow+\infty}C_{GLS} &= \left(\frac{1}{\sqrt{T}}Su_T\right)^2-\frac{1}{T}\sum_{t=1}^{T}u_t^2\\
\lim_{N\longrightarrow+\infty}D_{GLS} &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}}Su_{t-1}\right)^2
\end{aligned}
$$

It follows that

$$
\lim_{N\longrightarrow+\infty}T\widehat{\rho}_{GLS,N}=\frac{1}{2}\frac{E\left(\frac{1}{\sqrt{T}}Sv_{i,T}\right)^2-E\frac{1}{T}\sum_{t=1}^{T}v_{i,t}^2}{\frac{1}{T}\sum_{t=1}^{T}E\left(\frac{1}{\sqrt{T}}Sv_{i,t}\right)^2}
$$

and $\lim_{T\to+\infty}\lim_{N\longrightarrow+\infty} T\widehat{\rho}_{GLS,N}=0$. With T fixed, the denominator is the sum of a random variable that converges to 0 faster than $\sqrt{N}$, equal to

$$\frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{\sqrt{T}}Sv_{j,T}-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}Sv_{i,T}\right)^2-\frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}v_{j,T}-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}v_{i,T}\right)^2+O_p\left(\frac{1}{N}\right)$$

It can be reand as an empirical average of centered i.i.d. random variables that are L^2, a CLT applies.

When $(u_t)_{t\in\mathbb{Z}}\neq(\varepsilon_t)_{t\in\mathbb{Z}}$ or $(v_{i,t})_{t\in\mathbb{Z}}\neq(\eta_{i,t})_{t\in\mathbb{Z}}$, a feasible strategy consists of using the above test statistics and of computing its limit that involves nuisance parameters. To get nuisance parameter free test statistic, non parametric estimates can be computed and introduced in the expression. The estimate of ω is not of importance asymptotically, but for good power and size properties a judicious choice must be done. A natural candidate would be to consider the ratio of the long term variance of the processes $(v_{i,t})_{t\in\mathbb{Z}}$ and $(u_t)_{t\in\mathbb{Z}}$. In this case under DGP (I), we have

$$\begin{aligned}A_{OLS}-C_{GLS} &\Longrightarrow \frac{\omega}{\omega+N}\left(\sigma_\varepsilon^2 c(1)^2\left(W(1)^2-1\right)+2\sigma_\varepsilon\sigma_\eta c(1)\,d(1)\,W(1)\frac{1}{N}\sum_{i=1}^{N}B_i(1)\right)\\ &\quad+\sigma_\eta^2 d(1)^2\left(\frac{1}{N}\sum_{i=1}^{N}\left(B_i(1)^2-1\right)-\frac{N}{\omega+N}\left(\frac{1}{N}\sum_{i=1}^{N}B_i(1)\right)^2\right)\\ &\quad+\frac{\omega}{\omega+N}\left(\sigma_\varepsilon^2 c(1)^2-\gamma_u(0)\right)+\left(\sigma_\eta^2 d(1)^2-\gamma_v(0)\right)\end{aligned}$$

and

$$\begin{aligned}B_{OLS}-D_{GLS} &\Longrightarrow \frac{\omega}{\omega+N}\left(\sigma_\varepsilon^2 c(1)^2\int_0^1 W(s)^2\,ds+\sigma_\varepsilon c(1)\int_0^1 W(s)\left(\frac{\sigma_\eta d(1)}{N}\sum_{i=1}^{N}B_i(s)\right)ds\right)\\ &\quad+\sigma_\eta^2 d(1)^2\left(\frac{1}{N}\sum_{i=1}^{N}\int_0^1 B_i(s)^2\,ds-\frac{N}{\omega+N}\int_0^1\left(\frac{1}{N}\sum_{i=1}^{N}B_i(s)\right)^2 ds\right)\end{aligned}$$

whence using the CMT, the stated result.

Proof of Proposition 20: Under DGP (I), we have

$$\widehat{\rho}_{1,N}=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)\left(y_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}y_{j,t-1}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)^2}$$

and

$$\widehat{\rho}_{2,N}=\frac{\sum_{t=1}^{T}\left(\sum_{j=1}^{N}Sy_{j,t-1}\right)\left(\sum_{j=1}^{N}y_{j,t-1}\right)}{\sum_{t=1}^{T}\left(\sum_{j=1}^{N}Sy_{j,t-1}\right)^2}$$

Direct computations give

$$\begin{aligned}T\widehat{\rho}_{1,N} &= \frac{1}{2}\frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\left(Sy_{i,T}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,T}\right)^2-\frac{1}{T}\left(Sy_{i,0}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,0}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)^2}\\ &\quad-\frac{1}{2}\frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}y_{j,t-1}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\left(Sy_{i,t-1}-\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t-1}\right)^2}\end{aligned}$$

and

$$T\widehat{\rho}_{2,N}=\frac{1}{2}\frac{\left(\frac{1}{N\sqrt{T}}\sum_{j=1}^{N}Sy_{j,T}\right)^{2}-\left(\frac{1}{N\sqrt{T}}\sum_{j=1}^{N}Sy_{j,0}\right)^{2}-\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}y_{j,t-1}\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N\sqrt{T}}\sum_{j=1}^{N}Sy_{j,t-1}\right)^{2}}$$

Under assumptions **A-E** , when T goes to $+\infty$:

$$T\widehat{\rho}_{1,N}\Longrightarrow\frac{1}{2}\frac{\sigma_{\eta}^{2}d\left(1\right)^{2}\frac{1}{N}\sum_{i=1}^{N}\left(B_{i}\left(1\right)-\frac{1}{N}\sum_{j=1}^{N}B_{j}\left(1\right)\right)^{2}-\left(1-\frac{1}{N}\right)\gamma_{v}\left(0\right)}{\sigma_{\eta}^{2}d\left(1\right)^{2}\frac{1}{N}\sum_{i=1}^{N}\int_{0}^{1}\left(B_{i}\left(s\right)-\frac{1}{N}\sum_{j=1}^{N}B_{j}\left(s\right)\right)^{2}}$$

and

$$T\widehat{\rho}_{2,N}\Longrightarrow\frac{1}{2}\frac{\left(\sigma_{\varepsilon}c\left(1\right)W\left(1\right)+\sigma_{\eta}d\left(1\right)\frac{1}{N}\sum_{j=1}^{N}B_{j}\left(1\right)\right)^{2}-\left(\gamma_{u}\left(0\right)+\frac{1}{N}\gamma_{v}\left(0\right)\right)}{\int_{0}^{1}\left(\sigma_{\varepsilon}c\left(1\right)W\left(s\right)+\sigma_{\eta}d\left(1\right)\frac{1}{N}\sum_{j=1}^{N}B_{j}\left(s\right)\right)^{2}ds}$$

When N tends to $+\infty$, $\left(B_{j}\left(s\right)_{s\in[0,1]}\right)_{j\in\mathbb{N}}$ are i.i.d, from SLLN, CLT and results on moments of functional of Brownian motions from Lin, Levine and Chu (2002) we get that

$$\begin{aligned}\lim_{N\longrightarrow+\infty}\lim_{T\longrightarrow+\infty}T\widehat{\rho}_{1,N} &= \frac{\sigma_{\eta}^{2}d\left(1\right)^{2}-\gamma_{v}\left(0\right)}{\sigma_{\eta}^{2}d\left(1\right)^{2}}\\ \sqrt{N}\left(T\widehat{\rho}_{1,N}-\frac{\sigma_{\eta}^{2}d\left(1\right)^{2}-\gamma_{v}\left(0\right)}{\sigma_{\eta}^{2}d\left(1\right)^{2}}\right) &\Longrightarrow \mathcal{N}\left(0,2\right)\end{aligned}$$

Quite differently,

$$\lim_{N\longrightarrow+\infty}T\widehat{\rho}_{2,N}\Longrightarrow\frac{1}{2}\frac{\sigma_{\varepsilon}^{2}c\left(1\right)^{2}W\left(1\right)^{2}-\gamma_{u}\left(0\right)}{\sigma_{\varepsilon}^{2}c\left(1\right)^{2}\int_{0}^{1}W\left(s\right)^{2}ds}$$

which is the usual weak convergence result for Phillips-Perron type regression.

Proof of Proposition 25 : We consider the DGP :

$$\begin{aligned}y_{i,t} &= X\mu_{i}+r_{t}^{(c)}+\varepsilon_{t}+\eta_{i,t}\\ r_{t}^{(c)} &= r_{t-1}^{(c)}+w_{t}^{(c)}\end{aligned}$$

where X is a set of deterministic variables, $Vw_{t}^{(c)}=\sigma_{w(c)}^{2}$ and all the random terms are supposed to be gaussian. We want to test for $H_{0(c)}:\sigma_{w(c)}=0$ with a lagrange multiplier test statistic. Let C_{T} be the $(T\times T)$ lower triangular matrix whose coefficients are constant and equal to 1. If we stack all the subvectors μ_{i} in a large one denoted μ, we can write

$$y\sim\mathcal{N}\left(I_{NT}\otimes\mu,\sigma_{w(c)}^{2}C_{T}C_{T}^{\prime}\otimes J_{N}+\sigma_{\varepsilon}^{2}I_{T}\otimes J_{N}+\sigma_{\eta}^{2}I_{NT}\right)$$

If we denote $\omega=\frac{\sigma_{\eta}^{2}}{\sigma_{\varepsilon}^{2}}$ and $\xi_{(c)}=\frac{\sigma_{w(c)}^{2}}{\sigma_{\varepsilon}^{2}}$, $H_{0}:\xi_{(c)}=0$. Following Tanaka (1995), if $\widehat{\mu}$, $\widehat{\omega}$ and $\widehat{\sigma_{\varepsilon}^{2}}$ denote the maximum likelihood estimators of these parameters, the test statistic is derived as follows

$$\left.\frac{\partial Logl\left(\mu,\omega,\sigma_{\varepsilon}^{2},\xi_{(c)}\right)}{\partial\xi_{(c)}}\right|_{\widehat{\mu},\widehat{\omega},\widehat{\sigma_{\varepsilon}^{2}},\xi_{(c)}=0}=-\frac{1}{2}tr\left(\left(I_{T}\otimes J_{N}+\widehat{\omega}I_{NT}\right)^{-1}C_{T}C_{T}^{\prime}\otimes J_{N}\right)$$

$$-\frac{1}{2\widehat{\sigma_\varepsilon^2}}\left(y-I_{NT}\otimes\widehat{\mu}\right)'\left(I_T\otimes J_N+\widehat{\omega}I_{NT}\right)^{-1}C_TC_T'\otimes J_N\left(I_T\otimes J_N+\widehat{\omega}I_{NT}\right)^{-1}\left(y-I_{NT}\otimes\widehat{\mu}\right)$$

where $\left(I_T\otimes J_N+\widehat{\omega}I_{NT}\right)^{-1}=\frac{1}{\widehat{\omega}\widehat{\sigma_\varepsilon^2}}I_T\otimes\left(I_N-\frac{1}{N+\widehat{\omega}}J_N\right)$, then

$$\left.\frac{\partial Logl\left(\mu,\omega,\sigma_\varepsilon^2,\xi_{(c)}\right)}{\partial\xi_{(c)}}\right|_{\widehat{\mu},\widehat{\omega},\widehat{\sigma_\varepsilon^2},\xi_{(c)}=0}=-\frac{1}{2}\frac{N}{\left(N+\widehat{\omega}\right)\widehat{\sigma_\varepsilon^2}}\frac{T\left(T+1\right)}{2}$$

$$-\frac{1}{2\widehat{\sigma_\varepsilon^2}^3\left(\widehat{\omega}+N\right)^2}\left(y-I_{NT}\otimes\widehat{\mu}\right)'C_T\otimes e_NC_T'\otimes e_N\left(y-I_{NT}\otimes\widehat{\mu}\right)$$

When the DGP is (I), the second term takes the following form (for the other DGPs, we replace $y_{i,t}$ by the residual of its projection on the deterministic terms)

$$\frac{1}{2\widehat{\sigma_\varepsilon^2}}\left(\frac{N}{\widehat{\omega}+N}\right)^2\sum_{t=1}^{T}\left(\frac{\frac{1}{N}\sum_{j=1}^{N}Sy_{j,t}}{\widehat{\sigma_\varepsilon^2}}\right)^2$$

Divided by T^2, the above statistic is equal to

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}Sy_{i,t}\right)^2 &= \frac{1}{T^2}\sum_{t=1}^{T}\left(Su_t+\frac{1}{N}\sum_{i=1}^{N}Sv_{i,t}\right)^2\\ &= \frac{1}{T^2}\sum_{t=1}^{T}\left(Su_t\right)^2+\frac{2}{T^2}\sum_{t=1}^{T}Su_t\left(\frac{1}{N}\sum_{i=1}^{N}Sv_{i,t}\right)+\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}Sv_{i,t}\right)^2\end{aligned}$$

Under assumption **A-E**, when T goes to $+\infty$, the above expression weakly converges to

$$\sigma_\varepsilon^2c\left(1\right)^2\int_0^1W\left(s\right)^2ds+2\sigma_\varepsilon c\left(1\right)\sigma_\eta d\left(1\right)\int_0^1W\left(s\right)\left(\frac{1}{N}\sum_{i=1}^{N}B_i\left(s\right)\right)ds+\sigma_\eta^2d\left(1\right)^2\int_0^1\left(\frac{1}{N}\sum_{i=1}^{N}B_i\left(s\right)\right)^2ds$$

then if N goes to $+\infty$, its weak limit is given

$$\sigma_\varepsilon^2c\left(1\right)^2\int_0^1W\left(s\right)^2ds$$

due to the fact that the random variables $\left(B_i\left(s\right)_{s\in[0,1]}\right)_{i\in\mathbb{N}}$ are i.i.d. Similar arguments allows us to say that the same limit holds when first N goes to $+\infty$, then T. The same reasoning is valid if we replace $Sy_{i,t}$ by its residual after its projection on the appropriate set of deterministic functions, the Brownian motion being appropriately replaced by a Brownian bridge.

Proof of Proposition 24: We now consider the DGP :

$$\begin{aligned}y_{j,t} &= X\mu_j+r_{j,t}^{(i)}+\varepsilon_t+\eta_{j,t}\\ r_{j,t}^{(i)} &= r_{j,t-1}^{(i)}+w_{j,t}^{(i)}\end{aligned}$$

where X is a set of deterministic variables, $Vw_{i,t}^{(i)}=\sigma_{w(i)}^2$ and all the random terms are supposed to be gaussian. We want to test for $H_{0(i)}:\sigma_{w(i)}=0$ with a lagrange multiplier test statistic With the same notations as in the preceding proof, we can write

$$y\sim\mathcal{N}\left(I_{NT}\otimes\mu,\sigma_{w(i)}^2C_TC_T'\otimes I_N+\sigma_\varepsilon^2I_T\otimes J_N+\sigma_\eta^2I_{NT}\right)$$

If we denote $\omega = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$ and $\xi_{(i)} = \frac{\sigma_{w(i)}^2}{\sigma_\varepsilon^2}$, $H_0 : \xi_{(i)} = 0$. Using similar computations, if $\widehat{\mu}$, $\widehat{\omega}$ and $\widehat{\sigma_\varepsilon^2}$ denote the maximum likelihood estimators of these parameters, the test statistic is derived as follows

$$\frac{\partial Logl\left(\mu, \omega, \sigma_\varepsilon^2, \xi_{(i)}\right)}{\partial \xi_{(i)}}\Bigg|_{\widehat{\mu},\widehat{\omega},\widehat{\sigma_\varepsilon^2},\xi_{(i)}=0} = -\frac{1}{2}tr\left(\left(I_T \otimes J_N + \widehat{\omega} I_{NT}\right)^{-1} C_T C_T' \otimes I_N\right)$$

$$-\frac{1}{2\widehat{\sigma_\varepsilon^2}}\left(y - I_{NT} \otimes \widehat{\mu}\right)' \left(I_T \otimes J_N + \widehat{\omega} I_{NT}\right)^{-1} C_T C_T' \otimes I_N \left(I_T \otimes J_N + \widehat{\omega} I_{NT}\right)^{-1} \left(y - I_{NT} \otimes \widehat{\mu}\right)$$

where $\left(I_T \otimes J_N + \widehat{\omega} I_{NT}\right)^{-1} = \frac{1}{\widehat{\omega}\widehat{\sigma_\varepsilon^2}} I_T \otimes \left(I_N - \frac{1}{N+\widehat{\omega}} J_N\right)$, then

$$\frac{\partial Logl\left(\mu, \omega, \sigma_\varepsilon^2, \xi_{(i)}\right)}{\partial \xi_{(i)}}\Bigg|_{\widehat{\mu},\widehat{\omega},\widehat{\sigma_\varepsilon^2},\xi_{(i)}=0} = -\frac{1}{2}\frac{N\left(N + \widehat{\omega} - 1\right)}{\left(N + \widehat{\omega}\right)\widehat{\omega}\widehat{\sigma_\varepsilon^2}}\frac{T\left(T+1\right)}{2}$$

$$-\frac{1}{2\widehat{\sigma_\varepsilon^2}^3\widehat{\omega}^2}\left(y - I_{NT} \otimes \widehat{\mu}\right)' C_T \otimes I_N \left(I_T \otimes \left(I_N - \frac{2\widehat{\omega} + N}{\left(\widehat{\omega} + N\right)^2} J_N\right)\right) C_T' \otimes I_N \left(y - I_{NT} \otimes \widehat{\mu}\right)$$

When the DGP is (I), the second term takes the following form (for the other DGPs, we replace $y_{i,t}$ by the residual of its projection on the deterministic terms)

$$\frac{1}{2\widehat{\sigma_\varepsilon^2}}\sum_{j=1}^N \sum_{t=1}^T \left(\frac{Sy_{j,t} - \frac{2\widehat{\omega}+N}{(\widehat{\omega}+N)^2}\sum_{i=1}^N Sy_{i,t}}{\widehat{\sigma_\varepsilon^2}\widehat{\omega}}\right)^2$$

We therefore focus on the test statistic

$$\sum_{j=1}^N \sum_{t=1}^T \left(Sy_{j,t} - \frac{2\widehat{\omega} + N}{\left(\widehat{\omega} + N\right)^2}\sum_{i=1}^N Sy_{i,t}\right)^2 = \sum_{j=1}^N \sum_{t=1}^T \left(\frac{\widehat{\omega}^2}{\left(\widehat{\omega} + N\right)^2} Su_t + Sv_{j,t} - \frac{2\widehat{\omega} + N}{\left(\widehat{\omega} + N\right)^2}\sum_{i=1}^N Sv_{i,t}\right)^2$$

divided by NT^2 it is equal to

$$\frac{1}{T^2}\left(\frac{\widehat{\omega}^2}{\left(\widehat{\omega} + N\right)^2}\right)^2 \left(\sum_{t=1}^T \left(Su_t\right)^2 + 2\sum_{t=1}^T \left(Su_t \frac{1}{N}\sum_{j=1}^N Sv_{j,t}\right)\right)$$
$$+\frac{1}{N}\sum_{j=1}^N \frac{1}{T^2}\sum_{t=1}^T \left(Sv_{j,t} - \frac{2\widehat{\omega} + N}{\left(\widehat{\omega} + N\right)^2}\sum_{i=1}^N Sv_{i,t}\right)^2$$

We study the behavior of this expression when $(u_t)_{t\in\mathbb{Z}}$ and $(v_{i,t})_{t\in\mathbb{Z}}$ are not necessarily white noises. Under assumption **A-E**, when T goes to $+\infty$, the above expression weakly converges to

$$\left(\frac{\widehat{\omega}^2}{\left(\widehat{\omega} + N\right)^2}\right)^2 \left(\sigma_\varepsilon^2 c\left(1\right)^2 \int_0^1 W\left(s\right)^2 ds + 2\sigma_\varepsilon c\left(1\right)\sigma_\eta d\left(1\right)\int_0^1 W\left(s\right)\frac{1}{N}\sum_{j=1}^N B_j\left(s\right) ds\right)$$
$$+\frac{\sigma_\eta^2 d\left(1\right)^2}{N}\sum_{j=1}^N \int_0^1 \left(B_j\left(s\right) - \frac{2\widehat{\omega} + N}{\left(\widehat{\omega} + N\right)^2}\sum_{i=1}^N B_i\left(s\right)\right)^2 ds$$

which itself converges to $\frac{\sigma_\eta^2 d(1)^2}{2}$ when N goes to $+\infty$. The first part of the last formula multiplied by $\sqrt{N}$ is an $O_p\left(\frac{1}{N^3\sqrt{N}}\right)$ and also tends to 0 which eliminates the common component persistence. We can use a CLT for the set of i.i.d. random variables $\left(B_i\left(s\right)_{s\in[0,1]}\right)_{i\in\mathbb{N}}$ that gives

$$\sqrt{N}\left(\frac{1}{NT^2}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(Sy_{j,t}-\frac{2\widehat{\omega}+N}{\left(\widehat{\omega}+N\right)^2}\sum_{i=1}^{N}Sy_{i,t}\right)^2-\frac{\sigma_\eta^2 d\left(1\right)^2}{2}\right)\Longrightarrow\mathcal{N}\left(0,\frac{\sigma_\eta^4 d\left(1\right)^4}{3}\right)$$

If we now turn to the case where N goes first to $+\infty$, we have since $\left(Sv_{j,t}\right)_j$ are i.i.d. variables

$$\lim_{N\longrightarrow+\infty}\frac{1}{NT^2}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(Sy_{j,t}-\frac{2\widehat{\omega}+N}{\left(\widehat{\omega}+N\right)^2}\sum_{i=1}^{N}Sy_{i,t}\right)^2=E\frac{1}{T^2}\sum_{t=1}^{T}\left(Sv_{j,t}\right)^2$$

which converges to $\frac{\sigma_\eta^2 d(1)^2}{2}$ when T goes to $+\infty$. With T fixed, a CLT can be used and the variance of the asymptotic distribution is a fonction of T, σ_η^2 and the coefficients of $d\left(L\right)$.

Proof of Proposition 27: this is derived from Gregoir and Laroque (1994), Lemma 5.2.

Before proving Theorem 29, we introduce some notations and intermediary results. Starting from the elementary m-dimensional vector $y_{i,t}$, we write

$$y_{i,t}=\delta\gamma'\left(Sy_{i,t-1}-Sz_{t-1}^N\right)+\alpha\gamma' Sz_{t-1}^N+\sum_{j=1}^{p_1}\psi_j z_{t-j}^N+\sum_{j=1}^{p_2}\phi_j\left(y_{i,t-j}-z_{t-j}^N\right)+\varepsilon_t+\eta_{i,t}$$

We then place one vector next to another for a given date to get a $(m\times N)$ matrix y_t and a matrix equation

$$y_t=\delta\gamma'\left(Sy_{t-1}-Sz_{t-1}^N\right)+\alpha\gamma' Sz_{t-1}^N\otimes e_N'+\Theta w_t+\varepsilon_t\otimes e_N'+\eta_t$$

where $\Theta=\begin{pmatrix}\psi_1 & \cdots & \psi_{p_1} & \phi_1 & \cdots & \phi_{p_2}\end{pmatrix}$ and

$$w_t=\begin{pmatrix} z_{t-1}^N & \cdots & z_{t-1}^N\\ \vdots & & \vdots\\ z_{t-p_1}^N & \cdots & z_{t-p_1}^N\\ y_{1,t-1}-z_{t-1}^N & \cdots & y_{N,t-1}-z_{t-1}^N\\ \vdots & & \vdots\\ y_{1,t-p_2}-z_{t-p_2}^N & \cdots & y_{N,t-p_2}-z_{t-p_2}^N\end{pmatrix}$$

We repeat this operation for all the $(m\times N)$ matrices to end up with a $(m\times NT)$ one and the following equation

$$y=\delta\gamma'\left(Sy_{-1}-Sz_{-1}^N\right)+\alpha\gamma' Sz_{-1}^N\otimes e_N'+\Theta w+\varepsilon\otimes e_N'+\eta \tag{18}$$

where $\left(y,Sy_{-1}-Sz_{-1}^N,w,\eta\right)$ are $(m\times NT)$ matrices and $\left(Sz_{-1}^N,\varepsilon\right)$ are $(m\times T)$. In the sequel, we denote

$$
\begin{aligned}
Y &= \begin{pmatrix} \gamma'\left(Sy_{-1}-Sz_{-1}^{N}\right) \\ \gamma' Sz_{-1}^{N}\otimes e_N' \\ w \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2\otimes e_N' \\ Y_3 \end{pmatrix} \\
&= \begin{pmatrix} \begin{pmatrix} \gamma_1' \\ \gamma_2' \end{pmatrix}\left(Sy_{-1}-Sz_{-1}^{N}\right) \\ \gamma_3'\left(Sy_{-1}-Sz_{-1}^{N}\right) \\ \gamma_1' Sz_{-1}^{N}\otimes e_N' \\ \begin{pmatrix} \gamma_2' \\ \gamma_3' \end{pmatrix} Sz_{-1}^{N}\otimes e_N' \\ w \end{pmatrix} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21}\otimes e_N' \\ Y_{22}\otimes e_N' \\ Y_3 \end{pmatrix}
\end{aligned}
$$

and partitionned conformably the matrix coefficient

$$
vec\left(\begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix} \quad \delta_3 \quad \alpha_1 \quad \begin{pmatrix} \alpha_2 & \alpha_3 \end{pmatrix} \quad \Theta \right)
$$

Lemma 30 *When T goes to $+\infty$, (i) $\frac{1}{T\sqrt{T}}Y_{21}Y_{22}' = O_p\left(\frac{1}{\sqrt{T}}\right)$, (ii) $\frac{1}{T}Y_{21}Y_{21}' = O_p(1)$, (iii) there exists a $(s_1+s_2)\times(s_1+s_2)$ full rank matrix κ_2 such that $\frac{1}{T^2}Y_{22}Y_{22}' \Longrightarrow \kappa_2 \begin{pmatrix} \alpha_2' \\ \alpha_3' \end{pmatrix} \int_0^1 \widetilde{W}(s)\widetilde{W}(s)'\,ds\,($ where*

$$
\widetilde{W}(s) = \left(\Omega_{\varepsilon}^{1/2}W(s) + \frac{1}{N}\Omega_{\eta}^{1/2}\sum_{j=1}^{N}B_i(s)\right)
$$

and $\left(W(s)_{s\in[0,1]}, \left(B_i(s)_{s\in[0,1]}\right)_{i=1,..,N}\right)$ are $(N+1)$ independent m-dimensional standard Brownian motions, (iv)

$$
\frac{1}{T}\left(\varepsilon + \frac{1}{N}\eta I_T\otimes e_N\right)Y_{22}' \Longrightarrow \int_0^1 d\widetilde{W}(s)\,\widetilde{W}(s)'\,ds\begin{pmatrix} \alpha_2 & \alpha_3 \end{pmatrix}\kappa_2'
$$

and (v) $\frac{1}{T}Y_2Y_3' = O_p(1)$

Proof of Lemma 30 : we start from equation (10) in proposition 16:

$$
y_{i,t} = C_1\varepsilon_t + D_1\eta_{i,t} + (C_1 - D_1)\frac{1}{N}\sum_{j=1}^{N}\eta_{j,t} + (1-L)\zeta_{i,t}
$$

where $\zeta_{i,t}$ is a covariance stationary process. Applying S operator, we get that

$$
Sz_t^N = C_1\left(S\varepsilon_t + \frac{1}{N}\sum_{j=1}^{N}S\eta_{j,t}\right) + \frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{i,t}-\zeta_{i,0}\right)
$$

so that

$$
\begin{aligned}
Y_{21,t} &= \gamma_1' Sz_t^N = \gamma_1'\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{i,t}-\zeta_{i,0}\right) \\
Y_{22,t} &= \begin{pmatrix} \gamma_2' \\ \gamma_3' \end{pmatrix} Sz_t^N = \kappa_2\begin{pmatrix} \alpha_2' \\ \alpha_3' \end{pmatrix}\left(S\varepsilon_t + \frac{1}{N}\sum_{j=1}^{N}S\eta_{j,t}\right) + \begin{pmatrix} \gamma_2' \\ \gamma_3' \end{pmatrix}\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{i,t}-\zeta_{i,0}\right)
\end{aligned}
$$

with

$$\kappa_2 = \begin{pmatrix} \gamma_2' \\ \gamma_3' \end{pmatrix} \begin{pmatrix} \gamma_2 & \gamma_3 \end{pmatrix} \left(\begin{pmatrix} \alpha_2' \\ \alpha_3' \end{pmatrix} \left(I_p - \sum_{j=1}^{p_1} \psi_j \right) \begin{pmatrix} \gamma_2 & \gamma_3 \end{pmatrix} \right)^{-1}$$

Under **B'**, **C'** and **D**, when T goes to $+\infty$ and N is fixed, we have the following FCLT (see Phillips and Durlauf (1987))

$$\frac{1}{\sqrt{T}} Y_{22,[Ts]} \Longrightarrow \kappa_2 \begin{pmatrix} \alpha_2' \\ \alpha_3' \end{pmatrix} \left(\Omega_\varepsilon^{1/2} W(s) + \frac{1}{N} \Omega_\eta^{1/2} \sum_{j=1}^{N} B_j(s) \right)$$

where $\left(W(s), (B_i(s))_i\right)$ is described in the statement. Let us denote

$$\widetilde{W}(s) = \left(\Omega_\varepsilon^{1/2} W(s) + \frac{1}{N} \Omega_\eta^{1/2} \sum_{j=1}^{N} B_j(s) \right)$$

the following joint convergences hold

$$\begin{aligned} \frac{1}{T} Y_{21} Y_{22}' &= O_p(1) \\ \frac{1}{T} Y_{21} Y_{21}' &= O_p(1) \\ \frac{1}{T^2} Y_{22} Y_{22}' &\Longrightarrow \kappa_2 \begin{pmatrix} \alpha_2' \\ \alpha_3' \end{pmatrix} \int_0^1 \widetilde{W}(s) \widetilde{W}(s)' \, ds \begin{pmatrix} \alpha_2 & \alpha_3 \end{pmatrix} \kappa_2' \end{aligned}$$

Moreover, as $Y_{3,t}$ is a multivariate covariance stationary process, we have

$$\begin{aligned} \frac{1}{T} Y_{21} Y_3' &= O_p(1) \\ \frac{1}{T} Y_{22} Y_3' &= O_p(1) \end{aligned}$$

Lemma 31 *When T goes to $+\infty$, (i) $\frac{1}{NT\sqrt{T}} Y_{11} Y_{12}' = O_p\left(\frac{1}{\sqrt{T}}\right)$, (ii) $\frac{1}{NT} Y_{11} Y_{11}' = O_p(1)$, (iii) there exists a $s_1 \times s_1$ full rank matrix κ_1 such that $\frac{1}{NT^2} Y_{12} Y_{12}' \Longrightarrow \kappa_1 \delta_3' \Omega_\eta^{1/2} \frac{1}{N} \sum_{i=1}^{N} \int_0^1 \widetilde{B}_i(s) \widetilde{B}_i(s)' \, ds \Omega_\eta^{1/2}$ with*

$$\widetilde{B}_i(s) = B_i(s) - \frac{1}{N} \sum_{j=1}^{N} B_j(s)$$

where $\left(B_i(s)_{s \in [0,1]}\right)_{i=1,\dots N}$ are N independent m-dimensional standard Brownian motions, (iv)

$$\frac{1}{T\sqrt{N}} \eta Y_{12}' \Longrightarrow \frac{1}{\sqrt{N}} \Omega_\eta^{1/2} \sum_{j=1}^{N} \int_0^1 dB_j(s) \widetilde{B}_j(s) \Omega_\eta^{1/2} \delta_3 \kappa_1'$$

(v) $\frac{1}{TN} Y_1 Y_3' = O_p(1)$.

Proof of Lemma 31 : Again we start from equation (10) in proposition 16:

$$y_{i,t} = C_1 \varepsilon_t + D_1 \eta_{i,t} + (C_1 - D_1) \frac{1}{N} \sum_{j=1}^{N} \eta_{j,t} + (1 - L) \zeta_{i,t}$$

where $\zeta_{i,t}$ is a covariance stationary process. We then get that

$$Sy_{i,t} - Sz_t^N = D_1 \left(S\eta_{i,t} - \frac{1}{N} \sum_{j=1}^{N} S\eta_{j,t} \right) + \zeta_{i,t} - \zeta_{i,0} - \frac{1}{N} \sum_{j=1}^{N} (\zeta_{j,t} - \zeta_{j,0})$$

so that

$$\begin{aligned}
Y_{11,i,t} &= \begin{pmatrix} \gamma_1' \\ \gamma_2' \end{pmatrix} \left(Sy_{i,t} - Sz_t^N \right) \\
&= \begin{pmatrix} \gamma_1' \\ \gamma_2' \end{pmatrix} \left(\zeta_{i,t} - \zeta_{i,0} - \frac{1}{N} \sum_{j=1}^{N} (\zeta_{j,t} - \zeta_{j,0}) \right) \\
Y_{12,i,t} &= \gamma_3' \left(Sy_{i,t} - Sz_t^N \right) \\
&= \kappa_1 \delta_3' \left(S\eta_{i,t} - \frac{1}{N} \sum_{j=1}^{N} S\eta_{j,t} \right) + \gamma_3' \left(\zeta_{i,t} - \zeta_{i,0} - \frac{1}{N} \sum_{i=1}^{N} (\zeta_{i,t} - \zeta_{i,0}) \right)
\end{aligned}$$

with

$$\kappa_1 = \gamma_3' \gamma_3 \left(\delta_3' \left(I_p - \sum_{j=1}^{p_2} \phi_j \right) \gamma_3 \right)^{-1}$$

Under **B'**, **C'** and **D**, when T goes to $+\infty$ and N is fixed, we have the following FCLT (see Phillips and Durlauf (1987))

$$\frac{1}{\sqrt{T}} Y_{12,i,[Ts]} \Longrightarrow \kappa_1 \delta_3' \Omega_\eta^{1/2} \underbrace{\left(B_i(s) - \frac{1}{N} \sum_{j=1}^{N} B_j(s) \right)}_{\widetilde{B_i}(s)}$$

where $\left((B_i(s))_i \right)$ is described in the statement and the following joint convergences hold

$$\begin{aligned}
\frac{1}{TN} Y_{21} Y_{22}' &= O_p(1) \\
\frac{1}{TN} Y_{21} Y_{21}' &= O_p(1) \\
\frac{1}{NT^2} Y_{22} Y_{22}' &\Longrightarrow \kappa_1 \delta_3' \Omega_\eta^{1/2} \frac{1}{N} \sum_{i=1}^{N} \int_0^1 \widetilde{B}_i(s) \widetilde{B}_i(s)' ds \Omega_\eta^{1/2} \delta_3 \kappa_1'
\end{aligned}$$

On the other hand, we have

$$\frac{1}{T\sqrt{N}} \eta Y_{12}' = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} \eta_{j,t} Y_{12,t-1}' \right)$$

For N fixed, Phillips(1988) shows that when T goes to $+\infty$, the following weak convergence holds

$$\frac{1}{T} \sum_{t=1}^{T} \eta_{j,t} Y_{12,j,t-1}' \Longrightarrow \Omega_\eta^{1/2} \int_0^1 dB_j(s) \widetilde{B}_j(s) \Omega_\eta^{1/2} \delta_3 \kappa_1'$$

Moreover, as $Y_{3,t}$ is a multivariate covariance stationary process, we have

$$\begin{aligned}
\frac{1}{TN} Y_{21} Y_3' &= O_p(1) \\
\frac{1}{TN} Y_{22} Y_3' &= O_p(1)
\end{aligned}$$

Proof of Theorem 29: In a first step, we assume that we know $V\varepsilon_t = \Omega_\varepsilon$, $V\eta_{i,t} = \Omega_\eta$ and $\gamma = \begin{pmatrix} \gamma_1' & \gamma_2' & \gamma_3' \end{pmatrix}'$. We apply the *vec* operator on both sides of equation (18). We have to consider the heteroskedastic linear model

$$vec\,(y)) = \left(\begin{pmatrix} \gamma'\left(Sy_{-1} - Sz_{-1}^N\right) \\ \gamma' Sz_{-1}^N \otimes e_N' \\ w \end{pmatrix}' \otimes I_m\right) vec\begin{pmatrix} \delta & \alpha & \Theta \end{pmatrix} + vec\left(\varepsilon \otimes e_N' + \eta\right)$$

with $V\left(vec\left(\varepsilon \otimes e_N' + \eta\right)\right) = V = I_T \otimes \left(J_N \otimes \Omega_\varepsilon + I_N \otimes \Omega_\eta\right)$ that is such that

$$V^{-1} = I_T \otimes \left(I_N \otimes \Omega_\eta^{-1} - J_N \otimes \left(N\Omega_\eta + \Omega_\eta \Omega_\varepsilon^{-1}\Omega_\eta\right)^{-1}\right)$$

In the sequel, we set $\Sigma = \left(N\Omega_\eta + \Omega_\eta \Omega_\varepsilon^{-1}\Omega_\eta\right)^{-1}$. The GLS estimate is such that

$$\left(Y \otimes I_m V^{-1} Y' \otimes I_m\right) vec\,\widehat{\begin{pmatrix} \delta & \alpha & \Theta \end{pmatrix}} = Y \otimes I_m V^{-1} vec\,(y) \tag{19}$$

and

$$\left(Y \otimes I_m V^{-1} Y' \otimes I_m\right) \left(vec\,\widehat{\begin{pmatrix} \delta & \alpha & \Theta \end{pmatrix}} - vec\begin{pmatrix} \delta & \alpha & \Theta \end{pmatrix}\right) = Y \otimes I_m V^{-1} vec\left(\varepsilon \otimes e_N' + \eta\right) \tag{20}$$

The first matrix in the last equation (19) is the sum of two terms, on the one hand

$$YY' \otimes \Omega_\eta^{-1}$$

and on the other hand

$$-\left(Y\left(I_T \otimes e_N e_N'\right) Y'\right) \otimes \Sigma$$

but since $\forall t \in \{1, ..., T\}$, $\sum_{j=1}^N \left(Sy_{i,t} - Sz_t^N\right) = 0$, $\left(I_T \otimes e_N'\right) Y_1' = 0$ and $\left(Y_2 \otimes e_N\right) Y_1' = 0$, and moreover $\left(I_T \otimes e_N'\right)\left(Y_2' \otimes e_N\right) = NY_2'$, this matrix has the following structure

$$\begin{pmatrix} Y_1 Y_1' \otimes \Omega_\eta^{-1} & 0 & Y_1 \otimes I_m V^{-1} Y_3' \otimes I_m \\ 0 & NY_2 Y_2' \otimes \left(\Omega_\eta^{-1} - N\Sigma\right) & Y_2 \otimes I_m V^{-1} Y_3' \otimes I_m \\ Y_3 \otimes I_m V^{-1} Y_1' \otimes I_m & Y_3 \otimes I_m V^{-1} Y_2' \otimes I_m & Y_3 \otimes I_m V^{-1} Y_3' \otimes I_m \end{pmatrix}$$

Similarly, the second term of equation (20) is the sum of

$$\begin{aligned} Y \otimes \Omega_\eta^{-1} vec\left(\varepsilon \otimes e_N' + \eta\right) &= vec\left(\Omega_\eta^{-1}\left(\varepsilon \otimes e_N' + \eta\right) Y'\right) \\ &= vec\left(\Omega_\eta^{-1}\left(\varepsilon \otimes e_N' + \eta\right)\begin{pmatrix} Y_1' & Y_2' \otimes e_N & Y_3' \end{pmatrix}\right) \\ &= vec\left(\Omega_\eta^{-1}\begin{pmatrix} \eta Y_1' & \left(N\varepsilon + \eta I_T \otimes e_N\right) Y_2' & \left(\varepsilon \otimes e_N' + \eta\right) Y_3' \end{pmatrix}\right) \end{aligned}$$

and

$$-Y\left(\left(\left(I_T \otimes e_N e_N'\right) \otimes \Sigma\right) vec\left(\varepsilon \otimes e_N' + \eta\right)\right) = -\begin{pmatrix} 0 \\ NY_2 \otimes e_N' \\ Y_3\left(I_T \otimes e_N e_N'\right) \end{pmatrix} \otimes \Sigma vec\left(\varepsilon \otimes e_N' + \eta\right)$$

$$\begin{aligned} &= -vec\Sigma\left(\varepsilon \otimes e_N' + \eta\right)\begin{pmatrix} 0 & NY_2' \otimes e_N & \left(I_T \otimes e_N e_N'\right) Y_3' \end{pmatrix} \\ &= -vec\Sigma\begin{pmatrix} 0 & \left(N^2\varepsilon + N\eta I_T \otimes e_N\right) Y_2' & \left(\varepsilon \otimes e_N' + \eta\right)\left(I_T \otimes e_N e_N'\right) Y_3' \end{pmatrix} \end{aligned}$$

so that it is equal to

$$\begin{aligned} vec\,\Big(& \Omega_\eta^{-1}\eta Y_1' \quad \left(\Omega_\eta^{-1} - N\Sigma\right)\left(N\varepsilon + \eta I_T \otimes e_N\right) Y_2' \\ & \left(\Omega_\eta^{-1}\left(\varepsilon \otimes e_N' + \eta\right) - \Sigma\left(\varepsilon \otimes e_N' + \eta\right)\left(I_T \otimes e_N e_N'\right)\right) Y_3' \Big) \end{aligned}$$

Let us denote G_T the appropriate scaling matrix which is defined as follows

$$G_T = \begin{pmatrix} \frac{1}{\sqrt{TN}} I_{m-s_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{T\sqrt{N}} I_{s_1} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{TN}} I_{m-s_1-s_2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{T\sqrt{N}} I_{s_1+s_2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{TN}} I_{m(p_1+p_2)} \end{pmatrix}$$

We premultiply both terms of equation (20) and introduce the product $G_T G_T^{-1}$ between the matrix and the vector in the left hand side term. The transformed equation involves the matrix

$$\begin{pmatrix} YY_{11} & 0 & YY_{13} \\ 0 & YY_{22} & YY_{23} \\ YY_{13}' & YY_{23}' & YY_{33} \end{pmatrix}$$

where

$$\begin{aligned} YY_{11} &= \begin{pmatrix} \frac{1}{TN} Y_{11} Y_{11}' & \frac{1}{T\sqrt{TN}} Y_{11} Y_{12}' \\ \frac{1}{T\sqrt{TN}} Y_{12} Y_{11}' & \frac{1}{T^2 N} Y_{12} Y_{12}' \end{pmatrix} \otimes \Omega_\eta^{-1} \\ &= \begin{pmatrix} \frac{1}{TN} Y_{11} Y_{11}' & O_p\left(\frac{1}{\sqrt{T}}\right) \\ O_p\left(\frac{1}{\sqrt{T}}\right) & \frac{1}{T^2 N} Y_{12} Y_{12}' \end{pmatrix} \otimes \Omega_\eta^{-1} \end{aligned}$$

$$\begin{aligned} YY_{22} &= \begin{pmatrix} \frac{1}{T} Y_{21} Y_{21}' & \frac{1}{T\sqrt{T}} Y_{21} Y_{22}' \\ \frac{1}{T\sqrt{T}} Y_{22} Y_{21}' & \frac{1}{T^2} Y_{22} Y_{22}' \end{pmatrix} \otimes \left(\Omega_\eta^{-1} - N\Sigma\right) \\ &= \begin{pmatrix} \frac{1}{T} Y_{21} Y_{21}' & O_p\left(\frac{1}{\sqrt{T}}\right) \\ O_p\left(\frac{1}{\sqrt{T}}\right) & \frac{1}{T^2} Y_{22} Y_{22}' \end{pmatrix} \otimes \left(\Omega_\eta^{-1} - N\Sigma\right) \end{aligned}$$

$$YY_{33} = \frac{1}{TN} \left(Y_3 \otimes I_m V^{-1} Y_3' \otimes I_m\right)$$

$$\begin{aligned} Y_{13} &= \begin{pmatrix} \frac{1}{TN} Y_{11} \otimes I_m V^{-1} Y_3' \otimes I_m \\ \frac{1}{TN\sqrt{T}} Y_{12} \otimes I_m V^{-1} Y_3' \otimes I_m \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{TN} Y_{11} \otimes I_m V^{-1} Y_3' \otimes I_m \\ O_p\left(\frac{1}{\sqrt{T}}\right) \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} Y_{23} &= \begin{pmatrix} \frac{1}{TN} Y_{21} \otimes I_m V^{-1} Y_3' \otimes I_m \\ \frac{1}{TN\sqrt{T}} Y_{22} \otimes I_m V^{-1} Y_3' \otimes I_m \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{TN} Y_{21} \otimes I_m V^{-1} Y_3' \otimes I_m \\ O_p\left(\frac{1}{\sqrt{T}}\right) \end{pmatrix} \end{aligned}$$

Using these results, we get the following set of expansions : first, the estimates of the loading coefficients satisfy the two equations

$$\begin{pmatrix} \frac{1}{TN} Y_{11} Y_{11}' \otimes \Omega_\eta^{-1} & \frac{1}{TN} Y_{11} \otimes I_m V^{-1} Y_3' \otimes I_m \end{pmatrix} \begin{pmatrix} \sqrt{TN} vec\begin{pmatrix} \widehat{\delta}_1 - \delta_1 & \widehat{\delta}_2 - \delta_2 \end{pmatrix} \\ \sqrt{TN} vec\left(\widehat{\theta} - \theta\right) \end{pmatrix}$$

$$= \frac{1}{\sqrt{TN}} vec\left(\Omega_\eta^{-1}\eta Y_{11}'\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

and

$$\begin{pmatrix} \frac{1}{T}Y_{21}Y_{21}' \otimes \left(\Omega_\eta^{-1} - N\Sigma\right) & \frac{1}{TN}Y_{21}\otimes I_m V^{-1}Y_3' \otimes I_m \end{pmatrix} \begin{pmatrix} \sqrt{TN}vec\left(\widehat{\alpha_1} - \alpha_1\right) \\ \sqrt{TN}vec\left(\widehat{\theta} - \theta\right) \end{pmatrix}$$

$$= \frac{1}{\sqrt{TN}} vec\left(\left(\Omega_\eta^{-1} - N\Sigma\right)\left(N\varepsilon + \eta I_T \otimes e_N\right)Y_{21}'\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

Second, the estimates that are related to the non-stationary processes introduced in the regression and that should not be present under the null hypothesis are such that

$$\frac{1}{T^2N}Y_{12}Y_{12}' \otimes \Omega_\eta^{-1} T\sqrt{N} vec\left(\widehat{\delta_3}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) = \frac{1}{T\sqrt{N}} vec\left(\Omega_\eta^{-1}\eta Y_{12}'\right) \tag{21}$$

and

$$\frac{1}{T^2}Y_{22}Y_{22}' \otimes \left(\Omega_\eta^{-1} - N\Sigma\right) T\sqrt{N} vec\begin{pmatrix} \widehat{\alpha_2} & \widehat{\alpha_3} \end{pmatrix} + O_p\left(\frac{1}{\sqrt{T}}\right) \tag{22}$$

$$= \frac{1}{T\sqrt{N}} vec\left(\left(\Omega_\eta^{-1} - N\Sigma\right)\left(N\varepsilon + \eta I_T \otimes e_N\right)Y_{22}'\right)$$

Equation (21) implies that

$$vec\left(\Omega_\eta^{-1}\left(\left(T\sqrt{N}\widehat{\delta_3}\right)\frac{1}{T^2N}Y_{12}Y_{12}' - \frac{1}{T\sqrt{N}}\eta Y_{12}'\right)\right) = O_p\left(\frac{1}{\sqrt{T}}\right)$$

Ω_η^{-1} is a full rank matrix, so that

$$\left(\left(T\sqrt{N}\widehat{\delta_3}\right)\frac{1}{T^2N}Y_{12}Y_{12}' - \frac{1}{T\sqrt{N}}\eta Y_{12}'\right) = O_p\left(\frac{1}{\sqrt{T}}\right)$$

and since $\frac{1}{T^2N}Y_{12}Y_{12}' = O_p(1)$ and is almost surely invertible

$$T\sqrt{N}\widehat{\delta_3} = \frac{1}{T\sqrt{N}}\eta Y_{12}'\left(\frac{1}{T^2N}Y_{12}Y_{12}'\right)^{-1} + O_p\left(\frac{1}{\sqrt{T}}\right)$$

When T goes to $+\infty$, we obtain that

$$T\sqrt{N}\widehat{\delta_3} \Longrightarrow \frac{1}{\sqrt{N}}\Omega_\eta^{1/2}\sum_{j=1}^{N}\int_0^1 dB_j(s)\widetilde{B}_j(s)\Omega_\eta^{1/2}\delta_3\left(\delta_3'\Omega_\eta^{1/2}\frac{1}{N}\sum_{i=1}^{N}\int_0^1 \widetilde{B}_i(s)\widetilde{B}_i(s)'ds\Omega_\eta^{1/2}\delta_3\right)^{-1}\kappa_1^{-1}$$

and if we consider the test statistics introduced in Theorem 29, when γ, Ω_η and Ω_ε are known and N fixed, we have

$$Tr\left(TN\widehat{\delta_3}\left(\frac{1}{NT}Y_{12}Y_{12}'\right)\widehat{\delta}_3'\right)$$

whose weak limit when T tends to $+\infty$ is

$$Tr\left(\Omega_\eta^{1/2}\sum_{j=1}^{N}\int_0^1 dB_j\widetilde{B}_j'\Omega_\eta^{1/2}\delta_3\left(\delta_3'\Omega_\eta^{1/2}\frac{1}{N}\sum_{i=j}^{N}\int_0^1 \widetilde{B}_j\widetilde{B}_j{}'\Omega_\eta^{1/2}\delta_3\right)^{-1}\delta_3'\Omega_\eta^{1/2}\sum_{j=1}^{N}\int_0^1 \widetilde{B}_j dB_j'\Omega_\eta^{1/2}\right)$$

Similarly, equation (22) gives

$$vec\left(\sqrt{N}\left(\Omega_\eta^{-1}-N\Sigma\right)\left(T\left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right)\frac{1}{T^2}Y_{22}Y_{22}'-\frac{1}{T}\left(\varepsilon+\frac{1}{N}\eta I_T\otimes e_N\right)Y_{22}'\right)\right)=O_p\left(\frac{1}{\sqrt{T}}\right)$$

On the one hand,

$$\Omega_\eta^{-1}-N\Sigma=\Omega_\eta^{-1}-\left(\Omega_\eta+\frac{1}{N}\Omega_\eta\Omega_\varepsilon^{-1}\Omega_\eta\right)^{-1}$$

if a full rank matrix, on the other hand $\left(\frac{1}{T^2}Y_{22}Y_{22}'\right)=O_p(1)$ and almost surely invertible, so that when N is fixed, we get that

$$T\left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right)=\frac{1}{T}\left(\varepsilon+\frac{1}{N}\eta I_T\otimes e_N\right)Y_{22}'\left(\frac{1}{T^2}Y_{22}Y_{22}'\right)^{-1}+O_p\left(\frac{1}{\sqrt{T}}\right)$$

and following Lemma 30,

$$T\left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right)\Longrightarrow\int_0^1 d\widetilde{W}(s)\widetilde{W}(s)'ds\left(\begin{array}{cc}\alpha_2 & \alpha_3\end{array}\right)\left(\left(\begin{array}{c}\alpha_2'\\ \alpha_3'\end{array}\right)\int_0^1\widetilde{W}(s)\widetilde{W}(s)'ds\left(\begin{array}{cc}\alpha_2 & \alpha_3\end{array}\right)\right)^{-1}\kappa_2^{-1}$$

If we now consider the test statistics introduced in Theorem 29, when γ, Ω_η and Ω_ε are known and N fixed, we have

$$Tr\left(T\left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right)\left(\frac{1}{T}Y_{22}Y_{22}'\right)\left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right)\right)$$

whose weak limit when T tends to $+\infty$ is

$$Tr\left(\int_0^1 d\widetilde{W}\widetilde{W}'\left(\begin{array}{cc}\alpha_2 & \alpha_3\end{array}\right)\left(\left(\begin{array}{c}\alpha_2'\\ \alpha_3'\end{array}\right)\int_0^1\widetilde{W}\widetilde{W}'\left(\begin{array}{cc}\alpha_2 & \alpha_3\end{array}\right)\right)^{-1}\left(\begin{array}{c}\alpha_2'\\ \alpha_3'\end{array}\right)\int_0^1\widetilde{W}d\widetilde{W}'\right)$$

When γ is not known and is estimated by a principal component procedure as proposed in section 5. We have to take into account the estimation errors we introduce in using a two step approach. Let us denote

$$\widehat{Y}=\left(\begin{array}{c}\widehat{\gamma}'\left(Sy_{-1}-Sz_{-1}^N\right)\\ \widehat{\gamma}'Sz_{-1}^N\otimes e_N'\\ w\end{array}\right)=\left(\begin{array}{c}\widehat{Y}_1\\ \widehat{Y}_2\otimes e_N'\\ Y_3\end{array}\right)$$

Starting from (20), most of the preceding equations are unchanged except that Y must be replaced by $\widehat{Y}$ and $\varepsilon\otimes e+\eta$ by

$$\left(\varepsilon+\alpha_1\left(\widehat{Y}_{21}-Y_{21}\right)\right)\otimes e_N'+\eta+\left(\begin{array}{c}\delta_1\\ \delta_2\end{array}\right)\left(\widehat{Y}_{11}-Y_{11}\right)$$

The two key equations (22) and (21) that determine the asymptotic behavior of the estimates $\widehat{\delta_3}$, $\widehat{\alpha_2}$ and $\widehat{\alpha_3}$ have to be replaced by :

$$\begin{aligned}&\frac{1}{T^2N}Y_{12}Y_{12}'\otimes\Omega_\eta^{-1}T\sqrt{N}vec\left(\widehat{\delta_3}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)\\ &\quad=\frac{1}{T\sqrt{N}}vec\left(\Omega_\eta^{-1}\left(\eta+\left(\begin{array}{c}\delta_1\\ \delta_2\end{array}\right)\left(\widehat{Y}_{11}-Y_{11}\right)\right)\widehat{Y}_{12}'\right)\end{aligned}\tag{23}$$

and

$$\frac{1}{T^2}\widehat{Y}_{22}\widehat{Y}_{22}'\otimes\left(\Omega_\eta^{-1}-N\Sigma\right)T\sqrt{N}vec\left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)\tag{24}$$

$$= \frac{1}{T\sqrt{N}} vec \left(\left(\Omega_\eta^{-1} - N\Sigma \right) \left(N \left(\varepsilon + \alpha_1 \left(\widehat{Y}_{21} - Y_{21} \right) \right) \right. \right.$$

$$\left. \left. + \left(\eta + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} \left(\widehat{Y}_{11} - Y_{11} \right) \right) I_T \otimes e_N \right) Y_{22}' \right)$$

which leads to

$$T \begin{pmatrix} \widehat{\alpha_2} & \widehat{\alpha_3} \end{pmatrix} = \frac{1}{T} \left(\varepsilon + \alpha_1 \left(\widehat{Y}_{21} - Y_{21} \right) + \frac{1}{N} \eta I_T \otimes e_N \right) \widehat{Y}_{22}' \left(\frac{1}{T^2} \widehat{Y}_{22} \widehat{Y}_{22}' \right)^{-1} + O_p \left(\frac{1}{\sqrt{T}} \right) \tag{25}$$

and

$$T\sqrt{N} \widehat{\delta_3} = \frac{1}{T\sqrt{N}} \left(\eta + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} \left(\widehat{Y}_{11} - Y_{11} \right) \right) \widehat{Y}_{12}' \left(\frac{1}{T^2 N} \widehat{Y}_{12} \widehat{Y}_{12}' \right)^{-1} + O_p \left(\frac{1}{\sqrt{T}} \right) \tag{26}$$

The first step error induces nuisance parameters as by Proposition 27 and 31 and the related discussion in the text[4] :

$$\begin{aligned} \widehat{Y}_{11} - Y_{11} &= \begin{pmatrix} \widehat{\gamma}_1' - \gamma_1' \\ \widehat{\gamma}_2' - \gamma_2' \end{pmatrix} \left(Sy_{-1} - Sz_{-1}^N \right) \\ &= \begin{pmatrix} O_p \left(\frac{1}{\sqrt{T}} \right) \gamma_1' + O_p \left(\frac{1}{T} \right) \begin{pmatrix} \gamma_2' \\ \gamma_3' \end{pmatrix} \\ \widehat{\gamma}_{21}' O_p \left(\frac{1}{T} \right) + \left(O_p \left(\frac{1}{\sqrt{TN}} \right) \gamma_{21}' + O_p \left(\frac{1}{T\sqrt{N}} \right) \gamma_{22}' \right) \Theta'_{\widehat{\gamma_{1,\perp}}, \gamma_{1,\perp}} \begin{pmatrix} \gamma_2' \\ \gamma_3' \end{pmatrix} \end{pmatrix} \left(Sy_{-1} - Sz_{-1}^N \right) \end{aligned}$$

since $\widehat{\gamma_{21}}' \left(\widehat{\gamma_{1,\perp}}' - \Theta'_{\widehat{\gamma_{1,\perp}}, \gamma_{1,\perp}} \gamma_{1,\perp}' \right) + \left(\widehat{\gamma_{21}}' - \gamma_{21}' \Theta^{-1\prime}_{\widehat{\gamma_{1,\perp}}, \gamma_{1,\perp}} \right) \Theta'_{\widehat{\gamma_{1,\perp}}, \gamma_{1,\perp}} \gamma_{1,\perp}' = \widehat{\gamma}_2' - \gamma_2'$ so that with rough notations

$$\widehat{Y}_{11} - Y_{11} = O_p \left(\frac{1}{\sqrt{T}} \right) Y_{11,\gamma_1} + \left(O_p \left(\frac{1}{T} \right) + O_p \left(\frac{1}{\sqrt{TN}} \right) \right) Y_{11,\gamma_2} + \left(O_p \left(\frac{1}{T} \right) + O_p \left(\frac{1}{\sqrt{TN}} \right) \right) Y_{12}$$

and

$$\begin{aligned} \widehat{Y}_{21} - Y_{21} &= \left(\widehat{\gamma}_1' - \gamma_1' \right) Sz_{-1}^N \\ &= \left(O_p \left(\frac{1}{\sqrt{T}} \right) \gamma_1' + O_p \left(\frac{1}{T} \right) \begin{pmatrix} \gamma_2' \\ \gamma_3' \end{pmatrix} \right) Sz_{-1}^N \end{aligned}$$

or equivalently

$$\widehat{Y}_{21} - Y_{21} = O_p \left(\frac{1}{\sqrt{T}} \right) Y_{21} + O_p \left(\frac{1}{T} \right) Y_{22}$$

It follows that the cross-products

$$\frac{1}{T\sqrt{N}} \left(\widehat{Y}_{11} - Y_{11} \right) \widehat{Y}_{12}' = O_p \left(\sqrt{N} \right)$$

and

$$\frac{1}{T} \left(\widehat{Y}_{21} - Y_{21} \right) \widehat{Y}_{22}' = O_p (1)$$

that do remain in the asymptotic limit of the estimates, we have to get rid of them.

Premultiplying by δ_3' equation (26) we get first that

$$T \delta_3' \widehat{\delta}_3 = \frac{1}{TN} \delta_3' \eta \widehat{Y}_{12}' \left(\frac{1}{T^2 N} \widehat{Y}_{12} \widehat{Y}_{12}' \right)^{-1} + O_p \left(\frac{1}{\sqrt{T}} \right)$$

[4] In the sequel, the symbol $O_p \left(T^\kappa N^\varphi \right)$ represents a matrix with the appropriate dimensions.

which, when T goes to $+\infty$, converges to

$$T\delta_3'\widehat{\delta}_3 \Longrightarrow \frac{1}{N}\sum_{j=1}^{N}\int_0^1 dU_j(s)\,U_j'(s)\left(\frac{1}{N}\sum_{i=1}^{N}\int_0^1 U_i(s)\,U_i(s)'\,ds\right)^{-1}\kappa_1^{-1}$$

if we denote $\delta_3'\Omega_\eta^{1/2}\widetilde{B}_i(s) = U_i(s)$. On the one hand, $\left(\delta_3'\Omega_\eta^{1/2}B_i(s)\right)_i$ are i.i.d random multivariate variables, $(U_i(s))_i$ are the demeaned associated variables, $\frac{1}{N}\sum_{i=1}^N\int_0^1 U_i(s)\,U_i(s)'\,ds$ converges in probability. The same result applies to $\frac{1}{N}\sum_{j=1}^N\int_0^1 dU_j(s)\,U_j(s)'$. Let us denote $E\int_0^1 dU_j(s)\,U_j(s)'$ this second limit. Up to a multiplication by a full rank matrix $\delta_3'\Omega_\eta^{1/2}$, the diagonal terms are centered chi-square random variables and their mean is 0. The other terms are the covariance of mean zero and uncorrelated terms. This limit is a zero matrix. It follows that

$$\sqrt{N}\lim_{T\longrightarrow+\infty}\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)\otimes I_m vec\left(T\delta_3'\widehat{\delta}_3\right) = \sqrt{N}\left(\frac{1}{N}\sum_{j=1}^{N}\int_0^1 vec\left(dU_jU_j'\right)\right)$$

and

$$\sqrt{N}\left(\frac{1}{N}\sum_{j=1}^{N}\int_0^1 vec\left(dU_jU_j'\right)\right)\Longrightarrow\mathcal{N}\left(0,E\left(\int_0^1 U_jU_j'\right)\otimes\delta_3'\Omega_\eta\delta_3\right)$$

since

$$E\left(\int_0^1 U_j\otimes I_m dU_jdU_j'U_j'\otimes I_m\right) = E\left(\int_0^1 U_jU_j'\right)\otimes\delta_3'\Omega_\eta\delta_3$$

We conclude that

$$N\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)\otimes I_m vec\left(T\delta_3'\widehat{\delta}_3\right)$$
$$\left(\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)^{-1}\otimes\left(\delta_3'\Omega_\eta\delta_3\right)^{-1}\right)vec\left(T\delta_3'\widehat{\delta}_3\right)'\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)\otimes I_m$$

that is equal to

$$Nvec\left(T\delta_3'\widehat{\delta}_3\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)\right)$$
$$\left(\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)^{-1}\otimes\left(\delta_3'\Omega_\eta\delta_3\right)^{-1}\right)vec\left(T\delta_3'\widehat{\delta}_3\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)\right)'$$

or equivalently

$$Nvec\left(T\delta_3'\widehat{\delta}_3\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)\right)vec\left(\left(\delta_3'\Omega_\eta\delta_3\right)^{-1}T\delta_3'\widehat{\delta}_3\right)$$
$$= Tr\left(NT^2\delta_3'\widehat{\delta}_3\left(\frac{1}{T^2N}\widehat{Y}_{12}\widehat{Y}_{12}'\right)\widehat{\delta}_3{}'\delta_3\left(\delta_3'\Omega_\eta\delta_3\right)^{-1}\right)$$

converges when N goes to $+\infty$ to a $\chi_2(m-s_1-s_2)$. A consistent estimate of δ_3 is given by an estimate of $\begin{pmatrix}\delta_1' & \delta_2'\end{pmatrix}_\perp'$.

Premultiplying by $\begin{pmatrix}\alpha_2'\\ \alpha_3'\end{pmatrix} = \alpha_{1,\perp}'$ equation (25), we get that

$$T\alpha_{1,\perp}'\begin{pmatrix}\widehat{\alpha_2} & \widehat{\alpha_3}\end{pmatrix} = \frac{1}{T}\alpha_{1,\perp}'\left(\varepsilon+\frac{1}{N}\eta I_T\otimes e_N\right)\widehat{Y}_{22}'\left(\frac{1}{T^2}\widehat{Y}_{22}\widehat{Y}_{22}'\right)^{-1}+O_p\left(\frac{1}{\sqrt{T}}\right)$$

and when T goes to $+\infty$

$$T\alpha_{1,\perp}'\left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right) \Longrightarrow \alpha_{1,\perp}'\int_0^1 d\widetilde{W}(s)\widetilde{W}(s)'\,ds\alpha_{1,\perp}\left(\alpha_{1,\perp}'\int_0^1 \widetilde{W}(s)\widetilde{W}(s)'\,ds\alpha_{1,\perp}\right)^{-1}\kappa_2^{-1}$$

where

$$V\left(\alpha_{1,\perp}'\widetilde{W}(s)\right) = \alpha_{1,\perp}'\left(\Omega_\varepsilon + \frac{1}{N}\Omega_\eta\right)\alpha_{1,\perp}s$$

so that if we denote

$$W_*(s) = \left(\alpha_{1,\perp}'\left(\Omega_\varepsilon + \frac{1}{N}\Omega_\eta\right)\alpha_{1,\perp}\right)^{-1/2}\alpha_{1,\perp}'\widetilde{W}(s)$$

a $m - s_1$ multivariate standard Brownian motion, we get that by a Continuous Mapping Theorem

$$Trace\left\{T\left(\alpha_{1,\perp}'\left(\Omega_\varepsilon + \frac{1}{N}\Omega_\eta\right)\alpha_{1,\perp}\right)^{-1}\alpha_{1,\perp}'\left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right)\left(\frac{1}{T}\widehat{Y}_{22}\widehat{Y}_{22}'\right)\left(\begin{array}{c}\widehat{\alpha_2}' \\ \widehat{\alpha_3}'\end{array}\right)\alpha_{1,\perp}\right\}$$

converges to

$$Trace\left\{\int dW_*W_*'\left(\int W_*W_*'\right)^{-1}\int W_*dW_*'\right\}$$

A consistent estimate of $\alpha_{1,\perp}' = \left(\begin{array}{c}\alpha_2' \\ \alpha_3'\end{array}\right)$ is given by $\widehat{\alpha_{1,\perp}} = \left(\begin{array}{cc}\widehat{\alpha_2} & \widehat{\alpha_3}\end{array}\right)$.