Testing for Breaks in Cointegrated Panels with Common and Idiosyncratic Stochastic Trends

Chihwa Kao, Lorenzo Trapani and Giovanni Urga

CEA@Cass Working Paper Series
WP–CEA–02-2011
Testing for Breaks in Cointegrated Panels with Common and Idiosyncratic Stochastic Trends

Chihwa Kao
Syracuse University (USA)

Lorenzo Trapani
Cass Business School, London (UK)

Giovanni Urga*
Cass Business School, London (UK) and Bergamo University (Italy)

February 8, 2011

Abstract

In this paper, we develop tests for structural change in cointegrated panel regressions with common and idiosyncratic trends. We consider both the cases of observable and non-observable common trends, deriving a Functional Central Limit Theorem for the partial sample estimators under the null of no break. We show that test based on sup-Wald statistics are powerful versus breaks of size $O_p(1/\sqrt{T})$, also proving that power is present when the time of change differs across units and when only some units have a break. Our framework is extended to the case of cross correlated regressors and endogeneity. Monte Carlo evidence shows that the tests have the correct size and good power properties.

JEL codes: C23.

Keywords: Structural change, Panel cointegration, Common stochastic trends, Functional Central Limit Theorem.

*Corresponding author: g.urga@city.ac.uk. Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106, Bunhill Row, London, EC1Y 8TZ (U.K.) Tel. 00.44.(0)20.70408698; Fax. 00.44.(0)20.70408881; http://www.cass.city.ac.uk/faculty/g.urga http://www.cass.city.ac.uk/cea/index.html
1 Introduction

Since the seminal contributions by Perron (1989) and Rappoport and Reichlin (1989), the literature has produced a comprehensive set of results on the changepoint problem in a time series framework - we refer, inter alia, to the articles by Andrews (1993), Andrews and Ploberger (1994), Bai and Perron (1998), and Kejriwal and Perron (2008, 2010). Useful surveys can be found in Banerjee and Urga (2005) and Perron (2006). When extending the framework to a multivariate setting, the literature has shown that the cross sectional dimension can lead to better inference; for example, Bai, Lumsdaine and Stock (1998) show that the estimation of the changepoint in a VAR improves with the dimension of the VAR, due to the presence of cross sectional information. As pointed out by Qu and Perron (2007), a crucial condition is having nonzero correlations across equations, even when including equations without breaks.

Thus, a natural development to enhance the power of tests for structural breaks is to use panel data models, especially when cross sectional dependence is present. Despite the potential usefulness, the inferential theory on structural changes in panels is still underdeveloped. There are a few exceptions: Feng, Kao and Lazarova (2008) and Bai (2010) propose procedures for dating breaks in simple settings with no cross sectional dependence amongst units; Kim (2010a, 2010b) investigates the estimation of change points in panel time trend models with cross-sectional dependence; Breitung and Eickmeier (2010) propose a test for changes in the loadings of a panel factor model.

This paper fills the gap in the literature by proposing an estimation and testing framework for slope parameter instability in cointegrated panel regression; strong cross-sectional dependence is allowed for through the presence of common stochastic trends.

Basic model and extensions

We study a cointegrated panel with unit-specific variables (idiosyncratic shocks) and a set of possibly unobservable variables that are common across all units (common shocks):

\[ y_{it} = \alpha_i + \beta_F' F_t + \gamma_t x_{it} + u_{it}, \] (1)

where \( i = 1, ..., n \) and \( t = 1, ..., T \) and \( \beta_t \) and \( \gamma_t \) are \( R \times 1 \) and \( p \times 1 \) respectively. We assume that (1) is a cointegrating regression for all units \( i \), so that the vector \([F_t', x_{it}]'\) is \( I(1) \) and the error term \( u_{it} \) is stationary for all \( i \). The presence of (strong) cross-sectional dependence arises directly from the common shocks \( F_t \); due to the slope \( \beta \) being homogeneous across units, the term \( \beta' F_t \) represents an \( R \)-dimensional stochastic time effect. As well as having cross dependence due to the common shock \( F_t \), and as well as having homogeneous response to common shocks, we also consider dependence and heterogeneity in the \( x_{it} \)s as

\[ x_{it} = \Gamma_i G_t + w_{it}, \] (2)

2
where $G_t$ is a set of common factors that can be independent of the regressors $F_t$ or overlap with them, and $w_{it}$ is a unit specific (stationary or nonstationary) shock - see also Pesaran (2006) and Kapetanios, Pesaran and Yamagata (2011) for a similar (in spirit) framework. From (2), cross-dependence is accounted for directly (via $F_t$) and indirectly (via the factor structure in $x_{it}$). Heterogeneity in the response to common shocks is also allowed for through the possibly heterogeneous loadings $\Gamma$; also, the response to $F_t$ is allowed to be (indirectly) heterogeneous across individuals if $G_t$ contains $F_t$.

Model (1) encompasses a wide set of models in economics and finance which may be subject to breaks. Such a model may represent a situation whereby the decision $y_{it}$ of microeconomic agent $i$ is influenced by macroeconomic factors $F_t$ and by a set of individual specific characteristics, $\alpha_i$ and $x_{it}$. Examples that have been studied in the literature include, inter alia: demand for household food consumption (see e.g., Dynarski and Sheffrin, 1985, where households are assumed to have the same elasticity to food price, which is the common shock, and to permanent income, which is the idiosyncratic variable); firm size evolving according to a random walk, a case known in the literature as Gibrat’s law (see Sutton, 1997; Geroski et al, 2002); other examples can also be found in micro demand for investment, consumption, labour demand. Moreover, the forward rate unbiasedness hypothesis postulates that the forward rate is an unbiased predictor of the corresponding future spot rate. This hypothesis has been extensively tested for exchange rates (Baillie and Bollerslev, 1989; Liu and Maynard, 2005; Westerlund, 2007). Another example in finance are models for default intensity for firm $i$ at time $t$ expressed as function of common factors (such as U.S. 3-month T-bill and the trailing 1-year returns) and idiosyncratic covariates such as distance to default and trailing 1-year stock return of the firm $i$ (see Das et al., 2007).

Relevant is also the literature on output convergence where output for country $i$ at time $t$ depends on a set of common, to all $n$ countries, technological shocks/knowledge and heterogenous degrees of access to the technological knowledge (Pesaran, 2007; Phillips and Sul, 2007).

Finally, most of the results in this paper are derived assuming zero long run correlation between $[F_{it}', x_{it}', u_{it}]$. However, we show that our framework can be accommodated to allow for endogeneity. Whilst this involves modifying the estimation technique, i.e., from ordinary least squares (OLS) to fully-modified OLS (FMOLS), the limiting distribution of the test and the power versus local alternatives remain unaltered. Other estimation techniques have been proposed in similar contexts - see e.g. Bai, Kao and Ng (2009).

**Main results of this paper**

We focus our attention on testing for the constancy over time of $\theta_t = (\beta_t', \gamma_t')'$, thus developing tests for changes in the cointegration relationship between $y_{it}$ and $(x_{it}', F_t')'$. Considering, for simplicity, the alternative of only one abrupt change at (unknown) time $[Tr]$, the null is $H_0$:
\( \theta_t = \theta \) for all \( t \), whereas the alternative could be defined as

\[
H_a : \theta_t = \begin{cases} 
\theta_1 & \text{for } t = 1, \ldots, [Tr] \\
\theta_2 & \text{for } t = [Tr] + 1, \ldots, T
\end{cases},
\]

with \( \theta_1 \neq \theta_2 \).

This paper makes two contributions to the existing literature. First, we develop a Functional Central Limit Theorem (FCLT) for the partial sample estimators of \( \theta \), considering both the cases of observed and unobserved common shocks. Results are extended to the case of endogeneity by proving an FCLT for the partial sample FMOLS estimators. These results are of independent interest: ordinary large panels asymptotic theory (Phillips and Moon, 1999; Kao, 1999) cannot be applied in our framework due to the strong cross-sectional dependence introduced by the common shocks. Second, we show that tests based on Wald-type statistics (Andrews, 1993; Andrews and Ploberger, 1994) have nontrivial power versus local alternatives of order \( O_p(1/\sqrt{nT}) \), which provides further justification towards the use of panel models in order to enhance the power of tests. Although the tests are constructed under the alternative \( H_a \) of an abrupt and common change, we show that they have power versus other classes of alternatives, e.g. smooth parameter changes. Also, we prove that our tests, albeit designed for the common changepoint alternative \( H_a \), have nontrivial power versus alternatives where series have a break at potentially different points in time. This is a desirable property, since a break could be induced by a change common to all units, but each unit could have different levels of hysteresis and therefore respond with different lag. Also, we study the presence of power when only some units (say \( m_n \)) are subject to a change. We show that, in the extreme case of \( m_n \) finite, tests have power versus local alternatives shrinking as \( O_p \left( \frac{\sqrt{n}}{T} \right) \): when the panel contains many units that do not have a break, there is a loss of power with respect to the case of testing for one unit at a time.

The remainder of the paper is organized as follows. Section 2 introduces the model. Section 3 discusses the asymptotics; test statistics and their null distribution are in Section 4 defines the test statistic and it discusses the local power; extensions (including FMOLS) are in Section 5. In Section 6 we report the finite sample properties, i.e., size and power, of our proposed tests. Section 7 provides concluding remarks. Appendix A contains some preliminary lemmas; the proofs of the main results (test distribution under the null and under local-to-null alternatives) are in Appendix B.

NOTATION. We write integrals involving Brownian motions such as, e.g., \( \int_0^1 B(s) ds \) as \( \int B \) when there is no ambiguity over limits. We define \( \Omega^{1/2} \) to be any matrix such that \( \Omega = \left( \Omega^{1/2} \right) \left( \Omega^{1/2} \right)' \). We use \( \| \cdot \| \) to denote the Euclidean norm of a vector, \( \equiv \) to denote almost surely equality, \( \to \) to denote the ordinary limit, \( \overset{d}{\to} \) to denote convergence in distribution, \( \overset{p}{\to} \) to denote convergence in probability, \( \lfloor x \rfloor \) to denote the integer part, \( B = B(\Omega) \) to denote Brownian motion with covariance matrix \( \Omega \), and \( \bar{B} = B - \int B \) to denote the demeaned version of \( B \). We let \( M < \infty \) be a generic positive number which does not depend on \( n \) or \( T \).
2 Model and assumptions

Consider the panel model with common and idiosyncratic shocks

\[ y_{it} = \alpha_i + \beta_i F_t + \gamma_i x_{it} + u_{it}, \]  

with \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \). We let \( F_t = (F_{1t}, \ldots, F_{Rt})' \) be a \( R \times 1 \) vector of common stochastic trends

\[ F_t = F_{t-1} + \varepsilon_t, \]  

where \( \varepsilon_t \) is a \( 1 \times 1 \) vector of observable individual-specific regressors,

\[ x_{it} = x_{it-1} + \epsilon_{it}, \]  

and \([u_{it}, \varepsilon'_t, \epsilon'_it]'\) the error terms.

When common shocks \( F_t \) are not observable in (3), \( F_t \) can be estimated by a set of observable exogenous variables, \( z_{it} \), defined up to a factor-loading specification as

\[ z_{it} = \lambda'_i F_t + e_{it}, \]  

where \( \lambda_i \) is a vector of factor loadings and \( e_{it} \) is the error term. Whilst results are presented under the simplifying assumption that \( R \) is known, the number of common shocks \( R \) could be estimated using e.g. the criteria discussed in Bai (2004). Also, since panel \( z_{it} \) is employed solely for the estimation of \( F_t \), whilst panel \( y_{it} \) is used to estimate \( \beta \), the number of cross sectional units in the two panels need not be the same. Similarly, the cross sectional index \( i \) needs not refer to the same units: for example, in panel \( y_{it} \), \( i \) could index individuals, but as far as \( z_{it} \) is concerned, \( i \) could index different macro variables such as in Stock and Watson (1999, 2002, 2005).

Consider the following assumptions:

**Assumption 1:** Let \( \omega_{it} = [u_{it}, \varepsilon'_t, \epsilon'_it]' \). We assume that \( (a) \omega_{it} \) is a linear process across \( t \) with \( E \| \omega_{it} \|^4 < \infty \) for some \( \delta > 0 \), and a Beveridge-Nelson decomposition exists such that

\[ w_{it} = w^*_i + R_{it}, \]  

\[ \tilde{x}_{it} = \tilde{x}^*_i + R_{it}, \]  

\[ u_{it} = u^*_i + R_{it}, \]  

\[ e_{it} = e^*_i + R_{it}, \]  

where \( w^*_i \) and \( \tilde{x}^*_i \) are two random walks with long run covariance matrices \( \Omega_w \) and \( \Omega_x \) respectively,
and $u_{it}^*$ and $e_{it}^*$ are i.i.d. processes with variances $\sigma_u^2$, and $\sigma_e^2$ respectively. Also,

$$
\sup_i |R_{jt}\| = O_p\left(\frac{1}{\sqrt{T}}\right),
$$

for $j \in \{w, x, u, e\}$; (b) for a given $t$, $\{u_{it}\}$, $\{e_{it}\}$ are mutually independent across $i$ and independent of $\{\varepsilon_t\}$; (c) $\{x_{it}, F_t\}$ are not cointegrated and $\Omega_x$ and $\Omega_e$ are non singular; (d) the eigenvalues of $\Omega_e$ and the random matrix $\int B_eB'_e$ are distinct with probability 1.

**Assumption 2:** $\|\lambda_i\| \leq M$ and $\frac{1}{n} \sum_{i=1}^n \lambda_i \lambda'_i \rightarrow \Sigma_A$ as $n \rightarrow \infty$, where $\Sigma_A$ is non singular.

Assumption 1(a) is a standard requirement on the amount of serial dependence allowed for, and it enables the asymptotic theory developed by Phillips and Solo (1992) and Phillips and Moon (1999) to hold. The only requirement which is nonstandard is (11), which is needed as a sufficient condition in order to prove a panel functional central limit theorem (FCLT). Assumption 1(b) considers a framework where: (a) regressors are strictly exogenous and (b) no cross sectional dependence is allowed other than the one determined by the presence of the common regressors $F_t$. Particularly, 1(b) also rules out the presence of cross-sectional dependence among the idiosyncratic shocks $x_{i/ts}$. As discussed in the introduction, these restrictions are considered only for the purpose of simplicity of the exposition. We show that the main results of the paper (null distribution and local power of the test) still hold after relaxing each of these assumptions. We discuss the presence of cross-dependence among the idiosyncratic explanatory variables $x_{i/ts}$ and the presence of endogeneity in Section 5. The requirement that $\{e_{it}\}$ is independent of the other innovations considered in 1(b) is needed for the inference on the $F_t$s when they are not observable and it is a standard assumption - see Assumption D in Bai (2004, p. 141). Assumption 1(c) rules out cointegration among regressors. This too is a standard requirement in cointegration analysis - see e.g., Park and Phillips (1988) for discussion. Assumption 1(d) is a standard requirement in large panel factor literature, and it is needed in order to identify the factors $F_t$s in (6) when they are not observable.

Assumption 2 is also standard and it ensures that each factor has a nontrivial contribution towards $y_{it}$.

Let $\tilde{x}_{it} = x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it}$ and $w_t = F_t - \frac{1}{T} \sum_{t=1}^T F_t$. The following proposition is important for developing the asymptotics in this paper.

**Proposition 1** Let Assumption 1 hold. As $(n, T) \rightarrow \infty$, for all $r \in (0, 1)$

$$
\frac{1}{\sqrt{n}T^2} \sum_{i=1}^n \sum_{t=1}^{[T/r]} w_t \tilde{x}_{it}^r = O_p(1).
$$

Proposition 1 states that the asymptotic magnitude of the cross term $\sum_{i=1}^n \sum_{t=1}^{[T/r]} w_t \tilde{x}_{it}^r$ is $O_p(\sqrt{n}T^2)$. The $\sqrt{n}$-convergence is achieved since a CLT holds for the cross-sectional average
\[ \sum_{i=1}^{n} \tilde{x}_{it}, \text{ and this holds since the } \tilde{x}_{it}\text{s are i.i.d. among } i \text{ by Assumption 1(b) and because they have zero mean by construction.} \]

We now turn to partial sample estimation of \( \theta \), studying its asymptotics under the null of no structural change.

### 3 Partial sample estimation

This section contains the FCLT for the partial sample estimator (PSE) of \( \theta \) in (3).

When \( F_i \) is unobservable, we propose a two step approach. First, we derive an estimate of \( F_i \), say \( \hat{F}_i \), using (6). Second, we estimate \( \theta \) using OLS in (3) replacing \( F_i \) with \( \hat{F}_i \). The estimator \( \hat{F}_i \) can be obtained applying principal components to the \( z_{it} \); we refer to Bai (2004).

It is well known that \( F_i \) is identifiable only up to a transformation, say \( H'F_i \) where \( H \) is an \( R \times R \) matrix. Thus, using \( \hat{F}_i \) in (3) allows to estimate \( H^{-1}\beta \) rather than \( \beta \). However, as far as testing is concerned, knowledge of \( H'F_i \) and of \( H^{-1}\beta \) is the same as estimating \( F_i \) and \( \beta \). Hence, for simplicity, we assume \( H \) being a \( R \times R \) identity matrix in this paper.

Let \( \hat{w}_{t,r} = \hat{F}_i - \frac{1}{T} \sum_{t=1}^{[Tr]} \hat{F}_i \) and \( \hat{W}_{it,r} = [\hat{w}_{t,r}, \tilde{x}_{it}']' \); define, similarly, \( \hat{w}_{t,1-r} = \hat{F}_i - \frac{1}{T} \sum_{t=1}^{[Tr]} \hat{F}_i \) and \( \hat{W}_{it,1-r} = [\hat{w}_{t,1-r}, \tilde{x}_{it}']' \). Also, consider

\[ y_{it} = \alpha_i + \beta_i' \hat{F}_i + \gamma_i x_{it} + \nu_{it}, \tag{12} \]

where \( \nu_{it} = u_{it} + \beta_i' \left( F_i - \hat{F}_i \right) \). We define the PSE as

\[ \hat{\theta}_{1[Tr]} = \left[ \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{W}_{it,r} \hat{W}_{it,r}' \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{W}_{it,r} y_{it} \right], \tag{13} \]

\[ \hat{\theta}_{2[Tr]} = \left[ \sum_{i=1}^{n} \sum_{t=[Tr]+1}^{T} \hat{W}_{it,1-r} \hat{W}_{it,1-r}' \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=[Tr]+1}^{T} \hat{W}_{it,1-r} y_{it} \right], \tag{14} \]

for some \( r \in (0, 1) \). Under the null

\[ \hat{\theta}_{1[Tr]} - \theta = \begin{bmatrix} \hat{\beta}_{1[Tr]} - \beta \\ \hat{\gamma}_{1[Tr]} - \gamma \end{bmatrix} = \left[ \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{W}_{it,r} \hat{W}_{it,r}' \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{W}_{it,r} \nu_{it} \right] - \theta = \begin{bmatrix} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{w}_{t,r} \hat{w}_{t,r}' \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{x}_{it} \hat{w}_{t,r}' \\ \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{w}_{t,r} \hat{x}_{it} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{x}_{it} \hat{x}_{it}' \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{w}_{t,r} \nu_{it} \\ \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{x}_{it} \nu_{it} \end{bmatrix}, \]

and a similar expression can be derived for \( \hat{\theta}_{2[Tr]} - \theta \).
Before presenting the main result for the asymptotics of \( \hat{\theta}_{1[T]} \) and \( \hat{\theta}_{2[T]} \), we introduce some notation; let

\[
\sigma_{\zeta}^2 = \sigma_{\zeta}^2 + \sigma_{\Pi}^2,
\]

and

\[
\sigma_{\Pi}^2 = \sigma_{\zeta}^2 \left( \beta' Q_B \Sigma_A Q_B' \beta \right),
\]

and \( \frac{1}{T} \sum_{t=1}^{T} \tilde{w}_t \tilde{w}_t' \xrightarrow{d} \tilde{Q}_B \).

It holds that

**Theorem 1** Let Assumptions 1-2 hold. Then, as \( (n,T) \rightarrow \infty \) with \( \frac{n}{T} \rightarrow 0 \),

\[
\sqrt{nT} \left[ \hat{\theta}_{1[T]} - \theta \right] = \sqrt{nT} \left[ \hat{\beta}_{1[T]} - \beta \right] \xrightarrow{d} 0_{R \times p} \sigma_{\zeta} \left( J^T_0 \tilde{B}_e \tilde{B}_e' \right)^{-1/2} 0_{p \times R} \frac{6}{\sigma_{\zeta} \Omega^{-1/2}_e} \times \left[ \frac{Z_1}{\sqrt{6} C (r; 0)} \right],
\]

\[
\sqrt{nT} \left[ \hat{\theta}_{2[T]} - \theta \right] = \sqrt{nT} \left[ \hat{\beta}_{2[T]} - \beta \right] \xrightarrow{d} 0_{R \times p} \sigma_{\zeta} \left( J^T_0 \tilde{B}_e \tilde{B}_e' \right)^{-1/2} 0_{p \times R} \frac{6}{\sigma_{\zeta} \Omega^{-1/2}_e} \times \left[ \frac{Z_2}{\sqrt{6} C (1; r)} \right],
\]

uniformly in \( r \), where \( \tilde{B}_e (r) \) is a standard, demeaned \( R \)-dimensional Brownian motion, \( Z_1 \) and \( Z_2 \) are two independent \( R \)-dimensional Gaussian random variables and the stochastic process \( \frac{1}{\sqrt{6}} C (r; s) \) is Gaussian, independent of \( \tilde{B}_e (r) \), with independent increments, mean zero and covariance kernel given in (36).

**Remarks**

1.1 Theorem 1 is the building block to implement sequential testing for breaks. The process \( C (r; 0) \) is, in essence, a variance transformed Brownian motion. As far as \( \hat{\beta}_{1[T]} - \beta \) is concerned, it is \( \sqrt{nT} \) consistent, as in Phillips and Moon (1999) and Kao (1999). However, the limiting distribution is not normal, contrary to what typically found in cointegrated panels, but mixed normal. This is due to the shocks \( w_t \) being nonstationary and common across units, which results in the covariance matrix \( (nT)^{-1} \sum_{t=1}^{n} \sum_{t=1}^{T} w_t w_t' \) converging to the random matrix \( \int \tilde{B}_e \tilde{B}_e' \) rather than to a matrix of constants as in standard panel cointegration - see the proof of Lemma A.3.

1.2 Equation (36) in Lemma A.3 in Appendix describes the covariance structure of \( C (r; s) \); note that \( E [C (r; 0)]^2 = \frac{1}{6} r^2 \), and \( E [C (1; r)]^2 = \frac{1}{6} (1 - r)^2 \), which is essentially the same result as Lemma A.4 in Chiang et al. (2002). The covariance structure of the process is \( E [C (r; 0) C (s; 0)] = \frac{1}{6} (r \vee s)^2 \).
1.3 The estimates $\hat{\beta}_{1_{[Tr]}} - \beta$ and $\hat{\gamma}_{1_{[Tr]}} - \gamma$ are asymptotically independent. This result is a consequence of Proposition 1, whereby
\[
\frac{1}{nT^2} \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} w_t w_i' \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{x}_{it} w_i' \right] = \begin{bmatrix} O_p(1) & O_p\left(\frac{1}{\sqrt{n}}\right) \\ O_p\left(\frac{1}{\sqrt{n}}\right) & O_p(1) \end{bmatrix}. \tag{19}
\]
Also, $\hat{\theta}_{1_{[Tr]}} - \theta$ and $\hat{\theta}_{2_{[Tr]}} - \theta$ are independent, as a consequence of both $B(r)$ and $C(r; 0)$ having independent increments.

1.4 A technical note on Theorem 1. The FCLT is shown using a slightly different approach than ordinary panel CLT (Phillips and Moon, 1999). Here, the proof is based on showing that, as $T \to \infty$, the FCLT can hold and on working out the variance for each $r$. That $n \to \infty$ is only incidental to the proof. The proof essentially shows that (17) and (18) hold for all $r$, thereby proving convergence of the finite dimensional distributions; tightness is shown using conventional tightness arguments, based on bounding the remainder terms in the Beveridge-Nelson decomposition of $w_t$, $\tilde{x}_{it}$ and $u_{it}$ - see also the discussion in Phillips and Solo (1992).

As a consequence of Theorem 1, a similar result can also be shown to hold for the case of observable common shocks.

**Proposition 2** Let Assumption 1 hold. Then, as $(n, T) \to \infty$ with $n/T \to 0$, $\sqrt{nT} \left[ \hat{\theta}_{1_{[Tr]}} - \theta \right]$ and $\sqrt{nT} \left[ \hat{\theta}_{2_{[Tr]}} - \theta \right]$ converge to the same distributions as in (17) and (18) respectively, with $\sigma_\zeta$ replaced by $\sigma_u$.

The FCLT derived in Theorem 1 can now be used to derive the distribution of test statistics under the null of no break.

4 Testing

In this section, we consider three statistics: the supremum of the Wald statistic, $\text{Sup}W$, the average Wald statistic, $AveW$, and the logarithm of the Andrews-Ploberger exponential Wald statistic, $\text{Exp}W$ - see Andrews (1993), Andrews and Ploberger (1994).

**Assumption 3:** $r \in (0, 1)$.

Assumption 3 states that the fraction of $T$ at which the change point occurs, $r$, is bounded away from zero and one. Thus, our tests are designed to have power versus mid-sample alternatives, as it is typical in this literature (see Andrews, 1993, p.838).
Consider \( \hat{\theta}_{1[T]} \) and \( \hat{\theta}_{2[T]} \) defined in (13) and (14) respectively, and define, for \( j = 1, 2 \)
\[
\hat{\theta}^*_{j[T]} = \left[ \begin{array}{cc} \hat{\sigma}_{\xi,j}I_R & 0 \\ 0 & \hat{\sigma}_{u,j}I_p \end{array} \right]^{-1} \hat{\theta}_{j[T]},
\]
where \( \hat{\sigma}_{\xi,j} \) and \( \hat{\sigma}_{u,j} \) are consistent for \( \sigma_{\xi}^2 \) and \( \sigma_{u}^2 \) respectively under \( H_0 \). Compute the Wald-type statistic \( W ([T_T]) \) as
\[
W ([T_T]) = \left[ \hat{\theta}^*_{1[T]} - \hat{\theta}^*_{2[T]} \right] \left[ \begin{array}{ccc} \sum_{i=1}^{n} \sum_{i=1}^{[T_T]} \hat{W}_it \hat{W}_it^{-1} & \left( \sum_{i=1}^{n} \sum_{i=[T_T]+1}^{T} \hat{W}_it \hat{W}_it^{-1} \right)^{-1} \\ \left( \sum_{i=1}^{n} \sum_{i=[T_T]+1}^{T} \hat{W}_it \hat{W}_it^{-1} \right)^{-1} & \sum_{i=1}^{n} \sum_{i=1}^{[T_T]} \hat{W}_it \hat{W}_it^{-1} \end{array} \right]^{-1} \left[ \hat{\theta}^*_{1[T]} - \hat{\theta}^*_{2[T]} \right]. \quad (20)
\]
Before showing the limiting distribution for \( W ([T_T]) \) as \( n \) and \( T \) pass to infinity, some preliminary notation is necessary. Let \( s(r) = \left( \int_0^{r} \hat{B}_e B'_e \right)^{-1/2} Z_1 - \left( \int_1^{1} \hat{B}_e B'_e \right)^{-1/2} Z_2 \) and \( M (r) = \left[ \left( \int_0^{1} \hat{B}_e B'_e \right)^{-1} + \left( \int_1^{r} \hat{B}_e B'_e \right)^{-1} \right] \). Define
\[
J (r) = \left[ \frac{C(1,r) - C(r,0)}{\sqrt{r^2 + (1-r)^2}} \right] \left[ M (r) \right]^{-1/2} s(r). \quad (21)
\]
It holds that

**Theorem 2** Let Assumptions 1–3 hold. Then, under \( H_0 \), as \( (n, T) \to \infty \) with \( \frac{2}{T} \to 0 \)
\[
W ([T_T]) \xrightarrow{d} J (r)' J (r) = Q_R(r) + Q_p(r) \quad (22)
\]
where \( Q_R(r) \) and \( Q_p(r) \) are independent and defined as
\[
Q_R(r) = s(r)' \left[ M (r) \right]^{-1} s(r) \quad (23)
\]
and
\[
Q_p(r) = \left\| \frac{(1-r)^2 C (r; 0) - r^2 C (1; r)}{r^4 (1-r)^4 + r^2 (1-r)^4} \right\|^2 \quad (24)
\]

**Remarks**

2.1 Theorem 2 is an application of Theorem 1 and of the Continuous Mapping Theorem. For a given \( r \), \( Q_R(r) \sim \chi^2_R \) and \( Q_p(r) \sim \chi^2_p \); thus, when suitably normalized, the difference of the partial sample estimates has a chi-squared distribution for fixed \( r \) with \( R + p \) degrees of freedom.

2.2 Note that \( J (r)' J (r) \) is a variance transformed, tied-down Bessel process; when suitably normalized by its covariance kernel, as in (20), it has the same distribution as in Andrews (1993).
2.3 In principle, one can construct tests separately for $\beta$ and $\gamma$ using $Q_R(r)$ and $Q_p(r)$ since $Q_R(r)$ and $Q_p(r)$ are independent. Theorem 2 states that if one wants to test only for the constancy of $\beta$ it holds that $W([T\tau]) \xrightarrow{d} Q_R(r)$; if one is interested in testing merely for the constancy of $\gamma$ it holds that $W([T\tau]) \xrightarrow{d} Q_p(r)$.

2.4 Theorem 2 holds under more general conditions than Assumption 1. For example, the theorem still holds if one allows for the presence of (strong) cross-dependence among the $x_{it}$s. As we show in greater detail in Section 5, this is because the limiting distribution of $W([T\tau])$ follows from the asymptotic normality of $[\hat{\theta}'_{1[T\tau]}, \hat{\theta}'_{2[T\tau]}]'$, which still holds for cross correlated $x_{it}$s.

Theorem 2 is valid for any consistent estimators of $\sigma_u^2$ and $\sigma_\zeta^2$ (of course the choice of these estimators will affect the finite sample performance). Although here we propose to use estimators based on the pre- and post- break subsamples, alternatively one could use estimators based on the full sample - see also the discussion in Andrews (1993, p. 833). To estimate $\sigma_u^2$, one could compute

$$\hat{\sigma}_{u,1}^2 = \frac{1}{n(T\tau)} \sum_{i=1}^{n} \sum_{t=1}^{[T\tau]} \left[ y_{it} - \bar{y}_i - \hat{\theta}'_{1[T\tau]} \bar{W}_{it} \right]^2,$$  

which is consistent under $H_0$ (a similar definition would apply for $\hat{\sigma}_{u,2}^2$).

To find a consistent estimator of $\sigma_\zeta^2$, consider the case of $e_{it}$ being i.i.d. From (15) a possible choice is

$$\hat{\sigma}_{\zeta,1}^2 = \hat{\sigma}_{u,1}^2 + \hat{\sigma}_{\Pi,1}^2,$$

and similarly for $\hat{\sigma}_{\zeta,2}^2$. From (16), we have $\hat{\sigma}_{\Pi,1}^2 = \hat{\beta}'_{1[T\tau]} \hat{\sigma}_{u,1}^2 \hat{\beta}_{1[T\tau]}$ with

$$\hat{\sigma}_{\pi,1}^2 = \left[ \frac{1}{(T\tau)^2} \sum_{t=1}^{[T\tau]} \hat{w}_t \hat{w}_t' \right] \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{T\tau} \sum_{t=1}^{[T\tau]} \hat{e}_{it}^2 \right) \hat{\lambda}_i \hat{\lambda}_i' \right] \left[ \frac{1}{(T\tau)^2} \sum_{t=1}^{[T\tau]} \hat{w}_t \hat{w}_t' \right],$$  

where $\hat{\lambda}_i$ is a consistent estimate of $\lambda_i$ and $\hat{e}_{it}$ can be computed as $\hat{e}_{it} = z_{it} - \hat{\lambda}_i \hat{F}_i$. Therefore, we can provide an estimate for $\hat{\sigma}_\zeta^2$ for each subsample as

$$\hat{\sigma}_{\zeta,1}^2 = \hat{\sigma}_{u,1}^2 + \hat{\beta}'_{1[T\tau]} \hat{\sigma}_{\pi,1}^2 \hat{\beta}_{1[T\tau]}.$$  

If $e_{it}$ is serially correlated, a different formula should be used in (26), replacing $\frac{1}{T\tau} \sum_{t=1}^{[T\tau]} \hat{e}_{it}^2$ with e.g. some weighted sum-of-covariances estimator.

**Proposition 3** Suppose Assumptions 1-3 hold. Then, as $(n,T) \to \infty$, under $H_0$, it holds that $\hat{\sigma}_{u,1}^2 \xrightarrow{p} \sigma_u^2$ and $\hat{\sigma}_{\zeta,1}^2 \xrightarrow{p} \sigma_\zeta^2$, and similarly for $\hat{\sigma}_{u,2}^2$ and $\hat{\sigma}_{\zeta,2}^2$.

Following Andrews (1993) and Andrews and Ploberger (1994), we consider three test statis-
tics for the null of no break:

\begin{align*}
SupW & \equiv \max_{[T r^*] \leq [T r] \leq T - [T r^*]} W ([T r]), \\
AveW & \equiv \frac{1}{T} \sum_{[T r] = [T r^*]}^{T - [T r^*]} W ([T r]), \\
ExpW & \equiv \ln \left\{ \frac{1}{T} \sum_{[T r] = [T r^*]}^{T - [T r^*]} \exp \left[ \frac{1}{2} W ([T r]) \right] \right\},
\end{align*}

where \( r^* \) represents the fraction of the sample trimmed away from the beginning and the end of the sample. Using the continuous mapping theorem (CMT) we have the following result:

**Corollary 1** Suppose Assumptions 1-3 hold. Under \( H_0 \) as \( (n, T) \to \infty \) with \( \frac{T}{n} \to 0 \)

\begin{align*}
SupW & \xrightarrow{d} \sup_{r^* \leq r \leq 1 - r^*} \left[ J (r)^t J (r) \right], \\
AveW & \xrightarrow{d} \int_{r^*}^{1 - r^*} \left[ J (r)^t J (r) \right] dr, \\
ExpW & \xrightarrow{d} \ln \left\{ \int_{r^*}^{1 - r^*} \exp \left[ \frac{1}{2} J (r)^t J (r) \right] dr \right\},
\end{align*}

uniformly in \( r \).

Critical values for \( SupW, AveW, \) and \( ExpW \) can be taken from Andrews (1993) and Andrews and Ploberger (1994) in light of Remark 2.2. For example, when \( r^* = 0.15 \) and \( R = p = 1 \), the critical values of the 5% level for \( SupW, AveW, \) and \( ExpW \) are 11.79, 4.61, and 3.22 respectively.

### 4.1 Consistency of the test

In this section, we show that Wald-type tests have non-trivial power versus a general class of local-to-null alternatives of order \( O_p (1/\sqrt{nT}) \). Such alternatives include the case of abrupt change around one common changepoint, but they also include smooth transitions from one regime to another and the possibility of different changepoints for different units in the panel. The time series properties of the local alternatives considered here are in line with the findings in the literature, and particularly the results that the test has nontrivial power versus alternatives of order \( 1/T \) and the presence of power versus smooth transition changes. The cross-sectional properties of the local alternatives are found to be more general than those the test is designed for. The test is shown to have nontrivial power versus alternatives of order \( 1/\sqrt{n} \) (a consequence of the panel approach) and under the case whereby different units may undergo changes at different points in time (if any). We also investigate the power when only some units (possibly
a finite number) have a break, showing that when a finite number of units have a break, test
have power versus local alternatives of order \(O_p(\sqrt{n}/T)\).

We assume the following sequence of local alternatives:

\[
H^{(nT)}_a : \theta^{(nT)}_{it} = \theta + \frac{\sqrt{n}}{mnT} g_i \left( \frac{t}{T} \right),
\]

where \(g_i(\cdot)\) is a \((R + p) \times 1\) arbitrary, finite and non-zero function defined on the unit interval, and \(m_n\) is the number of units for which \(g_i(\cdot) \neq 0\) (i.e. the units that have a nontrivial break); \(m_n\) can be finite or pass to infinity as \(n \to \infty\).

The properties of \(g_i \left( \frac{t}{T} \right)\) are specified in the following assumption.

**Assumption 4:** The function \(g_i \left( \frac{t}{T} \right)\) is nonconstant and it belongs to the class of Riemann integrable functions and as \((n, T) \to \infty\), for all \(r\): (a) \(\frac{1}{m_nT^r} \sum_{i=1}^{m_n} \sum_{t=1}^{[Tr]} \|g_i(\cdot)\|^2 = O_p(1)\); (b) \(\frac{1}{m_nT^{r+1}} \sum_{i=1}^{m_n} \sum_{t=1}^{[Tr]} W_{it}W_{it}^t g_i(\frac{t}{T}) = O_p(1)\); (c) \(\frac{1}{m_nT^2} \sum_{i=1}^{m_n} \sum_{t=1}^{[Tr]} g_i(\frac{t}{T}) W_{it}W_{it}^t g_i(\frac{t}{T}) = O_p(1)\); and (d) \(\frac{1}{m_nT} \sum_{i=1}^{m_n} \sum_{t=1}^{[Tr]} W_{it}^t g_i(\frac{t}{T}) u_{it} = O_p(1)\).

Possible alternative functional forms for \(g_i(\cdot)\) include: a single step function, i.e., \(g_i(s) = 0\) if \(s < r\) and \(g_i(s) = \Delta \theta\) (finite) if \(s \geq r\), which represents a one-time change on \(\theta\) at \([Tr] = [Tr]\); multiple steps functions that represent multiple changes; time trending function \(g_i(\cdot) = t/T\).

Under the alternative, we assume that \(\theta\) can differ across individual. This has a twofold implication. First, having different \(g_i(\cdot)\) across \(i\) entails having, under the alternative, breaks of possibly different magnitude. The case whereby some units have a zero size break (i.e., no change) is taken into account also. Second, allowing for \(g_i(\cdot)\) to differ across \(i\) also allows to consider a specification, under the alternative, where the time of the break (and the presence of breaks itself) is not restricted to be the same across units. This case could be envisaged to take into account the presence of a common source of break but differently timed reactions due to different hysteresis across units. Of course, some units may not have any breaks at all, which is taken into account by allowing for \(m_n\) to be strictly smaller than \(n\).

We now turn to studying the asymptotic distribution of the Wald statistic under the sequence of local alternatives (28). Model (3) can be rewritten as \(y^{(nT)}_{it} = \alpha_i + W_{it}^t \theta^{(nT)}_i + u_{it}\). Similarly, when common shocks are replaced by their estimates \(\hat{W}_{it}\) we have \(y^{(nT)}_{it} = \alpha_i + \hat{W}_{it}^t \hat{\theta}^{(nT)}_i + \hat{u}_{it}\), with \(\hat{u}_{it} = u_{it} + \left(F_{it} - \hat{F}_i\right) \beta^{(nT)}_i\). Let \(\hat{\theta}^{(nT)}_1[Tr]\) and \(\hat{\theta}^{(nT)}_2[Tr]\) be the OLS estimators under the local alternative (28), and let \(\hat{\sigma}^2_{u,j}\) and \(\hat{\sigma}^2_{\hat{c},j}\) be consistent estimators for \(\sigma^2_u\) and \(\sigma^2_{\hat{c}}\) respectively under \(H^{(nT)}_a\). Define

\[
\hat{\theta}^{(nT)}_{j[Tr]} = \begin{bmatrix} \hat{\sigma}^2_{\hat{c},j} I_R & 0 \\ 0 & \hat{\sigma}^2_{u,j} I_p \end{bmatrix}^{-1} \hat{\theta}^{(nT)}_j[Tr].
\]
for $j = 1, 2$, the Wald statistics under the local alternative can be computed as

$$ W^{(nT)} ([T_r]) = \left[ \hat{\theta}_{1[T_r]} - \hat{\theta}_{2[T_r]} \right]' \begin{bmatrix} \left( \sum_{i=1}^{n} \sum_{t=T_{r}+1}^{T} \hat{W}_{it} \hat{W}_{it}' \right)^{-1} + \left( \sum_{i=1}^{n} \sum_{t=T_{r}+1}^{T} \hat{W}_{it} \hat{W}_{it}' \right)^{-1} \right]^{-1} \left[ \hat{\theta}_{1[T_r]} - \hat{\theta}_{2[T_r]} \right]. $$

(29)

The local asymptotic power for the Wald statistics is given in the following theorem:

**Theorem 3** Suppose Assumptions 1-4 hold. Then under $H^{(nT)}_u$, as $(n,T) \rightarrow \infty$ with $\frac{n}{T} \rightarrow 0$, it holds that $W^{(nT)} ([T_r]) \xrightarrow{d} [J(r) + d(r)]' [J(r) + d(r)]$, where $J(r)$ is defined in (21) and $d(r)$ is defined in Appendix.

**Remarks**

3.1 Consider the case $m_n = n$. The test has power versus alternatives shrinking as $O_p (1/\sqrt{nT})$; this is a direct consequence of the $\sqrt{nT}$ rate of convergence in Theorem 1. This finding is consistent with the analysis in Bai, Lumsdaine and Stock (1998), Qu and Perron (2007) and Bai (2010), where it is shown that the quality of the breakpoint estimates improves as the number of time series employed increases.

3.2 Theorem 3 shows that the test has some well-known time series properties, e.g., the presence of power versus “smooth” changes as opposed to abrupt changes for which it is designed for; this is consistent with the findings in Andrews (1993). The test also has some cross-sectional properties: albeit designed for the detection of common changepoints, the test exhibits nontrivial power versus breakpoints located at different times for different time series. Thus, cases whereby a common shock introduces breaks in all units but at different points in time due, e.g., to different levels of hysteresis of inertia across units are encompassed by the test.

3.3 When $m_n$ is $o(n)$, the test has a loss of power. This is not surprising, since only some of the cross sectional information (the one from the units which have a break) is actually relevant. As an extreme, consider the case when the number of units that do have a break is finite, i.e. $m_n = O(1)$. Theorem 3 shows that in this case the test has power versus alternatives of order $O_p(\sqrt{n}/T)$, thus being worse than in the univariate case where nontrivial power is attained versus local to null alternatives of order $O_p(1/T)$. Including cross sectional information is beneficial when there are breaks in the units, whereas it worsens the performance of the tests when many units do not have breaks.

Theorem 3 holds for any choice of the estimators $\hat{\sigma}^2_u$ and $\hat{\sigma}^2_\xi$, as long as they are consistent under $H^{(nT)}_u$. A possible estimator for $\sigma^2_u$ is $\hat{\sigma}^2_{u,1} = \frac{1}{m(T_r)} \sum_{i=1}^{n} \sum_{t=1}^{T_{r}} \left[ y_{it} - \hat{y}_{i} - \hat{\theta}_{1[T_r]} \hat{W}_{it} \right]^2$. To estimate $\sigma^2_\xi$ we propose $\hat{\sigma}^2_{\xi,1} = \hat{\sigma}^2_{u,1} + \hat{\beta}_{1[T_r]} \hat{\sigma}^2_{\pi,1} \hat{\beta}_{1[T_r]}$, where $\hat{\sigma}^2_{\pi,1}$ is defined in equation (26).
and \( \hat{\beta}_{1[T]}^{(nT)} \) is the partial sample OLS estimator for \( \beta \) under \( H_{u}^{(nT)} \); similar definitions can be computed for \( \hat{\sigma}_{u,2}^{2} \) and \( \hat{\sigma}_{\zeta,2}^{2} \).

The following proposition establishes consistency for \( \hat{\sigma}_{u,1}^{2} \) and \( \hat{\sigma}_{\zeta,1}^{2} \) under \( H_{u}^{(nT)} \):

\[ \text{Proposition 4} \quad \text{Suppose Assumptions 1-4 hold. Then under the local alternative hypotheses} \quad H_{a}^{(nT)} \quad \text{defined in equation (28), it holds that as} \quad (n,T) \rightarrow \infty, \quad \hat{\sigma}_{u,1}^{2} \overset{p}{\rightarrow} \sigma_{u}^{2} \quad \text{and} \quad \hat{\sigma}_{\zeta,1}^{2} \overset{p}{\rightarrow} \sigma_{\zeta}^{2}. \]

\[ \text{The same holds for} \quad \hat{\sigma}_{u,2}^{2} \quad \text{and} \quad \hat{\sigma}_{\zeta,2}^{2}. \]

5 Extensions

In this section, we consider two extensions: (a) the presence of cross-sectional correlation among the idiosyncratic shocks \( x_{it} \), and (b) the presence of endogeneity in the cointegration relationship. First, we prove that even though the asymptotic law of the OLS estimator changes when cross-dependence is considered among the \( x_{it}s \), however the asymptotic distribution (and local power) of Wald-type statistics is unaltered. Second, as far as the case of endogeneity is concerned, we develop an FMOLS estimator to accommodate for it, showing that, when using this estimator, results concerning the test remain unchanged.

Cross-sectional dependence among the \( x_{it}s \)

Theorem 1 shows an FCLT for \( \hat{\theta}_{1[T]} \) and \( \hat{\theta}_{2[T]} \), from which the limiting distribution of the Wald statistic \( W (\lfloor T \rfloor) \) can be inferred. We show that asymptotic mixed normality is preserved also when we allow for cross-dependence among the \( x_{it}s \), even though asymptotic orthogonality between \( \hat{\beta} - \beta \) and \( \hat{\gamma} - \gamma \) does not hold any more. Consider a simpler version of (2), viz. \( x_{it} = \Gamma_{i}G_{t} \), where \( G_{t} \) is a set of \( I (1) \) shocks that could contain some of the \( F_{it}s \), and let the following assumptions hold.

\[ \text{Assumption 5:} \quad G_{t} = G_{t-1} + \varepsilon_{t}^{G} \quad \text{with} \quad (a) \quad \{u_{it}\}, \quad \{\varepsilon_{i,t}, \varepsilon_{t}^{G}\}, \quad \text{and} \quad \{e_{it}\} \quad \text{mutually independent across} \quad i \quad \text{for all} \quad t; \quad (b) \quad \text{letting} \quad \omega_{2it} = \{u_{it}, \varepsilon_{i,t}, \varepsilon_{t}^{G}, e_{it}\}', \quad \omega_{2it} \quad \text{is a linear process across} \quad t \quad \text{with mean zero and finite 4th moment and an FCLT holds so that uniformly in} \quad r \quad \text{for all} \quad i \]

\[ \frac{1}{\sqrt{T}} \sum_{t=1}^{[T]} \omega_{2it} \overset{d}{\rightarrow} B_{2G} (r) = \left[ \begin{array}{c} B_{u} (r) \\ B_{e} (r) \\ B_{G} (r) \\ B_{e} (r) \end{array} \right], \]

where \( B_{G} (r) \) is a multivariate Brownian motion, whose elements have covariance matrices \( \sigma_{u}^{2}, \Omega_{\varepsilon}, \Omega_{G} \) and \( \sigma_{e}^{2} ; \) (c) a Beveridge-Nelson decomposition holds for \( \omega_{2it} \) with the remainder bounded as in equation (11).

\[ \text{Assumption 6:} \quad (a) \quad \{\Gamma_{i}\}_{i=1}^{n} \quad \text{is a random sequence independent of} \quad \{u_{it}, \varepsilon_{i,t}, \varepsilon_{t}^{G}, e_{it}\}' \quad \text{with} \quad (i) \quad n^{-1} \sum_{i=1}^{n} \Gamma_{i} \Gamma_{i}' \overset{p}{\rightarrow} \Sigma_{\Gamma} \quad \text{where} \quad \Sigma_{\Gamma} \quad \text{is non singular and} \quad (ii) \quad n^{-1} \sum_{i=1}^{n} \Gamma_{i} \overset{p}{\rightarrow} \hat{\Gamma}; \quad (b) \quad \text{as} \quad (n,T) \rightarrow \infty \]
holds that

\[ \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \Gamma_i G_t G_t' \Gamma_i' \sim \Sigma_1^{1/2} \left( \int_0^r \bar{B}_G \bar{B}_G' \right) \Sigma_1^{1/2} \]

and (under \( \frac{n}{T} \to 0 \))

\[ \frac{1}{\sqrt{nT^2}} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \Gamma_i G_t w_{i,t,r} \sim \bar{\Gamma} \int_0^r \bar{B}_G \bar{B}_G' \]

uniformly in \( r \).

Assumption 5 extends Assumption 1 to take account of the factor structure in the \( x_{it} \)s; for the sake of simplicity, the idiosyncratic component in (2), \( w_{it} \), is omitted here. Assumption 6(a)(ii) considers two possible, alternative cases, whereby the loadings \( \Gamma_i \) have either zero or nonzero mean. This will be shown to play an important role in the asymptotic variance of \( \hat{\theta}_{[Tr]} \).

The following result, which is the equivalent to Theorem 1, holds:

**Proposition 5** Suppose Assumptions 2, 5 and 6 hold. Under \( H_0 \), as \((n,T) \to \infty \) with \( \frac{n}{T} \to 0 \)

\[ \sqrt{nT} \left[ \hat{\theta}_{[Tr]} - \theta \right] \sim \sigma_\zeta \left[ \int_0^r \bar{B}_\varepsilon \bar{B}_\varepsilon' \bar{\Gamma} \int_0^r \bar{B}_\varepsilon \bar{B}_\varepsilon' \Sigma_1^{1/2} (\int_0^r \bar{B}_G \bar{B}_G') \Sigma_1^{1/2} \right]^{-1/2} \times Z, \]

uniformly in \( r \), with \( Z \) an \((R+p)\)-dimensional Gaussian random variable.

Proposition 5 shows, in essence, that the FCLT for \( \hat{\theta}_{[Tr]} - \theta \) is the same as \( \hat{\beta}_{[Tr]} - \beta \) for in Theorem 1 (up to adjusting the dimension from \( R \) to \( p+R \)). This entails that \( W (\{[Tr]\}) \) has, as \((n,T) \to \infty \) with \( \frac{n}{T} \to 0 \), the same distribution as \( Q_R (r) \) in (23), although the dimension changes to \( R+p \). Therefore, tests based on \( W (\{[Tr]\}) \) have the same properties under Assumptions 5 and 6 as under Assumptions 1; this includes power versus local alternatives and consistency of the estimator of \( \sigma_\zeta \).

Proposition 5 also has some important differences with Theorem 1, mainly in terms of the correlation between \( \hat{\beta} - \beta \) and \( \hat{\gamma} - \gamma \) and the presence of \( \sigma_\zeta \), instead of \( \sigma_u \), in the expression of \( \hat{\gamma} - \gamma \). The former is in general not zero, thus not having asymptotic independence, unless \( \bar{\Gamma} = 0 \); this, in essence, is because as \( T \to \infty \), \( T^{-2} \sum_{t=1}^T G_t F_t' \) does not converge to zero even if \( F_t \) and \( G_t \) are independent.

**Endogeneity and FMOLS estimation**

This section considers extending the framework to incorporate the case of endogeneity; we develop a FMOLS estimator and we show that the null distribution of the Wald-type test statistic is the same as in Theorem 2. This is based on showing that an FCLT holds for the partial sample FMOLS estimator, and thus all results derived above still hold in the case of endogeneity.
Assumption 7: (a) Assumption 1(a) holds; (b) a multivariate invariance principle for \( \omega_{it} \) holds, such that, as \( T \to \infty \) for all \( i \), \( \frac{1}{T} \sum_{t=1}^{T} \omega_{it} \xrightarrow{d} B_{i}(r) = \begin{bmatrix} B_{u,i}(r), B'_{\varepsilon}(r), B'_{e,i}(r), B_{e,i}(r) \end{bmatrix}' \), with long run covariance matrix given by \( \text{diag} \{ \Omega, \sigma^2 \} \), where \( \text{diag} \{ \cdot \} \) represents a block diagonal matrix and

\[
\Omega = \begin{bmatrix}
\Omega_u & \Omega_{ue} & \Omega_{ue} \\
\Omega_{eu} & \Omega_{e} & \Omega_{e} \\
\Omega_{eu} & \Omega_{e} & \Omega_{e}
\end{bmatrix},
\]

partitioned conformably with \( \omega_{it} \), where we assume \( \Omega_{ee} \) and \( \Omega_{ee} \) both equal to zero for simplicity; (c) \( \{u_{it}, \varepsilon'_t, e_{it}\} \) and \( \{e_{it}\} \) are two independent groups.

Assumption 7 substitutes Assumption 1, and it considers the presence of endogeneity, which is taken into account through the terms \( \Omega_{eu} \) and \( \Omega_{ee} \) in the long run covariance matrix: both the common and the idiiosyncratic shocks can be correlated with the error term \( u_{it} \). For simplicity, the long run covariance matrix of the \( B_{i}(r) \) is the same for all \( i \). Independence between \( \{e_{it}\} \) and \( \{u_{it}, \varepsilon_t, e_{it}\} \) is still required in order for the asymptotic theory for \( \hat{F}_{i} \) to hold.

Define also

\[
\Lambda = \lim_{T \to \infty} \frac{1}{T} E \begin{bmatrix}
\sum_{t} u_{it}u_{i0} & \sum_{t} u_{it}\Delta F'_{0} & \sum_{t} u_{it}\Delta x'_{i0} \\
\sum_{t} \Delta F_{i} u_{i0} & \sum_{t} \Delta F_{i} \Delta F'_{0} & \sum_{t} \Delta F_{i} \Delta x'_{i0} \\
\sum_{t} \Delta x_{it} u_{i0} & \sum_{t} \Delta x_{it} \Delta F'_{0} & \sum_{t} \Delta x_{it} \Delta x'_{i0}
\end{bmatrix} = \begin{bmatrix}
\Lambda_u & \Lambda_{ue} & \Lambda_{ue} \\
\Lambda_{eu} & \Lambda_e & \Lambda_{ee} \\
\Lambda_{eu} & \Lambda_{ee} & \Lambda_e
\end{bmatrix},
\]

and as before, for simplicity we assume \( \Lambda_{ee} \) and \( \Lambda_{ee} \) equal to zero. Assumption 7 entails (see e.g., Phillips and Durlauf, 1986)

\[
T \begin{bmatrix}
\sum_{t=1}^{T} \hat{W}_{it} \hat{W}'_{it}
\end{bmatrix}^{-1} \begin{bmatrix}
\sum_{t=1}^{T} \hat{W}_{it} \left[ u_{it} + \beta' (w_{it} - \hat{w}_{it}) \right]
\end{bmatrix} \xrightarrow{d} \left( \int B_{i} B'_{i} \right)^{-1/2} \sqrt{\Omega_u + \sigma^2 \left( \beta' \hat{Q}_{B} \Sigma \hat{Q}'_{B} \beta \right)} \times Z + \Lambda_{eu},
\]

for all \( i \). Therefore, the OLS estimator of \( \theta \) is no longer \( \sqrt{n}T \) consistent due to the persistence of the asymptotic bias terms \( \Lambda_{eu} \) and \( \Lambda_{eu} \) across units - see also the discussion in Phillips and Moon (1999, p. 1084).

Let \( \Omega_{(ee)} = \text{diag} \{ \Omega_{ee}, \Omega_{ee} \} \), \( \Lambda_{(ee)} = \text{diag} \{ \Lambda_{ee}, \Lambda_{ee} \} \), where \( \text{diag} \{ \cdot \} \) is understood as a block diagonal matrix with \( \Omega_{ee} \) and \( \Omega_{ee} \) on the main diagonal and similarly for \( \Lambda_{ee} \). Estimates of \( \Omega \) and \( \Lambda \) are based on \( \hat{\Gamma}_{i}(j) = \frac{1}{T} \sum_{t=1}^{T} W_{i,j+t} W'_{i,t} \), with \( W_{i,j}^{+} = \begin{bmatrix} \hat{u}_{it}, \Delta \hat{F}_{i}, \Delta x'_{i0} \end{bmatrix}' \). Estimates of \( \Omega_{i} \) and \( \Lambda_{i} \) can be calculated as \( \hat{\Omega}_{i} = \frac{1}{T} \sum_{j=1}^{T} w \left( \frac{1}{T} \right) \hat{\Gamma}_{i}(j) \) and \( \hat{\Lambda}_{i} = \frac{1}{T} \sum_{j=0}^{T-1} w \left( \frac{1}{T} \right) \hat{\Gamma}_{i}(j) \) respectively, where \( w(x) \) is a kernel and \( h \) is the bandwidth.

The averaged kernel estimators of \( \Omega \) and \( \Lambda \) are calculated as \( \hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} \hat{\Omega}_{i} \) and \( \hat{\Lambda} = \frac{1}{n} \sum_{i=1}^{n} \hat{\Lambda}_{i} \).
\( \frac{1}{n} \sum_{i=1}^{n} \hat{A}_i \). The following assumption characterizes the kernel \( w(x) \).

**Assumption 8:** (a) \( w(0) = 1, w(x) = w(-x), \int_{-\infty}^{\infty} w^2(x) \, dx < \infty \) and \( \lim_{x \to -\infty} |x|^{-q} [1 - w(x)] < \infty \), for some \( q \in (0, \infty) \); (b) as \( n, T \) and \( h \to \infty \), it holds that \( \frac{h}{T} \to 0 \) and \( \frac{h^2n}{T} \to \varepsilon > 0 \).

Assumption 8 is the same as Assumption 11 in Phillips and Moon (1999, p. 1085), and it ensures that

\[
\sqrt{n} \left( \hat{\Omega} - \Omega \right) = o_p(1),
\]
\[
\sqrt{n} \left( \hat{\Lambda} - \Lambda \right) = o_p(1).
\]

These are stronger than simple consistency and they are needed to prove the asymptotics of the FMOLS estimator (see Phillips and Moon, 1999, p. 1109). Note that \( n \) does not play a role in determining the value of the bandwidth \( h \).

Let

\[
\hat{A}^{+}_{(\varepsilon), u} = \hat{A}_{(\varepsilon), u} - \frac{1}{\sqrt{n}} \left( \hat{\Lambda}_{\varepsilon} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) \hat{\Omega}_{(\varepsilon), u}^{-1} \hat{\Omega}_{(\varepsilon), u};
\]

the partial sample estimates for \( \theta \) are defined as

\[
\hat{\theta}^{FM}_{1[T_r]} = \left[ \sum_{i=1}^{n} \sum_{t=1}^{[T_r]} \tilde{W}_{it} \tilde{W}_{it}' \right]^{-1} \times \left[ \sum_{i=1}^{n} \sum_{t=1}^{[T_r]} \tilde{W}_{it} y_{it} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{[T_r]} \left( \tilde{w}_{it} \Delta \hat{F}_{i} \right) \right],
\]

and

\[
\hat{\theta}^{FM}_{2[T_r]} = \left[ \sum_{i=1}^{n} \sum_{t=\lceil T_r \rceil+1}^{T} \tilde{W}_{it} \tilde{W}_{it}' \right]^{-1} \times \left[ \sum_{i=1}^{n} \sum_{t=\lceil T_r \rceil+1}^{T} \tilde{W}_{it} y_{it} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{t=\lceil T_r \rceil+1}^{T} \left( \tilde{w}_{it} \Delta \hat{F}_{i} \right) \right].
\]

This estimation procedure follows similar lines as e.g., in Phillips and Moon (1999, p. 1085); the only difference here is the presence of the normalization term \( \sqrt{n} \), due to the presence of common regressors across units.
The Wald statistic for the null of no structural change is defined as

\[ W^{FM}([Tr]) = \left[ \hat{\theta}^{FM}_{[Tr]} - \hat{\theta}^{FM}_{2[Tr]} \right]^{\prime} \left[ \sum_{i=1}^{n} \sum_{t=1}^{\lfloor Tr \rfloor} \tilde{W}_{it} \tilde{W}_{it}^{\prime} \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=\lfloor Tr \rfloor + 1}^{T} \tilde{W}_{it} \tilde{W}_{it}^{\prime} \right]^{-1} \left[ \hat{\theta}^{FM}_{1[Tr]} - \hat{\theta}^{FM}_{2[Tr]} \right], \]

where

\[ \hat{\theta}^{FM}_{j[Tr]} = \left[ \hat{\Omega}^2 I_R \quad 0 \\ 0 \quad \left( \hat{\Omega}_u - \hat{\Omega}_u \tilde{\Omega}_e \tilde{\Omega}^{-1}_e \tilde{\Omega}_u \right) I_p \right]^{-1} \hat{\theta}^{FM}_{j[Tr]}, \]

with \( \hat{\Omega}_\xi \) being an estimator of

\[ \Omega_\xi = \left( \Omega_u - \Omega_{ue} \Omega_e^{-1} \Omega_{zu} \right) + \sigma_e^2 \left( \beta' \hat{Q}_B \Sigma A \hat{Q}_B \beta \right), \]

constructed as \( \hat{\Omega}_\xi = \left( \hat{\Omega}_u - \hat{\Omega}_{ue} \hat{\tilde{\Omega}}_e \hat{\tilde{\Omega}}^{-1}_e \hat{\tilde{\Omega}}_u \right) + \sigma_e^2 \left( \beta' \hat{\tilde{\beta}} \sigma^2 \hat{\tilde{\beta}} \right), \)

where \( \beta' \sigma^2 \hat{\tilde{\beta}} \) is defined in (26).

It holds that

**Proposition 6** Let Assumptions 2, 7 and 8 hold. Under \( H_0 \), as \( (n,T) \to \infty \) with \( T \to 0 \), it holds that \( W^{FM}([Tr]) \xrightarrow{d} J (r) ' J (r). \)

Proposition states that the limiting distribution of \( W^{FM}([Tr]) \) is the same as the one in Theorem 2; thus, none of the results obtained so far, including Corollary 1 and the critical values for the statistics \( SupW^{FM}([Tr]), AveW^{FM}([Tr]), \) and \( ExpW^{FM}([Tr]), \) change.

### 6 Monte Carlo Simulations

In this section, we use synthetic data to assess the null rejection probabilities and the power properties of \( SupW ([Tr]), AveW ([Tr]), \) and \( ExpW ([Tr]). \) We consider combinations of \( n = \{20, 40, 60, 120, 240, 480\} \) and \( T = \{20, 40, 60, 120, 240, 480\}. \) The Monte Carlo experiments are based on the following design

\[
\begin{align*}
y_{it} &= \alpha_i + \beta_t F_t + \gamma \xi_{it} + u_{it}, \\
F_t &= F_{t-1} + \varepsilon_t, \\
x_{it} &= x_{it-1} + e_{it}, \\
z_{it} &= \lambda_t F_t + e_{it}.
\end{align*}
\]

We assume a single factor and one single idiosyncratic component, such that \( R = p = 1. \) Under the null hypothesis of no structural change, we set the values of the parameters \( \beta = 1 \) and \( \gamma = 1. \) Also, we generate \( \alpha_i \) and \( \lambda_t \) from i.i.d. \( N(0,1) \) and \( N(2,1) \) respectively.

We run a first set of experiments with i.i.d. data, reported in Tables 1a and 2a.
In order to assess power and size under serial correlation and endogeneity, we generate $[u_{it}, \varepsilon_{it}, \epsilon_{it}', \epsilon_{it}']$ as follows. We firstly create a Gaussian i.i.d. sequence $[\tilde{u}_{it}, \tilde{\varepsilon}_{it}, \tilde{\epsilon}_{it}', \tilde{\epsilon}_{it}']$, which ensures that the error term $e_{it}$ is independent of $\tilde{u}_{it}, \tilde{\varepsilon}_{it}$ and $\tilde{\epsilon}_{it}$. Contemporaneous correlation between $\tilde{u}_{it}, \tilde{\varepsilon}_{it}$ and $\tilde{\epsilon}_{it}$ is imposed by premultiplying $\tilde{e}_{t} = [\tilde{u}_{it}, \tilde{\varepsilon}_{it}, \tilde{\epsilon}_{it}']$ by the Choleski factor of the coefficient matrices $\Pi = \begin{bmatrix} I_{[Tr]} & 0 & \rho^F_{1,[Tr],n} \\ 0 & I_{np} & \rho^X_{1X_{np,n}} \\ \rho^F_{1n,[Tr]} & \rho^X_{1X_{n,np}} & I_{n} \end{bmatrix}$.

The coefficients $\rho^F$ and $\rho^X$ represent the correlation coefficients between $\tilde{u}_{it}$ and $\tilde{\varepsilon}_{it}$ and $\tilde{u}_{it}$ and $\tilde{\epsilon}_{it}$ respectively in the new vector $\tilde{e}_{t} = [\tilde{u}_{it}, \tilde{\varepsilon}_{it}, \tilde{\epsilon}_{it}']$; also, $I_{[Tr],n}$ is a $[Tr] \times n$ matrix of ones. Serial correlation is induced by creating $e_{t} = [u_{it}, \varepsilon_{it}', \epsilon_{it}']$ according to an ARMA(1,1) specification defined by

$$e_{t} = \rho_e e_{t-1} + \tilde{e}_{t} + \theta_e \tilde{e}_{t-1}. \quad (34)$$

We report results for $\{\rho^F, \rho^X, \rho_e, \theta_e\} = \{-0.4, -0.4, 0, 0.4\}$ in Tables 1b and 2b.

Size and power are evaluated at 5% level. For the purpose of size distortion assessment, we note that the critical values of the 5% level for $\sigma_{\bar{y}}$, $\sigma_{\bar{e}}$, and $\sigma_{\bar{\alpha}}$ are 11.79, 4.61, and 3.22 respectively, as derived by Andrews (1993) and Andrews and Ploberger (1994). Power assessment is conducted under the alternative hypothesis of structural change in $\theta = [\beta', \gamma']'$, where the break is located at the 40% of the sample. To control for the break magnitude, we set under $H_a$

$$\theta_l = \begin{cases} \theta & \text{for } t < [Tr] \\ (1 + c) \theta & \text{for } t \geq [Tr] \end{cases}$$

where $c$ is a scalar that defines the percentage change in the parameter values. We set $c = 0.5$.

All our results have been obtained using 10,000 iterations; when generating DGPs, the first 1,000 observations are discarded to avoid dependence on the initial conditions. All routines have been written using Gauss 6.0.

Table 1a contains empirical rejection frequencies of the test statistics, $\sigma_{\bar{y}}$, $\sigma_{\bar{e}}$, and $\sigma_{\bar{\alpha}}$, under the null that $\beta$ and $\gamma$ are stable over time. Overall, all three test statistics show good size as $n$ and $T$ increase. The three test statistics are undersized if $n$ and $T$ are small, though size distortion is significantly reduced as $T$ increases. Table 1b illustrates the null
rejection frequencies for tests constructed upon the FMOLS estimator. The test statistics tend to be slightly undersized for small samples, and using the FMOLS estimator leads to a more evident improvement in the sizes for all statistics for small samples, starting from $T = 20$. As samples get larger, results are substantially equivalent as those found when tests are conducted using the OLS estimator, with the only exception being the cases of $(n,T) = (20,120)$ and $(n,T) = (20,240)$ where the three statistics become substantially oversized.

Table 2a gives the power of the test statistics. All tests show very good power properties. The power gain is substantial as $T$ increases and more moderate for increasing sizes of $n$. This result is consistent with the $\sqrt{nT}$ asymptotics of the three tests, as reported in the paper. When using the FMOLS estimator (Table 2b), the power improves even for moderate samples ($n \geq 40$), which is consistent with the efficiency gain with respect to OLS. We note that the power improvement is more pronounced as $n$ increases rather than as $T$ increases. Even with $T = 20$, the power improves substantially as $n$ increases.

7 Conclusions

In this paper, we derived an inferential theory for testing for an unknown common change point in a cointegrated large panel regression. The model we considered contains unit specific regressors but it can also accommodate for the presence of common shocks, thereby allowing for strong cross-sectional dependence. We analyze both the cases of observable and unobservable common shocks, and we prove an FCLT for the partial sample estimators under various assumptions - i.i.d. data, strong cross dependence among idiosyncratic regressors, endogeneity.

For the purpose of testing, we study various transformations of Wald-type statistics, showing that under the null the limiting distributions are nuisance parameters free and depend only on the number of regressors. The distributions of the Wald-type statistics are functionals of the tied-down Bessel process, as found in the univariate stationary regression framework in Andrews (1993). The proposed tests are shown to have nontrivial power versus sequences of local alternatives of order $O_p(1/\sqrt{nT})$; the term $1/\sqrt{n}$ shows the usefulness of the panel approach. Although the tests are designed for the case of one abrupt change common to all units, we show that the tests have power versus smooth, transition-type alternatives (similarly to Andrews, 1993) and also versus heterogeneous changepoints. Monte Carlo evidence shows that tests have the correct size and good power properties, the power gain being substantial as $T$ increases and more moderate for increasing sizes of $n$, consistent with the $\sqrt{nT}$ asymptotics. However, when only some units have a break, our results show that the performance of tests actually becomes worse than in the one-unit-at-a-time case, as tests have power versus local alternatives shrinking at a rate $O_p(\sqrt{n}/T)$ in the extreme case of a finite number of units having a break.

An interesting development of the framework studied in this paper could be the extension to the multiple breaks case, following a similar approach as Kejriwal and Perron (2008, 2010). Also, our test statistics are based on taking the supremum of the Wald-type statistics over a trimmed
interval. Alternatively, the Wald-type statistics could be normalised to take the supremum over the whole sample. This approach is discussed in various contributions (we refer to Csorgo and Horvath, 1997, for a comprehensive review) in a time series setting; it would be interesting to extend it to a panel setting, analysing the role of \( n \to \infty \). This is an exciting research agenda for future work.

Acknowledgement

This is a revised version of an earlier article entitled “Modelling and Testing for Structural Breaks in Panel Cointegration Models with Common and Idiosyncratic Stochastic Trends”. We would like to thank participants at the International Conferences “Common Features in London” (Cass Business School, 16-17 December 2004), "New York Camp Econometrics I" (Saratoga Springs, February 4-5, 2006) and “Breaks and Persistence in Econometrics” (Cass Business School, 11-12 December 2006), "Factor Models in Economics and Finance" (Cass Business School, 4-5 December 2009), and econometrics seminars at Ohio State University, Academia Sinica, Bangor Business School, and Syracuse University for helpful comments. We benefited from discussions with Badi Baltagi, Robert De Jong, Long-Fei Lee, Zongwu Cai and Yupin Hu. The usual disclaimer applies.
References


### Table 1a. Empirical null rejection frequencies at 5% nominal level for tests based on SupW, AveW and ExpW. Data are generated as i.i.d.. The estimation method for \( \theta \) is OLS.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Size for SupW</th>
<th>Panel B: Size for AveW</th>
<th>Panel C: Size for ExpW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n \setminus T )</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>n \setminus T</td>
<td>20</td>
<td>0.0175</td>
<td>0.0375</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>0.0145</td>
<td>0.0236</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0149</td>
<td>0.0260</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0.0151</td>
<td>0.0287</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0.0172</td>
<td>0.0309</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0.0212</td>
<td>0.0285</td>
</tr>
</tbody>
</table>

Table 1a. Empirical null rejection frequencies at 5% nominal level for tests based on SupW, AveW and ExpW. Data are generated as i.i.d.. The estimation method for \( \theta \) is OLS.
## Panel A. Power for $SupW$

<table>
<thead>
<tr>
<th>$n \backslash T$</th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>120</th>
<th>240</th>
<th>480</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0715</td>
<td>0.2085</td>
<td>0.4723</td>
<td>0.7962</td>
<td>0.9972</td>
<td>1.0000</td>
</tr>
<tr>
<td>40</td>
<td>0.0850</td>
<td>0.2332</td>
<td>0.5129</td>
<td>0.9467</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>60</td>
<td>0.0932</td>
<td>0.3281</td>
<td>0.6545</td>
<td>0.9837</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>120</td>
<td>0.1340</td>
<td>0.5551</td>
<td>0.8512</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>240</td>
<td>0.2545</td>
<td>0.9640</td>
<td>0.9697</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>480</td>
<td>0.4195</td>
<td>0.9327</td>
<td>0.9996</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

## Panel B. Power for $AveW$

<table>
<thead>
<tr>
<th>$n \backslash T$</th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>120</th>
<th>240</th>
<th>480</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0859</td>
<td>0.2170</td>
<td>0.4477</td>
<td>0.8364</td>
<td>0.9983</td>
<td>1.0000</td>
</tr>
<tr>
<td>40</td>
<td>0.1055</td>
<td>0.2699</td>
<td>0.5523</td>
<td>0.9655</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>60</td>
<td>0.1172</td>
<td>0.3693</td>
<td>0.6917</td>
<td>0.9927</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>120</td>
<td>0.1686</td>
<td>0.6003</td>
<td>0.8921</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>240</td>
<td>0.2991</td>
<td>0.8142</td>
<td>0.9877</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>480</td>
<td>0.4700</td>
<td>0.9637</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

## Panel C. Power for $ExpW$

<table>
<thead>
<tr>
<th>$n \backslash T$</th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>120</th>
<th>240</th>
<th>480</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1009</td>
<td>0.2448</td>
<td>0.4993</td>
<td>0.8321</td>
<td>0.9980</td>
<td>1.0000</td>
</tr>
<tr>
<td>40</td>
<td>0.1224</td>
<td>0.2838</td>
<td>0.5628</td>
<td>0.9620</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>60</td>
<td>0.1329</td>
<td>0.3874</td>
<td>0.7006</td>
<td>0.9910</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>120</td>
<td>0.1806</td>
<td>0.6151</td>
<td>0.8899</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>240</td>
<td>0.3168</td>
<td>0.8131</td>
<td>0.9850</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>480</td>
<td>0.4894</td>
<td>0.9599</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 2a. Empirical power at 5% nominal level for tests based on $SupW$, $AveW$ and $ExpW$. Data are generated as i.i.d.. The estimation method for $\theta$ is OLS.
Table 1b. Empirical null rejection frequencies at 5% nominal level for tests based on $SupW$, $AveW$ and $ExpW$. Data are generated according to (34). The estimation method for $\theta$ is FMOLS.
Table 2b. Empirical power at 5% nominal level for tests based on $SupW$, $AveW$ and $ExpW$. Data are generated according to (34). The estimation method for $\theta$ is FMOLS.

<table>
<thead>
<tr>
<th>Panel A. Power for $SupW$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n \backslash T$</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>120</td>
</tr>
<tr>
<td>240</td>
</tr>
<tr>
<td>480</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Power for $AveW$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n \backslash T$</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>120</td>
</tr>
<tr>
<td>240</td>
</tr>
<tr>
<td>480</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Power for $ExpW$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n \backslash T$</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>120</td>
</tr>
<tr>
<td>240</td>
</tr>
<tr>
<td>480</td>
</tr>
</tbody>
</table>
Appendix A: Preliminary Lemmas

Henceforth, we use the following notation: \( C_nT = \min \{ \sqrt{T}, T \} \), \( F^0 = T^{-1} \sum_{t=1}^{T} F_t \), and \( F = T^{-1} \sum_{t=1}^{T} \hat{F}_t \).

**Lemma A.1** Under Assumptions 1 and 2, as \( (n,T) \to \infty \) and for all \( r \in (0,1) \)

(a) \( T^{-1} \sum_{t=1}^{T} \| \hat{F}_t - F_t \|^2 = O_p \left( C_n^{-2}T \right) \),
(b) \( T^{-1} \sum_{t=1}^{T} \| \hat{w}_t - w_t \|^2 = O_p \left( C_n^{-2}T \right) \),
(c) \( T^{-1} \sum_{t=1}^{T} w'_t (\hat{w}_t - w_t) = O_p \left( C_n^{-1}T \right) \).

**Proof.** Without loss of generality, we prove the Lemma for \( r = 1 \). Part (a) is taken from Lemma 1 in Bai (2004). Consider part (b):

\[
\frac{1}{T} \sum_{t=1}^{T} \| \hat{w}_t - w_t \|^2 = \frac{1}{T} \sum_{t=1}^{T} \left\| \left( \hat{F}_t - F_t \right) + \left( F - F^0 \right) \right\|^2 \\
\leq \frac{2}{T} \sum_{t=1}^{T} \left[ \| \hat{F}_t - F_t \|^2 + \| F - F^0 \|^2 \right] = I + II.
\]

Part (a) yields \( I = O_p \left( C_n^{-2}T \right) \); as far as \( II \) is concerned, \( \| F - F^0 \|^2 = \left\| T^{-1} \sum_{t=1}^{T} \left( \hat{F}_t - F_t \right) \right\|^2 \leq \left( T^{-1} \sum_{t=1}^{T} \| \hat{F}_t - F_t \|^2 \right) = O_p \left( C_n^{-2}T \right) \). Therefore \( II = O_p \left( C_n^{-2}T \right) \), and thus \( T^{-1} \sum_{t=1}^{T} \| \hat{w}_t - w_t \|^2 = O_p \left( C_n^{-2}T \right) \). To prove part (c), note

\[
\frac{1}{T} \sum_{t=1}^{T} w'_t (\hat{w}_t - w_t) = \frac{1}{T} \sum_{t=1}^{T} \left( F_t - F^0 \right)' \left[ \left( \hat{F}_t - F_t \right) + \left( F - F^0 \right) \right] \\
= \frac{1}{T} \sum_{t=1}^{T} F_t \left( \hat{F}_t - F_t \right) + \left( \frac{1}{T} \sum_{t=1}^{T} w'_t \right) \left( F - F^0 \right) \\
- \frac{1}{T} \sum_{t=1}^{T} \left( \hat{F}_t - F_t \right) = I + II + III.
\]

Lemma B.4(i) in Bai (2004) ensures that \( I = O_p \left( C_n^{-1}T \right) \). As far as \( II \) is concerned, we have \( II = \left[ T^{-1} \sum_{t=1}^{T} w'_t \right]' \left[ T^{-1} \sum_{t=1}^{T} \left( \hat{F}_t - F_t \right) \right] \leq \left[ T^{-1} \sum_{t=1}^{T} w'_t \right]' \left[ T^{-1} \sum_{t=1}^{T} \| \hat{F}_t - F_t \| \right] \); Assumption 1 entails \( T^{-1} \sum_{t=1}^{T} w_t = O_p \left( \sqrt{T} \right) \). From Lemma B.4(iii) in Bai (2004), \( T^{-1} \sum_{t=1}^{T} \| \hat{F}_t - F_t \| = O_p \left( T^{-1/2}C_n^{-1} \right) \); thus, \( II = O_p \left( C_n^{-1} \right) \). As far as \( III \) is concerned, note \( III = \left[ T^{-1} \sum_{t=1}^{T} F_t \right]' \left[ T^{-1} \sum_{t=1}^{T} \left( \hat{F}_t - F_t \right) \right] \), and therefore \( III = O_p \left( C_n^{-1} \right) \) similarly to \( II \).

**Lemma A.2** Under Assumptions 1 and 2, as \( (n,T) \to \infty \) and for all \( r \in (0,1) \).
(a) \( T^{-2} \sum_{t=1}^{[Tr]} \hat{w}_t \hat{w}_t' = T^{-2} \sum_{t=1}^{[Tr]} w_t w_t' + O_p \left( T^{-1/2} C_{nT}^{-1} \right) \), with \( T^{-2} \sum_{t=1}^{[Tr]} w_t w_t' = O_p \left( 1 \right) \);

(b) \( n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{w}_t w_{it} = n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} w_t w_{it} + O_p \left( C_{nT}^{-1} \right) \);

(c) \( n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{w}_t \left( F_t - \hat{F}_t \right) = \sqrt{n} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} w_t' \left( F_t - \hat{F}_t \right) + O_p \left( \sqrt{n} C_{nT}^{-2} \right) \), with \( n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} w_t \left( F_t - \hat{F}_t \right) = O_p \left( \sqrt{n} C_{nT}^{-2} \right) \);

(d) \( n^{-1/2} T^{-2} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} x_{it} \hat{w}_t' = n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} x_{it} w_t' + n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} x_{it} \left( \hat{w}_t - w_t \right)' = O_p \left( 1 \right) + O_p \left( T^{-1} C_{nT}^{-1} \right) \).

**Proof.** We prove the Lemma for \( r = 1 \). For part (a), note that

\[
\frac{1}{T^2} \sum_{t=1}^{T} \hat{w}_t \hat{w}_t' = \frac{1}{T^2} \sum_{t=1}^{T} \left( w_t + \hat{w}_t - w_t \right) \left( w_t + \hat{w}_t - w_t \right)'
\]

\[
= \frac{1}{T^2} \sum_{t=1}^{T} w_t w_t' + \frac{1}{T^2} \sum_{t=1}^{T} w_t \left( \hat{w}_t - w_t \right)'
\]

\[
+ \frac{1}{T^2} \sum_{t=1}^{T} \left( \hat{w}_t - w_t \right) w_t' + \frac{1}{T^2} \sum_{t=1}^{T} \left( \hat{w}_t - w_t \right) \left( \hat{w}_t - w_t \right)'
\]

\[
= I + II + III + IV.
\]

Assumption 1 ensures that \( I = O_p \left( 1 \right) \). As far as II and III are concerned, application of the Cauchy-Schwartz inequality and of Lemma A.1(a) ensures that they are bounded by

\[
II \leq \frac{1}{T^2} \left( \sum_{t=1}^{T} \left\| w_t \right\|^2 \right)^{1/2} \left( \sum_{t=1}^{T} \left\| \hat{w}_t - w_t \right\|^2 \right)^{1/2}
\]

\[
= \frac{1}{T^2} O_p \left( T \right) O_p \left( \sqrt{T} \right) = O_p \left( \frac{1}{\sqrt{TC_{nT}}} \right).
\]

As far as IV is concerned, using Lemma A.1(b) we have \( \left\| \sum_{t=1}^{T} \hat{w}_t - w_t \right\| \leq T^{-2} \sum_{t=1}^{T} \left\| \hat{w}_t - w_t \right\|^2 = O_p \left( T^{-1} C_{nT}^{-2} \right) \). Hence, \( T^{-2} \sum_{t=1}^{T} w_t w_t' = T^{-2} \sum_{t=1}^{T} w_t w_t' + O_p \left( T^{-1/2} C_{nT}^{-1} \right) + O_p \left( T^{-1} C_{nT}^{-2} \right) \).

Turning to part (b), \( n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{w}_t u_{it} = n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} w_t u_{it} + n^{-1/2} T^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \left( \hat{w}_t - w_t \right) u_{it} \). The Cauchy-Schwartz inequality and Lemma A.1(b) entail

\[
\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left( \hat{w}_t - w_t \right) u_{it} \leq \left( \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{w}_t - w_t \right\|^2 \right)^{1/2} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\| u_{it} \right\|^2 \right)^{1/2} = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

(35)
To prove (c) consider

\[
\frac{\sqrt{n}}{T} \sum_{i=1}^{T} \tilde{w}_t'(F_t - \tilde{F}_t) = \frac{\sqrt{n}}{T} \sum_{i=1}^{T} w_t'(F_t - \tilde{F}_t) + \frac{\sqrt{n}}{T} \sum_{i=1}^{T} (\tilde{w}_t - w_t)'(F_t - \tilde{F}_t) = I + II.
\]

From Lemma A.1(c) it follows that \( I = O_p\left(\sqrt{n}C_{nT}^{-1}\right) \). As far as \( II \) is concerned,

\[
II \leq \sqrt{n}\left(\frac{1}{T} \sum_{i=1}^{T} \|\tilde{w}_t - w_t\|^2\right)^{1/2} \left(\frac{1}{T} \sum_{i=1}^{T} \|F_t - \tilde{F}_t\|^2\right)^{1/2}
= \sqrt{n}O_p\left(\frac{1}{C_{nT}}\right)O_p\left(\frac{1}{C_{nT}}\right) = O_p\left(\frac{\sqrt{n}}{C_{nT}}\right).
\]

To prove part (d), let \( n^{-1/2} T^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{x}_it\tilde{w}_t + n^{-1/2} T^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{x}_it(\tilde{w}_t - w_t)' = I + II \). Proposition 1 (proved below) ensures that \( I = O_p(1) \). As far as \( II \) is concerned, note that the sequence \( \tilde{x}_it(\tilde{w}_t - w_t)' \) has zero mean and is conditionally i.i.d. across \( i \); thus, a CLT yields \( \sum_{i=1}^{n} \tilde{x}_it(\tilde{w}_t - w_t)' = O_p(\sqrt{n}) \); the rest of the proof, to show that \( \sum_{t=1}^{T} \tilde{x}_it(\tilde{w}_t - w_t)' = O_p(\sqrt{nC_{nT}^{-1}}) \), is similar to that of Lemma B.4(i) in Bai (2004), and thus passages are omitted.

**Lemma A.3** Let \( x_i = x_{i-1} + u_i^x \), with \( u_i^x \) a zero mean, unit variance MDS, and let \( u_i \) be an MDS independent of \( u_i^x \), with unit variance. Then, as \( T \to \infty \) for \( a < c < b < d \)

\[
E \left\{ \frac{1}{T} \sum_{i=\lceil Ta \rceil}^{\lceil Tb \rceil} \left[ x_i - \frac{1}{T(b-a)} \sum_{i=\lceil Ta \rceil}^{\lceil Tb \rceil} x_i \right] u_i \times \frac{1}{T} \sum_{i=\lceil Tc \rceil}^{\lceil Td \rceil} \left[ x_i - \frac{1}{T(d-c)} \sum_{i=\lceil Tc \rceil}^{\lceil Td \rceil} x_i \right] u_i \right\}
= \frac{(d-a)(b-c)}{(b-a)(d-c)} \left[ \frac{1}{6} (b-c)^2 \right].
\]

**Proof.** Let \( \tilde{x}_i \) be the demeaned version of \( x_i \) and note

\[
E \left[ \frac{1}{T^2} \sum_{i=\lceil Ta \rceil}^{\lceil Tb \rceil} \tilde{x}_i u_i \right] = E \left[ \frac{1}{T^2} \sum_{i=\lceil Tc \rceil}^{\lceil Td \rceil} \tilde{x}_i \tilde{x}_j u_i u_j \right] = E \left[ \frac{1}{T^2} \sum_{i=\lceil Ta \rceil}^{\lceil Tb \rceil} \tilde{x}_i^2 u_i^2 \right] = E \left( \frac{1}{T^2} \sum_{i=\lceil Ta \rceil}^{\lceil Tb \rceil} \tilde{x}_i^2 \right)
\]
due to the independence of $u_i$ and to $E(u_i^2) = 1$. We have

$$E\left(\frac{1}{T^2} \sum_{i=[Tc]}^{[Tb]} x_i^2\right) = E\left(\frac{1}{T^2} \sum_{i=[Ta]}^{[Tb]} x_i^2\right)$$

$$+ E\left[\frac{1}{T^2} \sum_{i=[Ta]}^{[Tb]} \left(\frac{1}{T} \sum_{i=[Ta]}^{[Tb]} x_i\right) \left(\frac{1}{T} \sum_{i=[Tc]}^{[Td]} x_i\right)\right]$$

$$- E\left[\frac{1}{T^2} \sum_{i=[Ta]}^{[Tb]} \left(\frac{1}{T} \sum_{i=[Tc]}^{[Td]} x_i\right) \left(\frac{1}{T} \sum_{i=[Tc]}^{[Td]} x_i\right)\right]$$

$$= I + II - III - IV.$$

As far as $I$ is concerned, it holds that, as $T \to \infty$, $I = T^{-2} \sum_{i=[Tc]}^{[Tb]} E(x_i^2) = T^{-2} \sum_{i=[Tc]}^{[Tb]} i \to \int_c^b u du = \frac{1}{2} (b^2 - c^2)$. Also

$$II = \frac{b - c}{(b-a)(d-c)} \frac{1}{T^3} \sum_{i=[Ta]}^{[Tb]} \sum_{j=[Tc]}^{[Td]} E(x_i x_j)$$

$$= \frac{b - c}{(b-a)(d-c)} \frac{1}{T^3} \left[\sum_{i=[Ta]}^{[Tb]} \sum_{j=[Tc]}^{[Td]} E(x_i x_j) + \sum_{i=[Tc]}^{[Td]} \sum_{j=[Tb]}^{[Tc]} E(x_i x_j)\right]$$

$$= \frac{b - c}{(b-a)(d-c)} \frac{1}{T^3} \left[T(d-c) \sum_{i=[Ta]}^{[Tc]} i + \sum_{i=[Tc]}^{[Td]} \sum_{j=[Tb]}^{[Tc]} (i \land j) + T(d-b) \sum_{i=[Tc]}^{[Td]} i\right]$$

$$\to \frac{b - c}{(b-a)(d-c)} (d-c) \int_a^c u du + \int_c^b \int_c^b (u \land v) \, du \, dv + (d-b) \int_c^b u du$$

$$= \frac{1}{2} \frac{(b-c)(d-c)}{(b-a)(d-c)} (c^2 - a^2) + \frac{(b-c)}{3} \frac{(b-c)}{(b-a)(d-c)} (b^3 - c^3)$$

$$+ \frac{1}{2} \frac{(b-c)(d-b)}{(b-a)(d-c)} (b^2 - a^2).$$
Similar passages yield

\[
III = \frac{1}{b-a} \frac{1}{T^3} \sum_{i=[Tc]}^{[Tb]} \sum_{j=[Ta]}^{[Tb]} E(x_i x_j) \\
= \frac{1}{b-a} \frac{1}{T^3} \left[ \sum_{i=[Ta]}^{[Tb]} \sum_{j=[Tc]}^{[Tb]} E(x_i x_j) + \sum_{i=[Tc]}^{[Tb]} \sum_{j=[Tc]}^{[Tb]} E(x_i x_j) \right] \\
= \frac{1}{b-a} \frac{1}{T^3} \left[ T(b-c) \sum_{i=[Ta]}^{[Tc]} i + \sum_{i=[Tc]}^{[Tb]} \sum_{j=[Tc]}^{[Tb]} (i \land j) \right] \\
\rightarrow \frac{1}{b-a} \left[ \int_a^c udu + \int_c^b \int_c^b (u \land v) dudv \right] \\
= \frac{1}{2b-a} (c^2 - a^2) + \frac{1}{3} \frac{b^3 - c^3}{b-c},
\]

\[
IV = \frac{1}{d-c} \frac{1}{T^3} \sum_{i=[Tc]}^{[Tb]} \sum_{j=[Tc]}^{[Td]} E(x_i x_j) \\
= \frac{1}{d-c} \frac{1}{T^3} \left[ \sum_{i=[Tc]}^{[Tb]} \sum_{j=[Tc]}^{[Td]} E(x_i x_j) + \sum_{i=[Tc]}^{[Tb]} \sum_{j=[Td]}^{[Tb]} E(x_i x_j) \right] \\
= \frac{1}{d-c} \frac{1}{T^3} \left[ \sum_{i=[Tc]}^{[Tb]} \sum_{j=[Tc]}^{[Td]} (i \land j) + T(d-b) \sum_{i=[Tc]}^{[Tb]} i \right] \\
\rightarrow \frac{1}{d-c} \left[ \int_c^b \int_c^b (u \land v) dudv + (d-b) \int_c^b udu \right] \\
= \frac{1}{3} \frac{b^3 - c^3}{d-c} + \frac{1}{2} \frac{d-b}{d-c} (b^2 - c^2).
\]

Putting everything together, (36) follows.
Appendix B: Proofs and Lemmas

Proof of Proposition 1. Define $\gamma_i[Tr] = T^{-2} \sum_{t=1}^{[Tr]} w_t x_{it}$. Then, the BN decompositions (7) and (8) in Assumption 1 yield

$$
\gamma_i[Tr] = \frac{1}{T^2} \sum_{t=1}^{[Tr]} w_t^* x_{it} + R_i[Tr] = \gamma_i[Tr] + R_i[Tr],
$$

with $R_i[Tr] = O_p(T^{-1/2})$ due to (11). Let $C$ be the $\sigma$-field associated with $\{w_t\}_{t=1}^T$ and consider $\gamma_i[Tr]$; conditioning on $C$, Assumption 1 ensures that $\gamma_i[Tr]$ is i.i.d. across $i$. Also, $E \left[ \gamma_i^* \gamma_i^* | C \right] = T^{-2} \sum_{t=1}^{[Tr]} w_t^* E(\tilde{x}_{it}) = 0$ for all $r$, since $\tilde{x}_{it}$ has zero mean by construction. Letting $I_i$ denote the $\sigma$-field defined as the union of $C$ and $\{\gamma_i^*, \ldots, \gamma_T^*\}$, we have $E \left[ \gamma_i^* | I_{i-1} \right] = E \left[ \gamma_i^* | C \right] = 0$. Therefore $\{\gamma_i[Tr], I_i\}$ is a zero mean MDS for all $r$. A Liapunov condition can be proved such that for some $\delta > 0$, $E \left[ \left\| \gamma_i[Tr] \right\|^{2+\delta} | C \right] < \infty$. This holds because

$$
\left\| \gamma_i[Tr] \right\|^{2+\delta} = \left\| \frac{1}{T^2} \sum_{t=1}^{[Tr]} w_t^* x_{it} \right\|^{2+\delta} \leq M \frac{1}{T^{2(2+\delta)}} \sum_{t=1}^{[Tr]} \left\| w_t^* \right\|^{2+\delta} \left\| x_{it} \right\|^{2+\delta}
$$

where the last equality holds in light of Assumption 1(a) and Theorem 5.2 in Park and Phillips (1999). Thus, $\left\| \gamma_i[Tr] \right\|^{2+\delta}$ is bounded by $O_p(1)$ for all $i$ as $T \to \infty$, and therefore an MDS CLT can be applied such that, for all $r$, $n^{-1/2} \sum_{i=1}^n \gamma_i[Tr] = O_p(1)$ as $(n, T) \to \infty$. Thus, as far as $\gamma_i[Tr]$ is concerned, we have

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_i[Tr] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_i[Tr] + \frac{1}{\sqrt{n}} \sum_{i=1}^n R_i[Tr] = O_p(1) + O_p \left( \sqrt{\frac{n}{T}} \right).
$$

Thus, as $(n, T) \to \infty$ under $n/T \to 0$, $n^{-1/2} \sum_{i=1}^n \sum_{t=1}^T w_t x_{it}^t = O_p(1)$. 

Lemma B.1 Under Assumptions 1-2, as $(n, T) \to \infty$ it holds that, for all $r \in (0, 1)$

(a) $\frac{1}{nT^2} \sum_{i=1}^n \sum_{t=1}^{[Tr]} W_{it} W_{it}^t \overset{d}{\rightarrow} \left[ \begin{array}{cc} f_0^r B_{x} B_{x}^t & 0 \\ 0 & \frac{1}{g} r^2 \Omega_{x} \end{array} \right]$, 

(b) $\frac{1}{nT^2} \sum_{i=1}^n \sum_{t=1}^{[Tr]} \hat{W}_{it} \hat{W}_{it}^t \overset{d}{\rightarrow} \left[ \begin{array}{cc} f_0^r \hat{B}_{x} \hat{B}_{x}^t & 0 \\ 0 & \frac{1}{g} r^2 \Omega_{x} \end{array} \right]$, 

35
Proof. Consider part (a). Proposition 1 implies that the off-diagonal terms of the matrix converge in probability to zero. Also, \( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} w_t w_t' = \frac{1}{T} \sum_{t=1}^{[Tr]} w_t w_t' \), and a standard FCLT yields the result. Finally, standard joint LLN - see Phillips and Moon (1999) - yields \( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \bar{x}_{it} \bar{x}_{it}' \overset{p}{\to} \frac{1}{6} \beta^2 \Omega_\epsilon \). To prove part (b), note

\[
\frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \bar{w}_{it} w_{it} \overset{d}{\to} \sigma_\epsilon \left( \int_0^T \bar{B}_\epsilon \frac{\sigma_\epsilon}{0_{n \times R}} \Omega_{\epsilon}^{1/2} \right) \times B(r),
\]

where the last equality holds from Lemma A2.(a). The desired result holds as long as \( (n, T) \to \infty \).

**Lemma B.2** Under Assumptions 1-2, as \( (n, T) \to \infty \) with \( \frac{n}{T} \to 0 \) it holds that, for all \( r \in (0, 1) \)

\[
\frac{1}{\sqrt{n^2}} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \bar{x}_{it} v_{it} \overset{d}{\to} \sigma_\epsilon \left( \int_0^T \bar{B}_\epsilon \frac{\sigma_\epsilon}{0_{n \times R}} \Omega_{\epsilon}^{1/2} \right) \times C(r; 0).
\]

**Proof.** The Lemma is an FCLT for cointegrated panel data. To prove it, note first that the covariance matrix is diagonal in light of Proposition 1; thus, we show separately that

\[
\frac{1}{\sqrt{n^2}} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \bar{w}_{it} v_{it} \overset{d}{\to} \sigma_\epsilon \left( \int_0^T \bar{B}_\epsilon \frac{\sigma_\epsilon}{0_{n \times R}} \Omega_{\epsilon}^{1/2} \right) \times B(r),
\]

\[
\frac{1}{\sqrt{n^2}} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \bar{x}_{it} v_{it} \overset{d}{\to} \sigma_\epsilon \Omega_{\epsilon}^{1/2} \times C(r; 0).
\]

Consider first (37), rewritten as

\[
\frac{1}{\sqrt{n^2}} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \bar{w}_{it} v_{it} = \frac{1}{T} \sum_{t=1}^{[Tr]} \bar{w}_t \left[ \sum_{i=1}^{n} \frac{1}{\sqrt{n}} (u_{it} + \beta' Q_{B,T} \lambda_i e_{it}) \right] + o_p (1);
\]

using the BN decomposition, this is equivalent to

\[
\frac{1}{T} \sum_{t=1}^{[Tr]} \bar{w}_t \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (u_{it} + \beta' Q_{B,T} \lambda_i e_{it}) \right] = \frac{1}{T} \sum_{t=1}^{[Tr]} \bar{w}_t \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (u_{it}^* + \beta' Q_{B,T} \lambda_i e_{it}^*) \right] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{w_i[T]} + o_p (1).
\]
Let \( u_{it}^* + \beta^t \bar{Q}_{B,T} \lambda e_{it}^* = v_{it}^* \); for all \( n \), conditional on \( C \), \( \tilde{w}_t^* \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it}^* \right) \) is a zero mean MDS with covariance matrix

\[
E \left[ \frac{1}{T} \sum_{t=1}^{[T]} \tilde{w}_t^* \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it}^* \right) \right]^2 = \frac{1}{T^2} \sum_{t=1}^{[T]} \tilde{w}_t^* \tilde{w}_t'^{*} E \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it}^* \right)^2
\]

so that, as \( n, T \rightarrow \infty \), \( E \left[ T^{-1} \sum_{t=1}^{[T]} \tilde{w}_t^* \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it}^* \right) \right]^2 = \sigma_{\xi}^2 \int_0^T \tilde{B}_e \tilde{B}_e'^* \), in light of Lemma B.1. Thus, for all \( r \), \( T^{-1} \sum_{t=1}^{[T]} \tilde{w}_t^* \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it}^* \right) \sim \sigma_{\xi} \frac{r}{\int_0^T \tilde{B}_e \tilde{B}_e'^*} \times Z \), where \( Z \) is a standard normal of dimension \( R \). Tightness follows if

\[
\sup_r \left| \frac{1}{T} \sum_{t=1}^{[T]} \tilde{w}_t \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it} \right) - \frac{1}{T} \sum_{t=1}^{[T]} \tilde{w}_t \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it}^* \right) \right| = o_p(1);
\]

however, this is equivalent to \( \sup_r \left| T^{-1} \sum_{t=1}^{[T]} \tilde{w}_t \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it} \right) - T^{-1} \sum_{t=1}^{[T]} \tilde{w}_t \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{it}^* \right) \right| = o_p(1) = \sup_r \left| n^{-1/2} \sum_{i=1}^{n} R_{uai, [T]} \right|. \) Neglecting the \( o_p(1) \) term and using (11)

\[
\sup_r \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{uai, [T]} \right| \leq \sup_r \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| R_{uai, [T]} \right| \right| \leq \sqrt{n} \sup_{i, r} \left| R_{uai, [T]} \right| = \sqrt{n} O_p \left( \frac{1}{\sqrt{T}} \right),
\]

which vanishes as \( n, T \rightarrow \infty \) under \( \frac{r}{T} \rightarrow 0 \).

As far as (38) is concerned, we have, applying Lemma A.2(d)

\[
\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \tilde{x}_{it} v_{it} = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \tilde{x}_{it} u_{it}^* + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{xui, [T]} + o_p(1).
\]

Let \( \xi_{nt}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{x}_{it} u_{it}^* \); \( \xi_{nt}^* \) is an MDS across \( t \). It holds that

\[
E \left( \frac{1}{T} \sum_{t=1}^{[T]} \xi_{nt}^* \right)^2 = \frac{1}{T^2} \sum_{t=1}^{[T]} \sum_{s=1}^{[T]} E \left( \xi_{nt}^* \xi_{ns}^* \right)
\]

\[
= \frac{1}{T^2} \sum_{t=1}^{[T]} E \left( \xi_{nt}^* \right) = \sigma_u^2 \frac{1}{T^2} \sum_{t=1}^{[T]} \frac{1}{n} \sum_{i=1}^{n} E \left( x_{it}^* - \frac{1}{T} \sum_{t=1}^{[T]} x_{it}^* \right)^2
\]

\[
= \sigma_u^2 \frac{1}{T^2} \sum_{t=1}^{[T]} E \left( x_{it}^* - \frac{1}{T} \sum_{t=1}^{[T]} x_{it}^* \right)^2.
\]
Applying Lemma A.3 with \( a = c = 0 \) and \( b = d = r \), it holds that \( \frac{1}{T^2} \sum_{t=1}^{[Tr]} E \left( x_{it}^* - \frac{1}{T} \sum_{i=1}^{[Tr]} x_{it}^* \right)^2 \rightarrow \Omega_{\epsilon} \frac{r^2}{T^2} \). That \( n \rightarrow \infty \) is only incidental to the main argument of the proof; thus, as \((n,T) \rightarrow \infty\), \( E \left( \frac{1}{T} \sum_{i=1}^{[Tr]} \xi_{it}^* \right)^2 = \frac{1}{b} \sigma_u^2 \Omega_{\epsilon} r^2 \). In order to prove tightness, similar arguments as above entail

\[
\sup_r \left| \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{l=1}^{[Tr]} \xi_{it} u_{it} - \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{l=1}^{[Tr]} \xi_{it}^* u_{it} \right| = \sup_r \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{xui,[Tr]} \right| + o_p(1) \leq \sqrt{n} \sup_{i,t} |R_{xui,[Tr]}| + o_p(1) = O_p \left( \sqrt{\frac{n}{T}} \right) + o_p(1);
\]

thus, an FCLT holds as \((n,T) \rightarrow \infty\) under \( \frac{T}{n} \rightarrow 0 \). The covariance structure of \( C(\cdot;\cdot) \) can be calculated using (36) in Lemma A.3, recalling that the long run variances of \( x_{it} \) and \( u_{it} \) are, respectively, \( \Omega_{\epsilon} \) and \( \sigma_u^2 \).

**Proof of Theorem 1.** The proof follows from Lemmas B.1 and B.2.

**Proof of Theorem 2.** To prove the theorem, note \((\hat{\theta}_{1[Tr]}^* - \hat{\theta}_{2[Tr]}^*) = (\hat{\theta}_{1[Tr]}^* - \theta) - (\hat{\theta}_{2[Tr]}^* - \theta)\), and therefore

\[
\sqrt{nT} \left( \hat{\theta}_{1[Tr]}^* - \hat{\theta}_{2[Tr]}^* \right) = \left[ I \quad -I \right] \left[ \begin{array}{c} \sqrt{nT} \left( \hat{\theta}_{1[Tr]}^* - \theta \right) \\ \sqrt{nT} \left( \hat{\theta}_{2[Tr]}^* - \theta \right) \end{array} \right],
\]

where \( I \) is \((R + p) \times (R + p)\) identity matrix. Theorem 1 and the continuous mapping theorem entail, for any consistent estimators \( \hat{\sigma}_u^2 \) and \( \hat{\sigma}_\epsilon^2 \)

\[
\sqrt{nT} \left( \hat{\theta}_{1[Tr]}^* - \hat{\theta}_{2[Tr]}^* \right) \overset{d}{\rightarrow} \left[ I \quad -I \right] \left[ \begin{array}{c} \left( \int_0^r B_0 B_0' \right)^{-1/2} Z_1 \\ \frac{\sqrt{8}}{T} \Omega_{\epsilon}^{-1/2} C(r;0) \\ \left( \int_r^1 B_r B_r' \right)^{-1/2} Z_2 \\ \frac{\sqrt{6}}{(1-r)^2} \Omega_{\epsilon}^{-1/2} C(1;r) \end{array} \right] = v(r).
\]

Using Lemma B.1(b) yields

\[
\left[ \begin{array}{c} \left( \frac{1}{nT^2} \sum_{i=1}^n \sum_{l=1}^{[Tr]} \hat{W}_{it} \hat{W}_{it}' \right)^{-1} \\ + \left( \frac{1}{nT^2} \sum_{i=1}^n \sum_{l=[Tr]+1}^T \hat{W}_{it} \hat{W}_{it}' \right)^{-1} \end{array} \right] \overset{d}{\rightarrow} \left[ \begin{array}{cc} \left( \int_0^r B_0 B_0' \right)^{-1} + \left( \int_r^1 B_r B_r' \right)^{-1} & 0 \\ 0 & \frac{5}{r^2} \Omega_{\epsilon}^{-1} + \frac{6}{(1-r)^2} \Omega_{\epsilon}^{-1} \end{array} \right] = G(r).
\]

38
Therefore, by the CMT and uniformly in $r$ we have

$$W ([Tr]) = \sqrt{\pi T} \left( \hat{\theta}_{1k} - \tilde{\theta}_{2k} \right)' \left[ \left( \frac{1}{\sqrt{\pi T}} \sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \bar{W}_{it} \bar{W}_{it}' \right)^{-1} + \left( \frac{1}{\sqrt{\pi T}} \sum_{i=1}^{n} \sum_{t=[Tr]+1}^{T} \bar{W}_{it} \bar{W}_{it}' \right)^{-1} \right]^{-1} \sqrt{\pi T} \left( \hat{\theta}_{1k} - \tilde{\theta}_{2k} \right)$$

$$\xrightarrow{d} v (r)' [G (r)]^{-1} v (r) = s (r)' [M (r)]^{-1} s (r) + \left[ \frac{1}{r^2} + \frac{1}{(1-r)^2} \right]^{-1} \times \left[ (1-r)^2 C (r; 0) - r^2 C (1; r) \right]' \left( \frac{1-r)^2 C (r; 0) - r^2 C (1; r) }{r^2 (1-r)^2} \right) = I + II.$$

Consider I. Conditioning on $C$, Theorem 1 yields, for fixed $r$, $s (r) | C \sim \left( (f_0 B_0 B_0')^{-1} + \left( f_0 B_0 B_0' \right)^{-1} \right)^{1/2} N [0, I_R]$, so that, conditioning on $C$, $s (r)' [M (r)]^{-1} s (r) | C \sim \chi_r^2$. This result does not depend on $C$ - i.e. it holds true for all the possible elements in the sigma-field $C$. Thus $s (r)' [M (r)]^{-1} s (r) = Q_R (r) \sim \chi_r^2$. As far as II is concerned, normality and independence of the increments of $C (r; 0)$ entail

$$(1-r)^2 C (r; 0) - r^2 C (1; r) \sim N \left\{ 0, \left[ (1-r)^4 r^2 + r^4 (1-r)^2 \right] I_p \right\}.$$ Therefore, for every $r$

$$II \sim \frac{r^2 (1-r)^2}{r^2 + (1-r)^2} \left\| N \left\{ 0, \left[ (1-r)^4 r^2 + r^4 (1-r)^2 \right] I_p \right\} \right\| \sim \chi_p^2.$$ Hence, for fixed $r$, $W ([Tr]) \xrightarrow{d} Q_R (r) + Q_p (r)$; this proves equation (2). Independence of $Q_R (r)$ and $Q_p (r)$ follows from the fact that $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically independent. This also proves that both $Q_p (r)$ and $Q_R (r)$ are nuisance parameters free.

**Proof of Proposition 3.** Consider $\hat{\sigma}_u^2$, defined as $\hat{\sigma}_u^2 = \frac{1}{\pi T} \sum_{t=1}^{n} \sum_{t=1}^{T} \left( y_{it} - \bar{y}_t - \hat{\theta} \bar{W}_{it} \right)^2$. Consistency of $\hat{\theta}$ under $H_0$ entails $\hat{\sigma}_u^2 \xrightarrow{P} \sigma_u^2$. As far as $\hat{\gamma}_c^2$ is concerned, we have $\hat{\gamma}_c^2 = \hat{\sigma}_u^2 + \hat{\beta}^2 \hat{\sigma}_u^2 \hat{\beta}$; under $H_0$, $\hat{\sigma}_u^2 = \sigma_u^2 + o_p (1)$ and likewise $\hat{\beta} = \beta + o_p (1)$. Also, it holds that $\hat{\sigma}_u^2 = \left( \frac{1}{T} \sum_{t=1}^{T} \hat{y}_t \hat{w}_t' \right) \left( \frac{1}{T} \sum_{t=1}^{T} \hat{w}_t \hat{w}_t' \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} \hat{y}_t \hat{w}_t' \right)$. From Lemma 2, it holds that $\frac{1}{T} \sum_{t=1}^{T} \hat{y}_t \hat{w}_t' = \frac{1}{T} \sum_{t=1}^{T} \hat{y}_t \hat{w}_t' + o_p (1)$. Since $\hat{\lambda}_i = \lambda_i + o_p (T^{-1})$ (see Bai, 2004), $T^{-1} \sum_{t=1}^{T} \hat{e}_i^2 \xrightarrow{P} \sigma_e^2$. Therefore, $\hat{\sigma}_e^2 = \sigma_e^2 \left( \hat{Q}_B \Sigma_{\lambda} \hat{Q}_B' \right) + o_p (1)$, whence $\hat{\sigma}_c^2 \xrightarrow{P} \sigma_c^2$. 

39
Proof of Theorem 3. Under the local alternative \( H_{a(nT)} \) the model can be rewritten as
\[
y_{it}^{(nT)} = \alpha_i + \widehat{W}_{it}^{(nT)} \theta + v_{it}
\]
and therefore
\[
\hat{\theta}_{1k}^{(nT)} = \left[ \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} \widehat{W}_{it}' \right]^{-1} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} y_{it}^{(nT)}
\]
and
\[
\hat{\theta}_{2k}^{(nT)}
\]
Hence
\[
\sqrt{nT} \left[ \hat{\theta}_{1k}^{(nT)} - \theta \right] = \left[ \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} \widehat{W}_{it}' \right]^{-1} \left[ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} v_{it} + \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} \sqrt{\frac{n}{m_n T}} \widehat{W}_{it}' g_i \left( \frac{t}{T} \right) \right]
\]
and similarly for the second partial sample, \( \hat{\theta}_{2k}^{(nT)} \)
\[
\sqrt{nT} \left[ \hat{\theta}_{2k}^{(nT)} - \theta \right]
\]
Combining these two results, it can be shown that
\[
\sqrt{nT} \left[ \hat{\theta}_{1k}^{(nT)} - \hat{\theta}_{2k}^{(nT)} \right] = \sqrt{nT} \left[ \hat{\theta}_{1[T]} - \hat{\theta}_{2[T]} \right] + \left( \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} \widehat{W}_{it}' \right)^{-1} \left[ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} v_{it} + \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} \sqrt{\frac{n}{m_n T}} \widehat{W}_{it}' g_i \left( \frac{t}{T} \right) \right]
\]
\[
- \left( \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{[T]} \widehat{W}_{it} \widehat{W}_{it}' \right)^{-1} \left[ \frac{1}{\sqrt{nT}} \sum_{i=1}^{m_n T} \sum_{t=1}^{[T]} \widehat{W}_{it} \sqrt{\frac{n}{m_n T}} \widehat{W}_{it}' g_i \left( \frac{t}{T} \right) \right]
\]
\[
= I + II.
\]
Consider I and II. It holds that

\[
\frac{1}{\sqrt{nT}} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \widehat{W}_{it} \frac{\sqrt{n}}{m_n T} \widehat{W}_{it} g_i \left( \frac{t}{T} \right)
\]

\[
= \frac{1}{m_n T^2} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \left[ W_{it} + \left( \widehat{W}_{it} - W_{it} \right) \right] \left[ W_{it} + \left( \widehat{W}_{it} - W_{it} \right) \right]^\prime g_i \left( \frac{t}{T} \right)
\]

\[
= \frac{1}{m_n T^2} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} W_{it} \widehat{W}_{it} g_i \left( \frac{t}{T} \right) + \frac{1}{m_n T^2} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \left( \widehat{W}_{it} - W_{it} \right)^\prime g_i \left( \frac{t}{T} \right)
\]

Assumption 4(b) states that \( a = O_p(1) \). Also, as far as \( b \) and \( c \) are concerned, it holds that

\[
\|b\| \leq \left( \frac{1}{m_n T^2} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \| \widehat{W}_{it} - W_{it} \|^2 \right)^{1/2} \left( \frac{1}{m_n T^2} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \left\| \widehat{W}_{it} g_i \left( \frac{t}{T} \right) \right\|^2 \right)^{1/2}
\]

As \( (n,T) \to \infty \), Lemma A.1.b ensures that \( \left( \frac{1}{T^2} \sum_{t=1}^{|T_r|} \| \widehat{W}_t - W_t \|^2 \right)^{1/2} = \left[ \frac{1}{T} O_p \left( \frac{1}{C_n T} \right) \right]^{1/2} = O_p(1) \); also, by Assumption 4(c), \( \frac{1}{\sqrt{nT}} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \left\| \widehat{W}_{it} g_i \left( \frac{t}{T} \right) \right\|^2 = O_p(1) \). Consequently, \( b \) and \( c \) are both \( o_p(1) \). Finally as far as \( d \) is concerned

\[
\|d\| \leq \left( \frac{1}{m_n T^2} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \| \widehat{W}_{it} - W_{it} \|^2 \right)^{1/2} \left( \frac{1}{m_n T^2} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \left\| g_i \left( \frac{t}{T} \right) \right\|^2 \right)^{1/2}
\]

\[
\leq \left( \frac{1}{T^2} \sum_{t=1}^{|T_r|} \| \widehat{w}_t - w_t \|^2 \right)^{1/2} \left( \frac{1}{m_n T^2} \sum_{i=1}^{m_n} \sum_{t=1}^{|T_r|} \left\| g_i \left( \frac{t}{T} \right) \right\|^2 \right)^{1/2}
\]

\[
= O \left( \frac{1}{T} \right) O_p \left( \frac{1}{C_n T} \right) O \left( \frac{1}{T} \right) O_p(1) = o_p(1),
\]

41
using Assumption 4(a). Combining these results we get

\[
\sqrt{nT}\left[\hat{\beta}^{(nT)}_{1k} - \hat{\beta}^{(nT)}_{2k}\right] = \sqrt{nT}\left[\hat{\beta}_{1|Tr} - \hat{\beta}_{2|Tr}\right] + \left(\frac{1}{nT^2}\sum_{i=1}^{n} \sum_{t=1}^{[Tr]} W_{it}W_{it}'\right)^{-1} \left[\frac{1}{mnT^2} \sum_{i=1}^{m} \sum_{t=1}^{[Tr]} W_{it}W_{it}'g_i\left(\frac{t}{T}\right)\right] - \left(\frac{1}{nT^2}\sum_{i=1}^{n} \sum_{t=[Tr]+1}^{T} W_{it}W_{it}'\right)^{-1} \left[\frac{1}{mnT^2} \sum_{i=1}^{m} \sum_{t=[Tr]+1}^{T} W_{it}W_{it}'g_i\left(\frac{t}{T}\right)\right] + o_p(1)
\]

Letting \(d(r)\) be the limit of \(D_{nT}\) as \((n, T) \to \infty\), \(\sqrt{nT}\left[\hat{\beta}^{(nT)}_{1k} - \hat{\beta}^{(nT)}_{2k}\right] \overset{d}{\to} |J(r) + d(r)|^{-1/2}\) uniformly in \(r\), with

\[
\Xi = \lim_{n,T \to \infty} \left[\left(\frac{1}{nT^2}\sum_{i=1}^{n} \sum_{t=1}^{[Tr]} \hat{W}_{it}\hat{W}_{it}'\right)^{-1} + \left(\frac{1}{nT^2}\sum_{i=1}^{n} \sum_{t=[Tr]+1}^{T} \hat{W}_{it}\hat{W}_{it}'\right)^{-1}\right].
\]

Upon assuming that both \(\hat{\sigma}_\xi\) and \(\hat{\sigma}_u\) are consistent for \(\sigma_\xi\) and \(\sigma_u\), Theorem 3 follows from the Continuous Mapping Theorem.

**Proof of Proposition 4.** We first consider the consistency of \(\hat{\sigma}^2_u\). Let \(\hat{y}_{it} = y_{it} - \bar{y}_i\); it holds that

\[
\hat{\sigma}^2_u = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[\hat{y}_{it} - \hat{W}_{it}\hat{\beta}_i\right]^2
\]

\[
= \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[\hat{y}_{it} - \hat{W}_{it}\hat{\beta}_i\right]^2 + \frac{1}{n} \sum_{i=1}^{m} \sum_{t=1}^{T} \frac{n}{mZ_i^2} g_i\left(\frac{t}{T}\right)\hat{W}_{it}\hat{W}_{it}'g_i\left(\frac{t}{T}\right) - 2\frac{1}{nT} \sum_{i=1}^{m} \sum_{t=1}^{T} \frac{n}{mZ_i^2} g_i\left(\frac{t}{T}\right)\hat{W}_{it}\hat{W}_{it}'g_i\left(\frac{t}{T}\right)
\]

\[
= I + II + III.
\]
As far as term \( I \) is concerned, we have

\[
I = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[ \tilde{y}_{it} - \hat{W}_{it}' \theta + \hat{W}_{it}' \left( \hat{\theta} - \theta \right) \right]^2
\]

\[
= \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ v_{it} + \hat{W}_{it}' \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} \hat{W}_{it}' \right]^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} v_{it} \right\}^2
\]

\[
= \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} v_{it}^2 - \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} v_{it} \hat{W}_{it}' \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} \hat{W}_{it}' \right]^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} v_{it}.
\]

As far as \( III \) is concerned, it holds that

\[
\frac{1}{nT} \sum_{i=1}^{m_n} \sum_{t=1}^{T} \left( \tilde{y}_{it} - \hat{W}_{it}' \hat{\theta} \right) \frac{\sqrt{n}}{m_n T} \hat{W}_{it} g_i \left( \frac{t}{T} \right)
\]

\[
= \frac{1}{nT} \sum_{i=1}^{m_n} \sum_{t=1}^{T} \left[ v_{it} + \hat{W}_{it}' \left( \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} \hat{W}_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} v_{it} \right] \frac{\sqrt{n}}{m_n T} \hat{W}_{it} g_i \left( \frac{t}{T} \right)
\]

\[
= \frac{1}{nT} \sum_{i=1}^{m_n} \sum_{t=1}^{T} \sqrt{n} \frac{1}{m_n T} g_i \left( \frac{t}{T} \right) \hat{W}_{it} v_{it}
\]

\[+ \frac{1}{nT} \left[ \sum_{i=1}^{m_n} \sum_{t=1}^{T} \sqrt{n} \frac{1}{m_n T} g_i \left( \frac{t}{T} \right) \hat{W}_{it} \hat{W}_{it}' \right] \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} \hat{W}_{it}' \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} v_{it} \right].
\]

Combining \( I \) and \( III \) with \( II \) we get

\[
\tilde{\sigma}_u^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} v_{it}^2
\]

\[- \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} v_{it} \hat{W}_{it}' \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} \hat{W}_{it}' \right]^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} v_{it}
\]

\[+ \frac{1}{nT} \sum_{i=1}^{m_n} \sum_{t=1}^{T} \sqrt{n} \frac{1}{m_n T} g_i \left( \frac{t}{T} \right) \hat{W}_{it} \hat{W}_{it}' g_i \left( \frac{t}{T} \right)
\]

\[- \frac{2}{nT} \sum_{i=1}^{m_n} \sum_{t=1}^{T} \sqrt{n} \frac{1}{m_n T} g_i \left( \frac{t}{T} \right) \hat{W}_{it} v_{it}
\]

\[- \frac{2}{nT} \left[ \sum_{i=1}^{m_n} \sum_{t=1}^{T} \sqrt{n} \frac{1}{m_n T} g_i \left( \frac{t}{T} \right) \hat{W}_{it} \hat{W}_{it}' \right] \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} \hat{W}_{it}' \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} v_{it} \right].
\]

Lemmas B.1 and B.2 ensure that

\[
\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} v_{it} \hat{W}_{it}' \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} \hat{W}_{it}' \right]^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{it} v_{it} = O_p \left( \frac{1}{nT} \right);
\]
from Lemma B.1 and Assumption 4(c) and 4(d) we have
\[
\frac{1}{nT} \sum_{t=1}^{nT} \frac{m_n}{T^2} g_t \left( \frac{t}{T} \right) \left( \frac{t}{T} \right)' \widehat{W}_t \widehat{W}_t' \left( \frac{t}{T} \right) = O_p \left( \frac{1}{m_n T} \right),
\]
\[
\frac{1}{nT} \sum_{t=1}^{nT} \frac{m_n}{T^2} g_t \left( \frac{t}{T} \right) \left( \frac{t}{T} \right)' \widehat{W}_t v_t = O_p \left( \frac{1}{\sqrt{nT}} \right); \]
finally, Lemmas B.1 and B.2 and Assumption 4(b) entail
\[
\frac{1}{nT} \left[ \sum_{i=1}^{m_n} \sum_{t=1}^{T} \frac{\sqrt{m_n} g_t \left( \frac{t}{T} \right) \left( \frac{t}{T} \right)' \widehat{W}_t \widehat{W}_t' \left( \frac{t}{T} \right) - \sum_{i=1}^{m_n} \sum_{t=1}^{T} \sum_{i=1}^{m_n} \sum_{t=1}^{T} \widehat{W}_t v_t }{\left( \sum_{i=1}^{m_n} \sum_{t=1}^{T} \sum_{i=1}^{m_n} \sum_{t=1}^{T} \widehat{W}_t v_t \right)^{-1}} \right] = O_p \left( \frac{1}{nT} \right).
\]
Hence, \( \hat{\sigma}_u^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} v_t^2 + o_p(1) \). It holds that
\[
\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[ \theta' \left( W_t - \widehat{W}_t \right) + u_{it} \right]^2
\]
\[
= \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it}^2 + \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[ \theta' \left( w_t - \hat{w}_t \right) \right]^2 + 2 \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \theta' \left( w_t - \hat{w}_t \right) u_{it}
\]
\[
= I + II + III.
\]
Assumption 1 and the LLN yield \( I = \sigma_u^2 + o_p(1) \). As far as \( II \) and \( III \) are concerned, we have \( \|II\| \leq \|\theta\|^2 \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left( \sum_{t=1}^{T} \left( w_t - \hat{w}_t \right)^2 \right)^{1/2} = o_p(1) \), using Lemma A.1(b). Also, \( \|III\| \leq \|\theta\|^2 \frac{1}{nT} \sum_{i=1}^{n} \left( \sum_{t=1}^{T} \left( w_t - \hat{w}_t \right)^2 \right)^{1/2} = o_p(1) \), after equation (35). Then under the local alternatives \( H_a^{(nT)} \) it holds that \( \hat{\sigma}_u^2 \overset{p}{\to} \sigma_u^2 \). We are now ready also to prove consistency of \( \hat{\sigma}_u^2 \). By definition, \( \hat{\sigma}_u^2 = \hat{\sigma}_u^2 + \hat{\beta}^{(nT)}' \hat{\sigma}_{\pi}^2 \hat{\beta}^{(nT)} \), and since \( \hat{\sigma}_u^2 \overset{p}{\to} \sigma_u^2 \) under \( H_a^{(nT)} \) (as it does not depend on \( H_a^{(nT)} \) being true or not), consistency of \( \hat{\beta}^{(nT)} \) entails \( \hat{\sigma}_u^2 \overset{p}{\to} \sigma_u^2 + \hat{\beta}^{(nT)}' \hat{\sigma}_{\pi}^2 \hat{\beta}^{(nT)} \)
\[
= \sigma_u^2.
\]
**Proof of Proposition 5.** Consider \( \hat{\beta}_{1[T]} - \beta \); the estimation error is
\[
\left[ \hat{\beta}_{1[T]} - \beta \right]
\]
\[
\left[ \hat{\gamma}_{1[T]} - \gamma \right]
\]
\[
= \left[ \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{w}_{t,r} \hat{u}_{t,r}' \Gamma_i \hat{G}_t \hat{u}_{t,r}' \Gamma_i'}{\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{w}_{t,r} \hat{G}_t \Gamma_i'} \right]^{-1} \left[ \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{w}_{t,r} v_t}{\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{G}_t \hat{u}_{t,r}'} \right].
\]
Consider the denominator. The Cauchy-Schwartz inequality entails, for all \( r \), \( \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{G}_t (\hat{w}_{t,r} - w_{t,r})' \leq \left( \sum_{i=1}^{n} \sum_{t=1}^{T} \|\hat{G}_t\|^2 \right)^{1/2} \left( \sum_{i=1}^{n} \sum_{t=1}^{T} \|\hat{w}_{t,r} - w_{t,r}\|^2 \right)^{1/2} = O_p(T) O_p \left( \sqrt{TC_{n^{-1}}} \right) \). Thus, using Assumption 6, as \( (n,T) \to \infty \) it holds that \( (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{G}_t \hat{w}_{t,r} \hat{u}_{t,r}' = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \Gamma_i \hat{G}_t \hat{u}_{t,r} + \)
secondly, we prove the consistency of the rescaling matrix \( \sqrt{\phi_2} \). Applying Lemma B.1(b), we have that
\[
\lim_{nT \to \infty} \frac{1}{nT^2} \sum_{i=1}^{n} \left[ \phi_1' w_{t,r} + \phi_2' \Gamma_i \right] v_{it} = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left[ \phi_1' w_{t,r} + \phi_2' \Gamma_i \right] v_{it} + o_p(1).
\]
An FCLT for \( \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left[ \phi_1' w_{t,r} + \phi_2' \Gamma_i \right] v_{it} \) follows from the proof of Theorem 1. By definition, the covariance matrix is
\[
\lim_{nT \to \infty} \frac{1}{nT} \sum_{i=1}^{n} \left[ \phi_1' w_{t,r} + \phi_2' \Gamma_i \right] v_{it} = \sigma^2 \left[ \phi_1' \int_0^r B_z B_z' \phi_1 + \phi_2' \Sigma^1/2 (\int_0^r B_G B_G') \Sigma^1/2 \phi_2 + 2 \phi_1' \phi_2' \right],
\]
using Lemma B.1(b) and Assumption 6. Thus, the finite-dimensional distributions are
\[
\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left[ \bar{w}_{t,r} v_{it} \right] \xrightarrow{d} \sigma^2 \left[ \int_0^r B_z B_z' \phi_1 + \phi_2' \Sigma^1/2 (\int_0^r B_G B_G') \Sigma^1/2 \right]^{1/2} \times Z.
\]
Tightness follows from Assumption 5(c), along the same lines as in the proof of Theorem 1.

**Proof of Proposition 6.** The proof follows similar lines to the proof of Theorem 9 in Phillips and Moon (1999) and is based on two steps: firstly, we show an FCLT for \( \sqrt{nT} \left[ \hat{\theta}^{FM}_{1[\tau]} - \theta \right] \); secondly, we prove the consistency of the rescaling matrix
\[
\left[ \begin{array}{c} \hat{\Omega} \bar{I}_R \\ 0 \end{array} \right] = \left[ \begin{array}{c} \hat{\Omega}_u \end{array} \right] \left( \Omega_{\hat{\Omega}_u} \hat{\Omega}_u^{-1} \hat{\Omega}_u \right) \left[ \begin{array}{c} \hat{I}_p \end{array} \right].
\]
Consider \( \hat{\theta}^{FM}_{1[\tau]} - \theta \). We have
\[
\sqrt{nT} \left[ \hat{\theta}^{FM}_{1[\tau]} - \theta \right] = \left[ \begin{array}{c} \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{[\tau]} \bar{W}_{it} \bar{W}'_{it} \\ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{[\tau]} \bar{W}_{it} u_{it} - \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{[\tau]} \left( \bar{w}_t \Delta \hat{F}'_t \sqrt{n} \Delta x'_{it} \right) \end{array} \right] \left( \hat{\Omega}^{-1} \hat{\Omega}_{(\hat{\Omega})u} \right) \left[ \begin{array}{c} \hat{I}_p \end{array} \right].
\]
Consider the denominator of $\tilde{\theta}_{1 \{Tr\}}^F - \theta$; Assumption 7 and the application of Lemma B.1(b) yield

$$\frac{1}{nT^2} \sum_{i=1}^n \sum_{t=1}^{|Tr|} \tilde{W}_{it} W_{it}^t \xrightarrow{d} \left[ \int_0^\tau B_\varepsilon B_\varepsilon^t 0 \begin{array}{c} 0 \\ \frac{1}{\theta^2} \Omega_\varepsilon \end{array} \right].$$

As far as the numerator of $\sqrt{nT} \left[ \tilde{\theta}_{1 \{Tr\}}^F - \theta \right]$ is concerned, we can write it as

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \left[ \hat{w}_t^* \left( u_{it}^* - n^{-1/2} \Omega_{e,u}^t \Omega_\varepsilon^{-1} \Delta \hat{F}_t^* \right) - \Lambda_{eu}^+ \right] + o_p(1) + O_p \left( \sqrt{\frac{n}{T}} \right), \quad (40)$$

where we used the fact that $\Omega_{(e)}$ is assumed diagonal. In (40), the remainder of order $o_p(1)$ comes from (30) and (31); the term of order $O_p \left( \sqrt{\frac{n}{T}} \right)$ comes from the BN decomposition assumed in Assumption 7(a), following similar lines as in the proof of Proposition 1 and Lemma B.2. From Assumption 7 and Assumption 1(a), $\hat{w}_t^* \left( u_{it}^* - n^{-1/2} \Omega_{e,u}^t \Omega_\varepsilon^{-1} \Delta \hat{F}_t^* \right) - \Lambda_{eu}^+$ and $x_{it}^* \left( u_{it}^* - \Omega_{e,u}^t \Omega_\varepsilon^{-1} \Delta x_{it}^* \right) - \Lambda_{eu}^+$ are (by construction) zero mean MDS with finite second moments. Using the same approach as in the proof of Lemma B.2, as $(n, T) \to \infty$ with $\frac{T}{n} \to 0$

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \left[ \hat{w}_t^* \left( u_{it}^* - \frac{1}{\sqrt{n}} \Omega_{e,u}^t \Omega_\varepsilon^{-1} \Delta \hat{F}_t^* \right) \right] \xrightarrow{d} \left[ \Omega_\varepsilon \left( \int_0^\tau B_\varepsilon B_\varepsilon^t \right)^{1/2} \begin{array}{c} 0_{R \times R} \\ 0 \end{array} \right] \times \left[ B(t) \begin{array}{c} C(r; 0) \end{array} \right],$$

uniformly in $r \in [0, 1]$. Combining this result with the denominator, the FCLT follows.

As far as the rescaling matrix is concerned, equation (30) and Proposition 3 entail $\hat{\Omega}_{\varepsilon} \xrightarrow{p} \Omega_{\varepsilon}$ and $\left( \hat{\Omega}_{\varepsilon} - \hat{\Omega}_{u,e} \Omega_{e}^{-1} \hat{\Omega}_{e,u} \right) \xrightarrow{p} \left( \Omega_{\varepsilon} - \Omega_{u,e} \Omega_{e}^{-1} \Omega_{e,u} \right)$ - see also Phillips and Moon (1999; p. 1109). The rest of the proof can be derived following the same lines as the proof of Theorem 2, and is therefore omitted.