Asymptotics for Panel Models with Common Shocks*

Chihwa Kao  
Syracuse University  
cdkao@maxwell.syr.edu

Lorenzo Trapani  
Cass Business School  
L.Trapani@city.ac.uk

Giovanni Urga  
Cass Business School and Universita’ di Bergamo  
G.Urga@city.ac.uk

July 30, 2010

Abstract

This paper develops a novel asymptotic theory for panel models with common shocks. We assume that contemporaneous correlation can be generated by both the presence of common regressors among units and weak spatial dependence among the error terms. Several characteristics of the panel are considered: cross-sectional and time-series dimensions can either be fixed or large; factors can either be observable or unobservable; the factor model can describe either a cointegration relationship or a spurious regression, and we also consider the stationary case. We derive the rate of convergence and the limit distributions for the ordinary least squares (OLS) estimates of the model parameters under all the aforementioned cases.

JEL Classification: C13, C23.

Keywords: Panel data, common shocks, cross-sectional dependence, asymptotics, joint limit, martingale difference sequence.

*This paper originated while Chihwa Kao was visiting the Centre for Econometric Analysis at Cass (CEA@Cass). We thank the Editor, Esfandiar Maasoumi, and Associate Editor and a Referee, for extremely constructive comments which have greatly improved the presentation of this paper. We wish to thank participants in the “1st Italian Congress of Econometrics and Empirical Economics” (Universita’ Ca’ Foscari, Venice, 24-25 January 2005), the 9th World Congress of the Econometric Society (London, 19-24 August 2005), in particular Peter Pedroni and Yixiao Sun, the North American Summer Meeting of the Econometric Society (Minneapolis, MN, 22-25 June 2006) and the 13th Conference on Panel Data (Cambridge, 7-9 July, 2006), the Monday Workshop of the Faculty of Finance at Cass Business School and the Lunch Seminars of the Department of Management and Information Technology at Bergamo University for useful comments. The paper has hugely benefited from conversations with Badi Baltagi and Jan Kiviet. All remaining errors are the authors’ sole responsibility.
1 Introduction

There is a growing body of literature dealing with limit theory for nonstationary panels. While the first generation of these contributions assumed independence across units (see for instance Phillips and Moon, 1999; Kao, 1999), in the second generation this assumption is relaxed, and hypothesis testing and estimation methods are evaluated assuming various degrees of cross-sectional dependence, e.g., see Bai (2003, 2004), Bai and Ng (2002, 2004), Stock and Watson (2002). We can distinguish the case where regressors are cross-sectionally dependent (see Donald and Lang, 2004; Moulton, 1990) from the case where it is the error terms across unit to be dependent (see for instance Bai and Kao, 2006; Moon and Perron, 2004) or both (see for instance Ahn, Lee, and Schmidt, 2001; Pesaran 2006).

The main aim of this paper is to propose a novel asymptotic theory for panel models where common shocks are present among the regressors, thereby introducing strong cross-sectional dependence. We generalize the asymptotics developed by Phillips and Moon (1999) and Andrews (2005) by employing and extending the theory for factor models in Bai (2003, 2004) and Bai and Ng (2004).

Phillips and Moon (1999) analyze nonstationary panels when both cross-sectional dimension \( n \) and time-series dimension \( T \) are large. They derive the seminal result that as \( n \to \infty \) a long-run average relationship between two nonstationary panel vectors exists even when the single units do not cointegrate. A similar result is also reported in Kao (1999). However, the asymptotics derived in Phillips and Moon (1999) is based on the assumption of cross-sectional independence though the authors point out that their results still hold when certain degree of weak cross-sectional dependence is allowed. Thus, the derivation of sequential and joint asymptotics with arbitrary dependence amongst units remains largely unexplored, and it is likely to lead to different asymptotics. Asymptotic normality may not hold, for example, when all or part of the regressors are common across cross-sectional units, and may result in mixed
asymptotic normality, as Andrews (2005) has demonstrated in a cross-sectional context. Andrews (2005, Theorem 4, p. 1567) proves that the presence of common factors among the cross-sectional units makes the limiting distribution of the OLS estimator of the regression slope mixed normal and not normal as in the classical regression analysis. Note that in this case mixed normality of the OLS estimator of the regression slope holds even if regressors are stationary, i.e., $I(0)$, and independent of errors. This finding is also obtained in our paper when studying the distribution limit for the OLS estimator of the regression slope for the fixed $T$ case (see equation 20 in Theorem 2 below), while when we consider the $T \to \infty$ case, not explored by Andrews (2005), we show that in the stationary case as $T \to \infty$ the OLS estimator of the regression slope is normally distributed.

1.1 Basic Model and Extensions

In this paper we consider the following panel regression model with common shocks

$$y_{it} = \alpha_i + \beta' F_t + u_{it}$$

(1)

$i = 1, ..., n, \ t = 1, ..., T$, where $\beta$ is a $k \times 1$ vector of slope parameters and the regressor $F_t = (F_{1t}, ..., F_{kt})'$ is a $k \times 1$ vector of common shocks,

$$F_t = F_{t-1} + \varepsilon_t.$$

Equation (1) could be either a spurious regression or a cointegration model depending on whether $u_{it}$ is $I(1)$ or $I(0)$, respectively. It is important to emphasize that, as far as the presence of $F_t$ is concerned, equation (1) represents a panel regression model with a set of regressors, $F_t$, which is common across units and with common slope coefficient $\beta$. Model (1) differs from a factor-loading specification as in Bai (2004) and Bai and Ng (2004), for example. Thus, in our setup $F_t$ is a genuine (observable or unobservable) regressor rather than a “common factor”. A framework which is similar in spirit to the one in this paper is in Stock and Watson (1999, 2002, 2005), where $y_{it}$ in (1) (with $n = 1$) is the time-series variable to be predicted and $z_i = (z_{i1}, z_{i2}, ..., z_{iT})'$ is an $n$-dimensional
multiple time-series of candidate predictors; also, model (1) resembles the panel cointegration model with global stochastic trends of Bai, Kao and Ng (2009), although (1) assumes having common $\beta$.

When common shocks are not observable, we assume that a set of exogenous variables, $z_{it}$, is observable such that

$$z_{it} = \lambda_i'F_t + e_{it}$$  \hspace{1cm} (2)$$

where $\lambda_i$ is a vector of factor loadings and $e_{it}$ is an idiosyncratic component. We assume throughout the paper, for the sake of the simplicity of the notation, that the number of the $z_{it}$s is the same as that of the $y_{it}$s. However, the panel dimensions of $y_{it}$ and the $z_{it}$ may be different, for example $y_{it}$ may refer to individuals while $z_{it}$ may index several macro variables.

To extend our results to the stationary panel model case, we also consider the first-differenced form of model (1),

$$\Delta y_{it} = \beta' \Delta F_t + \Delta u_{it}.$$  \hspace{1cm} (3)$$

Model (1) considers a very simple specification. However, it could be argued that a more complete and realistic framework should also embed a set of idiosyncratic shocks, i.e.,

$$y_{it} = \alpha_i + \beta' F_t + \gamma' x_{it} + u_{it}.$$  \hspace{1cm} (4)$$

For the sake of notational simplicity, the main results in the paper, reported in Section 3, are derived under the restrictive assumptions of no idiosyncratic shocks, i.e., under the constrain that $\gamma = 0$. However, in Section 4 we show that our main results concerning the asymptotics of the estimator of $\beta$ are still useful in presence of a more complicated specification as (4). This is obviously true when the regressors $F_t$ and $x_{it}$ are orthogonal. We also examine the case whereby the $x_{it}$ are allowed to be correlated with $F_t$ via the factor-loadings specification

$$x_{it} = \tau_i G_t + \omega_{it}$$  \hspace{1cm} (5)$$
where $G_t$ is a set of common factors that can be independent of the regressors $F_t$ or (fully or partly) overlap with them, and $\omega_{it}$ is a unit specific (stationary or nonstationary) shock. A similar framework that allows for cross-sectional dependence among the idiosyncratic regressors and dependence between the idiosyncratic regressors and the common regressors is in Pesaran (2006) and Kapetanios, Pesaran and Yamagata (2006), even though in our paper $F_t$ is a set of regressors and not nuisance parameters. Note that allowing for $x_{it}$ being dependent upon $F_t$ through some possibly heterogeneous loadings $\tau_i$ allows for the response of $y_{it}$ to $F_t$ being (indirectly) heterogeneous across individuals.

Models (1) and (4) are frequently employed for the purpose of forecasting (Stock and Watson, 1999, 2002, 2005), and they encompass a wide set of models in economics and finance. As a general interpretation, such models represent the decision of a microeconomic agent $i$ ($y_{it}$), being influenced by macroeconomic factors $F_t$ and by a set of individual specific characteristics, $\alpha_i$ and possibly $x_{it}$. Examples 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, labor 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, Duffie, Kapadia and Saita, 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). Considering (3), which represents a stationary panel regression with common shocks, the most natural application one may have in mind is to asset pricing models, such as the APT, where asset returns are explained by common factors (such as e.g., market return and powers thereof to represent coskewness and cokurtosis, macro factors, etc.); see Cochrane (2005) for a comprehensive review.

1.2 Main Results

Our asymptotic theory considers several features of the underlying model. First, we assume that contemporaneous correlation can be generated by both the presence of common regressors (e.g., macro shocks, aggregate fiscal and monetary policies) among units and weak spatial dependence among the error terms. Second, the common shocks can either be known or unobservable. Classical examples of observed common shocks are index models such as those used in international trade, labor economics, urban regional, public economics and finance literature. Most often, shocks are unknown, as in the cases of index extraction and indicators aggregation in economics, e.g., Quah and Sargent (1993), Forni and Reichlin (1998), and Bernanke and Boivin (2000). Third, regression model (1) may describe either a cointegration relationship or a spurious regression. Fourth, the time-series dimension $T$ and the cross-sectional dimension $n$ can be either fixed or large. We develop our limit theory by considering cases where the time-series dimension $T$ and the number of units $n$ are large and we also include the case of when either $n$ or $T$ is fixed.

An overview of the results derived in this paper is reported in Table 1.

[Insert Table 1 somewhere here]

As Table 1 shows, this paper provides a unified framework for the asymptotics of panels with common shocks. Particularly, results for the case of large $n$ and large $T$ with observable $F_t$ are novel and can be thought of as extensions
<table>
<thead>
<tr>
<th></th>
<th>$F_t$ known</th>
<th></th>
<th>$F_t$ unknown</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($n, T$)</td>
<td>C</td>
<td>($n, T$)</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>LD</td>
<td></td>
<td>LD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cointegration: $u_{it} \sim I(0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed $n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T \to \infty$</td>
<td>Yes</td>
<td>Mixed Normal (Eq. 11)</td>
<td>Yes</td>
<td>Non Standard (Eq. 37)</td>
</tr>
<tr>
<td>Fixed $T$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n \to \infty$</td>
<td>Yes</td>
<td>Mixed Normal (Eq. 13)</td>
<td>Yes</td>
<td>Mixed Normal (Eq. 13)</td>
</tr>
<tr>
<td>($n, T$) $\to \infty$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n/T \to 0$</td>
<td>Yes</td>
<td>Mixed Normal (Eq. 15)</td>
<td>$n/T \to 0$</td>
<td>Yes</td>
</tr>
<tr>
<td>$T/n \to 0$</td>
<td>Yes</td>
<td>Mixed Normal (Eq. 16)</td>
<td>$T/n \to 0$</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mixed Normal (Eq. 27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mixed Normal (Eq. 27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mixed Normal (Eq. 29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mixed Normal (Eq. 30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spurious Regression: $u_{it} \sim I(1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed $n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T \to \infty$</td>
<td>No</td>
<td>Non Standard (Eq. 12)</td>
<td>No</td>
<td>Non Standard (Eq. 38)</td>
</tr>
<tr>
<td>Fixed $T$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n \to \infty$</td>
<td>Yes</td>
<td>Non Standard (Eq. 14)</td>
<td>Yes</td>
<td>Non Standard (Eq. 14)</td>
</tr>
<tr>
<td>($n, T$) $\to \infty$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n/T \to 0$</td>
<td>Yes</td>
<td>Non Standard (Eq. 17)</td>
<td>$n/T \to 0$</td>
<td>Yes</td>
</tr>
<tr>
<td>$T/n \to 0$</td>
<td>Yes</td>
<td>Non Standard (Eq. 18)</td>
<td>$T/n \to 0$</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non Standard (Eq. 29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non Standard (Eq. 30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Differences: $\hat{\beta}^{FD}$ : $\Delta y_{it} = \beta' \Delta F_t + \Delta u_{it}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed $n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T \to \infty$</td>
<td>Yes</td>
<td>Normal (Eq. 19)</td>
<td>No</td>
<td>Degenerate (Eq. 39)</td>
</tr>
<tr>
<td>Fixed $T$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n \to \infty$</td>
<td>Yes</td>
<td>Mixed Normal (Eq. 20)</td>
<td>Yes</td>
<td>Mixed Normal (Eq. 20)</td>
</tr>
<tr>
<td>($n, T$) $\to \infty$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n/T \to 0$</td>
<td>Yes</td>
<td>Normal (Eq. 21)</td>
<td>$n/T \to 0$</td>
<td>Yes</td>
</tr>
<tr>
<td>$T/n \to 0$</td>
<td>Yes</td>
<td>Degenerate (Eq. 22)</td>
<td>$T/n \to 0$</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal (Eq. 31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Degenerate (Eq. 32)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of the asymptotic theory for panels derived by Phillips and Moon (1999) and Kao (1999), who consider a model with cross-sectional independence. Assuming cross-sectional dependence in the panel changes the asymptotic theory, and a typical feature (discussed in greater details hereafter) is the asymptotic distribution of estimates being no longer normal as opposed to the independence case. An important result here is the extension of the joint limit theory to the strong dependence case, and the development of a method of proof for the asymptotics of double sums involving common shocks. Thus, although our results are specific to model (1), the method of proof we follow can be extended to study the asymptotics of estimators and tests for different models. For example, the method of proof developed here extends readily to inferential theory for cointegrated panels with common factors (Westerlund, 2007) or it can be used to show the asymptotics of t-tests for long run parameters in mixed panels (Fuertes, 2008; Ng, 2008); other applications to models where common factors are treated either as nuisance parameters or are genuine observable regressors are possible.

Results obtained for the case whereby the common shocks \( F_t \) are not observable are also new. The asymptotic theory for the estimates of the common shocks \( F_t \) is based on previous work by Bai (2003, 2004) and Bai and Ng (2002, 2004), and extended to the case of finite \( n \). When common shocks are not observable, the estimated latent variables \( \hat{F}_t \) are used as generated regressors to estimate \( \beta \). This introduces a new error component in the regression equation. An important contribution of our paper is to study the impact of the estimation error when one needs to use an estimate of \( F_t \) in the regression model; see e.g., although in a nonparametric set-up, Connor, Hagmann and Linton (2007). Note that in Table 1, the “non standard” limiting distributions depend also on the assumptions made on the data generating process (DGP). Section 3 provides details on this.

Last, it is important to note that, as far as Theorems 1-4 are concerned, when both \( n \) and \( T \) are large the limits we derived are joint limits, which we obtain for \( (n, T) \to \infty \) under, as a restriction on the rate of expansion of \( n \) and
Although more details on the method of proof are provided in Theorem 9 in Appendix B and in the remarks and proofs (reported in Appendix C) of the other theorems, the derivation of the joint limit is carried out by conditioning on the \( \sigma \)-field generated by the common shocks \( F_t \). We show, in a similar spirit to Andrews (2005), that this entails that the quantities involved in the derivation of the asymptotics are martingale difference sequences (MDS), conditional on \( F_t \). For each of the cases considered here, we then prove a joint Liapunov condition, under \( (n, T) \to \infty \), which allows to apply the MDS central limit theorem (CLT) discussed in Hall and Heyde (1980) as \( (n, T) \to \infty \). The restriction on the rate of expansion \( n/T \) is derived using similar arguments as in Phillips and Moon (1999), based on the Beveridge-Nelson (BN) decomposition of the series involved in the calculations.

The remainder of the paper is organized as follows. Section 2 introduces and comments on the main assumptions. In Section 3, we report the asymptotic theory of the ordinary least square (OLS) estimators of \( \beta \) in models (1) and (3). We analyze both the cases of known factors (Section 3.1) and unknown factors (Section 3.2), and we distinguish the cases of large \( n \) and \( T \), finite \( T \) and large \( n \) and finite \( n \) and large \( T \). Section 4 considers a discussion of the asymptotics for the estimator of when the data are generated by (4). Some Monte Carlo evidence is reported in Section 5. Section 6 concludes. Appendix A reports and discusses a joint MDS CLT. The main proofs are reported in Appendix B, contained in the present paper. Other proofs and preliminary Lemmas are in Appendix C in an extended, working paper version of this paper, which can be found at http://www.cass.city.ac.uk/cea/research_papers/WorkingPapers2010/WP_CEA_01_2010.pdf

Notation is fairly standard. Throughout the paper, \( \|A\| \) denotes \( \sqrt{\text{tr}(AA^T)} \), \( \to \) the ordinary limit, \( \Rightarrow \) weak convergence, and \( \xrightarrow{p} \) convergence in probability. Stochastic processes such as \( B(r) \) on \([0,1]\) are usually written as \( B \), integrals such as \( \int_0^1 B(r) \, dr \) as \( \int B \) and stochastic integrals such as \( \int_0^1 B(r) \, dB(r) \) as \( \int B \, dB \). We let \( M < \infty \) be a generic positive number, not depending on \( T \) or \( n \).
2 Model and Assumptions

We assume that $y_{it}$ is generated as follows

$$y_{it} = \alpha_i + \beta' F_t + u_{it}$$

$$F_t = F_{t-1} + \varepsilon_t$$

$$z_{it} = \lambda'_i F_t + \epsilon_{it}$$

$i = 1, ..., n; t = 1, ..., T; \beta$ is a $k \times 1$ vector of slope parameters; $F_t = (F_{1t}, ..., F_{kt})'$ is a $k \times 1$ vector of common shocks; $u_{it}$ may be $I(1)$ or $I(0)$ (spurious regression or cointegration relationship); $z_{it}$ is a set of observed exogenous variables.

Define $B_{\varepsilon}$ as the Brownian motion associated with the partial sums of $\varepsilon_t$ with covariance matrix $\Omega_{\varepsilon\varepsilon}$ and $\bar{B}_{\varepsilon}(r)$ as the demeaned Brownian motion associated to the partial sums of $F_t$, i.e., $\bar{B}_{\varepsilon}(r) = B_{\varepsilon}(r) - \int_0^r B_{\varepsilon}(r) \, dr$. The following set of assumptions are used throughout the paper:

**Assumption 1:** (a) Either (i) (cointegration case) $u_{it} = D_i (L) \eta_{it}$, or (ii) (spurious regression case) $\Delta u_{it} = F_i (L) \eta_{it}$ with $F_i (1) \neq 0$ and such that $\sum_i u_{it} \sim I (1)$; for both cases, $\eta_{it} \overset{i.i.d.}{\sim} (0, \sigma^2_n)$ over $t$ and $i$, with $E |\eta_{it}|^8 < M$, $\sum_{j=0}^{\infty} j |D_{ij}| < M$, $\sum_{j=0}^{\infty} j |F_{ij}| < M$ and $D_i^2 (1) \sigma_n^2 > 0$, $F_i^2 (1) \sigma_n^2 > 0$; the two MA processes $u_{it} = D_i (L) \eta_{it}$ and $\Delta u_{it} = F_i (L) \eta_{it}$ are assumed to be invertible; (b) (time-series and cross-sectional correlation) letting $E (u_{it}u_{js}) = \tau_{ij,ts} = \tau_{ij,|t-s|}$ and $E (\Delta u_{it}\Delta u_{js}) = \gamma_{ij,ts} = \gamma_{ij,|t-s|}$, as $n \to \infty$ a law of large numbers (LLN) and a CLT hold for the quantities $n^{-1/2} \sum_{i} u_{it}$ and $n^{-1/2} \sum_{i} \Delta u_{it}$.

**Assumption 2:** $\varepsilon_t = C_0 (L) w_t$ where $C_0 (L) = \sum_{j=0}^{\infty} C_j L^j$; (a) $w_t \overset{i.i.d.}{\sim} (0, \Sigma_w)$ with $E \| w_t \|^{4+\delta} \leq M$ for some $\delta > 0$; (b) $Var (\Delta F_t) = \Sigma_{\Delta F} = \sum_{j=0}^{\infty} C_j \Sigma_w C'_j$ is a positive definite matrix; (c) $\sum_{j=0}^{\infty} j \| C_j \| < M$ and (d) $C (1)$ has full rank.

**Assumption 3:** $E \| F_0 \|^4 \leq M$ and $E |u_{it}|^4 \leq M$.

**Assumption 4:** The loadings $\lambda_i$ are non random quantities such that (a) $\| \lambda_i \| \leq M$; (b) either $n^{-1} \sum_{i=1}^{n} \lambda_i \lambda'_i = \Sigma_{\Lambda}$ if $n$ is finite, or $\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \lambda_i \lambda'_i = \Sigma_{\Lambda}$. 

10
Assumption 5: \( e_{it} = G_i(L)\nu_{it} \) where (a) \( \nu_{it} \overset{i.i.d.}{\sim} (0, \sigma^2_{\nu i}) \), \( E|\nu_{it}|^8 < M \), \( \sum_{j=0}^{\infty} j |G_{ij}| < M \) and \( G^2_i(1)\sigma^2_{\nu i} > 0 \); (b) \( E(\nu_{it}\nu_{jt}) = \tau_{ij} \) with \( \sum_{i=1}^{n} |\tau_{ij}| \leq M \) for all \( j \); (c) \( E |n^{-1/2} \sum_{i=1}^{n} [e_{it}e_{is} - E(e_{it}e_{is})]|^4 \leq M \) for every \( (t, s) \); (d) \( E \left[ n^{-1} \sum_{i=1}^{n} e_{it}e_{is} \right] = \gamma_{s-t}, |\gamma_{s-t}| \leq M \) for all \( s \) and \( T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} |\gamma_{s-t}| \leq M \); (e) \( E|\epsilon_{i0}|^4 \leq M \).

Assumption 6: \( \{\varepsilon_t\}, \{u_{it}\} \) and \( \{e_{it}\} \) are three independent groups; \( F_0 \) is independent of \( \{u_{it}\} \) and \( \{e_{it}\} \).

Assumption 1(a) considers the possibility that equation (1) is either a cointegration or a spurious regression. Processes \( u_{it} \) and \( \Delta u_{it} \) are assumed to be invertible MA processes as in Bai (2004) and Bai and Ng (2004), in a similar fashion to processes \( \varepsilon_t \) and \( e_{it} \). Assumption 1(b) also considers the presence of some, limited, cross-sectional dependence among the \( u_{it} \)s or the \( \Delta u_{it} \)s and therefore it rules out the possibility that all the cross-sectional dependence is taken into account by the common factors \( F_t \), e.g., see the related work by Conley (1999).

Even if it refers to a different framework (panel data with common shocks as opposed to factor models), we take a position similar to that in Bai (2003, 2004) and Bai and Ng (2002, 2004). Using the factor models terminology, this means having a model with an “approximate factor structure”, which differs from a strict common factor model where the \( u_{it} \) are assumed to be independent across \( i \), e.g., Chamberlain and Rothschild (1983) and Onatski (2005).

The amount of cross-sectional dependence we allow for in Assumption 1(b) is anyway limited, since we require that it allows a LLN and a CLT to hold for the (rescaled) sequences \( \sum_{i=1}^{n} u_{it} \) and \( \sum_{i=1}^{n} \Delta u_{it} \).

Assumption 2 allows for some weak serial correlation in the dynamics of \( \varepsilon_t \). This process can be described as invertible MA process, implied by the
absolute summability conditions. Both the short run and the long run variance of $\Delta F_t$ are positive definite (Assumptions 2(b) and 2(d), respectively). Note that Assumption 2(d) rules out the possibility that the (common) regressors $F_t$ in model (1) are cointegrated. This requirement is standard in cointegration to have non-degenerate limiting distributions.

Assumption 3 is a standard initial condition requirement. Assumption 4 serves to identify the factor loadings, which, merely for the purpose of a concise discussion, are assumed to be non random. This requirement could be relaxed, as in Bai (2003, 2004) and Bai and Ng (2004), assuming that the $\lambda_i$ are randomly generated and independent of $\varepsilon_t$ and $e_{it}$, and our results would keep holding. Assumption 4(b) ensures that the factor structure is identifiable. Note that it would be possible to relax this assumption by constraining the minimum eigenvalue of $\sum_{i=1}^{n} \lambda_i \lambda_i'$ to tend to infinity as $n \to \infty$, as pointed out by Onatski (2005). This structure would allow factors to be less pervasive than in our framework, thereby allowing the idiosyncratic component $e_{it}$ in equation (2) to have a greater impact in explaining the contemporaneous correlation among the $z_{it}$. Nonetheless, this would be made at the price of losing the possibility to model the $z_{it}$ as a serially correlated process, whilst in our framework some limited time-series and cross-sectional dependence in model (2) is allowed for as one could realize from Assumption 5. As pointed out in Bai (2003), the conditions in Assumption 5 are fairly general and allow for consistency and distribution results to hold for the principal component analysis (PCA).

Assumption 6 also rules out the existence of any form of dependence between factors $F_t$ and $u_{it}$. Therefore, it is a stronger requirement than the simple lack of correlation, and we need it in order to prove the main results in our paper.

The following definitions are employed throughout the paper. Let $h_i$ ($h_i^A$) and $h_{ij}$ ($h_{ij}^A$) be the long run variance for $u_{it}$ ($\Delta u_{it}$) and the long run covariance between processes $u_{it}$ and $u_{jt}$ ($\Delta u_{it}$ and $\Delta u_{jt}$) - we have $h_{ij} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} T_{ij,ts}$ and $h_{ij}^A = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \gamma_{ij,ts}$. Also, let $\bar{h} = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}$ and $\bar{h}^A = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}^A$. Last, the following variances arising from cross sectional aggregation of the $u_{it}$ and the $\Delta u_{it}$ are often used in our re-
sults: \( \tau_{ts} = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \tau_{ij,ts} \), and \( \gamma_{ts} = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij,ts} \).

### 3 Asymptotic Theory

The main objective of this paper is to derive the rate of convergence and limiting distribution of the OLS, \( \hat{\beta} \)

\[
\hat{\beta} = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} (F_i - \bar{F})(F_i - \bar{F})' \right]^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (F_i - \bar{F}) y_{it} \tag{7}
\]

in equation (1), where \( \bar{F} = T^{-1} \sum_{t=1}^{T} F_t \), and \( \hat{\beta}^{FD} \) when using equation (3),

\[
\hat{\beta}^{FD} = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta F_i \Delta F_i' \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta F_i \Delta y_{it} \right]. \tag{8}
\]

We consider several features of (1) and (3):

1. the shocks \( F_t \) can either be known or (more likely) unobservable;

2. the relationship described by equation (1) can be either a cointegration relationship or a spurious regression. As pointed out by Kao (1999) and Phillips and Moon (1999), convergence is obtained at rate \( \sqrt{n} \) in panel spurious regression models and \( \sqrt{nT} \) for panel cointegrated models;

3. the time series dimension \( T \) and the cross-sectional dimension \( n \) can be either fixed or large.

We first start with the exploration of the case of known common shocks (Section 3.1) and then move to the case of unknown common shocks (Section 3.2).

### 3.1 Observable \( F_t \)

When \( F_t \) is known we have:

\[
\hat{\beta} - \beta = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} W_i W_i' \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} W_i u_{it} \right]. \tag{9}
\]
where $W_t = F_t - \bar{F}$, and
\[
\hat{\beta}^{FD} - \beta = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta F_i \Delta F_i' \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta F_i \Delta u_{it} \right]. \tag{10}
\]

The convergence rate and the limiting distribution for $\hat{\beta}$ are stated in the following theorem.

**Theorem 1** Suppose Assumptions 1-6 hold, and define $Z \sim N(0, I_k)$, independent of the $F_t$s. For fixed $n$ and $T \to \infty$
\[
T \left( \hat{\beta} - \beta \right) = \frac{1}{n} \left( \int B_x B_x' \right)^{-1/2} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij} \right)^{1/2} \times Z, \tag{11}
\]

if equation (1) is a cointegration relationship, and
\[
\left( \hat{\beta} - \beta \right) = \left( \int \tilde{B}_x \tilde{B}_x' \right)^{-1} \left( \int \tilde{B}_x B_u \right) \left( \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij} \right)^{1/2}, \tag{12}
\]

if (1) is a spurious regression. For fixed $T$ and $n \to \infty$, we have
\[
\sqrt{n} \left( \hat{\beta} - \beta \right) = \left( \sum_{t=1}^{T} W_t W_t' \right)^{-1} \left( \sum_{t=1}^{T} \sum_{s=1}^{T} W_t W_s' \sigma_{ts} \right)^{1/2} \times Z, \tag{13}
\]

if (1) is a cointegration regression, whilst if it is a spurious relationship we have
\[
\sqrt{n} \left( \hat{\beta} - \beta \right) = \left( \sum_{t=1}^{T} W_t W_t' \right)^{-1} \left( \sum_{t=1}^{T} W_t \bar{u}_t \right), \tag{14}
\]

where $\bar{u}_t = \lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} u_{it}$.

Let equation (1) be a cointegration relationship; as $(n, T) \to \infty$ with $n/T \to 0$:
\[
\sqrt{n} T \left( \hat{\beta} - \beta \right) = \left[ \sqrt{h} \left( \int B_x B_x' \right)^{-1/2} \right] \times Z; \tag{15}
\]

as $(n, T) \to \infty$ with $T/n \to 0$:
\[
T^{3/2} \left( \hat{\beta} - \beta \right) = \left( \int \tilde{B}_x \tilde{B}_x' \right)^{-1} \Delta_1, \tag{16}
\]

where $\Delta_1$ is defined in the proof - see equation (61).

Let equation (1) be a spurious regression; as $(n, T) \to \infty$ with $n/T \to 0$:
\[
\sqrt{n} \left( \hat{\beta} - \beta \right) = \left( \int \tilde{B}_x \tilde{B}_x' \right)^{-1} \left( \int \tilde{B}_x B_u \right) \sqrt{h \Delta}; \tag{17}
\]
as \((n, T) \to \infty\) with \(T/n \to 0\):

\[
\sqrt{T} \left( \hat{\beta} - \beta \right) \Rightarrow \left( \int \bar{B}_z \bar{B}_z' \right)^{-1} \Delta_2,
\]

(18)

where \(\Delta_2\) is defined in the proof - see equation (64).

**Proof.** See Appendix B. ■

Equation (12) is a standard results in the literature. With respect to the speed of convergence, when \((n, T) \to \infty\) our results lead to the same rates of convergence as in Phillips and Moon (1999) and Kao (1999) for both the cointegration and the spurious regression case. Consistency is achieved under the spurious regression case as well, where the rate of convergence is \(\sqrt{n}\). Note that, irrespective of model (1) being a cointegration regression or a spurious regression, large \(n\) allows for consistency to hold.

For the case of \((n, T) \to \infty\) with \(n/T \to 0\), the rate of convergence for \(\hat{\beta}\) is the same as in Phillips and Moon (1999) under the case of contemporaneous independence across cross-sectional units. The limiting distributions in (15) and (17) are mixed normal rather than normal, as in Phillips and Moon (1999). The mixed normality is due to both \(F_t\) being nonstationary and common across cross-sectional units, as can be seen by considering equation (13) for \(T \to \infty\). The design matrix \((nT^2)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} F_i F_i' = T^{-2} \sum_{t=1}^{T} F_t F_t'\) converge in distribution to a random matrix, namely \(\int \bar{B}_z \bar{B}_z'\), rather than to a constant. Of course, \((nT^2)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} F_i F_i'\) would converge to a constant (in probability) if \(F_t\) were not common shocks, i.e., if \(F_t\) were replaced by \(F_{it}\). Theorem 1 also explores the case of a “short” panel, where \(T/n \to 0\). In this case, \(\hat{\beta}\) is still consistent, irrespective of whether (1) is a cointegration or a spurious regression, although consistency is achieved at a “slower” rate than in the case whereby \(n/T \to 0\). The limiting distributions, given in (16) and (18), are non standard, and they depend upon \(\Delta_1\) and \(\Delta_2\). As shown in Appendix C, these terms come from the bias arising from the BN decomposition of \(F_t\) and \(u_{it}\), and thus they depend upon the assumptions on the DGP of \(F_t\) and \(u_{it}\). A similar argument is discussed in Phillips and Moon (1999); of course, if \(F_t\) and \(u_{it}\) had initial values...
set equal to zero, and if they were i.i.d. processes, then (15) and (17) would hold for \((n,T) \to \infty\) for all combinations of \(n\) and \(T\).

The convergence rates and the limiting distributions for \(\hat{\beta}^{FD}\) are reported in the following theorem.

**Theorem 2** Suppose Assumptions 1-6 hold and define \(Z \sim N(0, I_k)\), independent of the \(\Delta F_t\)s.

For fixed \(n\) and \(T \to \infty\)

\[
\sqrt{T} \left( \hat{\beta}^{FD} - \beta \right) \Rightarrow n^{-1} \Sigma_{\Delta F}^{-1/2} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}^2 \right)^{1/2} \times Z. \tag{19}
\]

For \(T\) fixed and \(n \to \infty\), we have

\[
\sqrt{n} \left( \hat{\beta}^{FD} - \beta \right) \Rightarrow \left( \sum_{t=1}^{T} \Delta F_t \Delta F_t' \right)^{-1} \left( \sum_{t=1}^{T} \sum_{s=t}^{T} \Delta F_t \Delta F_s' \right) \times Z. \tag{20}
\]

When \((n,T) \to \infty\), under \(n/T \to 0\) it holds that

\[
\sqrt{nT} \left( \hat{\beta}^{FD} - \beta \right) \Rightarrow \left( \Sigma_{\Delta F}^{-1/2} \sqrt{h_{\Delta}} \right) \times Z; \tag{21}
\]

as \((n,T) \to \infty\), under \(T/n \to 0\), it holds that

\[
T \left( \hat{\beta}^{FD} - \beta \right) \overset{P}{\to} \Sigma_{\Delta F}^{-1} \Delta_3, \tag{22}
\]

where \(\Delta_3\) is defined in the proof - see equation (??).

**Proof.** See Appendix C.

Since the first differenced model is always stationary, irrespective of whether equation (1) is a cointegration equation or a spurious regression, one can always apply the CLT to obtain the limiting distribution of \(\hat{\beta}^{FD} - \beta\).

Equation (21) states that the limiting distribution of \(\hat{\beta}^{FD} - \beta\) is normal instead of mixed normal, despite the strong dependence across cross-sectional units. This can be seen from equation (20), which gives the limiting distribution for \(T\) fixed and \(n \to \infty\). The matrix \(T^{-1} \sum_{t=1}^{T} \Delta F_t \Delta F_t'\) is random for finite \(T\), but it converges to a constants as \(T \to \infty\) due to a LLN.
Equation (22) refers to a panel where \( T/n \to 0 \), and thus the number of cross-sectional units is “much larger” than the number of time-series observations. This case is similar to that found in Theorem 1: when \( T/n \to 0 \), the bias in the BN decomposition dominates, thereby making the limiting distribution non-standard and depending upon the assumptions on the DGP of \( F_t \) and \( u_{it} \). Of course, if one knows \( \Sigma^{-1}_{\Delta F} \Delta_3 \) (or could estimate it consistently at a rate \( \phi_{nT} \)) the remainder in the BN decomposition of \( \hat{\beta}^{FD} - (\beta + \Sigma^{-1}_{\Delta F} \Delta_3) \) has mean zero (or of order \( O_p(n^{1/2}) \)). Under additional assumptions, the bias in the BN decomposition for \( \hat{\beta} \) is of order \( O_p(n^{1/2}) \). Therefore, when normalized by \( \sqrt{nT} \), the bias is always dominated by the martingale approximation to \( \Sigma^{-1}_{\Delta F} \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta F_t \Delta u_{it} \) - resp., of order \( O_p(n^{1/2}) \). The former case implies that (21) holds as \( (n,T) \to \infty \), with no restriction on the rate of expansion of \( n \) and \( T \) as they pass to infinity. This is consistent with Phillips and Moon (1999, p. 1074, Remark (a)), and a similar argument could be in principle applied to Theorem 1.

### 3.2 Unobservable \( F_t \)

We turn now to the case when common shocks are unknown and thus they need to be estimated. The asymptotics of \( \hat{\beta} \) and \( \hat{\beta}^{FD} \) are affected by the errors in estimating shocks \( F_t \).

Let \( \hat{F}_t \) be an estimate of \( F_t \). Denote \( \hat{W}_t = \hat{F}_t - \sum_{t=1}^{T} \hat{F}_t \). Estimation of \( \beta \) using the model in levels or first differences respectively are now given by:

\[
\hat{\beta} = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_i \hat{W}'_t \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_i \hat{y}_{it} \right] \tag{23}
\]

and

\[
\hat{\beta}^{FD} = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta \hat{F}_i \Delta \hat{F}'_t \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta \hat{F}_i \Delta y_{it} \right] \tag{24}
\]

with estimation errors:

\[
\hat{\beta} - \beta = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_i \hat{W}'_t \right]^{-1} \left\{ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_t \left[ (\hat{W}_t - \hat{W}_t)' \beta + u_{it} \right] \right\} \tag{25}
\]
We assume that the number of common shocks \( k \) is known. This does not lead to any loss of generality since the distribution of the estimated shocks does not depend on whether \( k \) is known or estimated, and therefore the estimation error that arises from using \( \hat{k} \) instead of \( k \) does not play any role as long as \( \hat{k} \) is consistent, e.g., see Bai (2003, p. 143, note 5).

### 3.2.1 The case of \( n \) and \( T \) large

In this section, we estimate the common shocks \( F_t \) using the principal component estimator. This means minimizing either

\[
V_b(k) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (z_{it} - \lambda' F_i)^2
\]

when considering \( F_t \) in levels, or

\[
V_a(k) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (\Delta z_{it} - \lambda' \Delta F_i)^2
\]

when estimating shocks \( \Delta F_t \) from (2). Consider the \( T \times n \) matrix \( Z = (z_1, ..., z_T)' \), and the \( T \times k \) matrix of shocks \( F = (F_1, F_2, ..., F_T)' \). Then each objective function \( V_a(k) \) or \( V_b(k) \) can be minimized by concentrating out \( \lambda \) and obtaining estimates \( \Delta \hat{F} \) and \( \hat{F} \) using the normalizations \( \Delta \hat{F}' \Delta \hat{F} / T = I_k \) or \( \hat{F}' \hat{F} / T^2 = I_k \). The estimated shock matrices \( \Delta \hat{F} \) and \( \hat{F} \) are \( \sqrt{T} \) and \( T \) times respectively the eigenvectors corresponding to the \( k \) largest eigenvalues of the \( T \times T \) matrices \( \Delta Z \Delta Z' \) or \( ZZ' \).

In the context of unobservable common factors, the problem of identification arises. It is well known (see e.g., Bai, 2003, and Bai, 2004) that the solutions to the above minimization problems are not unique, e.g., when estimating shocks \( \Delta F_t \) and \( F_t \), these are not directly identifiable even though they are up to a transformation. This entails that whilst it is possible to “consistently” estimate
the space spanned by the common factors $\Delta F_t$ and $F_t$, it is not possible to estimate $\Delta F_t$ and $F_t$ themselves. Thus, $\hat{F}_t$ and $\Delta \hat{F}_t$ are, respectively, “consistent” estimators for $H'_1 F_t$ and $H'_1 \Delta F_t$, where $H_1$ is a non singular $k \times k$ matrix, meaning that $\hat{F}_t - H'_1 F_t$ and $\Delta \hat{F}_t - H'_1 \Delta F_t$ converge to zero in some sense. This makes it impossible to estimate $\beta$ consistently, although once again the space spanned by $\beta$ can be estimated consistently. Due to the rotational indeterminacy of the estimation of $F_t$ and $\Delta F_t$, the estimators $\hat{\beta}$ and $\tilde{\beta}^{FD}$ may “consistently” estimate $H_1^{-1} \beta$, so that $\hat{\beta} - H_1^{-1} \beta$ may converge to zero in some sense as $n$ and $T$ (or either) pass to infinity. Whilst there is no direct consistency result for $\beta$, being able to estimate the space it spans is sufficient for many purposes. For example, the quantity $\beta F_t$ can be consistently estimated by $\hat{\beta} \hat{F}_t$, and therefore predicting using (1) is feasible. Also, it is possible to carry out inference on e.g., the significance of the $F_t$s as regressors in (1) - see also a similar discussion in Bai (2003, p. 145). Likewise, it is impossible to consistently estimate the loadings $\lambda_i$ in (2), although it is possible to consistently estimate the space spanned, i.e. $\hat{\lambda}_i - H_2 \lambda_i$, where the rotation matrix $H_2$ is $n \times n$, can be shown to converge to zero in some sense.

The convergence rate and the limiting distribution for $\hat{\beta}$ are in the following theorem.

**Theorem 3** Suppose Assumptions 1-6 hold. Let equation (1) be a cointegration relationship; as $(n, T) \to \infty$ with $n/T \to 0$:

$$\sqrt{nT} \left( \hat{\beta} - H_1^{-1} \beta \right) \Rightarrow \left( H'_1 \int B_{\alpha} B_{\alpha}' H_1 \right)^{-1/2} h + \beta' H_1^{-1} \tilde{Q}_B \Gamma \tilde{Q}_B H_1^{-1} \beta \right]^{1/2} \times Z \tag{27}$$

where $Z \sim N_1 (0, I_k)$ independent of the $\sigma$-field generated by the common shocks $F_t$ and of the random matrix $\tilde{Q}_B$. Also

$$T^{-2} \sum_{t=1}^{T} \hat{W}_t W'_t \Rightarrow \tilde{Q}_B$$

and

$$\Gamma = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda'_j E (e_i e'_j) .$$
When \((n,T) \to \infty\) with \(T/n \to 0\), it holds that
\[
T^{3/2} \left( \hat{\beta} - H_1^{-1} \beta \right) \Rightarrow \left( H_1' \int \bar{B}_z \bar{B}_z' H_1 \right)^{-1} \Delta_1
\]
where \(\Delta_1\) is defined in (61).

Let equation (1) be a spurious regression; as \((n,T) \to \infty\) with \(n/T \to 0\):
\[
\sqrt{n} \left( \hat{\beta} - H_1^{-1} \beta \right) \Rightarrow \left( H_1' \int \bar{B}_z \bar{B}_z' H_1 \right)^{-1} \left( H_1' \int \bar{B}_z B_u \right) \sqrt{h^A}.
\]
As \((n,T) \to \infty\) with \(T/n \to 0\), it holds that
\[
\sqrt{T} \left( \hat{\beta} - H_1^{-1} \beta \right) \Rightarrow \left( H_1' \int \bar{B}_z \bar{B}_z' H_1 \right)^{-1} \Delta_2
\]
where \(\Delta_2\) is defined in (64).

**Proof.** See Appendix B. ■

The estimator \(\hat{\beta}\) is always consistent, even though \(T/n \to 0\) results in a slower rate of convergence and in a degenerate behavior of the numerator of \(\hat{\beta} - H_1^{-1} \beta\). When \(n/T \to \infty\), results for the case of (1) being a cointegrating regression are essentially the same as in equation (15) in Theorem 1. The only difference is that now the variance of \(\hat{\beta} - H_1^{-1} \beta\) is “inflated” by the non-negative random variable \(\beta' H_1^{-1} \hat{Q}_b \hat{Q}_b' H_1^{-1} \beta\), which arises from the estimation error when replacing \(F_t\) with \(\hat{F}_t\).

Notice the consequence of equation (1) being a spurious regression: as long as the number of cross sectional units \(n\) is “smaller” than \(T\), the classical \(\sqrt{n}\) consistency holds, and we have the same limiting distribution as in equation (17).

In both cases, the limiting distributions become non standard when \(T/n \to 0\).

The convergence rate and the limiting distribution for \(\hat{\beta}^{FD}\) are in the following theorem.

**Theorem 4** Suppose Assumptions 1-2 and 4-6 hold.

If \(\frac{n}{T} \to 0\)
\[
n \left( \beta^{FD} - H_1^{-1} \beta \right) \xrightarrow{P} (H_1' \Sigma_{\Delta F} H_1)^{-1} \times Z,
\]

(31)
where $Z$ is $N(0, Q)$ with

$$Q = \lim_{n,T \to \infty} \frac{1}{T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{u=1}^{T} \sum_{v=1}^{T} H_t' \Delta F_t \left( \Delta \hat{F}_s - H_t' \Delta F_s \right)' V^{-1} \beta \times$$

$$\beta' V^{-1} \left( \Delta \hat{F}_u - H_t' \Delta F_u \right)' \Delta F_t H_1 \times$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} E \{[e_{it}e_{is} - E(e_{it}e_{is})][e_{ju}e_{jv} - E(e_{ju}e_{jv})]\},$$

and $V$ is the probability limit of the diagonal matrix consisting of the first $k$ eigenvalues of $(nT)^{-1} \Delta Z \Delta Z'$ in decreasing order.

If $T/n \to 0$

$$T \left( \tilde{\beta}^{FD} - H_1^{-1} \beta \right) \overset{p}{\to} \tilde{h}_e V^{-1} \beta + (H_t' \Sigma_{\Delta F} H_1)^{-1} \Delta_3$$

(32)

where $\tilde{h}_e$ is the long-run variance of $\lim_{n \to \infty} n^{-1/2} \sum_{i=1}^{n} e_{it}$ and $\Delta_3$ is defined in (??).

**Proof.** See Appendix C. ■

Notice that in this case we have a degenerate limiting distribution when $T/n \to 0$, despite having a consistent estimate. The distribution limit depends on the bias arising in the BN decomposition, $\Delta_3$, but also on the presence of the error term $\Delta \hat{F}_t - H_t' \Delta F_t$.

3.2.2 The case of $T$ fixed and $n$ large

When $T$ is fixed and $n$ is large, consistent estimation of shocks is still possible, see Connor and Korajczyk (1986) and Bai (2003). However, the following restriction is necessary:

**Assumption 7:** $E(e_{it}e_{is}) = 0$ for all $t \neq s$.

Assumption 7 rules out the possibility of serial correlation in the DGP of the $e_{it}$, and therefore this is a constraint on Assumption 5(d). However, contemporaneous correlation and cross-sectional heteroscedasticity are preserved.

Under Assumptions 4-7, we know that shocks estimation is $\sqrt{n}$ consistent, i.e., we have both

$$\hat{F}_t - H_t' F_t = O_p \left(n^{-1/2}\right)$$
and
\[ \Delta \hat{F}_t - H_1' \Delta F_t = O_p \left( n^{-1/2} \right) \]
for all \( t \).

**Theorem 5** Suppose Assumptions 1-7 hold; then for every consistent estimator \( \hat{F}_t \) of \( H_1'F_t \) and for fixed \( T \) and \( n \to \infty \) we have the same result as in equation (14).

**Proof.** See Appendix C.

**Theorem 6** Suppose Assumptions 1-7 hold; then for every consistent estimator \( \Delta \hat{F}_t \) of \( H_1' \Delta F_t \) and for fixed \( T \) and \( n \to \infty \) we have the same result as in equation (20).

**Proof.** See Appendix C.

Theorems (5) and (6) do not anyway require \( \sqrt{n} \) consistency, since they ensure the consistency of \( \hat{\beta} \) and \( \hat{\beta}^{FD} \) for any consistent estimate of the shocks, irrespective of the rate of convergence. In both cases we have the same results as we would have if the \( F_t \)s were observable. Therefore, when \( T \) is fixed, having large \( n \) makes it indifferent to use observed or estimated shocks as long as shocks are estimated consistently.

**3.2.3 The case of \( n \) fixed and \( T \) large**

In what follows, we develop the inferential theory for the case when shocks are unknown and the cross-sectional dimension \( n \) is finite. Rewriting model (2) in the vector form, one gets:
\[ z_t = \Lambda F_t + e_t \]  
(33)
where \( z_t = (z_{1t}, ..., z_{nt})' \), \( e_t = [e_{1t}, ..., e_{nt}]' \), and \( \Lambda = (\lambda_1, \lambda_2, ..., \lambda_n)' \). One can estimate \( \Lambda \) using PCA. A feasible estimator of \( \Lambda \), \( \hat{\Lambda} \), is given by the \( \sqrt{n} \) times the eigenvectors corresponding to the \( k \) largest eigenvalues of \( Z'Z \). Notice that this estimator exploits the normalization \( \hat{\Lambda}'\hat{\Lambda}/n = I_k \), and it turns out to be computationally convenient for the case of \( n < T \). Henceforth, for sake of the
notation and without loss of generality, we modify Assumption 4 by assuming that \( n^{-1} \sum_{i=1}^{n} \lambda_i \lambda'_i = I_k \).

The following theorem characterizes consistency and limiting distribution of \( \hat{\Lambda} \).

**Proposition 1** Under Assumptions 3-6 we have

\[
T \left( \hat{\Lambda} - H_2 \Lambda \right) \Rightarrow \left[ I_n - n^{-1} H_2 \Lambda \left( H'_1 \int B_e' B_e \right) \Lambda' H_2' \right] \left( \int dW_e B_e' H_1 \right) \left( H'_1 \int B_e' B_e \right)^{-1} \left( \int dW_e B_e' H_1 \right) \Lambda H_2' + n^{-1} \left[ I_n - 2n^{-1} \Lambda H_2 \left( H'_1 \int B_e' B_e \right) \Lambda' H_2' \right] \Omega_e \Lambda H_2
\]

where \( W_e \) is the Wiener process associated to the partial sums of \( e_t \) and \( \Omega_e = E(e_t e'_t) \).

**Proof.** See Appendix B.  ■

Note that in this case we have a \( T \)-consistent estimate of \( \Lambda \), even though the PCA of \( F_t \) is not consistent, e.g., see Bai (2004) and Proposition 2 below, when \( n \) is finite.

Define the limiting distribution of \( T \left( \hat{\Lambda} - H_2 \Lambda \right) \) as \( D^1_{\Lambda} \), i.e., \( T \left( \hat{\Lambda} - H_2 \Lambda \right) \Rightarrow D^1_{\Lambda} \). Given the restriction \( \Lambda' \hat{\Lambda} / n = I_k \), the OLS estimator of \( F_t \), obtained regressing the \( z_t \) on the estimated loadings \( \hat{\Lambda} \), is

\[ \hat{F}_t = n^{-1} \hat{\Lambda}' z_t. \]

The following proposition characterizes (the inconsistency of) this estimator:

**Proposition 2** Consider \( \hat{F}_t = n^{-1} \hat{\Lambda}' z_t \), and also the first difference estimator, \( \Delta \hat{F}_t = n^{-1} \hat{\Lambda} \Delta z_t \). Then

\[
\max_{1 \leq t \leq T} \left\| \hat{F}_t - H'_1 F_t \right\| = O_p(1) \tag{35}
\]

and

\[
\max_{1 \leq t \leq T} \left\| \Delta \hat{F}_t - H'_1 \Delta F_t \right\| = O_p(1) \tag{36}
\]

uniformly in \( t \).
Proof. See Appendix C.

Proposition 2 states that the estimates of the shocks and of their first difference are inconsistent, in that the estimation error does not die out as $T \to \infty$. However this inconsistency has no impact on the consistency of $\hat{\beta}$ and $\hat{\beta}^{FD}$, though it affects their asymptotic distributions. See the proofs of Theorems 7 and 8.

The convergence rate and the limiting distribution for $\hat{\beta}$ are in the following theorem.

Theorem 7 For the estimator $\hat{\beta}$, we have:

$$T \left( \hat{\beta} - H_1^{-1} \beta \right) \Rightarrow \left[ H_1' \int \tilde{B}_t \bar{B}_t' H_1 \right]^{-1} \left\{ \int H_1' \tilde{B}_t d\bar{B}_t \left( \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij} \right)^{1/2} \right. $$

$$\left. - n^{-1} \left[ H_1' \int \tilde{B}_t d\bar{B}_t' H_2 \Delta 
\Delta^{-1} \beta + n^{-1} \bar{N} \bar{H}_2^2 \sum e H_2 \Delta \Delta^{-1} \beta \right] \right\} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}^2 \right)^{1/2}$$

(37)

where $\tilde{B}_t$ is the demeaned Brownian motion associated to the partial sums of $e_t$ and $\Sigma = \text{Var}(e_t)$. When this is a spurious relationship, one gets

$$\hat{\beta} - H_1^{-1} \beta \Rightarrow \left( H_1' \int \tilde{B}_t \bar{B}_t' H_1 \right)^{-1} \left( H_1' \int \tilde{B}_t d\bar{B}_t \right) \left( \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}^2 \right)^{1/2}$$

(38)

Proof. See Appendix C.

Note that even though common shocks cannot be estimated consistently, $\hat{\beta}$ is consistent when (1) is a cointegration relationship but inconsistent when (1) instead is a spurious regression. With respect to the case of observable shocks, shock estimation has an impact on the limit distribution of $\hat{\beta} - H_1^{-1} \beta$ when equation (1) is a cointegration regression - see equation (37) above. On the other hand, it does not affect the asymptotic distribution when equation (1) is a spurious regression - see equation (38).

Equations (37) and (38) show an important common feature of this theoretical framework. Only the numerators of equation (37) and (38) depend on whether equation (1) is a cointegrating or spurious regression, whilst the denominators are not affected. This is due to the fact (detailed in the proof)
that though $\hat{F}_t$ is not a consistent estimator for $F_t$, the quantity $\sum \hat{F}_t \hat{F}_t'$ is a consistent estimator for $\sum F_t F_t'$ for any consistent estimator of the loadings $\hat{\Lambda}$.

The convergence rate and the limiting distribution of $\hat{\beta}^{FD}$ are in the following theorem.

**Theorem 8** For the first difference estimator $\hat{\beta}^{FD}$, we have:

$$\hat{\beta}^{FD} - H_1^{-1} \beta - H_1^{-1} \beta + n \left[ H_1' \Sigma_{\Delta \epsilon} \right]^{-1} \left[ H_1' \Sigma_{\Delta F} H_1 \right] H_1^{-1} \beta$$

(39)

where $\Sigma_{\Delta \epsilon} = \text{Var} (\Delta e_t)$ and $\Sigma_{\Delta \epsilon} = \Lambda \left( H_1' \Sigma_{\Delta F} H_1 \right) \Lambda' + \Sigma_{\Delta \epsilon}$.

**Proof.** See Appendix C. ■

The estimator $\hat{\beta}^{FD}$ is inconsistent. This is due to the two terms $\sum \Delta F_t \Delta F_t'$ and $\sum \Delta e_t \Delta e_t'$ in the denominator having the same asymptotic magnitude, rather than to the common shock estimates being inconsistent. Also, this holds for any consistent estimator $\hat{\Lambda}$ (see proof in Appendix C).

### 4 Extensions

In this section, we consider two extensions of our basic framework:

(i) the case of model (2), where our basic framework (1) also contains some idiosyncratic shocks;

(ii) the case where model (1) is misspecified, and the common shocks $F_t$ actually have unit specific slopes, say $\beta_i$.

#### 4.1 The case of idiosyncratic shocks

Model (1) assumes that the DGP of $y_{it}$ depends only on a set of common shocks. In this section, we briefly consider the case where the model is augmented to take into account the presence of unit-specific variables. Even though the algebra becomes more tedious, all the results derived in the previous section still hold. The only novelty is the covariance between the common shocks $F_t$ and the unit specific regressors.
Recall the augmented model (4)

\[ y_{it} = \alpha_i + \beta' F_t + \gamma' x_{it} + u_{it} \]

with \( i = 1, \ldots, n, t = 1, \ldots, T \), where \( \beta \) and \( \gamma \) are \((k \times 1)\) and \((p \times 1)\) vectors of slope parameters. The idiosyncratic variables \( x_{it} \) are a \((p \times 1)\) vector of observable \( I(1) \) individual-specific regressors, defined as

\[ x_{it} = \tau_i G_t + \omega_{it} \quad \text{(40)} \]

where \( G_t \) is an \( R \times 1 \) vector of \( I(1) \) variables that may contain some elements of \( F_t \), and \( \tau_i \) is a \( p \times R \) matrix and

\[ \omega_{it} = \omega_{it-1} + \epsilon_{it} \quad \text{(41)} \]

with \( \omega_{it} \) assumed for simplicity i.i.d. across \( i \) and such that \( \{\epsilon_{it}\}, \{\epsilon_{it}\}, \{F_{it}\} \) and \( \{\omega_{it}\} \) are independent groups. Equation (40) considers the possible presence of "correlation" between \( F_t \) and \( x_{it} \). A similar framework that allows for cross dependence among the idiosyncratic regressors and dependence between the idiosyncratic regressors and the common regressors is considered in Pesaran (2006) and Kapetanios, Pesaran and Yamagata (2006). Thus, cross dependence is accounted for directly, via \( F_t \), and indirectly, via the factor structure in \( x_{it} \).

The impact of the presence of the \( x_{it} \)'s on the LS estimator of \( \beta \) will be discussed considering (for the purpose of a concise discussion) the case of observable \( F_t \). Let \( \tilde{x}_{it} = x_{it} - T^{-1} \sum_{t=1}^T x_{it} \); then,

\[
\begin{bmatrix}
\hat{\beta} - \beta \\
\hat{\gamma} - \gamma
\end{bmatrix} = \begin{bmatrix}
\sum_{i=1}^n \sum_{t=1}^T W_i W_t' & \sum_{i=1}^n \sum_{t=1}^T W_i \tilde{x}_{it}' \\
\sum_{i=1}^n \sum_{t=1}^T \tilde{x}_{it} W_t' & \sum_{i=1}^n \sum_{t=1}^T \tilde{x}_{it} \tilde{x}_{it}'
\end{bmatrix}^{-1} \begin{bmatrix}
\sum_{i=1}^n \sum_{t=1}^T W_i u_{it} \\
\sum_{i=1}^n \sum_{t=1}^T \tilde{x}_{it} u_{it}
\end{bmatrix}
\]

(42)

and

\[
\begin{bmatrix}
\hat{\beta}^{FD} - \beta \\
\hat{\gamma}^{FD} - \gamma
\end{bmatrix} = \begin{bmatrix}
\sum_{i=1}^n \sum_{t=1}^T \Delta F_i \Delta F_t' & \sum_{i=1}^n \sum_{t=1}^T \Delta F_i \Delta x_{it}' \\
\sum_{i=1}^n \sum_{t=1}^T \Delta x_{it} \Delta F_t' & \sum_{i=1}^n \sum_{t=1}^T \Delta x_{it} \Delta x_{it}'
\end{bmatrix}^{-1} \begin{bmatrix}
\sum_{i=1}^n \sum_{t=1}^T \Delta F_i \Delta u_{it} \\
\sum_{i=1}^n \sum_{t=1}^T \Delta x_{it} \Delta u_{it}
\end{bmatrix}
\]

(43)
Thus, the asymptotics of \( \hat{\beta} - \beta \) (and of \( \hat{\beta}^{FD} - \beta \)) depends on terms containing the \( x_{it} \)'s as well, due to the presence of the off-diagonal term \( \sum_{i=1}^{n} \sum_{t=1}^{T} \bar{x}_{it} \bar{W}_{it}' \) in the denominator of (42) and of \( \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta x_{it} \Delta F_{it}' \) in (43).

Depending on whether or not these quantities are asymptotically zero (i.e. \( F_t \) and \( x_{it} \) or their first differences are asymptotically orthogonal), the asymptotic distribution of \( \hat{\beta} \) and \( \hat{\beta}^{FD} \) may or may not change with respect to the results reported in Section 3. In order to investigate the cases whereby \( \hat{\beta} - \beta \) (or \( \hat{\beta}^{FD} - \beta \)) is orthogonal to \( \hat{\gamma} - \gamma \) (\( \hat{\gamma}^{FD} - \gamma \) respectively), consider the following preliminary assumption, where \( \tilde{G}_t = G_t - T^{-1} \sum_{t=1}^{T} G_t \) and similarly \( \tilde{W}_t \).

**Assumption 8:** (1) as \( n \to \infty \), \( n^{-1} \sum_{i=1}^{n} \tau_i = \bar{\tau} \); (2) as \( (n, T) \to \infty \), (i) \( \sum_{i=1}^{n} \sum_{t=1}^{T} \bar{x}_{it} \bar{x}_{it}' = O_p \left( nT^2 \right) \), (ii) \( \sum_{i=1}^{n} \sum_{t=1}^{T} W_{it} \bar{\omega}_{it} = O_p \left( \sqrt{nT^2} \right) \) and (iii) as \( T \to \infty \), \( T^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} G_t W_{it}' \Rightarrow \tilde{B}_G \tilde{B}_G' \); (3) as \( (n, T) \to \infty \), (i) \( \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta \bar{x}_{it} \Delta \bar{x}_{it}' = O_p \left( \sqrt{nT} \right) \) and (iii) as \( T \to \infty \), \( \sum_{t=1}^{T} \Delta F_t \Delta G_t' = O_p \left( T^\kappa \right) \) with \( \kappa = 1/2 \) or 1.

Assumption 8 requires some asymptotic results to hold with respect to the newly introduced variables \( x_{it} \), \( G_t \) and \( \omega_{it} \), and it could be expressed using some more primitive assumptions. For example, parts 2(i) and 3(i) could be shown using the same arguments as in Phillips and Moon (1999); likewise, 2(ii) and 3(ii) could be proved, under suitable assumptions, using similar derivations as for the proofs of Theorems 1 and 2. Note that a necessary condition in order for 3(ii) to hold is that \( \Delta \omega_{it} \) and \( \Delta F_t \) be uncorrelated. The result in 2(iii) could be proved using a FCLT argument. Note that 3(iii) accommodates both situations whereby \( F_t \) and \( G_t \) are independent of each other or overlap. The former case holds under \( \kappa = 1/2 \), which implies that a CLT is required to hold for the sequence \( \sum_{t=1}^{T} \Delta F_t \Delta G_t' \); the latter case is entailed by \( \kappa = 1 \), which requires a LLN to hold for \( \sum_{t=1}^{T} \Delta F_t \Delta G_t' \).

The following theorems provide a summary of the values of \( \sum_{i=1}^{n} \sum_{t=1}^{T} \bar{x}_{it} \bar{W}_{it}' \) and \( \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta x_{it} \Delta F_{it}' \) under various combinations of \( n \) and \( T \).
Proposition 3 Let Assumptions 8(1) and 8(2) hold, and define
\[ \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{x}_{it} W_t' = \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{\omega}_{it} W_t' + \sum_{i=1}^{n} \tau_i \sum_{t=1}^{T} \tilde{G}_t W_t' = a + b. \]
As \((n, T) \to \infty\), \(a = O_p (\sqrt{nT^2})\) and \(b = O_p (nT^2)\) with
\[ \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{G}_t W_t' \Rightarrow \tilde{\tau} \int \tilde{B}_t \tilde{B}_t'. \]

For fixed \(T\) and as \(n \to \infty\), \(a = O_p (\sqrt{n})\) and \(b = O_p (n)\) with \(n^{-1} b \Rightarrow \tilde{\tau} \sum_{t=1}^{T} \tilde{G}_t W_t'.\) For fixed \(n\) and as \(T \to \infty\), \(a = b = O_p (T^2)\).

Proof. See Appendix C. ■

Proposition 4 Let Assumptions 8(1) and 8(3) hold, and define
\[ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta x_{it} \Delta F_t' = \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta \omega_{it} \Delta F_t' + \sum_{i=1}^{n} \tau_i \sum_{t=1}^{T} \Delta G_t \Delta F_t' = a + b. \]
As \((n, T) \to \infty\), \(a = O_p (\sqrt{nT})\) and \(b = O_p (nT^\kappa)\). When \(\kappa = 1\), it holds that
\[ \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta G_t \Delta F_t' \to \tilde{\tau} E (\Delta G_t \Delta F_t'). \]

For fixed \(T\) and as \(n \to \infty\), \(a = O_p (\sqrt{n})\) and \(b = O_p (n)\) with \(n^{-1} b \Rightarrow \tilde{\tau} \sum_{t=1}^{T} \Delta G_t \Delta F_t'.\) For fixed \(n\) and as \(T \to \infty\), \(a = O_p (\sqrt{T})\) and \(b = O_p (T^\kappa)\).

Proof. See Appendix C. ■

Propositions 3 and 4 illustrate the cases when \(\hat{\beta} - \beta \) (\(\hat{\beta}^{FD} - \beta \)) is orthogonal to \(\hat{\gamma} - \gamma \) (\(\hat{\gamma}^{FD} - \gamma \)), thereby making the presence of the idiosyncratic shocks \(x_{it}\) irrelevant for the asymptotics of \(\hat{\beta} - \beta \) (or \(\hat{\beta}^{FD} - \beta \)). Orthogonality between \(\hat{\beta} - \beta \) and \(\hat{\gamma} - \gamma \) requires two necessary conditions to hold: \(n \to \infty\) and \(\tilde{\tau} = 0\). Note that when \(n\) is fixed, \(\hat{\beta} - \beta \) and \(\hat{\gamma} - \gamma \) cannot be orthogonal, irrespective of the presence of the common factors \(G_t\) in the DGP of \(x_{it}\). As far as \(\hat{\beta}^{FD} - \beta \) and \(\hat{\gamma}^{FD} - \gamma \) are concerned, \(n \to \infty\) and \(\tilde{\tau} = 0\) are only sufficient conditions but they are not necessary: if the \(x_{it}\)s are purely idiosyncratic variables, i.e. with no common factor structure (\(\tau_i = 0\)), or if the \(x_{it}\)s do have a common factor that is unrelated to \(F_t\) (i.e. \(\kappa = 1/2\)), then this suffices to have asymptotic orthogonality between \(\hat{\beta}^{FD} - \beta \) and \(\hat{\gamma}^{FD} - \gamma \). Similar results could be proved for the case of unobservable \(F_t\).
4.2 The case of heterogeneous slopes

In this section, we consider the case of heterogeneous slopes, i.e., the case whereby the coefficient of common shocks $F_t$ is unit specific. This entails that (1) now becomes

$$y_{it} = \alpha_i + \beta_i' F_t + u_{it}. \quad (44)$$

Although $\beta_i$ is different across units, the researcher could however be interested in estimating the average $\beta$; a typical case of this is in the literature on convergence (see e.g., Temple, 1999, p. 126). Thus, the model that will be used for estimation is

$$y_{it} = \alpha_i + \beta' F_t + v_{it}$$

where $v_{it} = u_{it} + (\beta_i - \beta)' F_t$; when unobservable common shocks are considered, the model becomes

$$y_{it} = \alpha_i + \beta' (H_1^i)^{-1} \tilde{F}_t + v_{it} \quad (45)$$

where now the error term $v_{it} = u_{it} + (\beta_i - \beta)' F_t - \beta' (\hat{F}_t - H_1^i F_t)$; it is important to note that, as it is well known from Phillips and Moon (1999), neglected heterogeneity introduces a further, nonstationary components in the error term, given by $(\beta_i - \beta)' F_t$. Thus, (45) is always a spurious regression. When first differenced data are used, (45) becomes

$$\Delta y_{it} = \beta' (H_1^i)^{-1} \Delta \tilde{F}_t + \Delta v_{it} \quad (46)$$

where $\Delta v_{it} = \Delta u_{it} + (\beta_i - \beta)' \Delta F_t - \beta' (H_1^i)^{-1} (\Delta \hat{F}_t - H_1^i \Delta F_t)$. For the sake of brevity, we will focus our attention to the case whereby $F_t$ is unobservable, thus analyzing the estimates of $\beta$ from (45) and (46). The
estimation errors are, respectively

\[
\hat{\beta} - H_1^{-1}\beta = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_t \hat{W}_t' \right]^{-1} \times \left\{ \sum_{i=1}^{n} \sum_{t=1}^{T} W_t \left[ u_{it} - \left( \hat{W}_t - H_1' W_t \right)' H_1^{-1} \beta + W_t' (\beta_i - \beta) \right] \right\}.
\]

\[
\hat{\beta}^{FD} - H_1^{-1}\beta = \left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta \hat{F}_t \Delta \hat{F}_t' \right]^{-1} \times \left\{ \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta \hat{F}_t \left[ \Delta u_{it} - \left( \Delta \hat{F}_t - H_1' \Delta F_t \right)' H_1^{-1} \beta + \Delta F_t' (\beta_i - \beta) \right] \right\}.
\]

Consider also the following assumption on the \( \beta_i \)'s.

**Assumption 9:** It holds that (i) \( \beta_i \overset{iid}{\sim} (\beta, \Sigma_\beta) \) with \( E \| \beta_i - \beta \|^{4+\delta} < \infty \) and (ii) \( \beta_i \) is independent of all other quantities.

Assumption 9 yields the usual asymptotic results for \( \{\beta_i\}_{i=1}^{n} \), such as a CLT and a LLN. Part (ii) entails that the long run average parameter \( \beta \) (see Phillips and Moon, 1999 and 2000) is genuinely \( E(\beta_i) \).

To illustrate the main point (summarized in a theorem hereafter), consider (47) - similar arguments hold for (48) and can thus be readily extended. Looking at the numerator, we know from Lemma 2 in Appendix A that, as \( (n,T) \to \infty \), \( \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_t \left( \hat{W}_t - H_1' W_t \right)' H_1^{-1} \beta = O_p \left( nTC_{nT}^{-1} \right) \). Also, as far as \( \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_t u_{it} \) is concerned, its magnitude depends on whether (44) is a cointegrated or a spurious regression, being of order \( O_p \left( \sqrt{nT} \right) + O_p \left( n\sqrt{T} \right) \) in the former case and of order \( O_p \left( \sqrt{nT^2} \right) + O_p \left( nT^{3/2} \right) \) in the latter - the terms \( O_p \left( n\sqrt{T} \right) \) and \( O_p \left( nT^{3/2} \right) \) come from the remainders in the BN decomposition. Neglecting the heterogeneity of the \( \beta_i \)'s entails that a further error term, \( \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_t W_t' (\beta_i - \beta) \), is present. To evaluate its magnitude, consider the
denominator as well
\[
\left[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_i \hat{W}'_t \right]^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_i W'_t (\beta_i - \beta) = \frac{1}{n} H_1^{-1} \sum_{i=1}^{n} (\beta_i - \beta) + \left[ \sum_{t=1}^{T} \hat{W}_t \right]^{-1} \left[ \sum_{t=1}^{T} \hat{W}_t (W_t - \bar{W}_t) \right] \left[ \frac{1}{n} \sum_{i=1}^{n} (\beta_i - \beta) \right]
\]
\[= I + II.\]

Assumption 9 yields \(I = O_p (n^{-1/2})\). Also, \(II = O_p (T^{-2}) O_p (TC_{nT}^{-1}) O_p (n^{-1/2})\), and thus \(II\) is always dominated. Therefore, \(\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_i W'_t (\beta_i - \beta) = O_p (\sqrt{T^2})\), thereby making the estimation error \(\hat{\beta} - \beta = O_p (n^{-1/2})\). This result is in line with the findings in Phillips and Moon (1999) and Moon and Phillips (2000), and it is consistent with the idea that neglecting heterogeneity creates an extra, nonstationary, error term, thereby making the model equivalent to a spurious regression. Thus, when (44) is a cointegrating equation, as \((n, T) \to \infty\) with \(n/T \to 0\), we have \(\hat{\beta} - H_1^{-1} \beta = O_p (n^{-1/2})\) and the limiting distribution of \(\hat{\beta} - H_1^{-1} \beta\) is driven by \(n^{-1/2} H_1^{-1} \sum_{i=1}^{n} (\beta_i - \beta)\). When (44) is a spurious regression, the numerator is driven by both \(n^{-1/2} H_1^{-1} \sum_{i=1}^{n} (\beta_i - \beta)\) and by \(n^{-1/2} T^{-2} \sum_{i=1}^{n} \sum_{t=1}^{T} H_1^{-1} W_t u_{it}\). However, Assumption 9(ii) entails independence (conditional on the common shocks \(F_t\)) between the two terms.

The following propositions summarize the asymptotics of \(\hat{\beta} - H_1^{-1} \beta\) and \(\hat{\beta}^{FD} - H_1^{-1} \beta\); results are presented in two cases, namely \(n \to \infty\) and fixed \(n\).

**Proposition 5** Let Assumptions 1-6 and 9 hold and assume \((n, T) \to \infty\) with \(n/T^2 \to 0\). Then,
\[
\sqrt{n} \left( \hat{\beta} - H_1^{-1} \beta \right) \Rightarrow H_1^{-1} \Sigma_{\beta}^{1/2} \times Z
\]
when (45) is a cointegration relationship, where \(Z \sim N(0, I_k)\), and
\[
\sqrt{n} \left( \hat{\beta}^{FD} - H_1^{-1} \beta \right) \Rightarrow H_1^{-1} \Sigma_{\beta}^{1/2} \times Z.
\]

Also, when (45) is a spurious regression under \(n/T \to 0\)
\[
\sqrt{n} \left( \hat{\beta} - H_1^{-1} \beta \right) \Rightarrow \left( H_1^{1/2} \int \tilde{B}_t \tilde{B}'_t H_1 \right)^{-1} \left( H_1^{-1} \int \tilde{B}_t B_a \right) \sqrt{n} + H_1^{-1} \Sigma_{\beta}^{1/2} \times Z
\]
where the two random variables are independent.
Proof. See Appendix C. ■

Proposition 6 Let Assumptions 1-6 and 9 hold and assume $T \to \infty$. Then

$\hat{\beta} - H_1^{-1} \beta = O_p(1)$ with $E \left( \beta - H_1^{-1} \beta \right) = 0$, Also, $\hat{\beta}^{FD} - H_1^{-1} \beta = O_p(1)$.

Proof. See Appendix C. ■

Propositions 5 and 6 characterize the asymptotics of $\hat{\beta} - H_1^{-1} \beta$ and $\hat{\beta}^{FD} - H_1^{-1} \beta$, under the cases of $n$ passing to infinity and being fixed respectively.

As Proposition 5 shows, neglecting heterogeneity always results in $\sqrt{n}$-consistency; this result is already well known in the nonstationary case, as proved by Phillips and Moon (1999), and it essentially follows form the fact that in the numerators of both (47) and (48), the terms that dominate are, respectively, $\sum_{i=1}^n \sum_{t=1}^T \hat{W}_i \hat{W}_t^t (\beta_i - \beta) = O_p(\sqrt{nt})$ and $\sum_{i=1}^n \sum_{t=1}^T \hat{F}_i \Delta \hat{F}_t^t (\beta_i - \beta) = O_p(\sqrt{nt})$. This is also shown in greater detail in the proofs. The main result is that the error arising from neglected heterogeneity dominates the common shock estimation error.

Proposition 6 states that, when $n$ is fixed, $\beta$ cannot be estimated consistently. Whilst this is in line with Theorems 7 and 8 for the case of spurious regression and for the first-differenced model respectively, it is now also the case for (44) being a cointegrated equation. However, albeit inconsistent, $\hat{\beta}$ is unbiased.

5 Conclusion

This paper developed limiting theory for the OLS estimator for panel models with common shocks, where contemporaneous correlation is generated by both the presence of common regressors (e.g., macro shocks, aggregate fiscal and monetary policies) among cross-sectional units and weak dependence among the error terms. We derived rates of convergence and limiting distributions under a comprehensive set of alternative characteristics of panels: several combinations of the cross-sectional dimension $n$ and the time series dimension $T$; shocks being either observable or unobservable; and stationary and nonstationary panel models, the latter representing either a cointegrating equation or a spurious regression.
When the common shocks are observable, the OLS estimator always provides consistent estimates of the $\beta$, the case of spurious regression with fixed $n$ being the only exception. Consistency holds for all possible combinations of the dimensions of $n$ and $T$, including the case of $n$ fixed, which so far has not been addressed in the literature on nonstationary panel factor models. We extend the study of consistency of OLS estimators to the case when the shocks are unobservable and we prove that consistency always holds, the cases of spurious regression and stationary regression when $n$ is fixed being the only exceptions. All the asymptotics for $(n, T) \to \infty$ has been derived in the joint limit, using an approach based on the application of a conditional MDS CLT.

A central result is represented by the limiting distributions derived under the strong cross-sectional dependence induced by the presence of common shocks. In this case, we obtained mixed normality as consequence of the common shocks being nonstationary; when shocks are stationary, normal distributions are obtained.

In this paper, we primarily consider a panel regression model with only latent shocks $F_t$ as regressors. As we discuss in Section 4, this formulation can be extended to a more general framework containing also idiosyncratic regressors, i.e., $y_{it} = \alpha_i + \beta' F_t + \gamma' x_{it} + u_{it}$, and with heterogeneous slopes, i.e., $y_{it} = \alpha_i + \beta'_i F_t + u_{it}$.

The results derived in this article are asymptotic, and therefore it would be important to assess the finite sample behavior of the estimates via Monte Carlo exercises. Another important extension is to relax the exogeneity hypothesis in Assumption 6(a). In this case, fully modified OLS (Phillips and Hansen, 1990) and/or instrumental variable estimators may be employed. These interesting issues are beyond the scope of the present paper, and we leave them for future studies.
Appendix A: Central Limit Theorem for Multi-index Martingale Difference Sequences

In this Appendix, we provide a joint CLT for MDS, based upon the theory in Hall and Heyde (1980). This is the building block employed to prove the joint limits in Theorems 1 to 4.

**Theorem 9** Consider the sequence of $k$-dimensional random variables $\{\xi_{iT}\}_{i=1}^n$. Let $C$ be an invariant $\sigma$-field such that, conditionally on $C$, (i) the $\xi_{iT}$ are independent across $i$; (ii) $E[\xi_{iT}|C] = 0$ for all $i$; (iii) for some $\delta > 0$ it holds that

$$E\|\xi_{iT}|C\|^{2+\delta} < \infty$$

as $T \to \infty$. Then, as $(n, T) \to \infty$, it holds that conditional on $C$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_{iT} \Rightarrow \mathcal{V}^{1/2} \times \mathcal{Z}$$

where $\mathcal{Z} \sim N(0, I_k)$ is independent of $\mathcal{V}$ a conditional variance and defined as

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left( \xi_{iT} \xi_{i'T} | C \right) \overset{p}{=} \mathcal{V}.$$  

**Proof.** In order to prove this theorem, we show that all the conditions required in Corollary 3.1 in Hall and Heyde (1980, p. 58) are satisfied, and thus this corollary can be applied here to prove (54). Consider the $\sigma$-field defined as $I_{n,i} = \{\xi_{1T}, \ldots, \xi_{iT}\} \cup C$; as $n$ expands, the $\sigma$-fields $I_{n,i}$ are nested since $I_{n,i} = I_i$ for any $i \leq n$, and therefore condition (3.21) in Hall and Heyde (1980, p. 58) holds. Henceforth, we therefore use the simpler notation $I_i$, thus suppressing the dependence on $n$. According to Assumptions (i) and (ii), the $\xi_{iT}$ are independent across $i$ conditional on $C$; this entails that $E[\xi_{iT}|I_{i-1}] = E[\xi_{iT}|C] = 0$, where the last equality holds by assumption. Thus, the $\xi_{iT}$ are (conditional on $C$) a zero mean martingale difference array. Equation (53) is essentially a conditional Liapunov condition, which requires conditional integrability of order $2 + \delta$. This also implies that $E\|\xi_{iT}|I_{i-1}\|^2 = E\|\xi_{iT}|C\|^2 < \infty$, and therefore
the conditional variance of $\xi_{iT}$ is finite. Also, the conditional Liapunov condition (53) is sufficient for a conditional Lindeberg condition to hold, whereby as $n \to \infty$ for some $\varepsilon > 0$ it holds that

$$\sum_{i=1}^{n} \left[ \xi_{iT}^2 d \left( \frac{|\xi_{iT}|}{\varepsilon} \right) \right] \overset{P}{\to} 0$$

where $d(\cdot)$ is the indicator function. Thus, all the assumptions required for Corollary 3.1 in Hall and Heyde (1980, p. 58) are satisfied, and a (cross-sectional) CLT holds for the sequence $\sum_{i=1}^{n} \xi_{iT}$, and as $n \to \infty$, the sequence $n^{-1/2} \sum_{i=1}^{n} \xi_{iT}$ converges to a normal random variable, whose asymptotic variance is given in (55).

Remarks

R1 Theorem 9 is a joint CLT for MDS sequences, of a similar type to those studied in Hall and Heyde (1980). The only difference is that the random sequence $\sum_{i=1}^{n} \xi_{iT}$ depends on two indexes, $n$ and $T$, both allowed to pass to infinity, which makes the result applicable in a panel data framework.

From the technical viewpoint, Theorem 9 lays out some sufficient conditions whereby Corollary 3.1 in Hall and Heyde (1980, p. 58) holds. A key role is played by the conditional Liapunov condition (53), which states that the sequence $\xi_{iT}$ is (conditionally upon $C$) integrable of order $2 + \delta$ as $T$ passes to infinity. This ensures that the MDS is square integrable, and that a Lindeberg condition holds for the $\xi_{iT}$ as $T \to \infty$; see also Phillips and Moon (1999, p. 1071) for the case of i.i.d. panel models.

R2 Theorem 9 can be applied to the panel models when $(n, T) \to \infty$ and in presence of e.g., strong cross-sectional dependence arising from the presence of a common factor structure. Cross sectional independence among $\xi_{iT}$ is not required, unlike in Phillips and Moon (1999), as long as the $\xi_{iT}$s are cross sectionally independent conditionally on some invariant $\sigma$-field $C$. When this is the case, the joint asymptotic theory follows from the MDS CLT.
A similar approach was employed to derive asymptotic results in cross-sectional regressions with common shocks by Andrews (2005). To the best of our knowledge, this is the first time that the MDS CLT is applied to study the asymptotics of multi-index sequences. Although the MDS CLT is essentially a cross-sectional result, the role played by $T$ in Theorem 9 is still quite evident, e.g., from (53).

Theorem 9 suggests a methodology to derive the joint asymptotics for panels with possibly strong cross-sectional dependence as $(n, T) \to \infty$, with no need to make appeal to sequential limit. To illustrate this, consider the double sum $\sum_i \sum_t \lambda_i f_t x_{it}$, where e.g., $x_{it}$ is i.i.d. across $i$, $\lambda_i$ is non random and $f_t$ is a variable common to all $i$ - this term could arise when studying the estimation error in a panel regression where the error term has a factor structure such as $y_{it} = \alpha + \beta x_{it} + u_{it}$ with $u_{it} = \lambda_i f_t$. The asymptotic theory derived in Phillips and Moon (1999) cannot be applied to $\sum_i \sum_t \lambda_i f_t x_{it}$, since the sequence $\lambda_i f_t x_{it}$ is not independent across $i$. The limiting distribution of $\sum_i \sum_t \lambda_i f_t x_{it}$ can be studied by applying Theorem 9 to the sequence $\xi_{iT} = s_T \sum_t f_t (\lambda_i x_{it})$, where $s_T$ is a suitable normalization, and by considering the $\sigma$-field generated by the $f_t$, say $C_f$. If $E [\xi_{iT} | C_f] = 0$ for $T \to \infty$ and if it can be proven that as $T \to \infty$ (53) holds, then Theorem 9 ensures that $(n^{-1/2} s_T) \sum_i \sum_t \lambda_i f_t x_{it}$ converges to a normal random variable with mean zero and variance $V$.

Equation (55) suggests a method to calculate $V$. Note that $V$ can be a constant or a random variable, depending on the assumptions on $\xi_{iT}$, and thus the limiting distribution of $n^{-1/2} \sum_i \xi_{iT}$ can be mixed normal, as already suggested by Andrews (2005) in the cross-sectional case.
Appendix B: Proofs

Henceforth, we define $C_{nT} = \min \{ \sqrt{n}, T \}$ and $\delta_{nT} = \min \{ \sqrt{n}, \sqrt{T} \}$. Also, in order to keep the notation simple, in this and in the other appendices, we set the rotation matrices $H_1$ and $H_2$ to be identity matrices of dimensions $k$ and $n$ respectively.

**Proof of Theorem 1.** Consider the estimation error $\hat{\beta} - \beta = |\sum_i \sum_t W_t W'_t|^{-1} |\sum_i \sum_t W_t u_{it}|$ as defined in (9).

Let us start with the denominator $\sum_i \sum_t W_t W'_t$. When $T \to \infty$ and $n$ is fixed, it holds that $\sum_i \sum_t W_t W'_t = O_p (T^2)$ irrespective of whether (1) is a spurious or a cointegrating regression from Assumptions 2 and 3, and

$$\frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{T} W_t W'_t = \int B_t B'_t. \quad (56)$$

As $n \to \infty$, and for fixed $T$, we have $\sum_i \sum_t W_t W'_t = O_p (n)$

$$\frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{T} W_t W'_t = \frac{1}{T^2} \sum_{t=1}^{T} W_t W'_t \quad (57)$$

whilst as both $n$ and $T$ are large we have $\sum_i \sum_t W_t W'_t = O_p (nT^2)$

$$\frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{T} W_t W'_t = \int B_t B'_t. \quad (58)$$

As far as the numerator is concerned, we derive the asymptotics with respect to three separate cases, following the same structure as in the theorem. We firstly derive the rate of convergence and the limiting distribution of $\sum_i \sum_t W_t u_{it}$ for the case when $T$ is large and $n$ is fixed; we then study the opposite case, when $T$ is fixed and $n$ is large; last, we analyze the case when both $T$ and $n$ are large.

**Case 1: large $T$ and fixed $n$.**

We firstly focus our attention to the case where equation (1) is a cointegration relationship. Denote

$$\xi_{it} = T^{-1} W_t \left( \sum_{i=1}^{n} u_{it} \right)$$

37
and

\[ \xi_{nT} = \sum_{t=1}^{T} \xi_{nt}. \]

Assumption 6 ensures that \( F_t \) and the \( u_{it} \)'s are independent. Also, according to Assumption 1(a), the process \( \sum_{i} u_{it} \) has covariance structure given by

\[ E \left[ \left( \sum_{i=1}^{n} u_{it} \right) \left( \sum_{i=1}^{n} u_{is} \right) \right] = \sum_{i=1}^{n} \sum_{j=1}^{n} \tau_{ij,ts}. \]

Then the absolutely summability condition on \( \tau_{ij,ts} \) over time implied in Assumption 1(b), and Assumptions 2 and 3 ensure that a functional central limit theorem (FCLT) holds such that

\[ \xi_{nT} \Rightarrow \int \bar{B}_{x} dW \]

where \( W \) is a Brownian motion with variance

\[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} \tau_{ij,ts} = \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}. \]

An alternative way to write the limiting distribution of \( \xi_{nT} \) is

\[ \xi_{nT} \Rightarrow \left( \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij} \right)^{1/2} \left( \int \bar{B}_{x} \bar{B}'_{x} \right)^{1/2} \times Z \]

where \( Z \sim N(0, I_{k}) \).

This entails that the rate of convergence of the numerator of \( \hat{\beta} - \beta \) is \( O_{p}(T) \); therefore, given equation (56) that ensures that the denominator of \( \hat{\beta} - \beta \) is \( O_{p}(T^{-2}) \), we have that \( \hat{\beta} - \beta = O_{p}(T^{-1}) \). As far as the distribution limit is concerned, we know, combining the asymptotics of \( \xi_{nT} \) with equation (56), we have that

\[ \left[ \frac{1}{T^{2}} \sum_{i=1}^{n} \sum_{t=1}^{T} W_{i}W'_{t} \right]^{-1} \left[ \frac{1}{T} \sum_{i=1}^{n} \sum_{t=1}^{T} W_{i}u_{it} \right] \Rightarrow \frac{1}{n} \left( \int \bar{B}_{x} \bar{B}'_{x} \right)^{-1/2} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij} \right)^{1/2} \times Z \]

which proves equation (11). Independence between \( Z \) and \( \bar{B}_{x} \) is ensured by Assumption 6.
We now consider the case when equation (1) is a spurious regression, i.e., \( u_{it} \sim I(1) \).

Define \( \xi_{nt}^S = T^{-2}W_t(\sum_{i=1}^n u_{it}) \) and \( \xi_{nT}^S = \sum_{t=1}^T \xi_{nt}^S \). The process \( \sum_{i=1}^n u_{it} \) is still a unit root process with long run variance given by \( \sum_{i=1}^n \sum_{j=1}^n h_{ij}^2 \).

Therefore, a FCLT, which follows from Assumptions 1(a), 2 and 3, ensures that \( \xi_{nT}^S = O_p(1) \). Together with (56), this proves that \( \hat{\beta} - \beta = O_p(1) \). As far as the limiting distribution is concerned, the asymptotics of the numerator of \( \hat{\beta} - \beta \) is given by

\[
\xi_{nT}^S = \frac{1}{T^2} \sum_{t=1}^T W_t \left( \sum_{i=1}^n u_{it} \right) \Rightarrow \left( \int \bar{B}_t B_u \right) \left( \sum_{i=1}^n \sum_{j=1}^n h_{ij}^2 \right)^{1/2}.
\]

Combining this with the asymptotics of the denominator given in (56), we get equation (12).

**Case 2: large \( n \) and fixed \( T \).**

Consider first the cointegration case. Define \( \tilde{\xi}_{nt} = W_t \left( n^{-1/2} \sum_{i=1}^n u_{it} \right) \) and \( \tilde{\xi}_{nT} = \sum_{t=1}^T W_t \left( n^{-1/2} \sum_{i=1}^n u_{it} \right) \).

Assumption 1(a) ensures that a CLT holds for \( n^{-1/2} \sum_{i=1}^n u_{it} \), so that as \( n \to \infty \) we have that, for every \( t \), \( n^{-1/2} \sum_{i=1}^n u_{it} \Rightarrow \tilde{u}_t \), where \( \tilde{u}_t \) is a normally distributed, zero mean random variable with, after Assumption 1(b)

\[
E[\tilde{u}_t \tilde{u}_s] = \tilde{\tau}_{ts}.
\]

Therefore, the quantities \( W_t \tilde{u}_t \) are mixed normals random variables (due to the randomness of \( W_t \)) and

\[
\tilde{\xi}_{nT} \sim N \left[ 0, \sum_{i=1}^T \sum_{s=1}^T W_t W_t' \tilde{\tau}_{ts} \right] = \left( \sum_{i=1}^T \sum_{s=1}^T W_t W_t' \tilde{\tau}_{ts} \right)^{1/2} \times Z
\]

where \( Z \sim N(0, I_k) \); Assumption 6 ensures independence between \( Z \) and the random variable \( \sum_{t=1}^T \sum_{s=1}^T W_t W_t' \tilde{\tau}_{ts} \).

Therefore, the rate of convergence of the numerator of \( \hat{\beta} - \beta \) is \( O_p(\sqrt{n}) \). Combining this with the rate of convergence of the denominator, given by (57),
we have that $\hat{\beta} - \beta = O_p\left(n^{-1/2}\right)$. As far as the distribution limit is concerned, combining the asymptotic law of $\hat{\xi}_{nT}$ with (57), we obtain (13).

Under the spurious regression case, define $\tilde{\xi}_{nT} = W_t \left(n^{-1/2} \sum_{t=1}^{n} u_{it}\right)$ and $\tilde{\xi}_{nT} = \sum_{t=1}^{T} \tilde{\xi}_{nt}$. Assumption 1(a) ensures the validity of the CLT for $n^{-1/2} \sum_{i=1}^{n} u_{it}$, so that uniformly in $t$ we have, as $n \to \infty$, $n^{-1/2} \sum_{i=1}^{n} u_{it} \Rightarrow \tilde{u}_t$. We have that $\tilde{\xi}_{nT} = O_p\left(1\right)$, and combining this with equation (57), we obtain $\hat{\beta} - \beta = O_p\left(n^{-1/2}\right)$. As far as the limiting distribution is concerned, since $\tilde{\xi}_{nT}$ is a finite sum, we have $\tilde{\xi}_{nT} \Rightarrow \sum_{t=1}^{T} W_t \tilde{u}_t$ as $n \to \infty$. Combining this with equation (57), we prove the validity of equation (14).

Case 3: large $n$ and large $T$.

The proof is largely based on Theorem 9.

Let us start with the case where equation (1) is a cointegration relationship. Define $\tilde{\xi}_{iT} = T^{-1} \sum_{t=1}^{T} W_t u_{it}$, and consider the BN decomposition for $W_t$ and $u_{it}$, given respectively by

\[
W_t \overset{a.s.}{=} W_t^* + W_0 + \tilde{w}_t,
\]
\[
u_{it} \overset{a.s.}{=} \nu_{it}^* + \tilde{\nu}_{it} + \tilde{\eta}_{it},
\]

where $W_t^* = C(1) \sum_{j=1}^{t} w_j$, $\nu_{it}^* = D_i(1) \eta_{it}$, $\tilde{w}_t = \sum_{j=0}^{\infty} \left(\sum_{i=0}^{j+1} C_i\right) w_{t-j}$ and $\tilde{\eta}_{it}$ defined similarly. Thus

\[
\tilde{\xi}_{iT} = \frac{1}{T} \sum_{t=1}^{T} W_t^* u_{it}^* + R_{iT} = \tilde{\xi}_{iT}^* + R_{iT},
\]

where, as far as the remainder $R_{iT}$ is concerned, it can be proved using similar arguments as Phillips and Moon (1999) that $R_{iT} = O_p\left(T^{-1/2}\right)$. Consider $\tilde{\xi}_{iT}^*$ and let $C$ be the $\sigma$-field generated by the $F_t$s. Then $E\left[\tilde{\xi}_{iT}^* \mid C\right] = T^{-1} \sum_{t=1}^{T} W_t^* E(u_{it}^*) = 0$ for all $i$ and $T$. Defining $I_i$ as the $\sigma$-field generated by $F_t$ and $\tilde{\xi}_{iT}^*$, it holds that $\left\{\tilde{\xi}_{iT}, I_i\right\}$ is an MDS since $E\left[\tilde{\xi}_{iT}^* \mid I_{i-1}\right] = E\left[\tilde{\xi}_{iT}^* \mid C\right] = 0$. For some constant $M_\delta$ and some $\delta > 0$, we have that, uniformly
\[ E \left\| \xi_{iT}^* \, C \right\|^{2+\delta} = E \left\| \frac{1}{T} \sum_{t=1}^{T} W_t^* u_{it}^* \, C \right\|^{2+\delta} \]
\[ \leq M_{\delta} \sum_{t=1}^{T} E \left\| \frac{1}{T} W_t^* u_{it}^* \, C \right\|^{2+\delta} \]
\[ = M_{\delta} \frac{1}{T^{2+\delta}} \sum_{t=1}^{T} \| W_t^* \|^{2+\delta} E |u_{it}^*|^{2+\delta}. \]

Since \( E |u_{it}^*|^{2+\delta} < \infty \) for all \( i \) and constant over \( t \) by Assumption 1, it holds that \( E \left\| \xi_{iT}^* \, C \right\|^{2+\delta} < \infty \) if \( T^{-(2+\delta)} \sum_{t=1}^{T} \| W_t^* \|^{2+\delta} \) is stochastically bounded as \( T \to \infty \). This is ensured by Theorem 5.3 in Park and Phillips (1999), which holds due to Assumption 2. Thus, a Liapunov condition holds whereby \( E \left\| \xi_{iT}^* \, C \right\|^{2+\delta} < \infty \) for some \( \delta > 0 \) - in light of Assumptions 1 and 2, this is ensured for up to \( \delta > 2 \). Thus the joint MDS CLT can be employed to get, for \( (n, T) \to \infty \)

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{iT}^* \Rightarrow \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \xi_{iT}^* \xi_{iT}^* \right) \right]^{1/2} \times Z \]

with \( Z \sim N\left(0, I_k\right) \) independent of \( E \left( \xi_{iT}^* \xi_{iT}^* \right) \). Thus

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{iT}^* \overset{\text{a.s.}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{iT}^* + O\left(\sqrt{\frac{n}{T}}\right) \]
\[ \Rightarrow \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \xi_{iT}^* \xi_{iT}^* \right) \right]^{1/2} \times Z, \]

under \( n/T \to 0 \). Last, \( n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \xi_{iT}^* \xi_{iT}^* \right) \) for \( (n, T) \to \infty \) is given by

\[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \xi_{iT}^* \xi_{iT}^* \right) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} W_t^* W_s^{\prime*} E \left( u_{it}^* u_{js}^* \right) \]
\[ = \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( u_{it}^* u_{js}^* \right) \right] \left[ \frac{1}{T^2} \sum_{t=1}^{T} W_t^* W_t^{\prime*} \right] \]
where the last equality holds because \( E(u_t^*u_{ts}^*) = 0 \) for all \( t \neq s \). As \((n, T) \to \infty\), the FCLT implied by Assumption 2 and the definition of \( \bar{h} \) entail \( n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \xi_{iT} \xi_{jT}^* | C \right) \Rightarrow \bar{h} \left( \int \tilde{B}_2 B_2^* \right) \), and therefore as \((n, T) \to \infty\) with \( n/T \to 0 \) we have

\[
\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} W_t u_{it} \Rightarrow \sqrt{\bar{h}} \left( \int \tilde{B}_2 B_2^* \right)^{1/2} Z.
\]

Note that in this proof the restriction whereby \( n/T \to 0 \) arises for the same reason as in Phillips and Moon (1999), i.e., from the fact that (either or both) \( W_t \) and \( u_{it} \) could be time dependent and the initial conditions \( W_0 \) and \( u_{i0} \) need not be zero. Combining this with equation (58), we get that \( \hat{\beta} - \beta = O_p \left( n^{-1/2}T^{-1} \right) \).

As far as the limiting distribution is concerned, combining the asymptotic law of \( \bar{s}_{nT} \) with equation (58), we have:

\[
\left[ \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t=1}^{T} W_t W_t^* \right]^{-1} \left[ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} W_t u_{it} \right] \Rightarrow \sqrt{\bar{h}} \left( \int \tilde{B}_2 B_2^* \right)^{-1/2} Z
\]

which corresponds to equation (15). When \( T/n \to 0 \), the term that dominates in (60) is the second one, and thus the order of magnitude of the numerator of \( \hat{\beta} - \beta \) is \( O_p \left( n \sqrt{T} \right) \). To prove (16), define

\[
\frac{\sqrt{T}}{n} \sum_{i=1}^{n} R_{iT} \overset{c}{\to} \Delta_1
\]

where \(^c\) denotes here convergence in some sense (e.g., in distribution or in probability) and \( R_{iT} \) is defined in (59). When normalizing \( \hat{\beta} - \beta \) by \( T^{3/2} \), the asymptotic law of the numerator is therefore given by the quantity \( \Delta_1 \) defined in (61). Combining this with equation (58), (16) follows immediately.

We now turn to the case when equation (1) is a spurious regression. Let \( \xi_{iT} = T^{-2} \sum_{t=1}^{T} W_t u_{it} \), and consider the BN decomposition of \( W_t \) and \( u_{it} \), whereby

\[
\begin{align*}
W_t &\overset{a.s.}{=} W_t^* + W_0 + \tilde{w}_t - \tilde{w}_t, \\
u_{it} &\overset{a.s.}{=} u_{it}^* + u_{i0} + \tilde{\eta}_{i0} - \tilde{\eta}_{i0},
\end{align*}
\]

with \( u_{it}^* = F_i(1) \sum_{j=1}^{t} \eta_{ij} \) and the other variables defined accordingly. It holds
that
\[
S_i^T = 1 \sum_{t=1}^{T} W_i^* u_{it} + R_i = S_i^T + R_i^T;
\]
(62)
again, it can be shown that \(R_i^T = O_p(T^{-1/2}).\) Conditioning on \(C\), it holds
that \(E \left[ S_i^T | C \right] = T^{-2} \sum_{t=1}^{T} W_i^* E(u_{it}) = 0\) for all \(i\) and \(T\). Defining \(I_i\) as the
\(\sigma\)-field generated by \(F_t\) and \((\tilde{\xi}_i^T, ..., \tilde{\xi}_i^T)\), it therefore holds that \(\{\tilde{\xi}_i^T, I_i\}\) is
an MDS since \(E \left[ \tilde{\xi}_i^T | I_i-1 \right] = E \left[ \tilde{\xi}_i^T | C \right] = 0\). A Liapunov condition can be
proved noting that for all \(i\) and some \(\delta > 0\)
\[
E \left[ \tilde{\xi}_i^T | C \right]^{2+\delta} \leq E \left[ \left( \frac{1}{T^2} \sum_{t=1}^{T} \|W_i^*\|^2 \right)^{1/2} \left( \frac{1}{T^2} \sum_{t=1}^{T} |u_{it}|^2 \right)^{1/2} \right]^{2+\delta}
\]
\[
= \left( \frac{1}{T^2} \sum_{t=1}^{T} \|W_i^*\|^2 \right)^{1+\delta/2} E \left[ \left( \frac{1}{T^2} \sum_{t=1}^{T} |u_{it}|^2 \right)^{1+\delta/2} \right].
\]
The FCLT and the Continuous Mapping Theorem (CMT) ensure that as \(T \to \infty\), both quantities are stochastically bounded. Thus, \(E \left[ \tilde{\xi}_i^T | C \right]^{2+\delta}\) is bounded
for all \(i\) and the (joint) MDS CLT can be employed to get, for \(n, T \to \infty\)
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\xi}_i^T \Rightarrow \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \tilde{\xi}_i^T \tilde{\xi}_j^T | C \right) \right]^{1/2} \times Z
\]
with \(Z \sim N(0, I)\) independent of \(E \left( \tilde{\xi}_i^T \tilde{\xi}_j^T | C \right)\). Thus
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\xi}_i^T \overset{a.s.}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\xi}_i^T + O \left( \frac{\sqrt{n}}{T} \right)
\]
(63)
\[
\Rightarrow \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \tilde{\xi}_i^T \tilde{\xi}_j^T | C \right) \right]^{1/2} \times Z
\]
under \(n/T \to 0\). Note that
\[
\frac{1}{n} \sum_{i=1}^{n} E \left( \tilde{\xi}_i^T \tilde{\xi}_j^T | C \right) = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} W_i^* W_s^* \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E(u_{it}^* u_{js}^*) \right]
\]
and an alternative representation for the random variable 
\[ n^{-1} \sum_{i=1}^{n} E \left( \xi_{iT}^{S_{iT}} \right) \right)^{1/2} \times Z 
\]

where \( \equiv D \) means equality in distribution. Thus, 
\[ n^{-1/2} \sum_{i=1}^{n} \xi_{iT}^{S_{iT}} \Rightarrow \left( \int \bar{B}_{z} \nu_{n} \right) \left( \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}^{A} \right)^{1/2} \]

This result, together with equation (58), proves that \( \bar{\beta} - \beta = O_{p} \left( n^{-1/2} \right) \). As far as the limiting distribution is concerned, combining this result with the one reported in equation (58), we get equation (17). When \( T/n \to 0 \), the term that dominates in (63) is the second one, and the order of magnitude of the numerator of \( \bar{\beta} - \beta \) is \( O_{p} \left( nT^{3/2} \right) \). To prove (18), define

\[ \frac{\sqrt{T}}{n} \sum_{i=1}^{n} R_{iT}^{S_{iT}} \Rightarrow \Delta_{2} \]

where again \( \Rightarrow \) denotes convergence in some sense, and \( R_{iT}^{S} \) is defined in (62). When normalizing \( \bar{\beta} - \beta \) by \( \sqrt{T} \), the asymptotic law of the numerator is therefore given by the quantity \( \Delta_{2} \) defined in (64). Combining this with equation (58), (18) follows immediately. \( \blacksquare \)

**Proof of Theorem 3.** According to equation (25)

\[ \hat{\beta} - \beta = \left( \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{i} \hat{W}_{i}' \right)^{-1} \left( \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{i} \left( W_{i} - \hat{W}_{i} \right)' \beta + u_{it} \right) \]

Let us first consider the denominator of this expression. Assumption 3 and Lemma 2.1 imply that

\[ \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_{i} \hat{W}_{i}' = O_{p} \left( nT^{2} \right) \]  

(65)

and

\[ (nT^{2})^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} W_{i} \hat{W}_{i}' \Rightarrow \int \bar{B}_{z} \hat{B}_{z}' \; ; \]  

(66)

this holds under both the cases of cointegration and spurious regression.
We now prove Theorem 3 for the case when equation (1) is a cointegration relationship. Following Theorem 9, define

\[
\hat{\xi}_{iT} = \frac{1}{T} \sum_{t=1}^{T} \hat{W}_t \left( (W_t - \hat{W}_t)' \beta + u_{it} \right)
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \hat{W}_t \left( (W_t - \hat{W}_t)' \beta + u_{it} \right) + \frac{1}{T} \sum_{t=1}^{T} (\hat{W}_t - W_t) (W_t - \hat{W}_t)' \beta
\]

\[
+ \frac{1}{T} \sum_{t=1}^{T} (\hat{W}_t - W_t) u_{it}
\]

\[
= \hat{\xi}_{iT} + \xi_{iT}^a + \xi_{iT}^b.
\]

Consider first \(\hat{\xi}_{iT}\); applying the BN decomposition to \(W_t\) and \(u_{it}\) with \(W_t = W_t^* + R_{WT}\) and \(u_{it} = u_{it}^* + R_{uiT}\), we have

\[
\hat{\xi}_{iT} = \frac{1}{T} \sum_{t=1}^{T} W_t^* \left( (W_t - \hat{W}_t)' \beta + u_{it}^* \right) + R_{iT}
\]

\[
= \hat{\xi}_{iT}^* + R_{iT} \tag{67}
\]

where following similar arguments as in Phillips and Solo (1992) it follows that \(R_{iT} = O_p \left( T^{-1/2} \right)\). Conditioning on the \(\sigma\)-field \(C\) generated by the \(F_t\), it holds that

\[
E \left[ \hat{\xi}_{iT}^* | C \right] = \frac{1}{T} \sum_{t=1}^{T} W_t^* E( u_{it}^* ) + \frac{1}{T} \sum_{t=1}^{T} W_t^* E \left[ (W_t - \hat{W}_t | C)' \beta \right] = O_p \left( \frac{1}{\sqrt{T}} \right)
\]

because \(E( u_{it} ) = 0\) and

\[
-E \left[ (W_t - \hat{W}_t | C \right] = \frac{1}{T^2} \sum_{s=1}^{T} \hat{W}_s \gamma_{s-t} \leq \frac{1}{T^2} \max_{1 \leq s \leq T} |\hat{W}_t| \left( \sum_{s=1}^{T} | \gamma_{s-t} | \right)
\]

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

Thus, letting \(\hat{I}_i\) the union between the \(\sigma\)-field generated by \(\{\hat{\xi}_{iT}, ..., \hat{\xi}_{nT}\}\) and
\(C, \{\tilde{\xi}_iT, \tilde{r}_i\}\) is an MDS as \(T \to \infty\). Also for some constant \(M_\delta < \infty\)

\[
E \left\| \xi_iT \right\|_C^{2+\delta} = E \left\| \frac{1}{T} \sum_{t=1}^{T} W_t^* \left[ (W_t - \hat{W}_i)^' \beta + u_{it}^* \right] \right\|_C^{2+\delta} \leq M_\delta \frac{1}{T^{2+\delta}} \sum_{t=1}^{T} \|W_t^*\|^{2+\delta} E \left\| (W_t - \hat{W}_i)^' \beta + u_{it}^* \right\|_C^{2+\delta} \leq M_\delta \frac{\|\beta\|}{T^{2+\delta}} \sum_{t=1}^{T} \|W_t^*\|^{2+\delta} E \left\| (W_t - \hat{W}_i) \right\|_C^{2+\delta} + M_\delta \frac{1}{T^{2+\delta}} \sum_{t=1}^{T} \|W_t^*\|^{2+\delta} E \|u_{it}^*\|^{2+\delta} = I + II.
\]

Considering \(II\), given that \(E |u_{it}^*|^{2+\delta} < \infty\) for all \(i\) and for some \(\delta > 0\), and since \(\sum_{t=1}^{T} \|W_t^*\|^{2+\delta} = O_p \left(T^{2+\delta}\right)\), it follows that \(II = O_p \left(1\right)\). As far as \(I\) is concerned, note that for some \(M'_0 < \infty\)

\[
T^{2(2+\delta)} E \left\| \left(W_t - \hat{W}_i\right) \right\|_C^{2+\delta} \leq M'_0 \sum_{s=1}^{T} \|\hat{F}_s\|^{2+\delta} E \left\| \frac{e'_s e_s}{n} \right\|^{2+\delta} + M'_0 \sum_{s=1}^{T} \|\hat{F}_s\|^{2+\delta} E \|\eta_{st}\|_C^{2+\delta} + M'_0 \sum_{s=1}^{T} \|\hat{F}_s\|^{2+\delta} E \|\xi_{st}\|_C^{2+\delta}
\]

and Assumption 2 ensures that \(E |\frac{e'_s e_s}{n}|^{2+\delta} < \infty\), \(E |\eta_{st}|_C^{2+\delta} < \infty\) and \(E \|\xi_{st}\|_C^{2+\delta} < \infty\). Since \(\sum_{s=1}^{T} \|\hat{F}_s\|^{2+\delta} = O_p \left(T^{2+\delta}\right)\), as \(T \to \infty\), \(II = o_p \left(1\right)\). Thus, \(E \left\| \tilde{\xi}_iT \right\|_C^{2+\delta} < \infty\) as \(T \to \infty\) and therefore an MDS-CLT can be applied to \(\tilde{\xi}_iT\). As far as \(\xi_iT^a\) and \(\xi_iT^b\) are concerned, note that

\[
\frac{1}{n} \sum_{i=1}^{n} \xi_iT^a = \frac{\sqrt{n}}{T} \sum_{t=1}^{T} \left(\hat{W}_t - W_t\right) \left(W_t - \hat{W}_t\right)^' \beta = O_p \left(\frac{\sqrt{n}}{C_{nT}^{\frac{1}{2}}}\right)
\]

according to Lemma 1.2(b); Lemma 2.2 ensures that \(n^{-1/2} \sum_{i=1}^{n} \xi_iT^b = O_p \left(C_{nT}^{-1}\right)\).
Thus, as \( (n, T) \to \infty \) with \( n/T \to 0 \)
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{iT} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{iT}^a + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{iT}^b + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{iT}^c \\
\Rightarrow \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \xi_{iT}^{a}, \xi_{jT}^{a} \right| C \right) \right] ^{1/2} \times Z
\]
with \( Z \sim N(0, I_k) \) is independent of \( E \left( \xi_{iT}^{a}, \xi_{jT}^{a} \right| C \) \). It holds that
\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( \xi_{iT}^{a}, \xi_{jT}^{a} \right| C \right) = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} W_t W_s \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( u_{it} u_{js} \right) \right] \\
+ \frac{2}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} W_t W_s E \left[ \left( W_t - \bar{W}_t \right)' C \right] \left[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left( u_{js}^a \right) \right] \\
+ \frac{n}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} W_t W_s \beta E \left[ \left( \bar{W}_t - W_t \right) \left( \bar{W}_s - \bar{W}_s \right)' C \right] \beta
\]
\[= I + II + III. \]

Note that, as in the proof of Theorem 1, \( I \Rightarrow \sqrt{n} \int \bar{B}_t \bar{B}_t'. \) Also, since \( E \left( u_{js}^a \right) = 0 \) for all \( j \) and \( s \), \( II = 0 \). As far as \( III \) is concerned, this is given by the variance of
\[
\frac{\sqrt{n}}{T} \sum_{t=1}^{T} W_t \left( \bar{W}_t - W_t \right)' \beta.
\]
Under \( n/T \to 0 \), Lemma 1.4 holds and
\[
\sqrt{n} \left( \bar{W}_t - W_t \right) = \frac{1}{T} \sum_{s=1}^{T} \bar{W}_s W_s' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda_i e_{it} + O_p \left( \frac{1}{\sqrt{nT}} \right).
\]
Thus,
\[
\frac{\sqrt{n}}{T} \sum_{t=1}^{T} W_t \left( \bar{W}_t - W_t \right)' \beta = \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \sum_{i=1}^{n} W_t e_{it} \lambda_i' \left( \frac{1}{T^2} \sum_{s=1}^{T} \bar{W}_s W_s' \right)' \beta + o_p \left( 1 \right).
\]
As far as \( (\sqrt{nT})^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n} W_t e_{it} \lambda_i' \) is concerned, similar arguments as in the proof of Theorem 1 lead, as \( (n, T) \to \infty \) under \( n/T \to 0 \), to
\[
\frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \sum_{i=1}^{n} W_t e_{it} \lambda_i' = \left[ \left( \int \bar{B}_t \bar{B}_t' \right) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j' E (e_{it} e_{jt}) \right]^{1/2} \times Z
\]
47
and \( \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j E(e_{it} e_{jt}) = \Gamma \) by definition. Since by definition \( T^{-2} \sum_{s=1}^{T} \hat{W}_s W'_s \Rightarrow \hat{Q}_B \), we have
\[
\frac{\sqrt{n}}{T} \sum_{t=1}^{T} W_t (\hat{W}_t - W_t)' \beta \Rightarrow \left( \int \tilde{B}_z \tilde{B}_z' \right)^{1/2} \left( \beta' \hat{Q}_B \Gamma \hat{Q}_B' \beta \right)^{1/2} \times Z.
\]
Therefore
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{iT} \Rightarrow \left( \int \tilde{B}_z \tilde{B}_z' \right)^{1/2} \sqrt{h + \beta' \hat{Q}_B \Gamma \hat{Q}_B' \beta} \times Z;
\]
combining this with (66), we obtain (27). Note that when \( T/n \to \infty \), the term that dominates in (68) is the second one, of magnitude \( O_p \left( \frac{1}{\sqrt{T}} \right) \). Thus, the numerator of \( \hat{\beta} - \beta \) is of order \( O_p \left( \frac{1}{\sqrt{T}} \right) \), and since the denominator is \( O_p \left( nT^2 \right) \) one has \( \hat{\beta} - \beta = O_p \left( T^{-3/2} \right) \). When normalizing \( \hat{\beta} - \beta \) by \( T^{3/2} \), the asymptotic law of the numerator is given by the quantity \( \Delta_1 \) defined in (61); (28) follows.

We now turn to the case where (1) is a spurious regression. Considering the numerator of \( \hat{\beta} - \beta \), we have
\[
\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{W}_t \left[ (W_t - \hat{W}_t)' \beta + u_{it} \right] = n \sum_{t=1}^{T} \hat{W}_t (W_t - \hat{W}_t)' \beta + \sum_{i=1}^{n} \sum_{t=1}^{T} W_t u_{it} + \sum_{i=1}^{n} \sum_{t=1}^{T} (\hat{W}_t - W_t) u_{it} = I + II + III.
\]
Lemma 1.3.(c) ensures that \( I = O_p \left( nTC_{nT}^{-1} \right) \). Equation (62) in the proof of Theorem 1 states that
\[
II = \sum_{i=1}^{n} \sum_{t=1}^{T} W_{it} u_{it}^* + T^2 \sum_{i=1}^{n} R_{iT}^S = O_p \left( \sqrt{nT^2} \right) + O_p \left( nT^{3/2} \right) = II_a + II_b.
\]
Last, \( III = O_p \left( \sqrt{nT^{3/2}C_{nT}^{-1}} \right) \) after Lemma 2.4. Thus, as \( (n, T) \to \infty \) with \( n/T \to 0 \), the asymptotics of the numerator of \( \hat{\beta} - \beta \) is driven by \( II_a \), and
therefore

\[ \sqrt{n} \left( \hat{\beta} - \beta \right) \xrightarrow{a.s.} \left[ \sum_{t=1}^{T} W_t W_t' \right]^{-1} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{T} W_t u_{it} \right] + o_p(1) \]

\[ \Rightarrow \sqrt{n} \left( \hat{B}_z B_z' \right)^{-1} \left( \int \hat{B}_z B_z \right) . \]

As far as the case whereby \((n, T) \to \infty\) with \(T/n \to 0\), note that the term that dominates the asymptotic law of the numerator of \(\hat{\beta} - \beta\) is now \(II_b\); thus, as \((n, T) \to \infty\) with \(T/n \to 0\),

\[ \sum_{i=1}^{n} \sum_{t=1}^{T} W_t \left( W_t - \hat{W}_t \right)^t \beta + u_{it} = O_p \left(nT^{3/2}\right), \]

and therefore \(\hat{\beta} - \beta = \left[O_p \left(nT^2\right)\right]^{-1} O_p \left(nT^{3/2}\right) = O_p \left(T^{-1/2}\right)\). Equation (30) follows from the definition of (64). □

**Proof of Proposition 1.** Let \(\tilde{F}_t\) be the principal component estimator for \(F_t\) as defined in Bai (2004). Then we know (see e.g. the proof of Lemma 3 in Bai, 2004) that \(T \left( \hat{\Lambda} - \Lambda \right)\) can be decomposed as

\[ T \left( \hat{\Lambda} - \Lambda \right) = \frac{1}{T} \left[ \sum_{t=1}^{T} e_t F_t' + \sum_{t=1}^{T} e_t \left( \tilde{F}_t - F_t \right)' + \Lambda \sum_{t=1}^{T} (F_t - \tilde{F}_t) \tilde{F}_t' \right] \left[ \frac{1}{T^2} \sum_{t=1}^{T} \tilde{F}_t \tilde{F}_t' \right]^{-1}. \]

As far as the denominator of this expression is concerned, let \(\Xi = \int B_z B_z'\). We have

\[ \sum_{t=1}^{T} \tilde{F}_t \tilde{F}_t' = \sum_{t=1}^{T} F_t F_t' + \sum_{t=1}^{T} (\tilde{F}_t - F_t) \tilde{F}_t' + \sum_{t=1}^{T} \tilde{F}_t (\tilde{F}_t - F_t)' + \sum_{t=1}^{T} (\tilde{F}_t - F_t) (\tilde{F}_t - F_t)' \]

where

\[ \sum_{t=1}^{T} F_t F_t' = O_p \left(T^2\right), \]

\[ \sum_{t=1}^{T} (\tilde{F}_t - F_t) \tilde{F}_t' = O_p \left(T\right), \]

and

\[ \sum_{t=1}^{T} (\tilde{F}_t - F_t) (\tilde{F}_t - F_t)' = O_p \left(T\right); \]

the last two equalities come directly from Lemma B.4(ii) and Lemma B.1 in Bai (2004). Therefore

\[ T^{-2} \sum_{t=1}^{T} \tilde{F}_t \tilde{F}_t' = T^{-2} \sum_{t=1}^{T} F_t F_t' + O_p \left(T^{-1}\right) \]
and
\[ T^{-2} \sum_{t=1}^{T} \tilde{F}_t \tilde{F}'_t \Rightarrow \xi. \]

As far as the numerator of equation (69) is concerned, we study each term. First of all we know that \( T^{-1} \sum_{t=1}^{T} e_t F'_t \Rightarrow \int dW_t B_x \). The limiting distribution of \( \sum_{t=1}^{T} e_t (\tilde{F}_t - F_t)' \) can be obtained from the following decomposition - see Bai (2004, p. 164) for details:

\[
\tilde{F}_t - F_t = T^{-2} \sum_{s=1}^{T} \tilde{F}'_s \gamma_n (s, t) + \sum_{s=1}^{T} \tilde{F}_s \zeta st + T^{-2} \sum_{s=1}^{T} \tilde{F}_s \eta st + T^{-2} \sum_{s=1}^{T} \tilde{F}_s \xi st
\]

where (as in Lemma 1) we let \( \gamma_n (s, t) = E (e'_t e_s / n) \), \( \zeta st = e'_t e_s / n - \gamma_n (s, t) \), \( \eta st = F'_s \Lambda' e_s / n \), \( \xi st = F'_t \Lambda' e_s / n \). Hence

\[
\frac{1}{T} \sum_{t=1}^{T} e_t (\tilde{F}_t - F_t) = T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s \gamma_n (s, t) + T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s \zeta st \\
+ T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s \eta st + T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s \xi st
\]

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

and

\[ I = O_p \left( T^{-1} \right) \text{ since } E \left\| e_t F'_s \gamma_n (s, t) \right\| \leq |\gamma_n (s, t)| \left( \max_{s,t} E \| e_t F'_s \| \right) \text{ and } \max_{s,t} E \| e_t F'_s \| = O_p (T); \]

\[ II = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s e'_t e_s - T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s \gamma_n (s, t) \]

we have

\[ n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s e'_t e_s = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t e'_t e_s F'_s \\
= n^{-1} T^{-1} \left( T^{-1} \sum_{t=1}^{T} e_t e'_t \right) \left( T^{-1} \sum_{s=1}^{T} e_s F'_s \right) = O_p \left( T^{-1} \right); \]

\[ III = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s F'_s \Lambda' e_t \] with

\[ n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F'_s F'_s \Lambda' e_t = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t e'_t \Lambda F'_s F'_s \\
= n^{-1} \left( T^{-1} \sum_{t=1}^{T} e_t e'_t \right) \Lambda \left( T^{-2} \sum_{s=1}^{T} F'_s F'_s \right) = O_p (1); \]

50
\[ IV = n^{-1} T^{-3} \sum_{t=1}^{T} \sum_{s=1}^{T} e_t \tilde{F}_s' F_t' \Lambda e_s \text{ and} \]
\[ n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t \tilde{F}_s' F_t' \Lambda e_s = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} e_t F_t' \Lambda' e_s \tilde{F}_s' \]
\[ = n^{-1} T^{-1} \left( T^{-1} \sum_{t=1}^{T} e_t F_t' \right) \Lambda' \left( T^{-1} \sum_{s=1}^{T} e_s \tilde{F}_s' \right) = O_p \left( T^{-1} \right). \]

Therefore the term that dominates is \( III \) and
\[ n^{-1} \left( T^{-1} \sum_{t=1}^{T} e_t e_t' \right) \Lambda \left( T^{-2} \sum_{s=1}^{T} F_s \tilde{F}_s' \right) \Rightarrow n^{-1} \Omega_e \Lambda Q. \]

Finally, as far as the term \( \Lambda \sum_{t=1}^{T} (F_t - \tilde{F}_t) \tilde{F}_t' \) in equation (69) is concerned, we have
\[ T^{-1} \Lambda \sum_{t=1}^{T} (F_t - \tilde{F}_t) \tilde{F}_t' = -T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{F}_s \tilde{F}_t' \gamma_n (s, t) - T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{F}_s \tilde{F}_t' \zeta_{st} \]
\[ -T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{F}_s \tilde{F}_t' \eta_{st} - T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{F}_s \tilde{F}_t' \xi_{st} \]
\[ = a + b + c + d. \]

We have that the terms \( a \) and \( b \) follow from the proof of Lemma B.4 in Bai, 2004):
\[ a = O_p \left( T^{-1} \right); \]
\[ b = O_p \left( T^{-1} \right), \]

the term
\[ c = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{F}_s \tilde{F}_t' F_s \Lambda e_t, \]
with
\[ n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{F}_s \tilde{F}_t' F_s \Lambda e_t = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{F}_s \tilde{F}_t' \eta_{st} \tilde{F}_t \]
\[ = n^{-1} \left( T^{-2} \sum_{s=1}^{T} \tilde{F}_s \tilde{F}_s' \right) \Lambda' \left( T^{-1} \sum_{t=1}^{T} e_t \tilde{F}_t \right) = O_p \left( 1 \right); \]

and
\[ d = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{F}_s \tilde{F}_t' F_s \Lambda e_s, \]
with
\[ n^{-1} T^{-3} \sum_{s,t=1}^{T} \bar{F}_s \tilde{F}_t \Lambda' e_s = n^{-1} T^{-3} \sum_{s=1}^{T} \sum_{t=1}^{T} \bar{F}_s e_s' \Lambda' \bar{F}_t \tilde{F}_t' \]
\[ = n^{-1} \left( T^{-1} \sum_{s=1}^{T} \bar{F}_s e_s' \right) \Lambda' \left( T^{-1} \sum_{t=1}^{T} \tilde{F}_t \bar{F}_t' \right) = O_p(1). \]

Thus the limiting distribution of \( \Lambda \sum_{t=1}^{T} (F_t - \bar{F}_t) \tilde{F}_t' \) is determined by \( c \) and \( d \).

and we have
\[ c = n^{-1} \left( T^{-2} \sum_{s=1}^{T} F_s \tilde{F}_s' \right) \Lambda' \left( T^{-1} \sum_{t=1}^{T} e_t \tilde{F}_t' \right) \]
\[ = n^{-1} \left( T^{-2} \sum_{s=1}^{T} F_s \tilde{F}_s' \right) \Lambda' \left( T^{-1} \sum_{t=1}^{T} e_t F_t \right) \]
\[ + n^{-1} \left( T^{-2} \sum_{s=1}^{T} F_s \tilde{F}_s' \right) \Lambda' \left[ T^{-1} \sum_{t=1}^{T} e_t \left( \bar{F}_t - F_t \right)' \right] \]
\[ \Rightarrow n^{-1} Q \Lambda' \left[ \int dW_x B_x' + n^{-1} \Omega_x \Lambda Q \right] \]

and
\[ d = n^{-1} \left( T^{-1} \sum_{s=1}^{T} \bar{F}_s e_s \right) \Lambda' \left( T^{-1} \sum_{t=1}^{T} F_t \bar{F}_t' \right) \Rightarrow n^{-1} \left[ \int B_x dW_x' + n^{-1} Q \Lambda' \Omega_x \right] \Lambda Q. \]

Combining the limiting distributions of all terms \( \sum_{t=1}^{T} \bar{F}_t \tilde{F}_t', \sum_{t=1}^{T} e_t F_t', \sum_{t=1}^{T} e_t \left( \bar{F}_t - F_t \right)' \) and \( \Lambda \sum_{t=1}^{T} (F_t - \bar{F}_t) \tilde{F}_t' \) in equation (69), we obtain equation (34).
References


