Inferential theory for heterogeneity and cointegration in large panels

Lorenzo Trapani*

University of Nottingham
e-mail: lorenzo.trapani@nottingham.ac.uk

Abstract: This paper provides an estimation and testing framework to assess the presence and the extent of slope heterogeneity and cointegration when the units are a mixture of spurious and/or cointegrating regressions. We propose two moment estimators for the degree of heterogeneity (measured by the dispersion of the slope coefficients around their average) and for the fraction of spurious regressions, which are found to be consistent in the whole parameter space. Based on these estimators, two tests for the null hypotheses of slope homogeneity and for cointegration are proposed. Monte Carlo simulations show that both tests have the correct size and satisfactory power.

JEL Codes: C12, C23.

Keywords: Large Panels, Heterogeneity, Cointegration, Spurious Regression, Method of Moments.

1. Introduction

This paper develops a methodology to test for slope heterogeneity and cointegration in large panels defined by the equation

$$y_{it} = \alpha_i + \beta_i' x_{it} + u_{it}, \tag{1.1}$$

with $i = 1, ..., n$ and $t = 1, ..., T$, and both y_{it} and x_{it} non-stationary for all i.
For each i, equation (1.1) could be cointegrating or not, i.e. u_{it} could be stationary or not. Also, it is well known that, in a non-stationary panel context, spurious regression and neglected heterogeneity are in a dual relationship (see Phillips and Moon (1999)). Indeed, when model (1.1) is wrongly replaced by the pooled version

$$y_{it} = \alpha_i + \beta' x_{it} + v_{it}, \tag{1.2}$$

this introduces the term $(\beta_i - \beta)' x_{it}$ in the error term $v_{it} = u_{it} + (\beta_i - \beta)' x_{it}$, thereby making it non-stationary by construction. Thus, v_{it} can have a unit root due to two non mutually exclusive reasons: genuine lack of cointegration in some of the units, or neglected slope heterogeneity. In this paper, we build on the possible presence of a unit root in v_{it} and on the source thereof, in order to propose two tests, for slope homogeneity and cointegration in equation (1.1) respectively. Our tests are based on the dispersion of the β_is around their average (denoted as σ^2_β), and on the fraction of units where u_{it} has a unit root (denoted as λ). Specifically, in order to test for poolability, a test for $H_0 : \sigma^2_\beta = 0$ is developed that is robust to any value of λ; the null hypothesis is *slope homogeneity*, i.e. $\beta_i = \beta$ for all i. In order to test for cointegration, a test for $H_0 : \lambda = 0$ is proposed, which is valid for any σ^2_β; thus, the null hypothesis is *cointegration*. Thus, the results in this paper may also be viewed as a tool to detect the presence (and the determinants) of a unit root in v_{it} in (1.2); in addition to the two tests, we also develop two estimators for λ and σ^2_β.

Both issues (testing for heterogeneity and cointegration) have been investigated to some extent by the literature. Tests for slope homogeneity are available in the context of stationary panels: examples include Phillips and Sul (2003), Kapetanios (2003) and Pesaran and Yamagata (2008), and have also been developed for cointegrated panels (see Mark and Sul, 2003 and Westerlund and Hess, 2011). However, existing tests designed for non-stationary panels require the assumption that all the units in the panel are cointegrated. Thus, currently available tests cannot be applied to the case where some of the units are (observationally equivalent to) spurious regressions, i.e. with u_{it} non-stationary. On the contrary, the test developed in this paper can be used irrespective of the u_{it}s being stationary, non-stationary, or a mixture of the two (and therefore to all possible values of λ in the interval $[0, 1]$).
Allowing for some units to be cointegrated (and simultaneously for others to be non cointegrated) is important in empirical applications. In addition to being a possible example of heterogeneity which makes the set-up more general and comprehensive, evidence of a "mixed panel", where some units are not cointegrated, is a well known stylized fact, which has been found in numerous contexts (see also the discussion in Fuertes, 2008, and Trapani, 2012). Indeed, it is not uncommon, at a disaggregate

level, to find evidence of a panel where some units are cointegrated but others are not. This can be due to various reasons, typically related to the presence of neglected nonlinearity in the adjustment mechanism. Examples of such nonlinearities, and, therefore, possible examples of empirical applications where our methodology could prove useful include: threshold effects such as fixed costs of adjustment, presence of target bands or no-arbitrage bands (Balke and Fomby, 1997; see also Lo and Zivot, 2001 for a survey); asymmetries in the adjustment process to a long-run equilibrium (Levy et al., 1997); and the possible presence of (non-stationary) measurement errors or omitted variables (Choi et al., 2008); or neglected breaks in the long run relationship (Gregory and Hansen, 1996). As well as a theoretical possibility, finding a mixed panel (and thus evidence against panel cointegration) is a well documented fact in empirical applications. This is the case e.g. in PPP studies (Balke and Wohar, 1998; Taylor, 2001; Coakley et al., 2005); in the Feldstein-Horioka puzzle literature (Coakley et al., 2004; Giannone and Lenza, 2009); and in the growth literature (Krueger and Lindahl, 2001), *inter alia*.

In the light of the discussion above, another potentially relevant contribution of this paper is a test for the null of cointegration. Having cointegration (instead of no cointegration) as the null can be viewed as more natural, since it usually is the working hypothesis of interest and it often corresponds to a well established economic theory (see e.g. Taylor, 2001, in the context of PPP). The test can be applied in the presence of slope heterogeneity and cross-sectional dependence in the u_{it}s. This complements the existing literature, where earlier contributions required cross sectional independence (e.g. McCoskey and Kao, 1998).

The inferential theory for σ_β^2 and for λ uses the Method of Moments, applied to data in levels. Consistency is shown on the whole parameter space $\left(\lambda, \sigma_\beta^2\right) \in \{[0,1] \times [0,+\infty)\}$. Indeed, we show that it is possible to estimate σ_β^2 even when $\lambda > 0$, without needing to know which units are cointegrated and which ones are spurious regressions. This would not be possible if one were to use a dispersion-type estimator based e.g. on the individual slope estimates, given that some of these would be inconsistent. As far as estimating λ is concerned, Ng (2008) proposes an estimation technique for the fraction of non-stationary units in a panel unit root context; however, the inferential theory is not fully developed. In this paper, we develop a consistent estimator of λ without imposing any restrictions on λ or on σ_β^2: again, the estimator can be employed without prior knowledge as to which units cointegrate or not, or whether slopes are homogeneous or not.

The remainder of the paper is organised as follows. Section 2 discusses the model and the main assumptions. Estimation of the degree of heterogeneity and of the fraction of spurious regressions, and the tests, are in Section 3; we present the main results for the case of a single regressor in (1.1), but in Section 3.3 we also report the generalisation of our results to the case - relevant for empirical applications - of a multiple regression. Monte Carlo evidence is in Section 4, and in Section 5 we present an application to the PPP hypothesis to illustrate the results in the paper. Section 6 concludes. Proofs of the main results are in Appendix A; further results and derivations are in Appendix B.

NOTATION. Throughout the paper, $\stackrel{d}{\to}$ and $\underset{H_0}{\stackrel{d}{\to}}$ denote weak convergence and weak convergence under the null H_0 respectively, and $\stackrel{p}{\to}$ convergence in probability; M denotes a generic finite constant that does not depend on n or T. Stochastic processes such as $W(r)$ on $[0,1]$ are usually written as W, integrals such as $\int_0^1 W(r)\,dr$ are written as $\int W$ and we define demeaned Brownian motions as $\bar{W} = W - \int W$. We extensively use the notation $\phi_{nT} = \min\left\{\sqrt{n}, \sqrt{T}\right\}$; $\lfloor \cdot \rfloor$ denotes the integer part of a number, and $\|\cdot\|$ is the Euclidean norm of a vector. Finally, the vectorization operator and the trace of a matrix A are defined as $vec(A)$ and $tr(A)$, respectively. Other notation is introduced later on in the paper.

2. Model and assumptions

We begin by considering the case of a single regressor in (1.1). Although this is an obvious limitation for practical applications, on the other hand it allows to present the main ideas of the paper without these being overshadowed by notation and mathematical details. In Section 3.3, we consider the extension of our estimators and tests to the case of a multiple regression.

Consider equation (1.2) with a single regressor:

$$y_{it} = \alpha_i + \beta x_{it} + v_{it},$$

where $v_{it} = u_{it} + (\beta_i - \beta) x_{it}$ and y_{it} and x_{it} are $I(1)$. We allow for strong cross sectional dependence among the regressors x_{it} by considering the following DGP

$$x_{it} = l_i' f_t + w_{it}, \tag{2.1}$$

with f_t an h-dimensional vector of common nonstationary factors (i.e. $f_t \sim I(1)$) and w_{it} assumed to be $I(1)$ also. Note that stationary common factors and stationary idiosyncratic shocks could be added to (2.1) without changing any of the arguments in the paper: in essence, this is because the asymptotics is driven by the nonstationary components only. Equation (2.1) considers the presence of common stochastic trends (and therefore of strong cross dependence) in the regressors, in a similar spirit to Kapetanios et al. (2011) (see also Bai et al., 2009).
As far as the error term u_{it} is concerned, we consider the following representation:

$$u_{it} = e_{it} + \sum_{j=-\infty}^{+\infty} c_{ij} \Delta x_{it+j} + g_i' h_t, \tag{2.2}$$

where h_t is a vector of common stationary factors to allow for cross dependence. The purpose of the term $\sum_{j=-\infty}^{+\infty} c_{ij} \Delta x_{it+j}$ in (2.2) is to take into account the possible presence of endogeneity in some or all the units: this way of modelling endogeneity is similar to Saikkonen (1991) and Choi et al. (2008). Finally, we consider the presence of both cointegrating and spurious regressions by defining

$$e_{it} = e_{\lambda,it} d_{\lambda,i} + e_{1-\lambda,it} \left(1 - d_{\lambda,i}\right), \tag{2.3}$$

with $d_{\lambda,i} = 1$ for $\lfloor n\lambda \rfloor$ units and zero for $n - \lfloor n\lambda \rfloor$ units. Let $\Delta e_{\lambda,it} = \varepsilon_{it}^e$ and $e_{1-\lambda,it} = \varepsilon_{it}^e$, where ε_{it}^e is stationary for all i. In essence, (2.3) entails that $e_{\lambda,it}$ is non-stationary and $e_{1-\lambda,it}$ is stationary. Since

$\lambda \in [0,1]$, three alternative cases are considered: *(a)* all units are cointegrated, *(b)* all units are spurious regressions, and *(c)* the panel is a mixture of cointegrated and spurious regressions (mixed panel).

Let $e_{it}^{w} = \Delta w_{it}$ and $e_{t}^{f} = \Delta f_{t}$, and define $\omega_{it} = [e_{it}^{w\prime}, \varepsilon_{it}^{e}]'$, $\tilde{\omega}_{it} = \left[\omega_{it}', e_{t}^{f\prime}, h_{t}'\right]'$. Consider the following assumptions.

Assumption 1: *(i)* ω_{it} is independent across i; *(ii)* l_i and g_i are either (a) nonstochastic with $\|l_i\| \leq M$, and $\|g_i\| \leq M$ for all i, or (b) stochastic and independent across i and of all other random variables with $E\|l_i\|^{8+\delta} \leq M$ and $E\|g_i\|^{4+\delta} \leq M$ for all i.

Assumption 2: *(i)* $\tilde{\omega}_{it}$ is a linear process with $\tilde{\omega}_{it} = \sum_{j=0}^{+\infty} \rho_{ij} \varepsilon_{it-j}^{\omega}$, where, for all i: (a) $\varepsilon_{it}^{\omega}$ is independent across t with $E(\varepsilon_{it}^{\omega}) = 0$, (b) $\sum_{j=0}^{+\infty} j^2 \|\rho_{ij}\| \leq M$, (c) $E\|e_{it}^{w}\|^{8+\delta} \leq M$, $E\left\|e_{t}^{f}\right\|^{8+\delta} \leq M$, $E\|\varepsilon_{it}^{e}\|^{4+\delta} \leq M$, and $E\|h_t\|^{4+\delta} \leq M$; (d) $E\|e_{i0}^{w}\|^{8+\delta} \leq M$, $E\left\|e_{0}^{f}\right\|^{8+\delta} \leq M$, $E\|\varepsilon_{i0}^{e}\|^{4+\delta} \leq M$, and $E\|h_0\|^{4+\delta} \leq M$;; *(ii)* letting the long-run variances of Δx_{it} and $\Delta e_{\lambda,it}$ be denoted as $\sigma_{x,i}^2$ and $\sigma_{e,i}^2$ respectively, it holds that they are non-stochastic with (a) $0 < M \leq \sigma_{x,i}^2 < \infty$ and (b) $0 < M \leq \sigma_{e,i}^2 < \infty$ for all i; *(iii)* it holds that, in equation (2.2), $\sum_{j=-\infty}^{+\infty} |c_{ij}| \leq M$ for all i; *(iv)* for $1 \leq t \leq T$, $\{\varepsilon_{it}^{e}\}$, $\{h_t\}$, $\left\{e_{t}^{f}\right\}$ and $\{e_{it}^{w}\}$ are mutually independent groups for all i.

Assumption 3: *(i)* β_i is independent across i with $E(\beta_i) = \beta$, $Var(\beta_i) = \sigma_{\beta}^2$, and $E|\beta_i|^{4+\delta} \leq M$ for some $\delta > 0$; *(ii)* $\{\beta_i\}$ is independent of all other random variables.

Assumption 1 considers the cross sectional properties of the innovations. Both ε_{it}^{e} and e_{it}^{w} are assumed to be cross sectionally independent. In essence, this is required in order for a cross sectional CLT to hold, and in principle it could be relaxed by replacing independence with some summability conditions. The presence of (strong) cross sectional dependence in the error term u_{it} is taken into account by introducing a common factor structure in (2.2). However, the common factors h_t are assumed to be stationary, otherwise u_{it} would be non-stationary by construction for all i - see also Ng (2008) - thus making $\lambda = 1$ by construction. This, inter alia, rules out the case where the y_{it}s are cross-unit cointegrated, which corresponds to having $h_t \sim I(1)$ and e_{it}, in (2.2), stationary for all i; we refer to the paper by Gengenbach et al. (2006) for a comprehensive discussion of the various sources of cointegration and nonstationarity in panels with common factors. We point out that, in principle, it would be possible to allow for common nonstationary factors, i.e. for $h_t \sim I(1)$, as long as these are "weak" - i.e. as long as $g_i = O\left(n^{-1/2}\right)$. In this paper, estimation of g_i and h_t is not required: this is essentially due to the fact that the relevant asymptotics is driven by the non-stationary components, and the stationary ones are smoothed away. Also, in (2.1), estimation of l_i and f_t is not required and results hold even if $l_i = 0$ for some or all the units.

Assumption 2 entails that time dependence and heteroskedasticity are allowed for; all innovations are assumed to be generated by a linear process. Note that, whilst endogeneity is allowed for in equation (2.2) through the term $\sum_{j=-\infty}^{+\infty} c_{ij} \Delta x_{it+j}$, estimation of the c_{ij}s, or using an estimator that takes into account endogeneity, is not required. Part *(iv)* of the assumption states that the innovations driving e_{it} and x_{it}

are independent at all leads and lags. This is tantamount to assuming that the (infeasible) dynamic regression method removes endogeneity, which corresponds to the assumption of strict exogeneity of the dynamic regression of Choi et al. (2008). As far as the notation is concerned, in part *(ii)* of the assumption we have defined the long-run variance of $\Delta e_{\lambda,it}$ as $\sigma^2_{e,i}$; based on (2.2), it is clear that $\sigma^2_{e,i}$ is the long-run variance of Δu_{it} for all the units which are spurious regressions, whereas, when u_{it} is stationary, Assumption 2 entails that the long-run variance of Δu_{it} is zero by construction.
Assumption 3 contains moment conditions for β_i, and it is based, essentially, on the so-called "random coefficient model" - see e.g. Kapetanios et al. (2011), who use a very similar setup.
Finally, we point out that the paper is based on the maintained assumption that y_{it} and x_{it} are $I(1)$ for all i. This can be verified by running a panel unit root test that takes cross sectional dependence into account - examples include Bai and Ng (2004) and Pesaran et al. (2013).

Let $\bar{x}_{it} = x_{it} - T^{-1}\sum_{t=1}^{T} x_{it}$ and $\bar{y}_{it} = y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$. The LSDV estimator for β in (1.2) is

$$\hat{\beta} = \left[\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^2\right]^{-1}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\bar{y}_{it}\right].$$

The quantities of interest, for the purpose of estimation and testing, are the fraction of spurious regressions, λ, and the degree of heterogeneity across coefficients, σ^2_β. In order to estimate them, we now present two statistics whose probability limits depend, linearly, on both λ and σ^2_β. In particular, we present a consistent estimator of the asymptotic variance of $\hat{\beta}$ (equation (2.4) below), and an inconsistent one (equation (2.5) below). The idea is that, once the probability limits of the two statistics are worked out and their dependence on λ and σ^2_β is explicitly expressed, inference can be based on either (or both) statistic, essentially by an application of the Method of Moments. Let $\hat{v}_{it} = \bar{y}_{it} - \hat{\beta}\bar{x}_{it}$ and define the two statistics

$$\hat{\psi}_1 = n\frac{\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\bar{x}_{it}\hat{v}_{it}\right]^2}{\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^2\right]^2}, \tag{2.4}$$

$$\hat{\psi}_2 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^2}. \tag{2.5}$$

Henceforth, we use the following notation. We define the σ-field associated with $\{f_t\}_{t=1}^{T}$ as $\mathcal{F}$; expectations conditional on $\mathcal{F}$ are denoted as $E_{\mathcal{F}}$. We let W_{xi} and W_{ei} denote the Brownian motions associated with the partial sums of x_{it} and e_{it} respectively, i.e. $T^{-3/2}\sum_{t=1}^{\lfloor Tr\rfloor} x_{it} \xrightarrow{d} W_{xi}(r)$ and similarly W_{ei}. We also define $\sigma^2_e = \lim_{n\to\infty}(n\lambda)^{-1}\sum_{i=1}^{n}\sigma^2_{e,i}d_{\lambda,i}$, using the convention that $\sigma^2_e = 0$ when $\lambda = 0$.

It holds that:

Proposition 1. *Let Assumptions 1-3 hold. As* $(n,T)\to\infty$

$$\hat{\psi}_1 \xrightarrow{p} \gamma_3^{-2}\left[\lambda\gamma_1+\sigma_\beta^2\gamma_2\right], \tag{2.6}$$

$$\hat{\psi}_2 \xrightarrow{p} \gamma_3^{-1}\left[\lambda\frac{\sigma_e^2}{6}+\sigma_\beta^2\gamma_3\right], \tag{2.7}$$

where γ_1*,* γ_2 *and* γ_3 *are short-hand notations for the limits of* $(n\lambda)^{-1}\sum_{i=1}^n E_{\mathcal{F}}\left[\left(\int\bar{W}_{xi}\bar{W}_{ei}\right)^2\right]d_{\lambda,i}$*,* $n^{-1}\sum_{i=1}^n E_{\mathcal{F}}\left[\left(\int\bar{W}_{xi}^2\right)^2\right]$ *and* $n^{-1}\sum_{i=1}^n E_{\mathcal{F}}\left(\int\bar{W}_{xi}^2\right)$ *respectively - we use the convention that* $\gamma_1=0$ *when* $\lambda=0$*. Further, as* $(n,T)\to\infty$ *with* $\frac{n}{T}\to 0$ *and* $\left(\lambda,\sigma_\beta^2\right)\neq(0,0)$

$$\sqrt{n}\left(\hat{\beta}-\beta\right)\xrightarrow{d}\gamma_3^{-1}\sqrt{\lambda\gamma_1+\sigma_\beta^2\gamma_2}\times Z, \tag{2.8}$$

with $Z\sim N(0,1)$ *independent of* $\mathcal{F}$ *. When* $\left(\lambda,\sigma_\beta^2\right)=(0,0)$*, it holds that* $\hat{\beta}-\beta=O_p\left(T^{-1}\right)$*.*

Equations (2.6) and (2.7) provide the probability limits for $\hat{\psi}_1$ and $\hat{\psi}_2$: as an ancillary result, it can be noted that $\hat{\psi}_1$ estimates the asymptotic variance of $\hat{\beta}$ consistently, whilst $\hat{\psi}_2$ is an inconsistent estimator. Equation (2.8) states that, as long as $\frac{n}{T}\to 0$ as $(n,T)\to\infty$, $\hat{\beta}$ is consistent for β at a rate $\sqrt{n}$. This result is typical of spurious panel regression (Kao, 1999; Phillips and Moon, 1999). An illustration of the rate of convergence for $\hat{\beta}$ when $\left(\lambda,\sigma_\beta^2\right)\neq(0,0)$ can be given by noting that the denominator of $\hat{\beta}-\beta$ is $O_p\left(nT^2\right)$; on the other hand, the term that dominates in the numerator is given by $\sum_{i=1}^n(\beta_i-\beta)\sum_{t=1}^T\bar{x}_{it}^2+\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^T\bar{x}_{it}\bar{e}_{it}$. Considering the first term, heuristically $\sum_{t=1}^T\bar{x}_{it}^2=O_p\left(T^2\right)$ by the FCLT, and the cross sectional independence of the β_is entails that the first term should have magnitude growing as $\sqrt{n}$, whence $\sum_{i=1}^n(\beta_i-\beta)\sum_{t=1}^T\bar{x}_{it}^2=O_p\left(\sqrt{n}T^2\right)$. Similarly, the second term is $O_p\left(\sqrt{n}T^2\right)$: in view of e_{it} being $I(1)$, the FCLT gives $\sum_{t=1}^T\bar{x}_{it}\bar{e}_{it}=O_p\left(T^2\right)$; the $\sqrt{n}$ rate follows from cross sectional independence of the e_{it}s. Thus, $\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^T\bar{x}_{it}\bar{e}_{it}=O_p\left(\sqrt{n}T^2\right)$. These results illustrate the dual relationship between neglected slope heterogeneity and spurious regression in a panel context.

When $\left(\lambda,\sigma_\beta^2\right)=(0,0)$, the rate of convergence of $\hat{\beta}$ is $O_p\left(T^{-1}\right)$ due to endogeneity: this introduces an asymptotic bias of order $O_p\left(T^{-1}\right)$ in each unit i, which does not get smoothed away by cross-sectional averaging. Under no endogeneity, Trapani (2012) shows that $\hat{\beta}-\beta=O_p\left(n^{-1/2}T^{-1}\right)$.

3. Inference on σ_β^2 and λ

This section discusses the consistent estimation of σ_β^2 and λ (Sections 3.1.1 and 3.2.1 respectively), providing estimators and rates of convergence in the whole parameter space $\left(\lambda,\sigma_\beta^2\right)$. Further, tests are proposed for the two null hypotheses $H_0:\sigma_\beta^2=0$ and $H_0:\lambda=0$ (Sections 3.1.2 and 3.2.2 respectively).

Building on Proposition 1, it is possible to construct estimators for σ_β^2 and λ using, with some modifications detailed below, $\hat{\psi}_1$ and $\hat{\psi}_2$ defined in (2.4) and (2.5). Estimation of σ_β^2 and λ (and testing for $H_0:\sigma_\beta^2=0$) is based on $\hat{\psi}_2$ only; conversely, in order to test for $H_0:\lambda=0$, we employ a test statistic based on the system of equations defined by (2.6) and (2.7). In order to present the results, we define the following Boolean variables, used throughout this section:

$$\begin{aligned} d_0 &= 1 \text{ if } \lambda = 0, \\ d_1 &= 1 \text{ if } \lambda = 1, \\ d_\sigma &= 1 \text{ if } \sigma_\beta^2 > 0, \end{aligned}$$

and zero otherwise.

3.1. Inference on σ_β^2

In this section, we present results on the estimation of σ_β^2 (Section 3.1.1), and on testing for $H_0 : \sigma_\beta^2 = 0$ (Section 3.1.2). In the whole section, λ is allowed to take any value in the interval $[0, 1]$: results are robust to λ and do not require its estimation.

3.1.1. Consistent estimation of σ_β^2

We define the estimator of σ_β^2, based on Proposition 1, as

$$\hat{\sigma}_\beta^2 = \hat{\psi}_2 - \frac{1}{6\hat{\gamma}_3}\frac{1}{n}\sum_{i=1}^{n}\tilde{\sigma}_{e,i}^2, \tag{3.1}$$

where

$$\hat{\gamma}_3 = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right]. \tag{3.2}$$

In (3.1), $\tilde{\sigma}_{e,i}^2$ is an estimator of the long run variance of Δu_{it}, and it can be calculated by using a weighted-sum-of-covariances estimator (see e.g. Andrews, 1991) based on the first-differenced residuals $\Delta\hat{u}_{it} = \Delta y_{it} - \tilde{\beta}_i \Delta x_{it}$, where $\tilde{\beta}_i$ is a consistent estimator of β_i. In order to accommodate for the possible presence of endogeneity, we suggest estimating β_i from the individual equations using the Feasible GLS (FGLS) estimator proposed in Choi et al. (2008).[1] Heuristically, it can be expected that $\tilde{\sigma}_{e,i}^2$ is consistent for $\sigma_{e,i}^2$ at a rate $\eta \in \left(0, \frac{1}{2}\right)$ for the units that are spurious regressions, i.e. $\tilde{\sigma}_{e,i}^2 = \sigma_{e,i}^2 + O_p\left(T^{-\eta}\right)$. On the other hand, for the units that are cointegrating regressions, $\tilde{\sigma}_{e,i}^2$ converges to zero by construction at a rate $T^{-\eta}$, viz. $\tilde{\sigma}_{e,i}^2 = O_p\left(T^{-\eta}\right)$. We postpone the discussion on how to actually compute $\tilde{\sigma}_{e,i}^2$ after Theorem 1.

The rates of convergence of $\hat{\sigma}_\beta^2$ are reported in the following theorem:

[1] The FGLS estimator of Choi et al. (2008) is based on estimating ρ in $\tilde{u}_{it} = \rho\tilde{u}_{it-1} + error$, where $\tilde{u}_{it}$ is the OLS regression error from (1.1), and then use the estimated ρ (say $\hat{\rho}$) to implement the Cochrane-Orcutt transformations $y_{it} - \hat{\rho}y_{it-1}$ and $x_{it} - \hat{\rho}x_{it-1}$. The slope β_i can be then estimated running OLS on the transformed data. Naturally, when the estimator is applied to an individual equation where cointegration is not present, the slope β_i merely represents the short-run correlation between the first differences of the y_{it}s and of the x_{it}s. The FGLS yields $\tilde{\beta}_i - \beta_i = O_p\left(T^{-1}\right)$ if unit i cointegrates, and $\tilde{\beta}_i - \beta_i = O_p\left(T^{-1/2}\right)$ if unit i is a spurious regression.

Theorem 1. *Let Assumptions 1-3 hold and assume that* $E\left|\tilde{\sigma}_{e,i}^2 - \sigma_{e,i}^2\right|^2 = O\left(T^{-2\eta}\right)$ *for each* i*. As* $(n,T) \to \infty$*, it holds that* $\hat{\sigma}_\beta^2 - \sigma_\beta^2 = o_p(1)$ *for all* $\left(\lambda, \sigma_\beta^2\right)$*, with*

$$\hat{\sigma}_\beta^2 - \sigma_\beta^2 = d_\sigma O_p\left(\frac{1}{\sqrt{n}}\right) + (1-d_0)\left[O_p\left(\frac{1}{T^\eta}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right] + (1-d_1)\, O_p\left(\frac{1}{T^\eta}\right).$$

Theorem 1 states that $\hat{\sigma}_\beta^2$ is consistent for σ_β^2 in the whole parameter space. Discontinuities are present on the boundary of the parameter space, as summarized below:

λ \ σ_β^2	0	>0
0	$O_p\left(\frac{1}{T^\eta}\right)$	$O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{T^\eta}\right)$
$\in (0,1]$	$O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{T^\eta}\right)$	$O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{T^\eta}\right)$

Table 1: rates of convergence for $\hat{\sigma}_\beta^2 - \sigma_\beta^2$ as $(n,T) \to \infty$.

In the case of either $\lambda \neq 0$ or $\sigma_\beta^2 \neq 0$, $\hat{\sigma}_\beta^2$ is consistent at a rate $1/\min\{\sqrt{n}, T^\eta\}$. It could be shown that $\sqrt{n}$-convergence would also be obtained if a dispersion-type estimator were employed, viz. $\tilde{\sigma}_\beta^2 = n^{-1}\sum_{i=1}^n \left(\hat{\beta}_i - \hat{\beta}\right)^2$, where $\hat{\beta}_i$ is a unit-specific slope estimate. However, this requires $\lambda = 0$: if $\lambda \neq 0$, some of the $\hat{\beta}_i$s will be inconsistent, thereby making $\tilde{\sigma}_\beta^2$ inconsistent also. Conversely, Theorem 1 states that $\hat{\sigma}_\beta^2$ estimates σ_β^2 consistently without imposing any restrictions on λ. A possible intuition is that the estimation error $\hat{\sigma}_\beta^2 - \sigma_\beta^2$ can be expressed in terms of cross-sectional averages - see equation (7.1). The rate $\sqrt{n}$ is a consequence of the cross-sectional independence of the β_is and of the e_{it}s.
As far as estimating $\sigma_{e,i}^2$ is concerned, as mentioned above this can be based on

$$\tilde{\sigma}_{e,i}^2 = \hat{\gamma}_{i0}^e + 2\sum_{j=1}^{h_T} \varphi\left(\frac{j}{h_T}\right)\hat{\gamma}_{ij}^e, \tag{3.3}$$

where

$$\hat{\gamma}_{ij}^e = T^{-1}\sum_{t=j+1}^{T} \Delta\hat{u}_{it}\Delta\hat{u}_{it-j}, \tag{3.4}$$

and $\varphi(\cdot)$ is a kernel. The rates in Table 1 indicate that convergence of $\tilde{\sigma}_{e,i}^2$ to $\sigma_{e,i}^2$ should be as fast as possible, i.e. the rate η should be as close to $\frac{1}{2}$ as possible. If the bandwidth h_T is chosen using the optimal selection rule of Andrews (1991), and $\varphi(\cdot)$ is chosen to be the quadratic spectral kernel (see Andrews, 1991), then it can be shown that, under our assumptions, $\eta = \frac{2}{5}$.
Theorem 1 provides rates of convergence - the limiting distribution of $\hat{\sigma}_\beta^2$ is in Theorem 7 in Appendix B, where it is also shown that $\hat{\sigma}_\beta^2$ has a bias of order $1/n$; the following corrected version can be employed as an alternative

$$\hat{\sigma}_{\beta,bc}^2 = \hat{\sigma}_\beta^2\left(1 - \frac{1}{n}\frac{\hat{\gamma}_2}{\hat{\gamma}_3^2}\right)^{-1}, \tag{3.5}$$

with

$$\hat{\gamma}_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2. \tag{3.6}$$

3.1.2. Testing for slope homogeneity - $H_0 : \sigma_\beta^2 = 0$

This section contains a test for slope homogeneity. The hypothesis testing framework is:

$$\begin{cases} H_0 : \sigma_\beta^2 = 0 \\ H_A : \sigma_\beta^2 > 0 \end{cases},$$

for $\lambda \geq 0$.
On the grounds of Theorem 1, we define the following test statistic for $H_0 : \sigma_\beta^2 = 0$

$$S_{nT}^{(\sigma)} = \hat{\gamma}_3 \times \hat{V}_\sigma^{-1/2} \times \sqrt{n}\hat{\sigma}_\beta^2, \tag{3.7}$$

where

$$\hat{V}_\sigma = \frac{1}{45}\frac{1}{n}\sum_{i=1}^{n}\left(\ddot{\sigma}_{e,i}^2\right)^2. \tag{3.8}$$

In (3.8), $\ddot{\sigma}_{e,i}^2$ is an estimator of the log-run variance $\sigma_{e,i}^2$; as Theorem 2 below indicates, this should be a consistent estimator of $\sigma_{e,i}^2$, with a slower rate of convergence than that of $\tilde{\sigma}_{e,i}^2$ employed in (3.1). Thus, based on (3.3), $\ddot{\sigma}_{e,i}^2$ can be computed as $\ddot{\sigma}_{e,i}^2 = \hat{\gamma}_{i0}^e + 2\sum_{j=1}^{h_T'}\varphi\left(h_T'^{-1}j\right)\hat{\gamma}_{ij}^e$, with bandwidth $h_T' = O\left(h_T^{1-\epsilon}\right)$ for $\epsilon > 0$.
The test statistic can also be computed using the bias-corrected version of $\hat{\sigma}_\beta^2$, $\hat{\sigma}_{\beta,bc}^2$, defined in (3.5), as $\tilde{S}_{nT}^{(\sigma)} = \hat{\gamma}_3 \times \hat{V}_\sigma^{-1/2} \times \sqrt{n}\hat{\sigma}_{\beta,bc}^2$. Under the null, $\hat{\sigma}_\beta^2$ should be "small", and therefore a test based on $S_{nT}^{(\sigma)}$ rejects for large values of $S_{nT}^{(\sigma)}$.
Let c_α be the $1-\alpha$ percentile of the standard normal distribution for some $\alpha \in [0,1]$, and define V_σ as the probability limit of $\hat{V}_\sigma$. It holds that:

Theorem 2. *Let Assumptions 1-3 hold; assume further that $E\left|\tilde{\sigma}_{e,i}^2 - \sigma_{e,i}^2\right|^2 = O\left(T^{-2\eta}\right)$ (and the same for $\ddot{\sigma}_{e,i}^2$) for each i. As $(n,T) \to \infty$ with $\frac{n}{T^{2\eta}} \to 0$ and for $\lambda \in (0,1]$, it holds that*

$$S_{nT}^{(\sigma)} \xrightarrow[H_0]{d} N(0,1). \tag{3.9}$$

As $(n,T) \to \infty$, if $\sqrt{n}\sigma_\beta^2 \to \infty$, it holds that $P\left[S_{nT}^{(\sigma)} > c_\alpha\right] = 1$.
If $\lambda = 0$, assuming that $E\left|\ddot{\sigma}_{e,i}^2 - \sigma_{e,i}^2\right|^2 = O\left(T^{-2\eta'}\right)$ for some $\eta' \in [0,\eta)$, as $(n,T) \to \infty$ with $\frac{n}{T^{2(\eta-\eta')}} \to 0$ it holds that $P\left[S_{nT}^{(\sigma)} \leq c_\alpha\right] = 1$ under H_0 and $P\left[S_{nT}^{(\sigma)} > c_\alpha\right] = 1$ if $\sqrt{n}\sigma_\beta^2 \to \infty$ for any $c_\alpha > 0$. The same results hold for $\tilde{S}_{nT}^{(\sigma)}$.

In view of (3.9), when $\lambda > 0$, $S_{nT}^{(\sigma)}$ is asymptotically normal under the null. If $\lambda = 0$, we show under which circumstances $S_{nT}^{(\sigma)} = o_p(1)$; although $S_{nT}^{(\sigma)}$ is degenerate in this case, the test can in principle still

be used. Indeed, when using the critical values of the standard normal distribution, the test has a zero probability of incurring in a Type I error; this theorem, essentially, states that the test can be employed with no prior knowledge on λ. Actually, the test should be employed under the assumption that $\lambda > 0$, using the critical values of the standard normal distribution. When $\lambda = 0$, such a choice yields a test which rejects the null when false with probability 1, and, when correct, with probability zero.
The test is shown to be consistent for any value of $\left(\lambda, \sigma^2_\beta\right)$. In the Appendix, we show that, under alternatives, the test statistic has a random drift term given by $\sqrt{n}\sigma^2_\beta \times \gamma_3 \times V_\sigma^{-1/2}$. Upon noting that, heuristically, V_σ increases as λ increases, this entails that the test based on $S^{(\sigma)}_{nT}$ becomes less capable of detecting heterogeneity as model (1.1) is contaminated with spuriousness. Also, $\hat{V}_\sigma$ increases as $\sigma^2_{e,i}$ increases; conversely, γ_3 increases with $\sigma^2_{x,i}$. Thus, the power of the test can be expected to be larger as the signal-to-noise ratio grows larger.

3.2. Inference on $\boldsymbol{\lambda}$

In this section, we present results on the estimation of λ (Section 3.2.1), and on testing for $H_0 : \lambda = 0$ (Section 3.2.2). For both results, the degree of slope heterogeneity σ^2_β can take any value in the interval $[0, +\infty)$.

3.2.1. Estimation of λ

We define the estimator of λ, based on Proposition 1, as

$$\hat{\lambda} = 6\hat{\gamma}_3\hat{\psi}^*_2 - 6 * \hat{\sigma}^2_\beta \times \frac{1}{nT^2}\sum_{i=1}^n\sum_{t=1}^T \frac{\bar{x}^2_{it}}{\hat{\sigma}^2_{e,i}}, \tag{3.10}$$

where

$$\hat{\psi}^*_2 = \frac{1}{\hat{\gamma}_3}\frac{1}{nT^2}\sum_{i=1}^n\sum_{t=1}^T \left(\frac{\hat{v}_{it}}{\hat{\sigma}_{e,i}}\right)^2. \tag{3.11}$$

In the calculation of $\hat{\lambda}$, $\hat{\sigma}^2_\beta$ is computed using the long run variance estimator $\tilde{\sigma}^2_{e,i}$ defined in (3.3), whose rate of convergence is $\tilde{\sigma}^2_{e,i} - \sigma^2_{e,i} = O_p\left(T^{-\eta}\right)$, and η is chosen to be as large as possible. Conversely, the estimators $\hat{\sigma}^2_{e,i}$ employed in (3.11) are defined in a different way, having a rate of convergence given by $\hat{\sigma}^2_{e,i} - \sigma^2_{e,i} = O_p\left(T^{-\varepsilon}\right)$ with $\varepsilon < \eta$. We postpone the discussion on the computation of $\hat{\sigma}^2_{e,i}$ after Theorem 3. Note that the construction of $\hat{\sigma}^2_{e,i}$ is a delicate issue, since no prior knowledge is usually available as to which units are spurious or cointegrating regressions; a similar issue is in Ng (2008). Thus, in equation (3.11), there could be some "divisions by zero" - this is essentially the reason why the choice of the rates of convergence in the estimation of the long-run variances (see also Theorem 2 above) requires a great deal of attention.

The rates of convergence of $\hat{\lambda}$ are reported in the following theorem.

Theorem 3. *Let Assumptions 1-3 hold and assume that* $E\left|\tilde{\sigma}_{e,i}^2-\sigma_{e,i}^2\right|^2=O\left(T^{-2\eta}\right)$ *and* $E\left|\hat{\sigma}_{e,i}^2-\sigma_{e,i}^2\right|^2=O\left(T^{-2\varepsilon}\right)$ *for each* i. *Let* s_T *be a sequence such that* $s_T^{-1}=o\left(T^{-\varepsilon}\right)$. *As* $(n,T)\rightarrow\infty$, *it holds that* $\hat{\lambda}-\lambda=o_p(1)$ *for every* $\left(\lambda,\sigma_\beta^2\right)$ *with*

$$\begin{aligned}\hat{\lambda}-\lambda &= (1-d_0)\left[O_p\left(\frac{1}{T^{\varepsilon}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)\right]\\ &\quad +d_\sigma\left[(1-d_0)\,O_p(1)+(1-d_1)\,o_p(s_T)\right]O_p\left(\frac{1}{\sqrt{n}}\right)\\ &\quad +(1-d_0)\,o_p\left(\frac{s_T}{n}\right)+(1-d_1)\,o_p\left(\frac{s_T}{T^{\eta}}\right)+d_\sigma(1-d_1)\,O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

Theorem 3 states that $\hat{\lambda}$ is consistent for λ, in the whole parameter space, with some discontinuities at the boundary.

λ \ σ_β^2	0	>0
0	$O_p\left(\frac{s_T}{T^{\eta}}\right)$	$O_p\left(\frac{s_T}{\sqrt{n}}\right)+O_p\left(\frac{s_T}{T^{\eta}}\right)$
$\in(0,1)$	$O_p\left(\frac{1}{T^{\varepsilon}}\right)+O_p\left(\frac{s_T}{\sqrt{n}}\right)+O_p\left(\frac{s_T}{T^{\eta}}\right)$	$O_p\left(\frac{1}{T^{\varepsilon}}\right)+O_p\left(\frac{s_T}{\sqrt{n}}\right)+O_p\left(\frac{s_T}{T^{\eta}}\right)$
1	$O_p\left(\frac{1}{T^{\varepsilon}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)$	$O_p\left(\frac{1}{T^{\varepsilon}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)$

Table 2: rates of convergence for $\hat{\lambda}-\lambda$ as $(n,T)\rightarrow\infty$.

As mentioned above the definitions of $\hat{\psi}_2^*$ in (3.11) contain some "divisions by zero" (asymptotically), which occur when $\sigma_{e,i}^2$ is estimated from a cointegrating regression. In this case, $\hat{\sigma}_{e,i}^2=O_p\left(T^{-\varepsilon}\right)$, which would entail that terms like $\frac{\hat{v}_{it}}{\hat{\sigma}_{e,i}}$ have asymptotic magnitude $O_p\left(T^{\varepsilon}\right)$. However, the presence of $\hat{\sigma}_{e,i}^2$ in the denominator is counterbalanced by the numerator converging to zero at a faster rate.

As Table 2 shows, the choice of the long run variance estimator $\hat{\sigma}_{e,i}^2$ plays an important role in determining the rate of convergence of $\hat{\lambda}-\lambda$. Based on (3.3), $\hat{\sigma}_{e,i}^2$ can be computed as

$$\hat{\sigma}_{e,i}^2=\hat{\gamma}_{i0}^e+2\sum_{j=1}^{h_{nT}}\left(1-\frac{j}{h_{nT}+1}\right)\hat{\gamma}_{ij}^e, \tag{3.12}$$

where the $\hat{\gamma}_{ij}^e$s are defined in (3.4); in (3.12), we use the Bartlett kernel, although other kernels could also be employed. As an example of how to select the bandwidth h_{nT}, consider the case $\lambda\in(0,1)$, where $\hat{\lambda}-\lambda=O_p\left(T^{-\varepsilon}\right)+O_p\left(s_T\phi_{\eta,nT}\right)$, where we use the short-hand notation $\phi_{\eta,nT}=\min\left\{\sqrt{n},T^{\eta}\right\}$ - recall from the previous section that η is chosen to be as high as possible, with typically $\eta=\frac{2}{5}$. The optimal ε which maximizes the rate of convergence of $\hat{\lambda}$ is the solution of $\min_\varepsilon\left[T^{-\varepsilon}+T^{\varepsilon+H}\phi_{\eta,nT}\right]$, where $H>0$ is an arbitrarily small number. It holds that:

$$T^{\varepsilon}=\sqrt{T^{H}\phi_{\eta,nT}}=T^{H/2}\min\left\{n^{1/4},T^{\eta/2}\right\}. \tag{3.13}$$

Equation (3.13) informs the choice of the bandwidth h_{nT} in (3.12). Based on Theorem 1 in Andrews (1991), the rate of convergence of $\hat{\sigma}_{e,i}^2$ is $O_p\left(\sqrt{\frac{h_{nT}}{T}}\right)$. This entails that h_{nT} can be chosen so as to satisfy $h_{nT} = O\left(T^{1-H}\phi_{\eta,nT}^{-1}\right)$. In Section 4, we run a set of simulations using $h_{nT} = T^{0.9}\phi_{\eta,nT}^{-1}$, under $\eta = \frac{2}{5}$; results do not seem to be strongly affected by different choices of h_{nT}, which serves as a guideline as to how to construct $\hat{\sigma}_{e,i}^2$. Note that, in this context, there is no need for an optimal estimator of $\sigma_{e,i}^2$ that minimizes the MSE of $\hat{\sigma}_{e,i}^2$: the requirement on the rate of $\hat{\sigma}_{e,i}^2 - \sigma_{e,i}^2$ is the one specified in (3.13), which states that the long run variances should be estimated at a rate that is neither too fast (otherwise the $\hat{\sigma}_{e,i}^2$s estimated from cointegrating regressions would make $\hat{\lambda}$ diverge), nor too slow (otherwise the $\hat{\sigma}_{e,i}^2$s estimated from spurious regressions would slow down the rate of convergence of $\hat{\lambda}$).

From (3.10), it can be noted that the bias correction is needed only when $\sigma_\beta^2 > 0$; indeed, when $\sigma_\beta^2 = 0$, the proof of Theorem 3 shows that implementing the bias correction would have the effect of making the convergence of $\hat{\lambda}$ to λ slower - in essence due to the estimation error of $\hat{\sigma}_\beta^2$. Thus, as an alternative, $\hat{\lambda}$ could be computed *after* testing $H_0 : \sigma_\beta^2 = 0$ (and, of course, after testing for $H_0 : \lambda = 0$); upon failing to reject the null, $\hat{\sigma}_\beta^2$ should be set to zero. The limiting distribution of $\hat{\lambda}$ for the case $\sigma_\beta^2 = 0$ (and $\hat{\sigma}_\beta^2$ set to zero) is in Theorem 8 in Appendix B.

3.2.2. Testing for cointegration - $H_0 : \lambda = 0$

This section contains a test for $H_0 : \lambda = 0$, i.e. for the null of panel cointegration. The hypothesis testing framework is

$$\left\{ \begin{array}{l} H_0 : \lambda = 0 \\ H_A : \lambda > 0 \end{array} \right. ,$$

for $\sigma_\beta^2 \geq 0$.

The estimator $\hat{\lambda}$ defined in (3.10) is affected by the estimation errors of $\hat{\sigma}_\beta^2$ and $\hat{\sigma}_{e,i}^2$. Thus, we do not base the test on $\hat{\lambda}$; rather, we define a test statistic, $\hat{\lambda}^\dagger$, which is proportional to λ but unaffected by $\hat{\sigma}_\beta^2$ and $\hat{\sigma}_{e,i}^2$, viz.

$$\tilde{S}_{nT}^{(\lambda)} = \hat{V}_\lambda^{-1/2} \times \sqrt{n}\left(\hat{\lambda}^\dagger - \hat{b}_{nT}^{S\lambda}\right), \tag{3.14}$$

where

$$\hat{\lambda}^\dagger = -\hat{\gamma}_3^3 \times \hat{\psi}_1 + \hat{\gamma}_2\hat{\gamma}_3 \times \hat{\psi}_2, \tag{3.15}$$

$$\hat{b}_{nT}^{S\lambda} = \hat{\gamma}_2 \times \frac{1}{nT}\sum_{i=1}^n \hat{\sigma}_{u,i}^2, \tag{3.16}$$

$$\hat{V}_\lambda = \frac{1}{n}\sum_{i=1}^n \left[\left(\hat{\beta}_i - \hat{\beta}\right)^2 - \hat{\sigma}_\beta^2\right]^2 \left[\hat{\gamma}_3\left(\frac{1}{T^2}\sum_{t=1}^T \bar{x}_{it}^2\right)^2 - \hat{\gamma}_2\left(\frac{1}{T^2}\sum_{t=1}^T \bar{x}_{it}^2\right)\right]^2; \tag{3.17}$$

in (3.16), $\hat{\sigma}_{u,i}^2 = T^{-1}\sum_{t=1}^T \hat{u}_{it}^2$ and $\hat{u}_{it} = \bar{y}_{it} - \hat{\beta}_i\bar{x}_{it}$, with the $\hat{\beta}_i$s being unit specific estimates, e.g. equation-by-equation OLS. Under the null, all units are stationary and therefore it can be expected that the $\hat{\beta}_i$s are superconsistent, and that $\hat{\sigma}_{u,i}^2$ is consistent for $Var\left(u_{it}\right)$.

Under the null λ should be "small", and therefore H_0 is rejected for large values of $\tilde{S}_{nT}^{(\lambda)}$. Define c_α to be the $1-\alpha$ percentile of the standard normal distribution, and V_λ as the probability limit of $\hat{V}_\lambda$. It holds that:

Theorem 4. *Let Assumptions 1-3 hold and assume that* $E\left|\tilde{\sigma}_{e,i}^2 - \sigma_{e,i}^2\right|^2 = O\left(T^{-2\eta}\right)$ *in the estimation of* $\hat{\sigma}_\beta^2$ *in (3.17). As* $(n,T) \to \infty$, *for* $\sigma_\beta^2 > 0$, *it holds that*

$$\tilde{S}_{nT}^{(\lambda)} \underset{H_0}{\overset{d}{\to}} N(0,1). \tag{3.18}$$

As $(n,T) \to \infty$, *if* $\sqrt{n}\lambda \to \infty$, *it holds that* $P\left[\tilde{S}_{nT}^{(\lambda)} > c_\alpha\right] = 1$.
If $\sigma_\beta^2 = 0$, *as* $(n,T) \to \infty$ *with* $\frac{\sqrt{n}}{T^{2-\eta}} \to 0$ *it holds that* $P\left[\tilde{S}_{nT}^{(\lambda)} \leq c_\alpha\right] = 1$ *under* H_0 *and* $P\left[\tilde{S}_{nT}^{(\lambda)} > c_\alpha\right] = 1$ *if* $\sqrt{n}\lambda \to \infty$ *for any* $c_\alpha > 0$.

Theorem 4 provides the null distribution and the consistency of the test, for any value of σ_β^2. No restrictions on the relative rate of divergence between n and T are needed when $\sigma_\beta^2 > 0$. When $\sigma_\beta^2 = 0$, under the null $\tilde{S}_{nT}^{(\lambda)} = o_p(1)$: the same considerations hold in this case as for $S_{nT}^{(\sigma)}$ when $\lambda = 0$. Specifically, the test can be applied with no prior knowledge as to whether $\sigma_\beta^2 > 0$, and should be applied under the assumption that $\sigma_\beta^2 > 0$ - when this is correct, the test follows a standard normal asymptotically, under the null. Conversely, when $\sigma_\beta^2 = 0$, the test rejects the null, when false, with probability 1 and, when correct, with probability zero.
In appendix, we show that the test statistic has, under alternatives, a drift proportional to $\left(\gamma_1\gamma_3 - \frac{\gamma_2}{6}\sigma_e^2\right) V_\lambda^{-1/2}$. This is inversely related to the signal-to-noise ratio, which entails that the test has more power, the lower the signal-to-noise ratio. Further, based on (3.17), it is also inversely related to the dispersion of the β_is around their average. Thus, as slope heterogeneity increases, tests based on $\tilde{S}_{nT}^{(\lambda)}$ are less capable of detecting failure of cointegration.

3.3. *Extension to multiple regressors*

In this section, we extend the theory to the case of k regressors. The presentation is aimed at empirical applications: the proofs of all the results reported here follow readily from modifying the arguments in the univariate case.
We begin by discussing the issue of testing, which essentially does not require any substantive modifications; we then move to the issue of estimation, where some modifications are required.

3.3.1. *Testing in the presence of multiple regressors*

Consider the following notation, which is employed throughout the whole section. We assume that the k-dimensional vector β_i is *i.i.d.* across i, with $E(\beta_i) = \beta$ and $Var(\beta_i) = \Sigma_\beta$; we also define

$$\hat{\Gamma}_3 = \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\bar{x}_{it}'.$$

The estimation error of the pooled OLS estimator for β in $y_{it} = \alpha_i + \beta' x_{it} + v_{it}$ is

$$\hat{\beta} = \beta + \left[\sum_{i=1}^{n}\sum_{t=1}^{T} \bar{x}_{it}\bar{x}_{it}'\right]^{-1}\left[\sum_{i=1}^{n}\sum_{t=1}^{T} \bar{x}_{it}\bar{x}_{it}'(\beta_i - \beta)\right] + \left[\sum_{i=1}^{n}\sum_{t=1}^{T} \bar{x}_{it}\bar{x}_{it}'\right]^{-1}\left[\sum_{i=1}^{n}\sum_{t=1}^{T} \bar{x}_{it}\bar{u}_{it}\right].$$

Let $Var\left(\hat{\beta}\right)$ denote the asymptotic covariance matrix of $\hat{\beta}$. It can be shown along the same lineas as (2.8) that

$$Var\left(\hat{\beta}\right) = \Gamma_3^{-1}(\lambda\Gamma_1 + \Gamma_2)\Gamma_3^{-1},$$

where Γ_2 is the limit of $n^{-1}\sum_{i=1}^{n} E_{\mathcal{F}}\left(\left(\int \bar{W}_{xi}\bar{W}_{xi}'\right)\Sigma_\beta\left(\int \bar{W}_{xi}\bar{W}_{xi}'\right)\right)$, Γ_3 is the limit of $\hat{\Gamma}_3$, and Γ_1 of $(n\lambda)^{-1}\sum_{i=1}^{n}\left(\int \bar{W}_{xi}\bar{W}_{ei}\right)\left(\int \bar{W}_{xi}\bar{W}_{ei}\right)' d_{\lambda,i}$.
As in the univariate case, we consider the following two estimators of $Var\left(\hat{\beta}\right)$:

$$\hat{\Psi}_1 = \hat{\Gamma}_3^{-1}\left\{\frac{1}{nT^4}\sum_{i=1}^{n}\left[\left(\sum_{t=1}^{T}\bar{x}_{it}\hat{v}_{it}\right)\left(\sum_{t=1}^{T}\bar{x}_{it}\hat{v}_{it}\right)'\right]\right\}\hat{\Gamma}_3^{-1}, \tag{3.19}$$

$$\hat{\Psi}_2 = \hat{\Gamma}_3^{-1}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2\right]. \tag{3.20}$$

As $(n,T)\to\infty$, similar calculations as in the proof of Proposition 1 yield

$$\hat{\Psi}_1 \xrightarrow{d} Var\left(\hat{\beta}\right),$$
$$\hat{\Psi}_2 \xrightarrow{d} \lambda\frac{\sigma_e^2}{6}\Gamma_3^{-1} + \Gamma_3^{-1}\Gamma_3^{\beta},$$

where Γ_3^{β} is the limit of $n^{-1}\sum_{i=1}^{n} E_{\mathcal{F}}\left(\int \bar{W}_{xi}'\Sigma_\beta\bar{W}_{xi}\right)$.
The estimating equation corresponding to (3.20) is

$$\hat{\Gamma}_3\hat{\Psi}_2 = \frac{1}{6}\left(\lambda\sigma_e^2\right)I_k + \left\{\hat{\Gamma}_3^{v}\left[vec\left(\Sigma_\beta\right)\right]\right\}I_k, \tag{3.21}$$

where $\hat{\Gamma}_3^{v} = \left(nT^2\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\bar{x}_{it}'\otimes\bar{x}_{it}'\right)$. Taking the trace, we obtain

$$\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2 = \hat{\Gamma}_3^{v}\left[vec\left(\Sigma_\beta\right)\right] + \frac{1}{6}\lambda\sigma_e^2. \tag{3.22}$$

Hence, after some manipulations based on exactly the same passages as in Section 3.1.2, the test statistic for the null of slope homogeneity ($H_0 : \Sigma_\beta = 0$) is

$$\breve{S}_{nT}^{(\sigma)} = \sqrt{n}\hat{V}_\sigma^{-1/2}\left[\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2 - \frac{1}{6n}\sum_{i=1}^{n}\tilde{\sigma}_{e,i}^2\right], \tag{3.23}$$

where $\tilde{\sigma}_{e,i}^2$s and $\hat{V}_\sigma$ are defined in (3.3) and (3.8) respectively.

Similarly, a test for the null of cointegration ($H_0 : \lambda = 0$) can be developed using the same logic as in Section 3.2.2, based on

$$\breve{S}_{nT}^{(\lambda)} = \sqrt{n}\hat{V}_{\lambda}^{-1/2}\left(\hat{\lambda}^{\dagger} - \hat{b}_{nT}^{S\lambda}\right), \tag{3.24}$$

where

$$\begin{aligned} \hat{\lambda}^{\dagger} &= \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2 - \hat{\Gamma}_3^v\left(\hat{\Gamma}_2^m\right)^{-1}\frac{1}{nT^4}\sum_{i=1}^{n} vec\left[\left(\sum_{t=1}^{T}\bar{x}_{it}\hat{v}_{it}\right)\left(\sum_{t=1}^{T}\bar{x}_{it}\hat{v}_{it}\right)'\right], \\ \hat{V}_{\lambda} &= \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\Gamma}_{3i}^v - \hat{\Gamma}_3^v\left(\hat{\Gamma}_2^m\right)^{-1}\left(\hat{\Gamma}_{3i}\otimes\hat{\Gamma}_{3i}\right)\right]\hat{K}_{\beta,i}\hat{K}_{\beta,i}'\left[\hat{\Gamma}_{3i}^v - \hat{\Gamma}_3^v\left(\hat{\Gamma}_2^m\right)^{-1}\left(\hat{\Gamma}_{3i}\otimes\hat{\Gamma}_{3i}\right)\right], \end{aligned}$$

with

$$\begin{aligned} \hat{\Gamma}_2^m &= \frac{1}{nT^4}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\bar{x}_{it}\bar{x}_{it}'\right)\otimes\left(\sum_{t=1}^{T}\bar{x}_{it}\bar{x}_{it}'\right), \\ \hat{\Gamma}_{3i} &= \frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{x}_{it}', \\ \hat{\Gamma}_{3i}^v &= \frac{1}{T^2}\sum_{t=1}^{T}\left(\bar{x}_{it}'\otimes\bar{x}_{it}'\right), \\ \hat{K}_{\beta,i} &= vec\left[\left(\hat{\beta}_i - \hat{\beta}\right)\left(\hat{\beta}_i - \hat{\beta}\right)' - \hat{\Sigma}_{\beta}\right]; \end{aligned}$$

recall that $\hat{b}_{nT}^{S\lambda}$ and $\hat{\beta}_i$ are defined in Section 3.2.2.
The asymptotics of $\breve{S}_{nT}^{(\lambda)}$ and $\breve{S}_{nT}^{(\sigma)}$ is in the following theorem, whose proof can be derived from the same arguments as in the proofs of Theorems 2 and 4.

Theorem 5. *Under the same assumptions as Theorem 2, it holds that, when* $\lambda \in (0, 1]$

$$\breve{S}_{nT}^{(\sigma)} \underset{H_0}{\overset{d}{\to}} N(0,1). \tag{3.25}$$

When $\lambda = 0$*, it holds that* $P\left[\breve{S}_{nT}^{(\sigma)} \leq c_{\alpha}\right] = 1$ *under* H_0*. Further, as* $(n, T) \to \infty$*, if* $\sqrt{n}\sigma_{\beta}^2 \to \infty$*, it holds that* $P\left[\breve{S}_{nT}^{(\sigma)} > c_{\alpha}\right] = 1$ *for any value of* λ*.*
Under the same assumptions as Theorem 4, it holds that, when $tr(\Sigma_{\beta}) > 0$

$$\breve{S}_{nT}^{(\lambda)} \underset{H_0}{\overset{d}{\to}} N(0,1). \tag{3.26}$$

When $tr(\Sigma_{\beta}) = 0$*, it holds that* $P\left[\breve{S}_{nT}^{(\lambda)} \leq c_{\alpha}\right] = 1$ *under* H_0*. Further, as* $(n, T) \to \infty$*, if* $\sqrt{n}\lambda \to \infty$*, it holds that* $P\left[\breve{S}_{nT}^{(\lambda)} > c_{\alpha}\right] = 1$ *for any value of* $tr(\Sigma_{\beta})$*.*

3.3.2. Estimation

We begin by considering (3.19), and the corresponding estimating equation

$$vec\left(\hat{\Gamma}_3\hat{\Psi}_1\hat{\Gamma}_3\right) = \Gamma_1^v + \hat{\Gamma}_2^m \left[vec\left(\Sigma_\beta\right)\right], \tag{3.27}$$

where the $k^2 \times 1$ vector Γ_1^v is defined as the limit of

$$vec\left[\frac{1}{n}\sum_{i=1}^n E_{\mathcal{F}}\left(\int \bar{W}_{xi}\bar{W}_{ei}\right)\left(\int \bar{W}_{xi}\bar{W}_{ei}\right)' d_{\lambda,i}\right].$$

There is no "natural" estimator of Γ_1^v; thus, we propose

$$\hat{\Gamma}_1^v = vec\left[\frac{1}{n}\sum_{i=1}^n \tilde{\sigma}_{e,i}^2\left(\frac{1}{T^2}\sum_{t=1}^T \bar{x}_{it}\bar{u}_{it}^*\right)\left(\frac{1}{T^2}\sum_{t=1}^T \bar{x}_{it}\bar{u}_{it}^*\right)'\right], \tag{3.28}$$

where recall that $\tilde{\sigma}_{e,i}^2$ is defined in (3.3); also, $\bar{u}_{it}^* = u_{it}^* - T^{-1}\sum_{t=1}^T u_{it}^*$, and the u_{it}^*s are simulated as *i.i.d.* $N(0,1)$, with independence holding over time and across units.

A comment on $\hat{\Gamma}_1^v$ is in order. The main difference with the univariate case is the presence of the $\bar{u}_{it}^*$s in (3.28). In order to understand how this works, note first that, intuitively, the $\tilde{\sigma}_{e,i}^2$s will converge to their limit (that is, to $\sigma_{e,i}^2 > 0$ for the units which are spurious regressions, and to 0 for those which are cointegrated), thus filtering out the cointegrated units, which do not play a role in the limit of $\hat{\Gamma}_1^v$. As far as the spurious regression units are concerned, heuristically it can be expected that, for each i

$$\tilde{\sigma}_{e,i}^2\left(\frac{1}{T^2}\sum_{t=1}^T \bar{x}_{it}\bar{u}_{it}^*\right)\left(\frac{1}{T^2}\sum_{t=1}^T \bar{x}_{it}\bar{u}_{it}^*\right)' \xrightarrow{d} \sigma_{e,i}^2\left(\int \bar{W}_{xi}\bar{W}_{ui}^*\right)\left(\int \bar{W}_{xi}\bar{W}_{ui}^*\right)', \tag{3.29}$$

where $\bar{W}_{ui}^*$ is a demeaned, standard Brownian motion. Averaging across i, it follows that

$$\frac{1}{n}\sum_{i=1}^n \tilde{\sigma}_{e,i}^2\left(\int \bar{W}_{xi}\bar{W}_{ui}^*\right)\left(\int \bar{W}_{xi}\bar{W}_{ui}^*\right)' \xrightarrow{d} (n\lambda)^{-1}\sum_{i=1}^n \sigma_{e,i}^2 E_{\mathcal{F}}\left[\left(\int \bar{W}_{xi}\bar{W}_{ui}^*\right)\left(\int \bar{W}_{xi}\bar{W}_{ui}^*\right)'\right] d_{\lambda,i}. \tag{3.30}$$

These arguments highlight the helpfulness of cross sectional averaging: in (3.29), the $\sigma_{e,i}\left(\int \bar{W}_{xi}\bar{W}_{ui}^*\right)$s are different from the $\left(\int \bar{W}_{xi}\bar{W}_{ei}\right)$s, but they have the same distribution by construction, and therefore the cross-sectional averages converge to the same limits.

It is now possible to estimate Σ_β; indeed, considering (3.27), the estimator is given by

$$vec\left(\breve{\Sigma}_\beta\right) = \left[\hat{\Gamma}_2^m\right]^{-1}\left[vec\left(\hat{\Gamma}_3\hat{\Psi}_1\hat{\Gamma}_3\right) - \hat{\Gamma}_1^v\right].$$

As far as the estimation of λ is concerned, the logic is the same as in Section 3.2.1. Let

$$\hat{\Psi}_2^* = \hat{\Gamma}_3^{-1}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\hat{v}_{it}^2}{\hat{\sigma}_{e,i}^2}\right],$$

$$\hat{\Gamma}_3^{*v} = \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\left(\bar{x}_{it}'\otimes\bar{x}_{it}'\right)}{\hat{\sigma}_{e,i}^2},$$

where $\hat{\sigma}_{e,i}^2$ is defined in (3.10). The estimator of λ is

$$\breve{\lambda} = 6\hat{\Gamma}_3\hat{\Psi}_2^* - 6\left\{\hat{\Gamma}_3^{*v}\left[vec\left(\hat{\Sigma}_\beta\right)\right]\right\}I_k. \tag{3.31}$$

Theorem 6. *Under the same assumptions as Theorem 1, it holds that*

$$\left\|vec\left(\breve{\Sigma}_\beta\right) - vec\left(\Sigma_\beta\right)\right\| = d_\sigma O_p\left(\frac{1}{\sqrt{n}}\right) + (1-d_0)\left[O_p\left(\frac{1}{T^\eta}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right] + (1-d_1)\,O_p\left(\frac{1}{T^\eta}\right).$$

Also, under the same assumptions as Theorem 3, $\breve{\lambda}$ has the same rate of convergence as $\hat{\lambda}$.

4. Small sample properties

This section contains Monte Carlo evidence on the size and power of tests for $H_0 : \sigma_\beta^2 = 0$ and $H_0 : \lambda = 0$. The DGP for each experiment is based on equations (1.1), (2.1) and (2.2)-(2.3):

$$\begin{aligned} y_{it} &= \alpha_i + \beta_i x_{it} + u_{it}, \\ x_{it} &= l_i f_t + w_{it}, \\ u_{it} &= e_{it} + g_i h_t, \\ e_{it} &= d_{\lambda,i}\left(e_{it-1} + \varepsilon_{it}^e\right) + \left(1 - d_{\lambda,i}\right)\left[\rho e_{it-1} + \varepsilon_{it}^e + \theta\varepsilon_{it-1}^e\right]. \end{aligned}$$

We set $(n,T) = \{(20,100),(20,200),(20,400),(50,50),(50,100),(50,200),(100,200)\}$. In order to avoid dependence on initial conditions, $T + 1000$ data have been generated, and the first 1000 observations have been discarded. We set the number of replications equal to 2000, so that the empirical null rejection frequencies reported in the tables have a 95% confidence interval of width ± 0.01.
The innovations $(\varepsilon_{it}^e, \Delta w_{it})$ are created as $i.i.d.$ Gaussian with $Var\left(\Delta w_{it}\right) = 1$ and $Var\left(\varepsilon_{it}^e\right) = \{0.5, 1, 2\}$ in order to assess the impact of different levels of the signal-to-noise ratio. We set $l_i \sim i.i.d.\ N(1,1)$ for all i and generate Δf_t as $i.i.d.\ N(0,1)$; we have tried using a multifactor structure, but results are virtually unchanged. Similarly, we have used $g_i \sim i.i.d.\ N(1,1)$ and $h_t \sim i.i.d.\ N(0,1)$, and again changing this into a multifactor structure has no impact on the results (as also predicted by the theory). Dynamics in the (stationary) error terms e_{it} is created using $\{\rho,\theta\} = \{0, 0.75\}\times\{-0.75, 0, 0.75\}$. Unreported experiments show that results are virtually unchanged when considering endogeneity as in (2.2). Finally, we have generated $\alpha_i \sim i.i.d.\ N(1,1)$ and $\beta_i \sim i.i.d.\ N\left(1, \sigma_\beta^2\right)$.
A final note on the estimation of the long-run variances $\sigma_{e,i}^2$: there are three estimators which need to

be considered for them. In all three cases, we have used a HAC-type estimator based on the Bartlett kernel. The bandwidth has been set equal to: $\lfloor T^{2/5} \rfloor$ for $\tilde{\sigma}^2_{e,i}$; $\lfloor T^{0.9} * T^{2/5} \rfloor$ for $\ddot{\sigma}^2_{e,i}$; and $\left\lfloor T^{0.9} * \phi^{-1}_{\eta,nT} \right\rfloor$ for $\hat{\sigma}^2_{e,i}$. All routines have been written using Gauss 10.

Estimation of σ^2_β and λ

We begin by evaluating the performance of the estimators of σ^2_β and λ - in particular, we consider $\hat{\sigma}^2_{\beta,bc}$ defined in (3.5) and $\hat{\lambda}$, defined in (3.10). We report bias and Root Mean Squared Error (RMSE); as far as $\hat{\sigma}^2_{\beta,bc}$ is concerned, these are computed as

$$\begin{aligned} bias &= \frac{1}{2000} \sum_{j=1}^{2000} \left[\hat{\sigma}^2_{\beta,bc}(j) - \sigma^2_\beta \right], \\ RMSE &= \sqrt{\frac{1}{2000} \sum_{j=1}^{2000} \left[\hat{\sigma}^2_{\beta,bc}(j) - \sigma^2_\beta \right]^2}, \end{aligned}$$

where $\hat{\sigma}^2_{\beta,bc}(j)$ is the value of $\hat{\sigma}^2_{\beta,bc}$ at the j-th replication; the same is done for $\hat{\lambda}$.
Results are based on $\left\{\lambda, \sigma^2_\beta\right\} \in \{0, 0.25, 0.75\} \times \{0, 0.25, 0.5\}$, under the set-up described above. As in the case of the other experiments in this section, we have also carried out simulations for $\lambda = 0.5$, which we do not report here to save space; results are anyway in line with the other cases reported here.

[Insert Tables 3a-4b somewhere here]

Considering $\hat{\sigma}^2_{\beta,bc}$, it can be noted that the bias and RMSE decline with T, mainly, and also with n, as can be expected from the theory - indeed, the RMSE declines as n increases in a more pronounced way than the bias. The impact of λ is quite apparent from the table: as λ increases, the estimator $\hat{\sigma}^2_{\beta,bc}$ improves. One exception is the slight increase in both bias and, especially, RMSE, as λ changes from 0 to 0.25. The presence of serial correlation has an impact on the performance of $\hat{\sigma}^2_{\beta,bc}$: the estimator performs clearly better when the error terms are uncorrelated, or when they have an autoregressive structure; conversely, they perform worse in the case of moving average structures. In this case, results are not sensitive to whether the MA root is positive or negative; a possible explanation could lie in the fact that the long-run variance estimators are based on estimated covariances, and these, in the case of moving average structures, decline to zero so that the HAC-type estimators considered here, which are truncated, are likely to miss some of the autocovariances.
Turning to $\hat{\lambda}$, the bias and RMSE decline as T increases but seem to be virtually unaffected by n - compare the results for $(n, T) = (20, 200)$, $(50, 200)$ and $(100, 200)$. Also, the estimator is not affected in any obvious way by the value taken by σ^2_β, whereas it can be noted that estimation becomes better as λ itself increases. The impact of error autocorrelation is also quite interesting, since $\hat{\lambda}$ behaves in the opposite way than $\hat{\sigma}^2_{\beta,bc}$. The performance of $\hat{\lambda}$ in the presence of moving average structures, at least

as T increases, is the same as one would have with uncorrelated errors; conversely, the presence of an autoregressive root worsens (dramatically, for small T) both the bias and the RMSE.

Testing for slope homogeneity - $H_0 : \sigma_\beta^2 = 0$

In addition to the settings described above, simulations were carried out with $\lambda \in \{0.25, 0.5, 0.75\}$. When evaluating power, we set $\sigma_\beta^2 \in \{0.1, 0.5\}$. Tests are based on using $\hat{\sigma}^2_{\beta,bc}$.

[Insert Table 5 somewhere here]

Table 5 contains the null rejection frequencies for the case $Var(\varepsilon_{it}^e) = 1$; virtually no changes were noted for different values of $Var(\varepsilon_{it}^e)$. In general, the test has a slight tendency to over-reject in small samples. As n and T increase (with $n \geq 50$ and $T \geq 200$), the size is very close to its nominal value and almost always within the confidence band $0.04 - 0.06$. The size is also robust to the value of λ, for moderately large samples: as λ changes, the empirical rejection frequencies change very little. Finally, dependence in the error term has almost no impact.

[Insert Table 6 somewhere here]

Table 6 report the power of the test under the alternative $\sigma_\beta^2 = 0.5$. The power decreases as $Var(\varepsilon_{it}^e)$ increases (as predicted by the theory), especially in small samples. The power increases as n increases, and also, albeit less so, as T increases. In small samples, it can be noted that the power decreases with λ, as predicted by the theory. When n and T increase, this becomes unnoticeable, and it is likely to be due to the fact that the number of units used for the test is actually $\lfloor n\lambda \rfloor$. Finally, the presence of serial dependence of the error term has little impact, especially as n and T increase.

Testing for cointegration - $H_0 : \lambda = 0$

Simulations have been carried out for $\sigma_\beta^2 \in \{0.25, 0.5, 0.75\}$, in order to assess the dependence of power and size on heterogeneity; also, $Var(\varepsilon_{it}^e)$ has been set to 0.5.

[Insert Table 7 somewhere here]

Table 7 shows that the test has a tendency to be conservative in finite samples, at least when σ_β^2 is small. This is in line with what predicted by the theory: Theorem 2 stipulates that, when $\sigma_\beta^2 = 0$, the test has zero probability of Type I error. This gradually vanishes, as n and T increase. Serial dependence in the error term has little effect, especially as n increases.

[Insert Table 8 somewhere here]

Table 8 reports the power of the test versus the alternative that $\lambda = 0.5$. The power increases clearly as n increases; the impact of T is less clear. The impact of nuisance parameters (σ_β^2 and $Var\left(\varepsilon_{it}^e\right)$) is as predicted by the theory. As far as σ_β^2 is concerned, as slope heterogeneity increases, the power of the test decreases significantly. As far as $Var\left(\varepsilon_{it}^e\right)$ is concerned, as the signal-to-noise ratio increases, the power also increases.

5. Empirical application

In order to illustrate our approach, we apply it to verify the PPP hypothesis. This application is also motivated by various empirical studies, which have found evidence of failure of cointegration at the unit-specific level, evidence of slope heterogeneity, and in general other mis-specification issues such as endogeneity (see e.g. Taylor, 2001 and Wagner, 2008).

We employ monthly observations, starting from April 1981 to December 2000 (so that $T = 237$), for $n = 21$ countries (the list is in Table 9, where the data and their source are also described). Based on the results in the previous section, we could expect that, with these sample sizes, the tests should have good power under the alternative and the correct size under the null - with, possibly, some tendency to be undersized when testing for $H_0 : \lambda = 0$ (at least for small values of σ_β^2) and to be oversized when testing for $H_0 : \sigma_\beta^2 = 0$ (at least for small values of λ).

Letting E_{it} be the exchange rate of country i with respect to the base country (in our case, the US, so that E_{it} is expressed as US dollars per unit of currency i), P_t^{US} the price of a representative basket of goods in the US, and finally P_{it} be the price of the same basket in country i, our model is

$$\ln E_{it} = \alpha_i + \beta \ln \frac{P_t^{US}}{P_{it}} + u_{it}. \tag{5.1}$$

Equation (5.1) is the so-called "strong form" of the PPP, where it is postulated that *(a)* all units are cointegrated (so that the PPP holds for each individual country) and *(b)* all units have the same slope β, which is equal to 1.

In Table 9, we have carried out tests for cointegration at an individual equation level, using an Engle-Granger procedure based on FM-OLS residuals; these results may be affected by the typical low power of such procedures (see e.g. Pedroni, 2004), and are reported only as a benchmark: however, as can be seen, the null of cointegration is rejected in 6 cases out of 21.

We have also included, in the table, the outcome of Kao (1999) test for the null of no panel cointegration. As can be noted, we found no evidence of cointegration at a panel level, thereby rejecting the PPP in the strong form - or, equivalently, rejecting the joint null that $\lambda = 0$ and $\sigma_\beta^2 = 0$. Of course, this approach too can be criticised due to numerous reasons, and it is reported here merely for comparison and reference - we refer to the analysis by Wagner (2008), where it is pointed out that conventional, first-generation panel unit root (and cointegration) tests may erroneously lead to rejection of PPP due to e.g. presence of strong cross-dependence, non-stationary common factors in the regressors (in this case, the prices), endogeneity, etc.

Based on the results above, a possible approach to studying the PPP at a panel level is based on the following "weak" form of the PPP

$$\ln E_{it} = \alpha_i + \beta_i \ln \frac{P_t^{US}}{P_{it}} + u_{it}, \tag{5.2}$$

testing whether slope homogeneity holds (that is, $\sigma_\beta^2 = 0$), and whether cointegration holds for all units (that is, $\lambda = 0$). We have implemented the tests proposed above with the same specifications as in the Monte Carlo exercise. Specifically, we have computed $S_{nT}^{(\sigma)}$ using the bias corrected estimator $\hat{\sigma}_{\beta,bc}^2$, although results do not change much when using $\hat{\sigma}_\beta^2$. Similarly, when computing $\hat{\lambda}$ - and when using equation (3.17) to compute $\tilde{S}_{nT}^{(\lambda)}$ - we have again used $\hat{\sigma}_{\beta,bc}^2$, and again using $\hat{\sigma}_\beta^2$ instead does not change the main findings.

All results are reported in Table 9.

[Insert Table 9 somewhere here]

As can be seen, the null of cointegration, $H_0 : \lambda = 0$, is rejected: the error terms in some of the units are observationally equivalent to non-stationary series. However, it can be noted that the point estimate is $\hat{\lambda} = 0.21$, which suggests that the PPP equation, in its weak form, is satisfied in around 5-6 out of 21 units. This result is very similar to the one implied by the individual specific tests. Furthermore, it should be noted that, as discussed in the Introduction, failure of cointegration in the micro equations (5.2) can be due to numerous reasons, such as neglected non-linearities, thus not necessarily implying the genuine lack of a long-run relationship between exchange rates and relative prices.
The null of slope homogeneity is also rejected; estimation shows that $\hat{\sigma}_\beta^2 = 0.6$ approximately. This estimate could be contrasted with the (reported) standard deviation of the individual estimates $\tilde{\beta}_i$, which has been carried out using the first-differenced version of (5.2) as per footnote 1, and which again appears very similar.

6. Conclusions

Testing for slope heterogeneity and cointegration are important issues in large, non-stationary panels. This paper develops an estimator for the degree of slope heterogeneity, that is robust to the case of having a mixture of cointegrated and spurious regressions. An estimator for the fraction of spurious regressions is also proposed. Hence, two tests are proposed for the null hypotheses of slope homogeneity and cointegration. Our methodology is valid under cross dependence, heteroskedasticity of errors and regressors, presence of common trends in the regressors, and endogeneity, which are typically found in empirical applications (see e.g. Wagner, 2008, in the context of PPP). Monte Carlo evidence shows that the tests have good size and power properties.

Some methodological and theoretical questions are still outstanding, at least partly. As Theorems 2 and 4 show, the two tests for $H_0 : \sigma_\beta^2 = 0$ and $H_0 : \lambda = 0$ can be applied separately, and directly from the outset; based on the result for the case of $\lambda = 0$ in Theorem 2, one should first test for $H_0 : \lambda = 0$, and subsequently decide how to compute $\hat{V}_\sigma$ in (3.8). Alternatively, a test for the null of homogeneous cointegration can be ran first and, upon rejecting the null, the two tests can be applied to identify the reason why homogeneous cointegration does not hold. As far as the model is concerned, one case that is left out from the basic model is the case of u_{it} containing $I(1)$ common factors. In this case, the framework implemented here could not be employed directly (essentially due to Proposition 1 failing; see Trapani, 2012, for details), and some pre-filtering of the common factors may be required. Similarly, we need to rule out the presence of linear trends in (1.1) and (2.1) - even in this case, Proposition 1 would fail, thus invalidating the main arguments in this paper, and therefore some detrending may be necessary prior to carrying out the analysis. Finally, inference is based mainly on $\hat{\psi}_2$, and also on $\hat{\psi}_1$ (at least as far as testing for $H_0 : \lambda = 0$ is concerned): albeit somewhat “natural”, these statistics are defined arbitrarily. In theory, it is possible to define other statistics similarly to $\hat{\psi}_1$ and $\hat{\psi}_2$, e.g. based on different estimators for β in (1.2). Several statistics could thence be combined, in a similar spirit to overidentified GMM. These issues are the subject of ongoing research by the author.

Acknowledgement

I wish to thank the Editors (Vasilis Sarafidis and Tom Wansbeek) and two anonymous Referees for very constructive comments which have improved the rigour and generality of our results. The usual disclaimer applies.

7. Appendix A: proofs of the main results

In this section, we present the proofs of the main results. Preliminary Lemmas, their proofs and the proof of Proposition 1, and further results of independent interest are in Appendix B. Henceforth, we use the following notation. The martingale approximation of a stationary process z_t (derived by applying the Beveridge-Nelson decomposition, BN henceforth; we refer to Phillips and Solo, 1992 for details) is denoted as z_t^*; similarly, the martingale approximation of a unit root process $S_t = \sum_{j=0}^{t} z_t$ is denoted as $S_t^* = \sum_{j=1}^{t} z_t^*$. Under Assumption 2, x_{it} and ε_{it}^{e} both admit a martingale approximation, and we refer to Phillips and Moon (1999) for details in a panel cointegration context. Recall also that $\phi_{nT} = \min\left\{\sqrt{n}, \sqrt{T}\right\}$; that $\mathcal{F}$ is the σ-field associated with $\{f_t\}_{t=1}^{T}$ and that $E_{\mathcal{F}}$ denotes expectation conditional on $\mathcal{F}$; and that the bar symbol above a series denotes demeaning, i.e. $\bar{z}_t = z_t - T^{-1}\sum_{t=1}^{T} z_t$. Where helpful, in the proofs we assume (with no loss of generality) that the first $\lfloor n\lambda \rfloor$ units are spurious regressions, and that the remaining ones are cointegrating regressions.

Proof of Theorem 1. We have

$$\begin{aligned}
\hat{\sigma}_{\beta}^{2} = \hat{\gamma}_3^{-1} \Bigg\{ & \frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)^2\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right) + \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{u}_{it}^2 \\
& + \left(\hat{\beta}-\beta\right)^2\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^2 + \frac{2}{n}\sum_{i=1}^{n}\left[(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)\right] \\
& - \left(\hat{\beta}-\beta\right)\frac{2}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right) - \left(\hat{\beta}-\beta\right)\frac{2}{n}\sum_{i=1}^{n}\left[(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\right] - \frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{\sigma}_{e,i}^2}{6}\Bigg\} \\
= {} & \Delta_{\sigma}^{-1}\left\{A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7\right\}. \qquad (7.1)
\end{aligned}$$

Consider A_1. We can write

$$A_1 = \sigma_{\beta}^2\hat{\gamma}_3 - \frac{1}{n}\sum_{i=1}^{n}\left[(\beta_i-\beta)^2 - \sigma_{\beta}^2\right]\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right) = A_{1,1} + A_{1,2}.$$

By construction, $\hat{\gamma}_3^{-1}A_{1,1} = \sigma_{\beta}^2$. As regards $A_{1,2}$, by the independence of the β_is, its variance is $n^{-2}\sum_{i=1}^{n} E\left[(\beta_i-\beta)^2 - \sigma_{\beta}^2\right]^2 E\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2$. By Assumption 3, $E\left[(\beta_i-\beta)^2 - \sigma_{\beta}^2\right]^2 < \infty$; further, by Assumptions 2*(i)* and similar arguments as in the proof of (8.17), it can be shown that $E\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2 < \infty$, so that putting all together, $\hat{\gamma}_3 A_1 = \sigma_{\beta}^2 + d_{\sigma}O_p\left(n^{-1/2}\right)$. Consider $A_2 + A_7$; it holds that

$$
\begin{aligned}
& A_2 + A_7 \\
= \; & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{u}_{it}^{*2}\right) - \frac{1}{6n}\sum_{i=1}^{n}\tilde{\sigma}_{e,i}^{2} + O_p\left(\frac{1}{\sqrt{T}}\right) \\
= \; & \frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{e}_{\lambda,it}^{*2} - \frac{\sigma_{e,i}^{2}}{6}\right) + \frac{1}{nT}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{1-\lambda,it}^{*2}\right) + \frac{1}{6n}\sum_{i=1}^{n}\left(\tilde{\sigma}_{e,i}^{2} - \sigma_{e,i}^{2}\right) \\
& + (1-d_0)\, O_p\left(\frac{1}{\sqrt{T}}\right) + (1-d_1)\, O_p\left(\frac{1}{T^{3/2}}\right) \\
= \; & A_{2,1} + A_{2,2} + A_{2,3} + (1-d_0)\, O_p\left(\frac{1}{\sqrt{T}}\right) + (1-d_1)\, O_p\left(\frac{1}{T^{3/2}}\right),
\end{aligned}
$$

where the $O_p\left(T^{-1/2}\right)$ term in the second line is the error term coming from the BN decomposition, similarly to the proof of Lemma 1. Consider $A_{2,1}$, and let $X_{1iT} = \left(T^{-2}\sum_{t=1}^{T}\bar{e}_{\lambda,it}^{*}\right)^2 - \sigma_{e,i}^{2}/6$. The sequence X_{1iT} is independent across i; by direct calculation, it can be shown that $E\left(X_{1iT}\right) = O\left(T^{-1}\right)$; further, by similar (indeed, easier) passages as in the proofs of (8.16) and (8.17), it holds that $E\left(X_{1iT}^{2}\right) < \infty$; thus, $E\left[\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(X_{1iT} - E\left(X_{1iT}\right)\right)\right]^2 = O(n)$, whence $A_{2,1} = (1-d_0)\,\left[O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right)\right]$. Consider now $A_{2,2}$; by Assumption 2, it follows immediately that $A_{2,2} = (1-d_1)\, O_p\left(T^{-1}\right)$. Finally, consider $A_{2,3}$; by having assumed $E\left|\tilde{\sigma}_{e,i}^{2} - \sigma_{e,i}^{2}\right|^2 = O\left(T^{-2\eta}\right)$, it follows that $A_{2,3} = O_p\left(T^{-\eta}\right)$. Thus, putting all together, we have $A_2 + A_7 = (1-d_0)\, O_p\left(n^{-1/2}\right) + O_p\left(T^{-\eta}\right)$. As regards A_3, it is of the same order as $\left(\hat{\beta} - \beta\right)^2$, so that, by Proposition 1, $A_3 = O_p\left(\phi_{nT}^{-2}\right)$. Turning to A_4, it is bounded by the square root of its variance, given by $n^{-2}\sigma_{\beta}^{2}\sum_{i=1}^{n} E\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)^2$. By using the proof of (8.16), this term is shown to be $O\left(n^{-1}\right)$ for the units that are spurious regressions; as regards the units that cointegrate, we have $E\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)^2 = O\left(T^{-2}\right)$, which is shown in the proof of Proposition 2 in Trapani (2013). Thus, putting all together, $A_4 = d_\sigma\, O_p\left(n^{-1/2}\right)\left[(1-d_0)\; O_p(1) + (1-d_1)\, O_p\left(T^{-1}\right)\right]$. As regards A_5, note that

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right) = \; & \frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right) + \frac{1}{nT}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{u}_{1-\lambda,it}^{*}\right) \\
& + (1-d_0)\, O_p\left(\frac{1}{\sqrt{T}}\right) + (1-d_1)\, O_p\left(\frac{1}{T^{3/2}}\right);
\end{aligned}
$$

again, the terms of order $O_p\left(T^{-1/2}\right)$ and $O_p\left(T^{-3/2}\right)$ arise from the BN decomposition. By the cross-sectional independence of $\bar{e}_{\lambda,it}^{*}$, it can be shown that the first term is $O_p\left(n^{-1/2}\right)$; the second one is $O_p\left(T^{-1}\right)$ by (8.9), so that $A_5 = (1-d_0)\, O_p\left(\phi_{nT}^{-1}\right) + (1-d_1)\, O_p\left(T^{-1}\right)$. Finally, turning to A_6, the variance of $n^{-1}\sum_{i=1}^{n}\left[\left(\beta_i - \beta\right)\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{2}\right)\right]$ is given by $\sigma_{\beta}^{2}\; n^{-2}\sum_{i=1}^{n} E\left[\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{2}\right)^2\right] = O\left(n^{-1}\right)$; hence, $A_6 = d_\sigma O_p\left(n^{-1/2}\phi_{nT}^{-1}\right)$. Putting all together, the Theorem follows. Note that the convergence rate for the case $\lambda = \sigma_{\beta}^{2} = 0$ can be derived following the same passages as above, and recalling that $\hat{\beta}-\beta=O_p\left(T^{-1}\right)$ by Proposition 1. □

Proof of Theorem 2. We start by considering the case $\lambda>0$. Based on the proof of Theorem 1, under the null that $\sigma_\beta^2=0$, the term that dominates in (7.1) is $\hat{\gamma}_3^{-1}\left(A_2+A_7\right)$, which is $O_p\left(n^{-1/2}\right)$. We have

$$\begin{aligned}\sqrt{n}\left(A_2+A_7\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[\frac{1}{T^2}\sum_{t=1}^{T}\bar{e}_{\lambda,it}^{*2}-\frac{1}{6}\sigma_{e,i}^2\right]+O_p\left(\frac{\sqrt{n}}{T^\eta}\right)+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor n\lambda\rfloor}Y_{aiT}+O_p\left(\frac{\sqrt{n}}{T^\eta}\right)+o_p(1).\end{aligned}$$

By direct calculation it follows that $E\left(Y_{aiT}\right)$ is $O_p\left(T^{-1}\right)$. Consider the sequence $\bar{Y}_{aiT}=Y_{aiT}-E\left(Y_{aiT}\right)$; by virtue of Assumption 1*(i)*, this is an *i.i.d.* sequence across i. Further, by direct calculation (see also Baltagi et al., 2008), it holds that $E\left(\bar{Y}_{aiT}^2\right)=\frac{\sigma_{e,i}^4}{45}$. Also, it can be shown that $E\left|\bar{Y}_{aiT}\right|^{2+\delta}<\infty$, by similar passages as in (8.16) and (8.17). This entails that equation (3.20) in Phillips and Moon (1999) holds; thus, by Theorem 2 of Phillips and Moon (1999), as $(n,T)\to\infty$ with $\frac{n}{T^{2\eta}}\to 0$, $\left(\sum_{i=1}^{\lfloor n\lambda\rfloor}\frac{\sigma_{e,i}^4}{45}\right)^{-1/2}$ $\sum_{i=1}^{\lfloor n\lambda\rfloor}\bar{Y}_{aiT}\xrightarrow{d}N(0,1)$. The consistency of $\hat{V}_\sigma$ holds by construction: since $E\left|\tilde{\sigma}_{e,i}^2-\sigma_{e,i}^2\right|^2=O\left(T^{-2\eta}\right)$ with $\sigma_{e,i}^2=0$ for $i=\lfloor n\lambda\rfloor+1$,..., n, we have

$$\hat{V}_\sigma=\frac{1}{45n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sigma_{e,i}^4+O_p\left(T^{-2\eta}\right). \tag{7.2}$$

Putting all together, (3.9) follows. Consider now the limiting distribution when $\sigma_\beta^2>0$. The passages above can be readily adapted to show that, in such case, $\sqrt{n}\hat{\sigma}_\beta^2=\sqrt{n}\sigma_\beta^2+O_p(1)+O_p\left(\sqrt{n}T^{-\eta}\right)$, so that

$$\sqrt{n}\hat{V}_\sigma^{-1/2}\left(A_{1,1}+A_2+A_7\right)=\sqrt{n}\gamma_3V_\sigma^{-1/2}\sigma_\beta^2+\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor n\lambda\rfloor}\bar{Y}_{aiT}+O_p\left(\frac{\sqrt{n}}{T^\eta}\right). \tag{7.3}$$

By construction, the random variable γ_3 is positive almost surely, so that $S_{nT}^{(\sigma)}\xrightarrow{p}\infty$ when $\sqrt{n}\sigma_\beta^2\to\infty$. This holds for any value of $\lambda\in[0,1]$.
Finally, consider the behaviour of $S_{nT}^{(\sigma)}$ under the null when $\lambda=0$. In this case, by Theorem 1, $\hat{\sigma}_\beta^2$ $=O_p\left(T^{-\eta}\right)$; also, by (7.2), under the assumption that $E\left|\tilde{\sigma}_{e,i}^2-\sigma_{e,i}^2\right|^2=O\left(T^{-2\eta'}\right)$, we have $\hat{V}_\sigma=$ $O_p\left(T^{-2\eta'}\right)$, so that $\hat{V}_\sigma^{-1/2}=O_p\left(T^{\eta'}\right)$. Therefore, $S_{nT}^{(\sigma)}=O_p\left(\sqrt{n}T^{\eta'-\eta}\right)$; thus, by (7.3), $S_{nT}^{(\sigma)}\to\infty$ when $\sigma_\beta^2>0$ and $(n,T)\to\infty$ with $\frac{\sqrt{n}}{T^{\eta'-\eta}}\to 0$; therefore $P\left[S_{nT}^{(\sigma)}>c_\alpha\right]=1$ for any $c_\alpha<\infty$. □

Proof of Theorem 3. The proof is similar to that of Theorem 1, and therefore passages will be omitted when possible to save space. We have

$$\hat{\lambda}=6\left\{\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{\bar{u}_{it}}{\hat{\sigma}_{e,i}}\right)^2+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\beta_i-\beta}{\hat{\sigma}_{e,i}}\right)^2\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\right.$$

$$+\left(\hat{\beta}-\beta\right)^2 \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{\sigma}_{e,i}^2}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)+2\left(\hat{\beta}-\beta\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\beta_i-\beta}{\hat{\sigma}_{e,i}}\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)$$

$$+2\left(\hat{\beta}-\beta\right)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{\sigma}_{e,i}}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\frac{\bar{u}_{it}}{\hat{\sigma}_{e,i}}\right)$$

$$-2\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\beta_i-\beta}{\hat{\sigma}_{e,i}}\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\frac{\bar{u}_{it}}{\hat{\sigma}_{e,i}}\right)-\hat{\sigma}_{\beta}^2\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{\sigma}_{e,i}^2}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\Bigg\}$$

$$= \Delta_{\lambda}^{-1}\left\{B_1+B_2+B_3+B_4+B_5+B_6+B_7\right\}. \tag{7.4}$$

Consider B_1

$$\begin{aligned}
B_1 &= \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{\bar{u}_{it}^*}{\hat{\sigma}_{e,i}}\right)^2+(1-d_0)\,O_p\left(\frac{1}{\sqrt{T}}\right)+(1-d_1)\,O_p\left(\frac{1}{T^{3/2}}\right)\\
&= \frac{1}{nT^2}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^{T}\left(\frac{\bar{e}_{\lambda,it}^*}{\hat{\sigma}_{e,i}}\right)^2+\frac{1}{nT^2}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\sum_{t=1}^{T}\left(\frac{\bar{u}_{1-\lambda,it}^*}{\hat{\sigma}_{e,i}}\right)^2\\
&\quad+(1-d_0)\,O_p\left(\frac{1}{\sqrt{T}}\right)+(1-d_1)\,O_p\left(\frac{1}{T^{3/2}}\right)\\
&= B_{12,1}+B_{12,2}+(1-d_0)\,O_p\left(\frac{1}{\sqrt{T}}\right)+(1-d_1)\,O_p\left(\frac{1}{T^{3/2}}\right),
\end{aligned}$$

where the term $(1-d_0)\,O_p\left(T^{-1/2}\right)+(1-d_1)\,O_p\left(T^{-3/2}\right)$ is the remainder from the BN decomposition. Consider $B_{12,1}$; it holds that

$$B_{12,1}=\frac{1}{nT^2}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^{T}\left(\frac{\bar{e}_{\lambda,it}^*}{\sigma_{e,i}}\right)^2+O_p\left(\frac{1}{T^{\varepsilon}}\right). \tag{7.5}$$

In order to explain the presence of the $O_p\left(T^{-\varepsilon}\right)$ term in (7.5), let M be a constant and note that

$$\begin{aligned}
&E\left|\frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{\bar{e}_{\lambda,it}^*}{\hat{\sigma}_{e,i}}\right)^2\right]-\frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{\bar{e}_{\lambda,it}^*}{\sigma_{e,i}}\right)^2\right]\right|\\
\leq\ & M\frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(\frac{\sigma_{e,i}^2-\hat{\sigma}_{e,i}^2}{\sigma_{e,i}^4}\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{e}_{\lambda,it}^{*2}\right)+o_p\left(T^{-\varepsilon}\right)\\
\leq\ & M\frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[E\left(\frac{\sigma_{e,i}^2-\hat{\sigma}_{e,i}^2}{\sigma_{e,i}^4}\right)^2\right]^{1/2}\left[E\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{e}_{\lambda,it}^{*2}\right)^2\right]^{1/2}=O_p\left(T^{-\varepsilon}\right)O_p(1),
\end{aligned}$$

by using the Cauchy-Schwartz inequality, and the fact that $E\left|\hat{\sigma}_{e,i}^2-\sigma_{e,i}^2\right|^2=O\left(T^{-2\varepsilon}\right)$. Hence

$$B_{12,1}=\frac{\lambda}{6}-\frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{\bar{e}_{\lambda,it}^{*}}{\sigma_{e,i}}\right)^2-\frac{1}{6}\right]+O_p\left(\frac{1}{T^{\varepsilon}}\right)=\frac{\lambda}{6}+O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\frac{1}{T^{\varepsilon}}\right);$$

where the $O_p\left(n^{-1/2}\right)$ term follows from the proof of Theorem 1. Therefore, $6\times B_{12,1}=\lambda+O_p\left(n^{-1/2}\right)$ $+\,O_p\left(T^{-\varepsilon}\right)$. As regards $B_{12,2}$, passages are very similar and therefore omitted when possible. Using the Cauchy-Schwartz inequality

$$E\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{\bar{u}_{1-\lambda,it}^{*}}{\hat{\sigma}_{e,i}}\right)^2\right]\leq\frac{T^{\varepsilon+H}}{T}\left\{E\left[\left(\frac{1}{T^{\varepsilon+H}\hat{\sigma}_{e,i}^2}\right)^2\right]\right\}^{1/2}\left\{E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{1-\lambda,it}^{*2}\right)^2\right]\right\}^{1/2};\qquad(7.6)$$

by Assumption 2*(i)* and by the fact that $T^{\varepsilon+H}\hat{\sigma}_{e,i}^2\to\infty$ as $T\to\infty$ for any arbitrarily small $H>0$, is $o_p\left(T^{\varepsilon+H-1}\right)$. Thus, $B_{12,2}=(1-d_1)\;o_p\left(T^{\varepsilon+H-1}\right)$. Turning to B_2+B_7, let $Z_{1iT}=T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^2$; we have:

$$\begin{aligned}B_2+B_7 &= \frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[(\beta_i-\beta)^2-\sigma_\beta^2\right]\left(\frac{Z_{1iT}}{\hat{\sigma}_{e,i}^2}\right)+\left(\hat{\sigma}_\beta^2-\sigma_\beta^2\right)\frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(\frac{Z_{1iT}}{\hat{\sigma}_{e,i}^2}\right)\\ &\quad+\frac{T^{\varepsilon+H}}{n}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\left[(\beta_i-\beta)^2-\sigma_\beta^2\right]\left(\frac{Z_{1iT}}{T^{\varepsilon+H}\hat{\sigma}_{e,i}^2}\right)+\left(\hat{\sigma}_\beta^2-\sigma_\beta^2\right)\frac{T^{\varepsilon+H}}{n}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\left(\frac{Z_{1iT}}{T^{\varepsilon+H}\hat{\sigma}_{e,i}^2}\right)\\ &= B_{2,1}+B_{2,2}+B_{2,3}+B_{2,4}.\end{aligned}$$

Consider $B_{2,1}$; in the proof of Theorem 1, the same term (modulo the presence of $\hat{\sigma}_{e,i}^2$) is called $A_{1,2}$, and we show that it is $d_\sigma\;(1-d_0)\;O_p\left(n^{-1/2}\right)$. As far as $B_{2,2}$ is concerned, its order of magnitude is the same as that of $\hat{\sigma}_\beta^2-\sigma_\beta^2$, whence we write compactly $B_{2,2}=(1-d_0)\;O_p\left(\left|\hat{\sigma}_\beta^2-\sigma_\beta^2\right|\right)$. Similar passages, in light of (7.6), yield $B_{2,3}=d_\sigma\,(1-d_1)\,o_p\left(T^{\varepsilon+H}n^{-1/2}\right)$ and $B_{2,4}=(1-d_1)\;o_p\left(T^{\varepsilon+H}\right)\;O_p\left(\left|\hat{\sigma}_\beta^2-\sigma_\beta^2\right|\right)$. Therefore $B_2=d_\sigma\;O_p\left(n^{-1/2}\right)\left[O_p\left(1\right)+\,o_p\left(T^{\varepsilon+H}\right)\right]+O_p\left(\left|\hat{\sigma}_\beta^2-\sigma_\beta^2\right|\right)\left[O_p\left(1\right)+\,o_p\left(T^{\varepsilon+H}\right)\right]$. Considering B_3, similar considerations as for B_2, and Proposition 1, entail $B_3=\left[(1-d_0)\,O_p\left(1\right)\,+\,(1-d_1)\,o_p\left(T^{\varepsilon+H}\right)\right]$ $O_p\left(\phi_{nT}^{-2}\right)$. Turning to B_4, define $Z_{2iT}=(\beta_i-\beta)\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)$; then $\frac{1}{2}B_4=n^{-1}\left(\hat{\beta}-\beta\right)\sum_{i=1}^{\lfloor n\lambda\rfloor}\frac{1}{\hat{\sigma}_{e,i}}Z_{2iT}$ $+\;n^{-1}\left(\hat{\beta}-\beta\right)\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\frac{1}{\hat{\sigma}_{e,i}}Z_{2iT}\;=\;B_{4,1}+B_{4,2}$. By the cross sectional independence holding by Assumption 3, the variance of $B_{4,1}$ is

$$\sum_{i=1}^{\lfloor n\lambda\rfloor}\frac{1}{\sigma_{e,i}^2}EZ_{2iT}^2+o\left(1\right),$$

so that $B_{4,1}=d_\sigma\;(1-d_0)\;O_p\left(n^{-1/2}\phi_{nT}^{-1}\right)$. Also, $B_{4,2}$ is bounded by the square root of $\left(\hat{\beta}-\beta\right)T^{\frac{\varepsilon+H}{2}}$ $n^{-1}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}E\left|\frac{1}{T^{\frac{\varepsilon+H}{2}}\hat{\sigma}_{e,i}}Z_{2iT}\right|$. Given that $E\left|\frac{1}{T^{\frac{\varepsilon+H}{2}}\hat{\sigma}_{e,i}}Z_{2iT}\right|\leq\left[E\left(T^{\varepsilon+H}\hat{\sigma}_{e,i}^2\right)^{-1}\right]^{1/2}\left[E\left(Z_{2iT}\right)^2\right]^{1/2}$ $=\;o_p\left(1\right)$, we have $B_{4,2}\;=\;d_\sigma\;\;(1-d_1)\;\;O_p\left(n^{-1/2}T^{(\varepsilon+H)/2}\phi_{nT}^{-1}\right)$, whence $B_4\;=\;d_\sigma\;\left[(1-d_0)\,O_p\left(1\right)\;+\right.$ $\left.(1-d_1)\,o_p\left(T^{(\varepsilon+H)/2}\right)\right]\;O_p\left(n^{-1/2}\phi_{nT}^{-1}\right)$. As regards B_5, let $Z_{3iT}\;=\;T^{-2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}$. As shown in Theorem 1, in view of Assumption 1 and Lemma 1, it holds that $E\left(Z_{3iT}\right)=O\left(T^{-1/2}\right)$ for $i=1,...,\lfloor n\lambda\rfloor$ and $E\left(Z_{3iT}\right)=O\left(T^{-3/2}\right)$ for $i=\lfloor n\lambda\rfloor+1,...,n$. Thence, $\frac{1}{2}B_5=n^{-1}\left(\hat{\beta}-\beta\right)\sum_{i=1}^{\lfloor n\lambda\rfloor}\frac{1}{\hat{\sigma}_{e,i}^2}Z_{3iT}+n^{-1}\left(\hat{\beta}-\beta\right)\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\frac{1}{\hat{\sigma}_{e,i}^2}Z_{3iT}=B_{5,1}+B_{5,2}$. Consider $B_{5,1}$:

$$\begin{aligned}B_{5,1} &= \frac{\left(\hat{\beta}-\beta\right)}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\frac{1}{\sigma_{e,i}^2}\left[Z_{3iT}-E\left(Z_{3iT}\right)\right]+\frac{\left(\hat{\beta}-\beta\right)}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\frac{1}{\sigma_{e,i}^2}E\left(Z_{3iT}\right)+\frac{\left(\hat{\beta}-\beta\right)}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(\frac{1}{\hat{\sigma}_{e,i}^2}-\frac{1}{\sigma_{e,i}^2}\right)Z_{3iT}\\ &= B_{5,1,1}+B_{5,1,2}+B_{5,1,3}.\end{aligned}$$

By the cross sectional independence of the $e_{\lambda,it}$s (Assumption 1), it follows that $B_{5,1,1}=(1-d_0)\,O_p\left(n^{-1/2}\phi_{nT}^{-1}\right)$; moreover, $B_{5,1,2}\leq\left(\hat{\beta}-\beta\right)\max_{1\leq i\leq\lfloor n\lambda\rfloor}\left|E\left(Z_{3iT}\right)\right|n^{-1}\sum_{i=1}^{\lfloor n\lambda\rfloor}\frac{1}{\sigma_{e,i}^2}=(1-d_0)\,O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Finally, $B_{5,1,3}$ is bounded by $\left(\hat{\beta}-\beta\right)\left[\left|E\left(Z_{3iT}\right)\right|^2\right]^{1/2}\left[E\left|\hat{\sigma}_{e,i}^2-\sigma_{e,i}^2\right|^2\right]^{1/2}=(1-d_0)\;O_p\left(T^{-\varepsilon}\phi_{nT}^{-1}\right)$; thus, $B_{5,1}=(1-d_0)\;O_p\left(\phi_{nT}^{-1}\right)\left[O_p\left(n^{-1/2}\right)+O_p\left(T^{-\varepsilon}\right)\right]$. As regards $B_{5,2}$, similar arguments as in the proof of $B_{3,2}$ yield that $B_{5,2}$ is bounded by

$$\frac{\left(\hat{\beta}-\beta\right)}{n}T^{\varepsilon+H}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}E\left|\frac{1}{T^{\varepsilon+H}\hat{\sigma}_{e,i}^2}Z_{3iT}\right|\leq\frac{\left(\hat{\beta}-\beta\right)}{n}T^{\varepsilon+H}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\sqrt{E\left|\frac{1}{T^{\varepsilon+H}\hat{\sigma}_{e,i}^2}\right|^2}\sqrt{E\left|Z_{3iT}\right|^2},$$

whence $B_{5,2}=(1-d_1)\;o_p\left(T^{\varepsilon+H-1}\phi_{nT}^{-1}\right)$. Thus, $B_5=(1-d_0)\;O_p\left(\phi_{nT}^{-1}\right)\left[O_p\left(n^{-1/2}\right)+O_p\left(T^{-\varepsilon}\right)\right]+(1-d_1)\;o_p\left(T^{\varepsilon+H-1}\right)\;O_p\left(\phi_{nT}^{-1}\right)$. Essentially the same passages yield $B_6=d_\sigma\;\left[(1-d_0)\,O_p\left(1\right)+(1-d_1)\,o_p\left(T^{\varepsilon+H}\right)\right]\;O_p\left(n^{-1/2}\right)$. Putting all together, the results follow. □

Proof of Theorem 4. Consider the expansion of $\hat{\lambda}^\dagger-\hat{b}_{nT}^{S\lambda}$

$$\begin{aligned}\hat{\lambda}^\dagger-\hat{b}_{nT}^{S\lambda} &= -\hat{\gamma}_3\times\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)^2+\hat{\gamma}_2\times\left(\frac{1}{nT}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{it}^2-\frac{1}{nT}\sum_{i=1}^{n}\hat{\sigma}_{u,i}^2\right)\\ &-\hat{\gamma}_3\times\frac{1}{n}\sum_{i=1}^{n}\left(\beta_i-\beta\right)^2\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2+\hat{\gamma}_2\times\frac{1}{n}\sum_{i=1}^{n}\left(\beta_i-\beta\right)^2\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\\ &-\left(\hat{\beta}-\beta\right)^2\left[\hat{\gamma}_3\times\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2-\hat{\gamma}_2\times\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\right]\\ &-2\left(\hat{\beta}-\beta\right)\hat{\gamma}_3\times\frac{1}{n}\sum_{i=1}^{n}\left(\beta_i-\beta\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2+2\hat{\gamma}_3\times\frac{1}{n}\sum_{i=1}^{n}\left(\beta_i-\beta\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)\\ &-2\left(\hat{\beta}-\beta\right)\hat{\gamma}_3\times\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)+2\left(\hat{\beta}-\beta\right)\hat{\gamma}_2\times\frac{1}{n}\sum_{i=1}^{n}\left(\beta_i-\beta\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\\ &-2\hat{\gamma}_2\times\frac{1}{n}\sum_{i=1}^{n}\left(\beta_i-\beta\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)+2\left(\hat{\beta}-\beta\right)\hat{\gamma}_2\times\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)\end{aligned}$$

$$= C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7 + C_8 + C_9 + C_{10} + C_{11}. \tag{7.7}$$

The proof of the theorem is based on very similar arguments as in the previous proofs, and passages are omitted where possible. By (8.6), $C_1 = O_p\left(T^{-2}\right)$. As regards C_2, note that $(nT)^{-1} \sum_{i=1}^n \left(T^{-1} \sum_{t=1}^T \bar{u}_{it}^2\right) - (nT)^{-1} \sum_{i=1}^n \hat{\sigma}_{u,i}^2 = (nT)^{-1} \sum_{i=1}^n \left(T^{-1} \sum_{t=1}^T \bar{u}_{it}^2\right) - (nT)^{-1} \sum_{i=1}^n T^{-1} \sum_{t=1}^T \hat{u}_{it}^2$, and consider $(nT)^{-1} \sum_{i=1}^n T^{-1} \sum_{t=1}^T \left(u_{it}^2 - \hat{u}_{it}^2\right) = n^{-1} \sum_{i=1}^n \left(\hat{\beta}_i - \beta_i\right)^2 \left(T^{-2} \sum_{t=1}^T x_{it}^2\right) -2\,(nT)^{-1} \sum_{i=1}^n \left(\hat{\beta}_i - \beta_i\right) \left(T^{-1} \sum_{t=1}^T x_{it} u_{it}\right) = C_{2,1} + C_{2,2}$. Under $\lambda = 0$, Assumption 2 entails that $E\left|\hat{\beta}_i - \beta_i\right|^4 = O_p\left(T^{-4}\right)$, so that $C_{2,1}$ is bounded by $\left[E\left(\hat{\beta}_i - \beta_i\right)^4\right]^{1/2} \left[E\left(T^{-2} \sum_{t=1}^T x_{it}^2\right)^2\right]^{1/2} = O\left(T^{-2}\right)$; similarly, $C_{2,2}$ is bounded by $T^{-1} \left[E\left(\hat{\beta}_i - \beta_i\right)^2\right]^{1/2} \left[E\left(T^{-1} \sum_{t=1}^T x_{it} u_{it}\right)^2\right]^{1/2} = O\left(T^{-2}\right)$. Therefore, $C_2 = O_p\left(T^{-2}\right)$. By construction, $C_5 = 0$. Essentially the same arguments as in the proofs of the previous theorems entail $C_6 = d_\sigma O_p\left(n^{-1/2} \phi_{nT}^{-1}\right)$, and the same for C_9; similarly, C_7 and C_8 (and C_{10}, C_{11}) are $d_\sigma O_p\left(n^{-1/2} T^{-1}\right)$ and $O_p\left(T^{-1} \phi_{nT}^{-1}\right)$ respectively. Thus, under $H_0 : \lambda = 0$, the term that dominates is $C_3 + C_4 = O_p\left(n^{-1/2}\right)$; indeed

$$\begin{aligned} C_3 + C_4 &= \frac{1}{n} \sum_{i=1}^n \left(\beta_i - \beta\right)^2 \left[\hat{\gamma}_2 \left(\frac{1}{T^2} \sum_{t=1}^T \bar{x}_{it}^2\right) - \hat{\gamma}_3 \left(\frac{1}{T^2} \sum_{t=1}^T \bar{x}_{it}^2\right)^2\right] \\ &= \frac{1}{n} \sum_{i=1}^n \left[\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right] Z_{4iT} + \sigma_\beta^2 \frac{1}{n} \sum_{i=1}^n \left[\hat{\gamma}_2 \left(\frac{1}{T^2} \sum_{t=1}^T \bar{x}_{it}^2\right) - \hat{\gamma}_3 \left(\frac{1}{T^2} \sum_{t=1}^T \bar{x}_{it}^2\right)^2\right] + O_p\left(\frac{1}{n}\right), \end{aligned}$$

where $Z_{4iT} = \gamma_2 \left(T^{-2} \sum_{t=1}^T \bar{x}_{it}^2\right) - \gamma_3 \left(T^{-2} \sum_{t=1}^T \bar{x}_{it}^2\right)^2$. By construction, the second term is identically zero; the $O_p\left(n^{-1}\right)$ term arises from the fact that $\hat{\gamma}_2 - \gamma_2$ and $\hat{\gamma}_3 - \gamma_3$ are both $O_p\left(n^{-1/2}\right)$. Thus, the asymptotic distribution of $\sqrt{n}\left(C_3 + C_4\right)$ is determined by $n^{-1/2} \sum_{i=1}^n \left[\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right] Z_{4iT}$. By Assumption 3, $\left[\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right] Z_{4iT}$ has mean zero, and, conditional upon $\mathcal{F}$, it is independent across i, for all T. A Liapunov condition (conditional upon $\mathcal{F}$) holds, since for every i, $E_{\mathcal{F}} \left|\left[\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right] Z_{4iT}\right|^{2+\delta} = E\left|\left[\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right]\right|^{2+\delta} E_{\mathcal{F}} \left|Z_{4iT}\right|^{2+\delta}$, which is almost surely finite by Assumption 3 and by adapting (8.16)-(8.17). Also, $E_{\mathcal{F}} \left[\left(\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right) Z_{4iT}\right]^2 = E\left[\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right]^2 E_{\mathcal{F}}\left(Z_{4iT}^2\right)$; under $\sigma_\beta^2 > 0$ and Assumption 2*(ii)*, this is almost surely greater than zero for all i and uniformly in T, so that, as $(n, T) \to \infty$, $\sum_{i=1}^n E\left[\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right]^2 E_{\mathcal{F}}\left(Z_{4iT}^2\right) > 0$ almost surely. The proof is now almost identical to that of Proposition 1: a conditional version of the Lindeberg CLT can be applied with

$$\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\left(\beta_i - \beta\right)^2 - \sigma_\beta^2\right] Z_{4iT}}{\sqrt{\left(\kappa_\beta - \sigma_\beta^4\right) \frac{1}{n} \sum_{i=1}^n E_{\mathcal{F}}\left(Z_{4iT}^2\right)}} \xrightarrow{d} Z, \tag{7.8}$$

where $Z \sim N(0,1)$ independent of $\mathcal{F}$; the definition of V_λ comes directly from (7.8). Under H_0, the β_is are estimated superconsistently for each i, and $\hat{\sigma}_{\beta}^{2}$ is consistent by Theorem 1; thus, after very similar passages as in the rest of the paper (particularly, see the proof of equation (8.13)), it can be shown that $\hat{V}_{\lambda} \xrightarrow{p} V_{\lambda}$. The desired result follows from putting everything together.

If $\sigma_{\beta}^{2}=0$, in (7.7) it holds that $C_3=C_4=C_6=C_7=C_9=C_{10}=0$; also, recall that when $\lambda=\sigma_{\beta}^{2}=0$, we have $\hat{\beta}-\beta=O_p\left(T^{-1}\right)$ by Proposition 1; combining this with the same arguments as in the proof of Theorem 3, we have $C_5=C_8=C_{11}=O_p\left(T^{-2}\right)$. These results and the previous ones entail that when $\lambda=\sigma_{\beta}^{2}=0$, it holds that $\hat{\lambda}^{\dagger}-\hat{b}_{nT}^{S\lambda}=O_p\left(T^{-2}\right)$. Recalling that, when $\lambda=\sigma_{\beta}^{2}=0$, $\hat{\sigma}_{\beta}^{2}=O_p\left(T^{-\eta}\right)$, it can be readily shown that $\hat{V}_{\lambda}$ is bounded by $O_p\left(T^{-2\eta}\right)$. Therefore, under $\lambda=\sigma_{\beta}^{2}=0$, $\tilde{S}_{nT}^{(\lambda)}=\sqrt{n}\,O_p\left(T^{-2}\right)$ $O_p\left(T^{\eta}\right)=o_p\left(1\right)$ as $\left(n,T\right)\rightarrow\infty$ and $\frac{\sqrt{n}}{T^{2-\eta}}\rightarrow 0$. This entails that, for any $c_{\alpha}>0$, $P\left[\tilde{S}_{nT}^{(\lambda)}>c_{\alpha}\right]=0$ under H_0.

Under the alternative, similar passages as in the proof of Theorem 3 and of Proposition 1 yield that the term that dominates is $\sqrt{n}\lambda\left(\frac{\gamma_2}{6}\sigma_e^2-\gamma_1\gamma_3\right)V_{\lambda}^{-1/2}+\sqrt{n}\left(C_3+C_4\right)$, whence (3.18). □

8. Appendix B: further proofs and results

We start by reporting two preliminary Lemmas.

Lemma 1. *Let Assumptions 1 and 2 hold. As $(n,T)\to\infty$, it holds that:*

$$\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^2 = \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^{*2}+O_p\left(\frac{1}{\sqrt{T}}\right)=O_p(1), \tag{8.1}$$

$$\frac{1}{nT^4}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2 = \frac{1}{nT^4}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)^2+O_p\left(\frac{1}{\sqrt{T}}\right) = O_p(1), \tag{8.2}$$

$$\frac{1}{(n\lambda)T^2}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^{T}\bar{u}_{\lambda,it}^2 = \frac{1}{(n\lambda)T^2}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^{T}\bar{e}_{\lambda,it}^{*2}+O_p\left(\frac{1}{\sqrt{T}}\right) = O_p(1), \tag{8.3}$$

$$\frac{1}{[n(1-\lambda)]T^2}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\sum_{t=1}^{T}\bar{u}_{1-\lambda,it}^2 = O_p\left(\frac{1}{T}\right), \tag{8.4}$$

$$\frac{1}{(n\lambda)T^4}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{\lambda,it}\right)^2 = \frac{1}{(n\lambda)T^4}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)^2+O_p\left(\frac{1}{\sqrt{T}}\right) = O_p(1), \tag{8.5}$$

$$\frac{1}{[n(1-\lambda)]T^4}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\left(\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{1-\lambda,it}\right)^2 = O_p\left(\frac{1}{T^2}\right), \tag{8.6}$$

$$\frac{1}{nT^2}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}g_i'\bar{h}_t\right)^2 = O_p\left(\frac{1}{T}\right), \tag{8.7}$$

$$\frac{1}{nT^4}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\bar{x}_{it}g_i'\bar{h}_t\right)^2 = O_p\left(\frac{1}{T^2}\right). \tag{8.8}$$

Proof of Lemma 1. Equation (8.1) follows, under Assumption 2, by the proof of Lemma 13 in Phillips and Moon (1999), where, letting $R_{iT}=T^{-2}\left(\sum_{t=1}^{T}\bar{x}_{it}^2-\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)$, it is shown that $E\left|R_{iT}\right|=O\left(T^{-1/2}\right)$ for each i; Assumption 2*(i)* also ensures that $E\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)=O(1)$. Equation (8.2) follows from the same logic.

Turning to (8.3), note that, for all $i=1,...,\lfloor n\lambda\rfloor$, $T^{-2}\sum_{t=1}^{T}\left[\bar{u}_{\lambda,it}\right]^2 - T^{-2}\sum_{t=1}^{T}\left[\bar{e}_{\lambda,it}\right]^2 = T^{-2}\sum_{t=1}^{T}\left(\sum_{j=-\infty}^{+\infty}c_{ij}\Delta x_{it+j}\right)^2 + 2T^{-2}\sum_{t=1}^{T}\bar{e}_{\lambda,it}\left(\sum_{j=-\infty}^{+\infty}c_{ij}\Delta x_{it+j}\right) = I + II$. As regards I, it is bounded by $T^{-2}\sum_{t=1}^{T}\left(\sum_{j=-\infty}^{+\infty}c_{ij}^2\right)E\left(\Delta x_{it+j}\right)^2 = O\left(T^{-1}\right)$ for each i, by Assumption 2*(i)* and 2*(iii)*. Similarly, using the Cauchy-Schwartz inequality, II is bounded by $\left\{T^{-2}\sum_{t=1}^{T}E\left(\bar{e}_{\lambda,it}\right)^2\right\}^{1/2}$ $\left\{T^{-2}\sum_{t=1}^{T}E\left(\sum_{j=-\infty}^{+\infty}c_{ij}\Delta x_{it+j}\right)^2\right\}^{1/2}$. Similarly to (8.1), Assumption 2 yields $E\left|T^{-2}\sum_{t=1}^{T}E\left(\bar{e}_{\lambda,it}^2\right)\right.$

$- T^{-2}\sum_{t=1}^{T} E\left(\bar{e}_{\lambda,it}^{*2}\right)\Big| = O\left(T^{-1/2}\right)$; by virtue of Assumption 2*(i)*, $E\left(\bar{e}_{\lambda,it}^{*2}\right) = O\left(t\right)$, whence $T^{-2}\sum_{t=1}^{T} E\left(\bar{e}_{\lambda,it}^{*2}\right)$ $= O\left(1\right)$. Hence, $T^{-2}\sum_{t=1}^{T} E\left(\bar{e}_{\lambda,it}^{2}\right) = O\left(1\right)$. Also, as shown above, $T^{-2}\sum_{t=1}^{T} E\left(\sum_{j=-\infty}^{+\infty} c_{ij}\Delta x_{it+j}\right)^2$ $= O\left(T^{-1}\right)$, which entails that $E\left(II\right) = O\left(T^{-1/2}\right)$; this is not the sharpest bound, but it suffices for our purposes. Putting all together, (8.3) follows.

As regards (8.4), by stationarity and Assumption 2*(i)* we have $T^{-1}\sum_{t=1}^{T} E\left(\bar{u}_{1-\lambda,it}^{2}\right) = O\left(1\right)$ for all i, from which the desired result follows. Considering (8.5), note first that, for all i, $T^{-4}\left(\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{\lambda,it}\right)^2 -$ $T^{-4}\left(\sum_{t=1}^{T}\bar{x}_{it}\bar{e}_{\lambda,it}\right)^2 = T^{-4}\left[\sum_{t=1}^{T}\bar{x}_{it}\left(\sum_{j=-\infty}^{+\infty} c_{ij}\Delta x_{it+j}\right)\right]^2 + 2T^{-4}\left(\sum_{t=1}^{T}\bar{x}_{it}\bar{e}_{\lambda,it}\right)\sum_{t=1}^{T}\bar{x}_{it}\left(\sum_{j=-\infty}^{+\infty} c_{ij}\Delta x_{it+j}\right)$ $= I + II$. Consider I; using the Cauchy-Schwartz inequality, (8.2) and the proof of (8.3), it follows that $E\left(I\right) = O\left(T^{-1/2}\right)$. Also, considering II, $E\left|T^{-4}\left(\sum_{t=1}^{T}\bar{x}_{it}\bar{e}_{\lambda,it}\right)\ \sum_{t=1}^{T}\bar{x}_{it}\left(\sum_{j=-\infty}^{+\infty} c_{ij}\Delta x_{it+j}\right)\right|$ $\leq \left\{E\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}\bar{e}_{\lambda,it}\right)^2\right\}^{1/2}\ \left\{E\left[T^{-2}\sum_{t=1}^{T}\bar{x}_{it}\left(\sum_{j=-\infty}^{+\infty} c_{ij}\Delta x_{it+j}\right)\right]^2\right\}^{1/2}$. Assumption 2 entails $E\left|\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}\bar{e}_{\lambda,it}\right)^2 - \left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)^2\right| = O\left(T^{-1/2}\right)$. Also, by using Assumption 2*(iv)*, $E\left|\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)^2\right|$ $\leq T^{-2}\sum_{t=1}^{T} E\left(\bar{x}_{it}^{*}\right)^2 E\left(\bar{e}_{\lambda,it}^{*}\right)^2$. Given that, by Assumption 2*(i)*, $E\left(\bar{x}_{it}^{*2}\right)$ and $E\left(\bar{e}_{\lambda,it}^{*2}\right)$ are both $O\left(t\right)$, it follows that $E\left|\left(T^{-2}\sum_{t=1}^{T}\ \bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)^2\right| = O\left(1\right)$. Therefore, $E\left(II\right) = O\left(T^{-1/2}\right)$. Hence, (8.5) follows.

Turning to (8.6), Assumption 2 implies $E\left|\left(T^{-1}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{1-\lambda,it}\right)^2 - \left(T^{-1}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{u}_{1-\lambda,it}^{*}\right)^2\right| = O\left(T^{-1/2}\right)$. Further, $E\left(T^{-1}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{u}_{1-\lambda,it}^{*}\right)^2 = T^{-2}\ \sum_{t=1}^{T}\ \sum_{s=1}^{T}\ E\left(\bar{x}_{it}^{*}\bar{x}_{is}^{*}\right)\ E\left(\bar{u}_{1-\lambda,it}^{*}\bar{u}_{1-\lambda,is}^{*}\right) = T^{-2}\ \sum_{t=1}^{T}$ $E\left(\bar{x}_{it}^{*2}\right)\ E\left(\bar{u}_{1-\lambda,it}^{*2}\right) = O\left(1\right)$. Hence, (8.6) follows. Finally, (8.7) and (8.8) can be shown using the same approach as for (8.4) and (8.6) respectively. □

Lemma 2. *Let Assumptions 1-3 hold. As $(n, T) \to \infty$, it holds that:*

$$\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{1-\lambda,it} = O_p\left(nT\right), \tag{8.9}$$

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}g_i'\bar{h}_t = O_p\left(nT\right), \tag{8.10}$$

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\left[\left(\beta_i-\beta\right)\bar{x}_{it}+\bar{u}_{\lambda,it}d_{\lambda,i}\right] = \left(d_\sigma+1-d_0\right)O_p\left(\sqrt{n}T^2\right) + \left(1-d_0\right)O_p\left(nT^{3/2}\right). \tag{8.11}$$

Proof of Lemma 2. Consider (8.9). By the proof of (8.6), it follows that $E\left|T^{-2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{1-\lambda,it}\right| =$ $O\left(T^{-1}\right)$, which yields (8.9). The same logic yields (8.10). Turning to (8.11), $\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\left[\left(\beta_i-\beta\right)\bar{x}_{it}\right.$ $\left.+\ \bar{u}_{\lambda,it}d_{\lambda,i}\right] = \sum_{i=1}^{n}\left(\beta_i-\beta\right)\left(\sum_{t=1}^{T}\bar{x}_{it}^{2}\right) + \sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{\lambda,it}d_{\lambda,i} = I + II$. Consider I; using Assumption 3, it is bounded by the square root of $T^4\sum_{i=1}^{n} E\left[\left(\beta_i-\beta\right)^2\right] E\left[\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{2}\right)^2\right]$. Equation (8.2) entails $E\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{2}\right)^{2}=O\left(1\right)$ for each i, so that $I=O_{p}\left(\sqrt{n}T^{2}d_{\sigma}\right)$. Considering II, note that, by (8.5) and (8.6), $II=\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}+O_{p}\left[nT^{3/2}\left(1-d_{0}\right)\right]+O_{p}\left(nT\right)$. By Assumption 1*(i)*, the first term is bounded by the square root of $\sum_{i=1}^{\lfloor n\lambda\rfloor}E\left(\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)^{2}$; since $E\left(\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)^{2}=O\left(T^{2}\right)$ for every i in view of (8.5), $\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}=O_{p}\left(\sqrt{n}T^{2}\right)$. Putting all together, (8.11) follows. □

Proof of Proposition 1. We start by reporting a result on the rate of convergence of $\hat{\beta}$ when $\left(\lambda,\sigma_{\beta}^{2}\right)\neq(0,0)$. Recall that $\hat{\beta}-\beta=\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^{2}\right]^{-1}\left\{\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\left[\bar{u}_{it}+\left(\beta_{i}-\beta\right)\bar{x}_{it}\right]\right\}$. By equation (8.1), $\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^{2}=O_{p}\left(nT^{2}\right)$; also, by Lemma 2, $\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\left[\bar{u}_{it}+\left(\beta_{i}-\beta\right)\bar{x}_{it}\right]=O_{p}\left(\sqrt{n}T^{2}\right)+O_{p}\left(nT^{3/2}\right)$ when either λ or σ_{β}^{2} are different from zero. Thus

$$\hat{\beta}-\beta=O_{p}\left(\frac{1}{\sqrt{n}}\right)+O_{p}\left(\frac{1}{\sqrt{T}}\right)=O_{p}\left(\phi_{nT}^{-1}\right). \tag{8.12}$$

In order to prove (2.6) and (2.7), we start by showing that, as $(n,T)\to\infty$

$$\frac{1}{nT^{2}}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^{2}-\frac{1}{n}\sum_{i=1}^{n}E_{\mathcal{F}}\left(\int\bar{W}_{xi}^{2}\right)=o_{p}\left(1\right). \tag{8.13}$$

The proof is based on the same logic as in Phillips and Moon (1999), with the only difference that in our context the x_{it}s are not independent, but conditionally independent. As a preliminary result, recall that, by (8.1), $\left(nT^{2}\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^{2}=\left(nT^{2}\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^{*2}+o_{p}\left(1\right)$ and let $Q_{0iT}=T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}$. Conditional on $\mathcal{F}$, this is an independent sequence across i for all T. As $T\to\infty$, Assumption 2 and the Continuous Mapping Theorem entail that $Q_{0iT}^{2}\overset{d}{\to}Q_{0i}^{2}=\left(\int\bar{W}_{xi}^{2}\right)^{2}$ (see e.g. Phillips and Durlauf, 1986). Also, based on Assumption 2*(i)*, it can be verified by direct calculation that $E\left(Q_{0iT}^{2}\right)\to E\left(Q_{0i}^{2}\right)$ as $T\to\infty$. This entails that Q_{0iT} is uniformly square integrable in T (see Theorem 5.4 in Billingsley, 2013). By adapting, to the conditional case, the definition of joint convergence in probability (see Phillips and Moon, 1999), $Q_{0nT}=n^{-1}\sum_{i=1}^{n}Q_{0iT}$ converges jointly in probability to $\mu_{Q}=n^{-1}\sum_{i=1}^{n}E_{\mathcal{F}}\left(Q_{0i}\right)$ conditional on $\mathcal{F}$ if and only if $P\left[\left|Q_{0nT}-\mu_{Q}\right|>\varepsilon\right|\mathcal{F}]=0$ almost surely as $(n,T)\to\infty$ for an arbitrary $\mathcal{F}$ -measurable random variable ε, positive almost surely. By Chebychev's inequality, it suffices to show that $E_{\mathcal{F}}\left(Q_{0nT}-\mu_{Q}\right)^{2}\to0$ almost surely. Let $\mu_{Q,T}=n^{-1}\sum_{i=1}^{n}E_{\mathcal{F}}\left(Q_{0iT}\right)$; we have $E_{\mathcal{F}}\left(Q_{0nT}-\mu_{Q}\right)^{2}\leq2E_{\mathcal{F}}\left(Q_{0nT}-\mu_{Q,T}\right)^{2}+2\left(\mu_{Q,T}-\mu_{Q}\right)^{2}=I+II$. Consider I; by conditional independence, $I=n^{-2}\sum_{i=1}^{n}E_{\mathcal{F}}\left(Q_{0iT}-E_{\mathcal{F}}\left(Q_{0iT}\right)\right)^{2}$; since $E\left(Q_{0iT}^{2}\right)<\infty$ (see the proof of (8.17) below), $E_{\mathcal{F}}\left(Q_{0iT}-E_{\mathcal{F}}\left(Q_{0iT}\right)\right)^{2}$ is finite almost surely, so that it can be shown that $I\to0$ almost surely. Also, by the square integrability of Q_{0iT} (and, therefore, of $E_{\mathcal{F}}\left(Q_{0iT}\right)$), it follows that $II\to0$ almost surely.

We now turn to proving (2.6) and (2.7). Considering the numerators of the two expressions, we have

$$\begin{aligned}
&\frac{1}{nT^4}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\bar{x}_{it}\hat{v}_{it}\right)^2 \\
=& \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)^2+\frac{1}{n}\sum_{i=1}^{n}\left[(\beta_i-\beta)^2\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2\right]+\left(\hat{\beta}-\beta\right)^2\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2 \\
&+2\frac{1}{n}\sum_{i=1}^{n}\left[(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)\right]-2\left(\hat{\beta}-\beta\right)\frac{1}{n}\sum_{i=1}^{n}\left[(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)^2\right] \\
&-2\left(\hat{\beta}-\beta\right)\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)\right] \qquad (8.14)\\
=& a_1+a_2+a_3+a_4+a_5+a_6,
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2 \\
=& \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{u}_{it}^2+\frac{1}{nT^2}\sum_{i=1}^{n}\left[(\beta_i-\beta)^2\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\right]+\left(\hat{\beta}-\beta\right)^2\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right) \\
&+2\frac{1}{n}\sum_{i=1}^{n}\left[(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right)\right]-2\left(\hat{\beta}-\beta\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}\right) \\
&-2\left(\hat{\beta}-\beta\right)\frac{1}{n}\sum_{i=1}^{n}\left[(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^2\right)\right] \qquad (8.15)\\
=& b_1+b_2+b_3+b_4+b_5+b_6.
\end{aligned}$$

Consider (8.14). We have

$$a_1=\frac{1}{nT^4}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[\sum_{t=1}^{T}\bar{x}_{it}\left(\bar{u}_{\lambda,it}+g_i'\bar{h}_t\right)\right]^2+\frac{1}{nT^4}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\left[\sum_{t=1}^{T}\bar{x}_{it}\left(\bar{u}_{1-\lambda,it}+g_i'\bar{h}_t\right)\right]^2=a_{1,1}+a_{1,2}.$$

By (8.5), (8.6) and (8.8), $a_{1,1}=O_p(1)$ and $a_{1,2}=O_p\left(T^{-2}\right)$. As regards the limit of $a_{1,1}$, it can be derived by using similar arguments as in the proof of (8.13). The term that dominates is $n^{-1}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)^2$. Let $Q_{1iT}=\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)^2$; conditional on $\mathcal{F}$, this is an independent sequence across i for all T. As $T\to\infty$, Assumption 2 entails that $Q_{1iT}^2\xrightarrow{d}Q_{1i}^2=\left(\int\bar{W}_{xi}\bar{W}_{ei}\right)^4$ (see e.g. Phillips and Durlauf (1986)). Assumption 2*(i)* and standard calculations yield $E\left(Q_{1iT}^2\right)\to E\left(Q_{1i}^2\right)$ as $T\to\infty$. This entails that Q_{1iT} is a uniformly square integrable sequence in T. Using the same passages as in the proof of (8.13), it follows that

$$a_{1,1}-\frac{1}{n}\sum_{i=1}^{\lfloor n\lambda\rfloor}E_{\mathcal{F}}\left(\int\bar{W}_{xi}\bar{W}_{ei}\right)^{2}=o_{p}\left(1\right).$$

Consider now a_2; the term that dominates is $Q_{2iT}=(\beta_i-\beta)^2\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)^2$. Conditional on $\mathcal{F}$, Q_{2iT} is independent across i with finite, nonzero mean as long as $\sigma_\beta^2\neq 0$, for all T. By Assumptions 2 and 3, $Q_{2iT}^2\xrightarrow{d}Q_{2i}^2=(\beta_i-\beta)^2\left(\int\bar{W}_{xi}^2\right)^2$; further, standard calculations yield $E\left(Q_{2iT}^2\right)\xrightarrow{d}E\left(Q_{2i}^2\right)$, so that Q_{2iT} is a uniformly square integrable sequence in T. Again, by the same logic as for the proof of (8.13), $a_2-\sigma_\beta^2\,n^{-1}\sum_{i=1}^{n}E_{\mathcal{F}}\left(\int\bar{W}_{xi}^2\right)^2=o_p(1)$. Turning to a_3, by (8.2) and (8.12), $a_3=O_p\left(\phi_{nT}^{-2}\right)$. As regards a_4, let $a_4=n^{-1}\sum_{i=1}^{n}Q_{3iT}$; by Assumption 3, $E\left(Q_{3iT}\right)=0$ and $E\left(Q_{3iT}Q_{3jT}\right)=0$ for all $i\neq j$. Therefore, $E\left[\left(n^{-1}\sum_{i=1}^{n}Q_{3iT}\right)^2\right]=n^{-2}\sum_{i=1}^{n}E\left[\left(Q_{3iT}\right)^2\right]$. It holds that $E\left[\left(Q_{3iT}\right)^2\right]<\infty$; the proof follows similar lines as that of (8.16)-(8.17) below for details on how to prove this. Thus, $a_4=O_p\left(n^{-1/2}\right)$. Similar passages yield $a_5=O_p\left(n^{-1/2}\phi_{nT}^{-1}\right)$. Finally, turning to a_6, by by (8.1), (8.5), (8.6) and (8.8) we have $a_6=n^{-1}\sum_{i=1}^{n}Q_{4iT}+O_p\left(T^{-1/2}\right)$, with $Q_{4iT}=\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)\left(T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right)$. By Assumptions 1 and 2*(i)*, $E\left(Q_{4iT}\right)=0$ and $E\left(Q_{4iT}Q_{4jT}\right)=0$ for all $i\neq j$, whence similar passages as for a_4 yield $a_6=O_p\left(\phi_{nT}^{-2}\right)$. Putting all together, (2.6) is verified.

Turning to (2.7), passages are fairly similar to those for the proof of (2.6) and omitted when possible. Consider (8.15); we have

$$b_1=\frac{1}{nT^4}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{t=1}^{T}\left(\bar{u}_{\lambda,it}+g_i'\bar{h}_t\right)^2+\frac{1}{nT^4}\sum_{i=\lfloor n\lambda\rfloor+1}^{n}\sum_{t=1}^{T}\left(\bar{u}_{1-\lambda,it}+g_i'\bar{h}_t\right)^2=b_{1,1}+b_{1,2}.$$

As $(n,T)\to\infty$, (8.4) and (8.7) entail $b_{1,2}=O_p\left(T^{-1}\right)$. As regards $b_{1,1}$, by (8.3) and (8.7) the term that dominates is $n^{-1}\sum_{i=1}^{\lfloor n\lambda\rfloor}T^{-2}\sum_{t=1}^{T}\left(\bar{e}_{\lambda,it}^{*}\right)^2$. Under Assumption 2, Lemma 13 in Phillips and Moon (1999) can be applied directly, with $b_{1,1}\xrightarrow{p}\lambda\lim_{n\to\infty}(n\lambda)^{-1}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sigma_{e,i}^2/6$. Similar passages as for a_2 entail $b_2-\sigma_\beta^2\,n^{-1}\sum_{i=1}^{n}E_{\mathcal{F}}\left(\int\bar{W}_{xi}^2\right)=o_p(1)$. Equations (8.1) and (8.12) entail $b_3=O_p\left(\phi_{nT}^{-2}\right)$. As regards b_4, b_5 and b_6, the same arguments as above yield $b_4=O_p\left(n^{-1/2}\right)$, $b_5=O_p\left(\phi_{nT}^{-2}\right)$ and $b_6=O_p\left(n^{-1/2}\phi_{nT}^{-1}\right)$. Putting all together, (2.7) follows.

Finally, we turn to proving (2.8). The denominator is studied in (8.13). As regards the numerator, by (8.9) and (8.10), $\left(\sqrt{n}T^2\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\left[\bar{u}_{it}+(\beta_i-\beta)\bar{x}_{it}\right]=\left(\sqrt{n}T^2\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}$ $\left[(\beta_i-\beta)\bar{x}_{it}+\bar{u}_{\lambda,it}d_{\lambda,i}\right]+O_p\left(\frac{\sqrt{n}}{T}\right)$; also, by similar arguments as in the proof of Lemma 1, $\left(\sqrt{n}T^2\right)^{-1}$ $\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}\left[(\beta_i-\beta)\bar{x}_{it}+\bar{u}_{\lambda,it}d_{\lambda,i}\right]=\left(\sqrt{n}T^2\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{x}_{it}^{*}\left[(\beta_i-\beta)\bar{x}_{it}^{*}+\bar{e}_{\lambda,it}^{*}d_{\lambda,i}\right]+O_p\left(\sqrt{\frac{n}{T}}\right)$. Let $Q_{5iT}=T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\left[(\beta_i-\beta)\bar{x}_{it}^{*}+\bar{e}_{\lambda,it}^{*}d_{\lambda,i}\right]$. We prove a conditional CLT for the sequence Q_{5iT}; this is based on following the same passages as in the proof of Theorem 2 in Phillips and Moon (1999), with the only difference of having a conditional version of the Lindeberg condition. First, by Assumptions 1-3, $E\left(Q_{5iT}\right)=0$ and the Q_{5iT}s are independent across i for all T conditional on $\mathcal{F}$. Further, a Liapunov condition holds conditionally on $\mathcal{F}$; indeed, for every i, $E_{\mathcal{F}}\left|Q_{5iT}\right|^{2+\delta}\leq 2^{1+\delta}\,E\left|\beta_i-\beta\right|^{2+\delta}$ $E_{\mathcal{F}}\left|T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right|^{2+\delta}+2^{1+\delta}\,E_{\mathcal{F}}\left|T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right|^{2+\delta}$, after using the C_r-inequality, the Cauchy-

Schwartz inequality and Assumption 3, which also ensures that $E\left|\beta_i-\beta\right|^{2+\delta}<\infty$. For every i

$$E\left|\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right|^{2+\delta}\leq\left[\frac{1}{T^{3+\delta}}\sum_{t=1}^{T}E\left|\bar{x}_{it}^{*}\right|^{4+2\delta}\right]^{1/2}\left[\frac{1}{T^{3+\delta}}\sum_{t=1}^{T}E\left|\bar{e}_{\lambda,it}^{*}\right|^{4+2\delta}\right]^{1/2}, \tag{8.16}$$

by convexity and the Cauchy-Schwartz inequality. By Assumption 2*(i)*, $E\left|\bar{e}_{\lambda,it}^{*}\right|^{4+2\delta}=O\left(t^{2+\delta}\right)$, whence $T^{-(3+\delta)}\sum_{t=1}^{T}E\left|\bar{e}_{\lambda,it}^{*}\right|^{4+2\delta}<\infty$; also, for every i

$$\frac{1}{T^{3+\delta}}\sum_{t=1}^{T}E\left|\bar{x}_{it}^{*}\right|^{4+2\delta}\leq\left(E\left|l_i\right|^{4+2\delta}\right)\frac{2^{1+\delta}}{T^{3+\delta}}\sum_{t=1}^{T}E\left|\bar{f}_{t}^{*}\right|^{4+2\delta}+\frac{2^{1+\delta}}{T^{3+\delta}}\sum_{t=1}^{T}E\left|\bar{w}_{it}^{*}\right|^{4+2\delta}=I+II; \tag{8.17}$$

again, both I and II can be shown to be finite on account of Assumption 2*(i)*, in the same way as (8.16). Equations (8.16) and (8.17) entail that $E_{\mathcal{F}}\left|T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right|^{2+\delta}$ and $E_{\mathcal{F}}\left|T^{-2}\sum_{t=1}^{T}\bar{x}_{it}^{*}\bar{e}_{\lambda,it}^{*}\right|^{2+\delta}$ are both finite almost surely. Finally, Assumption 2*(ii)* ensures that $E_{\mathcal{F}}\left(Q_{5iT}^2\right)>0$ almost surely for all i, uniformly in T, so that $\sum_{i=1}^{n}E_{\mathcal{F}}\left(Q_{5iT}^2\right)>0$ almost surely for large n. Thus, the conditional version of the Lindeberg version of the CLT can be applied (see Billingsley, 2008, Theorem 27.2, for a proof of the Lindeberg CLT; see also Theorem 7 in Rao, 2009, and the comments thereafter for the conditional version of the CLT); the only difference is that limits are taken as $(n,T)\to\infty$, with $T\to\infty$ being incidental to the main argument of the proof. Hence, as $(n,T)\to\infty$

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Q_{5iT}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}E_{\mathcal{F}}\left(Q_{5iT}^2\right)}}\xrightarrow{d}Z,$$

with $Z\sim N(0,1)$ independent of $\mathcal{F}$. By the proof of (2.6), as $(n,T)\to\infty$, we have

$$\frac{1}{n}\sum_{i=1}^{n}E_{\mathcal{F}}\left(Q_{5iT}^2\right)-\left\{\lambda\frac{1}{n\lambda}\sum_{i=1}^{\lfloor n\lambda\rfloor}E_{\mathcal{F}}\left[\left(\int\bar{W}_{xi}\bar{W}_{ei}\right)^2\right]+\sigma_\beta^2\frac{1}{n}\sum_{i=1}^{n}E_{\mathcal{F}}\left[\left(\int\bar{W}_{xi}^2\right)^2\right]\right\}=o_p(1),$$

which completes the proof of (2.8). Finally, when $\left(\lambda,\sigma_\beta^2\right)=(0,0)$, $\hat{\beta}-\beta=O_p\left(T^{-1}\right)$ by virtue of (8.9), (8.10) and (8.13). □

Proof of Theorem 6. . The consistency (and the speed of convergence) of $\check{\lambda}$ can be shown in the same way as in the proof of Theorem 3, after showing the corresponding results for $\check{\Sigma}_\beta$. In turn, the only part of the proof of this which differs from Theorem 1 is to show that $\hat{\Gamma}_1^v=\Gamma_1^v+O_p\left(T^{-\eta}\right)$, where Γ_1^v is the limit of $n^{-1}\sum_{i=1}^{n}E_{\mathcal{F}}\left(\int\bar{W}_{xi}\bar{W}_{ei}\right)\left(\int\bar{W}_{xi}\bar{W}_{ei}\right)'d_{\lambda,i}$. Letting $G_{i,T}=T^{-2}\sum_{t=1}^{T}\bar{x}_{it}\bar{u}_{it}^{*}$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\sigma}_{e,i}^2G_{i,T}G_{i,T}'=\frac{1}{n}\sum_{i=1}^{n}\sigma_{e,i}^2G_{i,T}G_{i,T}'d_{\lambda,i}+\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\sigma}_{e,i}^2-\sigma_{e,i}^2\right)G_{i,T}G_{i,T}',$$

and by the same arguments as above, we have that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\sigma}_{e,i}^{2}-\sigma_{e,i}^{2}\right)G_{i,T}G_{i,T}^{\prime}=d_{0}O_{p}\left(T^{-\eta}\right).$$

Also

$$\frac{1}{n}\sum_{i=1}^{n}\sigma_{e,i}^{2}G_{i,T}G_{i,T}^{\prime}d_{\lambda,i}\overset{d}{\rightarrow}\frac{1}{n}\sum_{i=1}^{n}\sigma_{e,i}^{2}E_{\mathcal{F}}\left(\int\bar{W}_{xi}\bar{W}_{ui}^{*}\right)\left(\int\bar{W}_{xi}\bar{W}_{ui}^{*}\right)^{\prime}d_{\lambda,i},$$

where W_{ui}^{*} is the Brownian motion associated to u_{it}^{*}. Conditional on $\mathcal{F}$, it holds that $\int\bar{W}_{xi}\bar{W}_{ei}d_{\lambda,i}$ has the same distribution as $\sigma_{e,i}\int\bar{W}_{xi}\bar{W}_{ui}^{*}d_{\lambda,i}$, so that

$$\frac{1}{n}\sum_{i=1}^{n}\sigma_{e,i}^{2}G_{i,T}G_{i,T}^{\prime}d_{\lambda,i}\overset{d}{\rightarrow}\frac{1}{n}\sum_{i=1}^{n}E_{\mathcal{F}}\left(\int\bar{W}_{xi}\bar{W}_{ei}\right)\left(\int\bar{W}_{xi}\bar{W}_{ei}\right)^{\prime}d_{\lambda,i}$$

obtains immediately. The desired result now follows from adapting the proof of Theorem 1. □

Limiting distribution of $\hat{\sigma}_{\beta}^{2}$ and $\hat{\lambda}$

We present two ancillary results that complement the consistency results in Theorems 1 and 3: the limiting distributions of $\hat{\sigma}_{\beta}^{2}$ (Theorem 7) and of $\hat{\lambda}$ (Theorem 8).

Theorem 7. *Let Assumptions 1-3 hold. As $(n,T)\rightarrow\infty$ with $\frac{n}{T^{2\eta}}\rightarrow 0$, it holds that, for $\sigma_{\beta}^{2}>0$*

$$\sqrt{n}\left(\hat{\sigma}_{\beta}^{2}-\sigma_{\beta}^{2}\right)\overset{d}{\rightarrow}\gamma_{3}^{-1}\times V_{\sigma\beta}^{1/2}\times Z, \tag{8.18}$$

where $V_{\sigma\beta}=V_{\sigma}+(1-d_{0})\,V_{\sigma\beta,1}$, with V_{σ} defined in (3.8) and

$$V_{\sigma\beta,2}=\left(\kappa_{\beta}-\sigma_{\beta}^{4}\right)\gamma_{2}+4\sigma_{\beta}^{2}\gamma_{1},$$

$\kappa_{\beta}=E\left[(\beta_{i}-\beta)^{4}\right]$, and $Z\sim N(0,1)$ independent of $\mathcal{F}$.

Theorem 8. *Let the assumptions of Theorem 3 hold. As $(n,T)\rightarrow\infty$ with $\frac{n}{T^{2\varepsilon}}\rightarrow 0$, it holds that, for $\sigma_{\beta}^{2}=0$ and having set $\lambda_{hc}=0$ in (3.10)*

$$\sqrt{n}\left(\hat{\lambda}-\lambda\right)\overset{d}{\rightarrow}N\left(0,\frac{4}{5}\lambda\right). \tag{8.19}$$

In both cases, results are presented for the unrestricted case. However, the two parameters are restricted, since $\sigma_{\beta}^{2}\geq 0$ and $\lambda\in[0,1]$. Thus, we propose the following truncated estimators: $\hat{\sigma}_{\beta,trunc}^{2}=\hat{\sigma}_{\beta}^{2}\times I_{[0,+\infty)}\left(\hat{\sigma}_{\beta}^{2}\right)$ and $\hat{\lambda}_{trunc}=\hat{\lambda}\times I_{[0,1]}\left(\hat{\lambda}\right)$. Let v_{σ}^{2} and v_{λ}^{2} denote the asymptotic variances defined in (8.18) and (8.19) respectively. Under the assumptions of the two Theorems, the densities of the two limiting distributions are given by

$$f_{\hat{\sigma}^2_\beta}(x) = \sqrt{\frac{n}{2\pi v^2_\sigma}}\left[1-\Phi\left(-\frac{\sqrt{n}\sigma^2_\beta}{v_\sigma}\right)\right]^{-1}\exp\left\{-\frac{1}{2}\frac{n\left(x-\sigma^2_\beta\right)^2}{v^2_\sigma}\right\}I_{[0,+\infty)}(x),$$

$$f_{\hat{\lambda}}(x) = \sqrt{\frac{n}{2\pi v^2_\lambda}}\left[\Phi\left(\frac{\sqrt{n}(1-\lambda)}{v_\lambda}\right)-\Phi\left(-\frac{\sqrt{n}\lambda}{v_\lambda}\right)\right]^{-1}\exp\left\{-\frac{1}{2}\frac{n(x-\lambda)^2}{v^2_\lambda}\right\}I_{[0,1]}(x),$$

where Φ denotes the cumulative distribution of the standard normal - see e.g. Johnson et al. (1995).

Proof of Theorem 7. Based on (7.1) and the passages thereafter, as $(n,T)\to\infty$ with $\frac{n}{T^{2\eta}}\to 0$, the terms that dominate have magnitude $O_p\left(n^{-1/2}\right)$, and they are A_1, A_2+A_7, and A_4; thus, the limiting distribution of $\sqrt{n}\left(\hat{\sigma}^2_\beta-\sigma^2_\beta\right)$ is given by $\gamma_3^{-1}\ n^{-1/2}\sum_{i=1}^n\left(\bar{Y}_{aiT}+Y_{biT}+Y_{ciT}\right)+O_p\left(\frac{\sqrt{n}}{T^\eta}\right)$, where

$$Y_{biT} = 2(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^T\bar{x}^*_{it}\bar{e}^*_{\lambda,it}\right),$$

$$Y_{ciT} = \left[(\beta_i-\beta)^2-\sigma^2_\beta\right]\left(\frac{1}{T^2}\sum_{t=1}^T\bar{x}^{*2}_{it}\right),$$

and $\bar{Y}_{aiT}$ is defined in the proof of Theorem 2. In order to prove a CLT for $\sqrt{n}\left(\hat{\sigma}^2_\beta-\sigma^2_\beta\right)$, let $Y_{\beta iT}=\phi_a\bar{Y}_{aiT}+\phi_bY_{biT}+\phi_cY_{ciT}$ for some nonzero ϕ_a, ϕ_b and ϕ_c.and consider $n^{-1/2}\sum_{i=1}^n Y_{\beta iT}$. The sequence $Y_{\beta iT}$ has mean zero and is independent across i, conditional on $\mathcal{F}$, for all T. Also, a Liapunov condition holds conditional on $\mathcal{F}$. This can be shown by noting that $E\left|Y_{\beta iT}\right|^{2+\delta}\leq 2^{1+\delta}\left(\left|\phi_a\right|^{2+\delta}E\left|\bar{Y}_{aiT}\right|^{2+\delta} + \left|\phi_b\right|^{2+\delta}E\left|Y_{biT}\right|^{2+\delta}+\left|\phi_c\right|^{2+\delta}\ E\left|Y_{ciT}\right|^{2+\delta}\right)$. In the proof of Theorem 2, it is shown that $E\left|\bar{Y}_{aiT}\right|^{2+\delta}<\infty$; by Assumption 3 and (8.16)-(8.17), it can similarly be shown that $E\left|Y_{biT}\right|^{2+\delta}<\infty$ and $E\left|Y_{ciT}\right|^{2+\delta}<\infty$; hence, $E\left|Y_{\beta iT}\right|^{2+\delta}<\infty$ almost surely. Finally, as in the previous proofs, Assumption 2*(ii)* ensures that the asymptotic variance is non zero. Thus, as $(n,T)\to\infty$ under $\frac{n}{T^{2\eta}}\to 0$

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_{\beta iT}}{\sqrt{V_{\sigma\beta,abc}}}\xrightarrow{d} Z, \tag{8.20}$$

where $Z\sim N(0,1)$ independent of $\mathcal{F}$ and

$$V_{\sigma\beta,abc}=\frac{1}{n}\sum_{i=1}^n E_{\mathcal{F}}\left[\phi_a^2\bar{Y}^2_{aiT}+(1-d_0)\left(\phi_b^2Y^2_{biT}+\phi_c^2Y^2_{ciT}\right)\right]$$
$$+2(1-d_0)\frac{1}{n}\sum_{i=1}^n E_{\mathcal{F}}\left(\phi_a\phi_c\bar{Y}_{aiT}Y_{ciT}+\phi_a\phi_b\bar{Y}_{aiT}Y_{biT}+\phi_b\phi_bY_{biT}Y_{ciT}\right).$$

By Assumption 3, $E_{\mathcal{F}}\left(\bar{Y}_{aiT}Y_{biT}\right)=E_{\mathcal{F}}\left(\bar{Y}_{aiT}Y_{ciT}\right)=0$. Also, Assumption 2*(iv)* entails that $E_{\mathcal{F}}\left(Y_{biT}Y_{ciT}\right)=0$. Putting all together, (8.18) follows.

We now analyse the asymptotic bias of $\hat{\sigma}_\beta^2$. Among the higher order terms in (7.1), there are some terms of order $O_p\left(n^{-1}\right)$, namely A_3 and A_6. Consider the following component of $A_3 + A_6$

$$-\hat{\gamma}_3^{-2}\left[\frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)\right]^2 = -\gamma_3^{-2}\left[\frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)\right]^2 + o_p\left(\frac{1}{n}\right);$$

we have

$$\begin{aligned}
& E_{\mathcal{F}}\left\{\frac{1}{\gamma_3^2}\left[\frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)\right]^2\right\} \\
= & \frac{1}{\gamma_3^2}E_{\mathcal{F}}\left[\gamma_3^{-2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\beta_i-\beta)(\beta_j-\beta)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{jt}^{*2}\right)\right] \\
= & \frac{1}{n^2\gamma_3^2}\sum_{i=1}^{n}E_{\mathcal{F}}\left[(\beta_i-\beta)^2\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)^2\right] \\
= & \left(\frac{\sigma_\beta^2}{n\gamma_3^2}\right)\frac{1}{n}\sum_{i=1}^{n}E_{\mathcal{F}}\left[\left(\frac{1}{T^2}\sum_{t=1}^{T}\bar{x}_{it}^{*2}\right)^2\right];
\end{aligned}$$

this can be estimated by $n^{-1}\hat{\sigma}_\beta^2\hat{\gamma}_2\hat{\gamma}_3^{-2}$, whence the bias correction. □

Proof of Theorem 8. Passages are similar to those in the proof of Theorem 7, and thus we only outline the main idea of the proof. Consider (7.4); when $\sigma_\beta^2 = 0$ and $\hat{\sigma}_\beta^2$ is set to zero, the only terms that are present are B_1 and B_5 (which is dominated). Thus, the asymptotics of $\sqrt{n}\left(\hat{\lambda}-\lambda\right)$ is governed by

$$\sqrt{n}\left(\hat{\lambda}-\lambda\right) = 6\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{\bar{e}_{\lambda,it}^{*}}{\sigma_{e,i}}\right)^2-\frac{1}{6}\right]+O_p\left(\frac{\sqrt{n}}{T^{\varepsilon}}\right).$$

We have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{\bar{e}_{\lambda,it}^{*}}{\sigma_{e,i}}\right)^2-\frac{1}{6}\right] = \sqrt{\lambda}\frac{1}{\sqrt{n\lambda}}\sum_{i=1}^{\lfloor n\lambda\rfloor}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{\bar{e}_{\lambda,it}^{*}}{\sigma_{e,i}}\right)^2-\frac{1}{6}\right]\xrightarrow{d}\sqrt{\lambda}\times N\left(0,\frac{1}{45}\right),$$

which follows from the same passages as in the proof of Theorem 2. Putting all together, the theorem follows. □

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 817–858.

Bai, J., C. Kao, and S. Ng (2009). Panel cointegration with global stochastic trends. *Journal of Econometrics 149*(1), 82–99.

Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica 72*(4), 1127–1177.

Balke, N. S. and T. B. Fomby (1997). Threshold cointegration. *International Economic Review*, 627–645.

Balke, N. S. and M. E. Wohar (1998). Nonlinear dynamics and covered interest rate parity. *Empirical Economics 23*(4), 535–559.

Baltagi, B. H., C. Kao, and L. Liu (2008). Asymptotic properties of estimators for the linear panel regression model with random individual effects and serially correlated errors: the case of stationary and non-stationary regressors and residuals. *The Econometrics Journal 11*(3), 554–572.

Billingsley, P. (2008). *Probability and measure.* John Wiley & Sons.

Billingsley, P. (2013). *Convergence of probability measures.* John Wiley & Sons.

Choi, C.-Y., L. Hu, and M. Ogaki (2008). Robust estimation for structural spurious regressions and a Hausman-type cointegration test. *Journal of Econometrics 142*(1), 327–351.

Coakley, J., R. Flood, A.-M. Fuertes, and M. P. Taylor (2005). Long-run purchasing power parity and general relativity. *Journal of International Money and Finance 28*(2), 293–316.

Coakley, J., A.-M. Fuertes, and F. Spagnolo (2004). Is the Feldstein-Horioka puzzle history? *The Manchester School 72*(5), 569–590.

Fuertes, A.-M. (2008). Sieve bootstrap t-tests on long-run average parameters. *Computational Statistics & Data Analysis 52*(7), 3354–3370.

Gengenbach, C., F. C. Palm, and J.-P. Urbain (2006). Cointegration testing in panels with common factors. *Oxford Bulletin of Economics and Statistics 68*(1), 683–719.

Giannone, D. and M. Lenza (2009). The Feldstein-Horioka fact. Technical report, National Bureau of Economic Research.

Gregory, A. W. and B. E. Hansen (1996). Practitioners corner: tests for cointegration in models with regime and trend shifts. *Oxford bulletin of Economics and Statistics 58*(3), 555–560.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous univariate distributions.* Wiley, New York.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics 90*(1), 1–44.

Kapetanios, G. (2003). Determining the poolability of individual series in panel datasets. Technical report.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics 160*(2), 326–348.

Krueger, A. B. and M. Lindahl (2001). Education for growth: Why and for whom? *Journal of Economic Literature 39*(4), 1101–1136.

Levy, D., M. Bergen, S. Dutta, and R. Venable (1997). The magnitude of menu costs: direct evidence from large us supermarket chains. *The Quarterly Journal of Economics 112*(3), 791–824.

Lo, M. C. and E. Zivot (2001). Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics 5*(4), 533–576.

Mark, N. C. and D. Sul (2003). Cointegration vector estimation by panel dols and long-run money demand. *Oxford Bulletin of Economics and Statistics 65*(5), 655–680.

McCoskey, S. and C. Kao (1998). A residual-based test of the null of cointegration in panel data. *Econometric Reviews 17*(1), 57–84.

Ng, S. (2008). A simple test for nonstationarity in mixed panels. *Journal of Business & Economic Statistics 26*(1), 113–127.

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. *Econometric Theory 20*(3), 597–625.

Pesaran, M. H., L. V. Smith, and T. Yamagata (2013). Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics 175*(2), 94–115.

Pesaran, M. H. and T. Yamagata (2008). Testing slope homogeneity in large panels. *Journal of Econometrics 142*(1), 50–93.

Phillips, P. C. and S. N. Durlauf (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies 53*(4), 473–495.

Phillips, P. C. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*(5), 1057–1111.

Phillips, P. C. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics*, 971–1001.

Phillips, P. C. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal 6*(1), 217–259.

Rao, B. P. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics 61*(2), 441–460.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory 7*(1), 1–21.

Taylor, A. M. (2001). Potential pitfalls for the purchasing-power-parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica 69*(2), 473–498.

Trapani, L. (2012). On the asymptotic t-test for large nonstationary panel models. *Computational*

Statistics & Data Analysis 56(11), 3286–3306.

Trapani, L. (2013). On bootstrapping panel factor series. *Journal of Econometrics 172*(1), 127–141.

Wagner, M. (2008). On PPP, unit roots and panels. *Empirical Economics 35*(2), 229–249.

Westerlund, J. and W. Hess (2011). A new poolability test for cointegrated panels. *Journal of Applied Econometrics 26*(1), 56–88.

	(n,T)	(20, 100)			(20, 200)			(20, 400)		
			σ^2_β			σ^2_β			σ^2_β	
(ρ,θ)	λ	0	0.25	0.50	0	0.25	0.50	0	0.25	0.50
(0, 0)	0	−0.173	−0.173	−0.178	−0.126	−0.128	−0.128	−0.101	−0.100	−0.101
		(0.177)	(0.179)	(0.186)	(0.129)	(0.131)	(0.145)	(0.103)	(0.103)	(0.129)
	0.25	−0.095	−0.094	−0.097	−0.076	−0.068	−0.081	−0.066	−0.061	−0.070
		(0.147)	(−0.152)	(0.225)	(0.131)	(0.144)	(0.152)	(0.133)	(0.131)	(0.141)
	0.50	−0.030	−0.021	−0.027	−0.022	−0.011	−0.010	−0.027	−0.035	−0.047
		(0.164)	(0.177)	(0.189)	(0.161)	(0.161)	(0.205)	(0.154)	(0.168)	(0.206)
	0.75	0.039	0.035	0.035	0.031	0.026	0.034	0.012	0.017	0.012
		(0.196)	(0.208)	(0.208)	(0.205)	(0.205)	(0.257)	(0.197)	(0.198)	(0.288)
(0.75, 0)	0	−0.173	−0.173	−0.178	−0.126	−0.128	−0.128	−0.101	−0.100	−0.101
		(0.177)	(0.179)	(0.186)	(0.129)	(0.131)	(0.145)	(0.103)	(0.103)	(0.129)
	0.25	−0.210	−0.207	−0.209	−0.176	−0.170	−0.181	−0.153	−0.148	−0.156
		(0.240)	(0.243)	(0.298)	(0.210)	(0.214)	(0.225)	(0.193)	(0.190)	(0.200)
	0.50	−0.104	−0.095	−0.108	−0.092	−0.090	−0.095	−0.082	−0.076	−0.083
		(0.181)	(0.191)	(0.232)	(0.185)	(0.184)	(0.231)	(0.181)	(0.177)	(0.198)
	0.75	−0.007	0.033	0.024	0.000	−0.018	−0.020	−0.008	−0.004	−0.022
		(0.184)	(0.238)	(0.233)	(0.200)	(0.194)	(0.293)	(0.187)	(0.217)	(0.269)
(0, 0.75)	0	−0.255	−0.253	−0.256	−0.189	−0.187	−0.185	−0.152	−0.154	−0.146
		(0.260)	(0.258)	(0.265)	(0.193)	(0.192)	(0.198)	(0.155)	(0.159)	(0.163)
	0.25	−0.162	−0.151	−0.149	−0.118	−0.118	−0.122	−0.103	−0.103	−0.102
		(0.197)	(0.192)	(0.205)	(0.167)	(0.165)	(0.227)	(0.153)	(0.161)	(0.173)
	0.50	−0.072	−0.054	−0.073	−0.049	−0.065	−0.058	−0.060	−0.062	−0.047
		(0.174)	(0.173)	(0.227)	(0.164)	(0.166)	(0.207)	(0.152)	(0.171)	(0.183)
	0.75	0.025	0.030	0.034	0.009	−0.003	−0.001	−0.018	−0.003	−0.011
		(0.189)	(0.206)	(0.254)	(0.186)	(0.196)	(0.244)	(0.178)	(0.209)	(0.262)
(0, −0.75)	0	−0.287	−0.285	−0.291	−0.207	−0.205	−0.204	−0.163	−0.163	−0.163
		(0.293)	(0.291)	(0.305)	(0.211)	(0.210)	(0.217)	(0.166)	(0.168)	(0.169)
	0.25	−0.184	−0.178	−0.187	−0.132	−0.134	−0.139	−0.114	−0.118	−0.099
		(0.219)	(0.219)	(0.235)	(0.173)	(0.181)	(0.189)	(0.165)	(0.161)	(0.183)
	0.50	−0.087	−0.080	−0.082	−0.077	−0.068	−0.072	−0.052	−0.066	−0.064
		(0.182)	(0.190)	(0.236)	(0.166)	(0.178)	(0.205)	(0.164)	(0.172)	(0.186)
	0.75	0.019	0.020	0.013	0.019	0.019	−0.001	−0.005	−0.016	−0.007
		(0.171)	(0.205)	(0.268)	(0.186)	(0.201)	(0.247)	(0.182)	(0.202)	(0.239)

Table 3a: bias and RMSE for $\hat{\sigma}^2_{\beta,bc}$

(ρ,θ)	λ	$(50,50)$ σ^2_β = 0	0.25	0.50	$(50,100)$ σ^2_β = 0	0.25	0.50	$(50,200)$ σ^2_β = 0	0.25	0.50	$(100,200)$ σ^2_β = 0	0.25	0.50
$(0,0)$	0	-0.165	-0.166	-0.165	-0.162	-0.160	-0.164	-0.117	-0.118	-0.116	-0.095	-0.096	-0.096
		(0.167)	(0.168)	(0.170)	(0.163)	(0.161)	(0.167)	(0.118)	(0.120)	(0.122)	(0.116)	(0.117)	(0.123)
	0.25	-0.078	-0.077	-0.075	-0.090	-0.092	-0.082	-0.067	-0.073	-0.069	-0.054	-0.054	-0.054
		(0.101)	(0.104)	(0.118)	(0.110)	(0.114)	(0.131)	(0.093)	(0.099)	(0.115)	(0.078)	(0.082)	(0.088)
	0.50	0.027	0.028	0.017	-0.024	-0.019	-0.015	-0.019	-0.025	-0.012	-0.011	-0.009	-0.009
		(0.089)	(0.104)	(0.138)	(0.097)	(0.098)	(0.110)	(0.094)	(0.101)	(0.109)	(0.065)	(0.068)	(0.084)
	0.75	0.109	0.113	0.109	0.042	0.048	0.055	0.027	0.033	0.016	0.015	0.021	0.014
		(0.161)	(0.165)	(0.176)	(0.120)	(0.130)	(0.162)	(0.109)	(0.123)	(0.141)	(0.086)	(0.094)	(0.111)
$(0.75,0)$	0	-0.165	-0.166	-0.165	-0.162	-0.160	-0.164	-0.117	-0.118	-0.116	-0.095	-0.096	-0.096
		(0.167)	(0.168)	(0.170)	(0.163)	(0.161)	(0.167)	(0.118)	(0.120)	(0.122)	(0.116)	(0.117)	(0.123)
	0.25	-0.148	-0.147	-0.143	-0.196	-0.198	-0.188	-0.162	-0.167	-0.166	-0.145	-0.135	-0.135
		(0.162)	(0.165)	(0.172)	(0.208)	(0.211)	(0.216)	(0.175)	(0.181)	(0.190)	(0.162)	(0.164)	(0.167)
	0.50	-0.022	-0.022	-0.011	-0.083	-0.091	-0.083	-0.081	-0.077	-0.081	-0.057	-0.060	-0.059
		(0.102)	(0.100)	(0.128)	(0.127)	(0.132)	(0.164)	(0.120)	(0.123)	(0.137)	(0.101)	(0.104)	(0.111)
	0.75	0.079	0.088	0.089	0.017	0.009	0.014	0.000	-0.002	-0.002	0.001	0.002	0.005
		(0.134)	(0.151)	(0.171)	(0.111)	(0.122)	(0.144)	(0.109)	(0.117)	(0.137)	(0.081)	(0.079)	(0.099)
$(0,0.75)$	0	-0.228	-0.230	-0.230	-0.234	-0.235	-0.234	-0.175	-0.176	-0.172	-0.164	-0.165	-0.165
		(0.230)	(0.233)	(0.235)	(0.236)	(0.238)	(0.241)	(0.176)	(0.178)	(0.180)	(0.195)	(0.196)	(0.200)
	0.25	-0.124	-0.121	-0.120	-0.149	-0.153	-0.148	-0.117	-0.107	-0.118	-0.098	-0.095	-0.097
		(0.141)	(0.142)	(0.148)	(0.163)	(0.170)	(0.171)	(0.134)	(0.127)	(0.143)	(0.117)	(0.117)	(0.123)
	0.50	-0.016	-0.012	-0.009	-0.047	-0.060	-0.060	-0.046	-0.048	-0.042	-0.039	-0.038	-0.035
		(0.095)	(0.092)	(0.108)	(0.101)	(0.113)	(0.139)	(0.102)	(0.106)	(0.133)	(0.077)	(0.079)	(0.102)
	0.75	0.102	0.094	0.094	0.022	0.028	0.030	0.018	0.026	0.011	0.011	0.017	0.010
		(0.152)	(0.153)	(0.174)	(0.111)	(0.118)	(0.144)	(0.109)	(0.130)	(0.142)	(0.080)	(0.088)	(0.117)
$(0,-0.75)$	0	-0.288	-0.288	-0.283	-0.267	-0.268	-0.267	-0.191	-0.193	-0.192	-0.180	-0.179	-0.178
		(0.291)	(0.291)	(0.290)	(0.269)	(0.270)	(0.273)	(0.193)	(0.194)	(0.197)	(0.211)	(0.210)	(0.211)
	0.25	-0.172	-0.170	-0.169	-0.175	-0.174	-0.164	-0.121	-0.128	-0.124	-0.110	-0.110	-0.112
		(0.184)	(0.185)	(0.195)	(0.181)	(0.188)	(0.198)	(0.134)	(0.145)	(0.161)	(0.129)	(0.130)	(0.136)
	0.50	-0.046	-0.040	-0.041	-0.071	-0.067	-0.073	-0.053	-0.051	-0.055	-0.044	-0.038	-0.043
		(0.099)	(0.107)	(0.131)	(0.119)	(0.114)	(0.150)	(0.110)	(0.116)	(0.127)	(0.082)	(0.082)	(0.095)
	0.75	0.083	0.074	0.090	0.021	0.021	0.018	0.007	0.009	0.021	0.005	0.008	0.009
		(0.135)	(0.135)	(0.174)	(0.116)	(0.115)	(0.143)	(0.112)	(0.117)	(0.150)	(0.079)	(0.087)	(0.099)

Table 3b: bias and RMSE for $\hat{\sigma}^2_{\beta,bc}$

(n,T)		(20, 100)			(20, 200)			(20, 400)		
			λ			λ			λ	
(ρ,θ)	σ_β^2	0	0.25	0.75	0	0.25	0.75	0	0.25	0.75
(0, 0)	0	0.334	0.205	−0.044	0.226	0.118	−0.076	0.166	0.076	−0.092
		(0.334)	(0.211)	(0.077)	(0.226)	(0.126)	(0.098)	(0.166)	(0.085)	(0.106)
	0.25	0.325	0.201	−0.050	0.221	0.108	−0.077	0.160	0.068	−0.097
		(0.326)	(0.207)	(0.081)	(0.221)	(0.118)	(0.098)	(0.160)	(0.081)	(0.107)
	0.50	0.314	0.133	−0.067	0.200	0.099	−0.092	0.144	0.063	−0.130
		(0.314)	(0.154)	(0.096)	(0.210)	(0.110)	(0.115)	(0.144)	(0.076)	(0.147)
(0.75, 0)	0	0.495	0.339	0.018	0.344	0.218	−0.035	0.265	0.155	−0.055
		(0.496)	(0.342)	(0.068)	(0.345)	(0.221)	(0.066)	(0.265)	(0.158)	(0.072)
	0.25	0.489	0.332	0.009	0.339	0.210	−0.034	0.260	0.150	−0.061
		(0.490)	(0.336)	(0.065)	(0.340)	(0.213)	(0.063)	(0.260)	(0.153)	(0.077)
	0.50	0.477	0.293	−0.002	0.323	0.202	−0.068	0.247	0.141	−0.082
		(0.478)	(0.299)	(0.068)	(0.324)	(0.205)	(0.095)	(0.247)	(0.144)	(0.098)
(0, 0.75)	0	0.331	0.221	−0.029	0.223	0.127	−0.062	0.164	0.081	−0.083
		(0.331)	(0.223)	(0.062)	(0.223)	(0.131)	(0.078)	(0.164)	(0.085)	(0.093)
	0.25	0.326	0.215	−0.033	0.218	0.123	−0.062	0.160	0.078	−0.092
		(0.326)	(0.218)	(0.068)	(0.218)	(0.126)	(0.081)	(0.160)	(0.082)	(0.101)
	0.50	0.302	0.209	−0.052	0.213	0.099	−0.076	0.148	0.064	−0.104
		(0.303)	(0.212)	(0.087)	(0.214)	(0.107)	(0.098)	(0.148)	(0.069)	(0.115)
(0, −0.75)	0	0.346	0.235	−0.018	0.232	0.136	−0.058	0.170	0.085	−0.084
		(0.347)	(0.237)	(0.057)	(0.232)	(0.139)	(0.078)	(0.170)	(0.089)	(0.093)
	0.25	0.341	0.230	−0.023	0.228	0.134	−0.068	0.166	0.080	−0.091
		(0.342)	(0.232)	(0.063)	(0.228)	(0.137)	(0.086)	(0.166)	(0.083)	(0.101)
	0.50	0.327	0.217	−0.045	0.213	0.122	−0.080	0.151	0.080	−0.104
		(0.328)	(0.220)	(0.081)	(0.214)	(0.125)	(0.097)	(0.152)	(0.084)	(0.113)

Table 4a: bias and RMSE for $\hat{\lambda}$

	(n,T)	$(50,50)$			$(50,100)$			$(50,200)$			$(100,200)$		
		λ			λ			λ			λ		
(ρ,θ)	σ_β^2	0	0.25	0.75	0	0.25	0.75	0	0.25	0.75	0	0.25	0.75
$(0,0)$	0	0.429	0.257	−0.044	0.313	0.188	−0.064	0.220	0.114	−0.078	0.215	0.110	−0.074
		(0.430)	(0.259)	(0.066)	(0.313)	(0.190)	(0.074)	(0.220)	(0.117)	(0.086)	(0.215)	(0.112)	(0.075)
	0.25	0.429	0.249	−0.048	0.310	0.185	−0.068	0.217	0.111	−0.080	0.214	0.107	−0.074
		(0.429)	(0.251)	(0.070)	(0.310)	(0.187)	(0.078)	(0.217)	(0.114)	(0.087)	(0.214)	(0.108)	(0.078)
	0.50	0.421	0.241	−0.050	0.299	0.177	−0.071	0.214	0.088	−0.079	0.210	0.108	−0.079
		(0.422)	(0.244)	(0.074)	(0.300)	(0.180)	(0.083)	(0.214)	(0.094)	(0.086)	(0.210)	(0.110)	(0.083)
$(0.75,0)$	0	0.568	0.389	0.013	0.460	0.299	−0.021	0.333	0.203	−0.047	0.332	0.197	−0.053
		(0.569)	(0.391)	(0.049)	(0.460)	(0.300)	(0.042)	(0.333)	(0.204)	(0.056)	(0.332)	(0.198)	(0.057)
	0.25	0.567	0.383	0.006	0.458	0.297	−0.108	0.331	0.199	−0.046	0.331	0.196	−0.053
		(0.567)	(0.385)	(0.047)	(0.458)	(0.298)	(0.039)	(0.331)	(0.200)	(0.057)	(0.331)	(0.197)	(0.057)
	0.50	0.560	0.376	0.001	0.448	0.286	−0.025	0.329	0.190	−0.052	0.327	0.194	−0.054
		(0.562)	(0.378)	(0.058)	(0.449)	(0.288)	(0.047)	(0.329)	(0.192)	(0.061)	(0.327)	(0.195)	(0.059)
$(0,0.75)$	0	0.422	0.273	−0.023	0.312	0.200	−0.049	0.215	0.122	−0.062	0.211	0.123	−0.060
		(0.422)	(0.275)	(0.051)	(0.312)	(0.202)	(0.059)	(0.215)	(0.123)	(0.069)	(0.211)	(0.124)	(0.064)
	0.25	0.421	0.272	−0.021	0.311	0.200	−0.048	0.214	0.121	−0.065	0.210	0.121	−0.060
		(0.421)	(0.274)	(0.048)	(0.312)	(0.201)	(0.058)	(0.214)	(0.122)	(0.072)	(0.210)	(0.122)	(0.063)
	0.50	0.414	0.261	−0.027	0.304	0.197	−0.056	0.209	0.115	−0.073	0.206	0.120	−0.064
		(0.415)	(0.263)	(0.060)	(0.304)	(0.198)	(0.066)	(0.209)	(0.117)	(0.079)	(0.206)	(0.121)	(0.068)
$(0,-0.75)$	0	0.447	0.299	−0.006	0.327	0.212	−0.039	0.225	0.130	−0.062	0.221	0.130	−0.055
		(0.447)	(0.300)	(0.046)	(0.327)	(0.213)	(0.054)	(0.225)	(0.131)	(0.069)	(0.221)	(0.131)	(0.059)
	0.25	0.445	0.296	−0.004	0.326	0.212	−0.042	0.223	0.127	−0.060	0.219	0.129	−0.057
		(0.446)	(0.297)	(0.046)	(0.326)	(0.213)	(0.055)	(0.223)	(0.128)	(0.067)	(0.219)	(0.130)	(0.061)
	0.50	0.437	0.292	−0.015	0.320	0.208	−0.053	0.213	0.123	−0.069	0.217	0.126	−0.058
		(0.438)	(0.295)	(0.056)	(0.320)	(0.210)	(0.067)	(0.214)	(0.125)	(0.077)	(0.217)	(0.126)	(0.063)

Table 4b: bias and RMSE for $\hat{\lambda}$

(ρ, θ)	λ	(n, T)	$(20, 100)$	$(20, 200)$	$(20, 400)$	$(50, 50)$	$(50, 100)$	$(50, 200)$	$(100, 200)$
	0.25		0.095	0.073	0.067	0.061	0.055	0.058	0.060
$(0, 0)$	0.50		0.066	0.054	0.061	0.056	0.050	0.050	0.043
	0.75		0.059	0.060	0.055	0.063	0.058	0.053	0.045
	0.25		0.097	0.079	0.065	0.060	0.059	0.055	0.060
$(0.75, 0)$	0.50		0.071	0.053	0.055	0.063	0.052	0.053	0.051
	0.75		0.052	0.055	0.052	0.057	0.057	0.058	0.055
	0.25		0.096	0.072	0.062	0.063	0.055	0.054	0.059
$(0, 0.75)$	0.50		0.064	0.056	0.060	0.059	0.058	0.053	0.058
	0.75		0.056	0.055	0.048	0.048	0.056	0.051	0.051
	0.25		0.095	0.072	0.061	0.061	0.058	0.040	0.042
$(0, -0.75)$	0.50		0.064	0.055	0.061	0.047	0.060	0.060	0.056
	0.75		0.056	0.059	0.058	0.044	0.050	0.049	0.047

Table 5: size for $H_0 : \sigma_\beta^2 = 0$

(ρ, θ)	λ	(n, T)	$(20, 100)$	$(20, 200)$	$(20, 400)$	$(50, 50)$	$(50, 100)$	$(50, 200)$	$(100, 200)$
			0.958	0.935	0.907	1.000	0.999	0.998	1.000
	0.25		1.000	1.000	1.000	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.602	0.611	0.650	0.909	0.916	0.964	0.983
$(0, 0)$	0.50		0.998	0.974	0.981	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.355	0.312	0.348	0.612	0.615	0.754	0.868
	0.75		0.973	0.963	0.963	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.505	0.705	0.632	1.000	1.000	0.994	1.000
	0.25		1.000	1.000	1.000	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.502	0.234	0.504	0.945	0.949	0.961	0.991
$(0.75, 0)$	0.50		1.000	0.994	1.000	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.430	0.548	0.649	0.669	0.746	0.719	0.948
	0.75		0.825	1.000	0.981	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.865	0.742	0.851	1.000	0.986	0.945	1.000
	0.25		1.000	1.000	0.999	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.507	0.497	0.535	0.985	0.934	0.948	0.997
$(0, 0.75)$	0.50		1.000	0.997	0.999	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.552	0.510	0.465	0.667	0.762	0.748	0.959
	0.75		0.968	0.996	0.998	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.954	0.940	0.913	1.000	0.976	1.000	1.000
	0.25		1.000	1.000	0.999	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.413	0.427	0.425	0.978	0.911	0.899	0.987
$(0, -0.75)$	0.50		0.966	0.996	0.999	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000
			0.276	0.316	0.395	0.678	0.629	0.677	0.940
	0.75		1.000	0.996	0.997	1.000	1.000	1.000	1.000
			1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 6: power for $H_0 : \sigma_\beta^2 = 0$ (simulated under $H_A : \sigma_\beta^2 = 0.5$). The three figures reported for each value of λ correspond to $\sigma_e^2 \in \{0.5, 1, 2\}$ respectively.

(ρ, θ)	σ^2_β \ (n, T)	$(20, 100)$	$(20, 200)$	$(20, 400)$	$(50, 50)$	$(50, 100)$	$(50, 200)$	$(100, 200)$
$(0, 0)$	0.25	0.034	0.039	0.043	0.024	0.043	0.037	0.039
	0.50	0.053	0.045	0.047	0.046	0.047	0.053	0.045
	0.75	0.048	0.047	0.045	0.043	0.043	0.048	0.047
$(0.75, 0)$	0.25	0.038	0.039	0.040	0.028	0.041	0.040	0.040
	0.50	0.054	0.044	0.051	0.059	0.053	0.049	0.041
	0.75	0.046	0.043	0.050	0.052	0.047	0.042	0.039
$(0, 0.75)$	0.25	0.039	0.037	0.041	0.024	0.039	0.039	0.038
	0.50	0.058	0.048	0.049	0.053	0.050	0.057	0.044
	0.75	0.047	0.047	0.043	0.046	0.049	0.044	0.049
$(0, -0.75)$	0.25	0.037	0.041	0.039	0.022	0.038	0.038	0.042
	0.50	0.046	0.047	0.049	0.054	0.043	0.059	0.048
	0.75	0.053	0.045	0.048	0.047	0.044	0.043	0.041

Table 7. Size for $H_0 : \lambda = 0$, based on $\tilde{S}^{(\lambda)}_{nT}$.

(ρ, θ)	σ^2_β || (n,T)	$(20,100)$	$(20,200)$	$(20,400)$	$(50,50)$	$(50,100)$	$(50,200)$	$(100,200)$
		0.489	0.407	0.452	0.592	0.543	0.576	0.803
	0.25	0.743	0.788	0.803	0.893	0.903	0.965	0.999
		0.806	0.903	0.851	0.976	0.954	0.968	0.998
		0.145	0.106	0.123	0.346	0.403	0.482	0.603
$(0,0)$	0.50	0.408	0.502	0.543	0.803	0.834	0.856	0.802
		0.713	0.801	0.764	0.972	0.923	0.943	0.993
		0.079	0.064	0.098	0.106	0.182	0.199	0.293
	0.75	0.482	0.372	0.451	0.668	0.573	0.584	0.781
		0.701	0.734	0.786	0.934	0.884	0.879	0.997
		0.346	0.404	0.398	0.685	0.704	0.728	0.894
	0.25	0.798	0.832	0.856	0.899	0.972	0.998	1.000
		0.894	0.898	0.912	0.999	0.989	0.993	1.000
		0.274	0.306	0.341	0.408	0.461	0.462	0.603
$(0.75,0)$	0.50	0.701	0.679	0.743	0.805	0.798	0.781	0.816
		0.843	0.871	0.873	0.984	0.973	0.988	0.990
		0.204	0.308	0.294	0.342	0.401	0.399	0.502
	0.75	0.532	0.497	0.523	0.561	0.487	0.563	0.538
		0.698	0.716	0.773	0.914	0.871	0.904	0.911
		0.396	0.431	0.427	0.714	0.723	0.841	0.932
	0.25	0.803	0.856	0.823	0.876	0.836	0.875	1.000
		0.879	0.932	0.872	0.899	0.934	0.978	0.999
		0.302	0.333	0.351	0.503	0.562	0.515	0.589
$(0,0.75)$	0.50	0.516	0.476	0.502	0.696	0.773	0.699	0.734
		0.887	0.894	0.888	0.962	0.988	0.966	0.961
		0.201	0.234	0.241	0.406	0.451	0.376	0.528
	0.75	0.356	0.416	0.442	0.401	0.602	0.702	0.773
		0.684	0.703	0.698	0.901	0.932	0.915	0.998
		0.216	0.381	0.451	0.667	0.698	0.682	0.703
	0.25	0.814	0.825	0.873	0.993	0.968	0.974	0.998
		0.833	0.900	0.888	0.981	0.974	0.996	0.995
		0.378	0.333	0.327	0.478	0.561	0.588	0.758
$(0,-0.75)$	0.50	0.614	0.781	0.798	0.903	0.942	0.896	0.993
		0.687	0.764	0.800	0.887	0.923	0.955	1.000
		0.305	0.274	0.342	0.516	0.588	0.617	0.754
	0.75	0.781	0.689	0.782	0.831	0.918	0.936	0.935
		0.853	0.899	0.784	0.924	0.873	0.922	1.000

Table 8. Power for $\lambda = 0$ (under $H_A : \lambda = 0.5$), based on $\tilde{S}^{(\lambda)}_{nT}$. The three figures reported for each value of σ^2_β correspond to the power of the test for $\sigma^2_e = \{0.5, 1, 2\}$ respectively.

Individual specific results

Country	t-stat	p-value	$\hat{\beta}_i$
Austria	−2.72	0.006*	0.783
Belgium	−3.78	0.002*	0.785
Canada	−2.33	0.019*	0.945
Denmark	−2.04	0.039*	0.926
Finland	−1.60	0.102	1.057
France	−2.71	0.007*	0.880
Germany	−2.42	0.015*	0.695
Greece	−2.11	0.033*	1.062
India	−2.63	0.008*	2.580
Ireland	−2.50	0.001*	0.997
Italy	−3.36	0.008*	1.621
Japan	−1.79	0.067	0.401
Korea	−2.38	0.017*	1.533
Netherlands	−2.01	0.042*	0.582
Norway	−1.95	0.048*	1.208
Portugal	−1.44	0.138	1.130
South Africa	−1.18	0.217	3.412
Spain	−2.07	0.036*	1.675
Sweden	0.087	0.703	1.385
Switzerland	−2.07	0.036*	0.761
UK	−0.81	0.361	1.176

Panel-based results

Kao's (1999) test	test stat.	p-value
	−1.329	0.092
Estimation		
$\hat{\sigma}_\beta$	0.594	
$\hat{\sigma}_{\beta,bc}$	0.624	
$\hat{\lambda}$	0.213	
st. dev. of individual $\tilde{\beta}_i$s	0.584	
Testing	test stat.	p-value
$H_0 : \sigma_\beta = 0$	5.24	0.000*
$H_0 : \lambda = 0$	2.21	0.014*

Table 9. Cointegration tests and estimation of β_i for the unit specific equations $\ln E_{it} = \alpha_i + \beta_i \ln \frac{P_t^{US}}{P_{it}} + u_{it}$; the values reported are the Engle-Granger based t-ADF statistics (and the associated p-values) for a unit root test on residuals; a "*" next to the p-value indicates rejection of the null of no cointegration at the 5% level. Data have been downloaded from the Federal Reserve Economic Data (FRED). In the table, we report Kao's (1999) test for the null of no cointegration;