Journal of Economic Theory and Econometrics, Vol. 35, No. 4, Dec. 2024, 55–86

Tests of the Null of Cointegration Using Integrated and Modified OLS Residuals*

Cheol-Keun Cho†

Abstract This study develops a KPSS (Kwiatkowski *et al.*, 1992)-type cointegration test utilizing residuals from integrated and modified ordinary least squares (IMOLS) estimation. The test statistic, denoted by $KPSS^{Fb}$ has a pivotal null limit distribution under fixed-b assumption. The proposed test demonstrates reasonable performance in terms of size and power when the Andrews' AR(1) plug-in data-dependent (*DD*) bandwidth is employed and fixed-b critical values are used. Additionally, two modified IMOLS residuals are proposed to obtain alternative data-dependent bandwidths. In the simulation experiment, these bandwidths deliver improved power properties for the proposed test.

Keywords Data-dependent bandwidth, cointegration, fixed-b, IMOLS, KPSS.

JEL Classification C12, C22.

*I thank the Editor and the two anonymous reviewers for their constructive comments and suggestions. An earlier version of the paper was presented at the 2019 International Symposium on Econometric Theory and Applications (SETA). This research was funded by the University of Ulsan under Grant No. 2023-0373.

†Deparment of Economics, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan, Republic of Korea 44610. E-mail: cheolcho@ulsan.ac.kr.

Received October 1, 2024, Revised December 23, 2024, Accepted December 27, 2024

1. INTRODUCTION

This study addresses residual-based cointegration test with the presence of a cointegrating relationship taken as the null hypothesis.[1] A fixed-b inferential method is developed within the integrated and modified OLS (Vogelsang and Wagner, 2014) framework. A KPSS (Kwiatkowski *et al.*, 1992)-type statistic is used as the test statistic, and its pivotal fixed-b limit distribution (Kiefer and Vogelsang, 2005) is employed to obtain fixed-b critical values for hypothesis testing.

To construct the test statistic, a kernel nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimator is required to estimate the long-run variance (LRV). However, it is well-known that statistical inference using test statistics scaled by a HAC estimator often suffers from size distortions, particularly when the persistence of the underlying time series is high. See Müller (2014), Müller (2005), Caner and Kilian (2001), and Gabriel (2003). The choice of bandwidth required for constructing a HAC estimator is central to this problem. The simulation results in the present paper indicate that the proposed test suffers from greater size distortions as the ratio ($b \equiv M/T$) of bandwidth (M) to the sample size (T) becomes smaller. Conversely, the power of the test decreases as b increases. A data-dependent (DD) bandwidth scheme, such as the AR(1) plug-in method proposed by Andrews (1991), is commonly used in practice to select an appropriate bandwidth. However, in the context of cointegration tests, this data-driven bandwidth often delivers low power (see Xiao and Phillips (2002), Choi and Ahn (1995), and Gabriel (2003)). The DD scheme, when applied to residuals from the existing methods—such as OLS or Fully Modified OLS (FMOLS) residuals—delivers too large a bandwidth of order $O_p(T)$, with M/T not shrinking to zero. Consequently, the test is not consistent, as consistency requires $M/T \to 0$ as $T \to \infty$. For a detailed analysis, see Xiao and Phillips (2002). An important finding in this study is that the DD bandwidth calculated with IMOLS residuals is $O_p\left(T^{1/3}\right)$ under non-cointegration, yielding a bandwidth ratio ($b \equiv M/T$) that shrinks to zero as T increases. This unique property can lead to reasonable power for the test. The rest of the paper is organized as follows: Section 2 presents the model, assumptions, and a brief review of the IMOLS estimation. Section 3 develops the asymptotic theory and investigates the limit behavior of the DD bandwidth calculated with the IMOLS residual. Section 4 investigates the performance of the $KPSS^{Fb}$ test via Monte Carlo sim-

[1]Many existing cointegration tests take the presence of cointegrating relationship as the null hypothesis. See Shin (1994), Harris and Inder (1994), and Xiao and Phillips (2002).

ulations. Some modified IMOLS residuals that can be used to obtain alternative *DD* bandwidths are proposed. Section 5 concludes the study. The proofs and additional results are available in the Appendix.

2. MODEL SETUP AND ASSUMPTIONS

Consider the following model:

$$y_t = f_t'\delta + x_t'\beta + u_t, \quad (1)$$
$$x_t = x_{t-1} + v_t,$$

for $t = 1,...,T$, where $f_t = \left(1,t,...,t^{p-1}\right)'$ is a $p \times 1$ deterministic regressor vector and x_t is a $k \times 1$ stochastic regressor vector. Both y_t and x_t are $I(1)$ processes, and x_t has no drift.[2]

Assumption 1 (*H_0*: Cointegration). Let $\eta_t = (u_t, v_t')'$ and assume that a functional central limit theorem holds:

$$T^{-1/2}\sum_{t=1}^{[rT]}\eta_t \Rightarrow B(r) = \begin{pmatrix} B_u(r) \\ B_v(r) \end{pmatrix} = \Omega^{1/2}W(r), \;\; r \in [0,1],$$

where $W(r) = (w_{uv}(r), w_v'(r))'$ is a $(k+1)\times 1$ vector of standard Brownian motion with

$$\Omega = \lim_{T\to\infty}\left(\text{var}\left(T^{-1/2}\sum_{t=1}^{T}\eta_t\right)\right) := \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix},$$

where Ω is a p.d. and equals $\sum_{j=-\infty}^{\infty} E\left(\eta_t\eta_{t-j}'\right)$ if η_t is assumed to be stationary.

Under Assumption 1, y_t and x_t are cointegrated up to the deterministic trend f_t. As in Vogelsang and Wagner (2014), the following Cholesky form of $\Omega^{1/2}$ is used:

$$\Omega^{1/2} = \begin{bmatrix} \sigma_{uv} & \lambda_{uv} \\ \mathbf{0} & \Omega_{vv}^{1/2} \end{bmatrix},$$

where $\sigma_{uv}^2 = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and $\lambda_{uv} = \Omega_{uv}\left(\Omega_{vv}^{-1/2}\right)'$. In addition, define

$$\Sigma = E\left(\eta_t\eta_t'\right) = \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix}.$$

[2]Theorem 3 in the Appendix provides additional results for the case where x_t has a drift, extending the results in Hansen (1992b).

When there is no cointegrating relationship (i.e., with u_t being an $I(1)$ series), u_t can be represented by the sum of its differences (say $\{\varepsilon_t\}$), and (1) becomes a spurious regression.

Assumption 2 (H_1: Non-cointegration). Let $\widetilde{\eta}_t = (\varepsilon_t, v_t')'$ and assume that a functional central limit theorem holds for $\{\widetilde{\eta}_t\}$:

$$T^{-1/2}\sum_{t=1}^{[rT]}\widetilde{\eta}_t \Rightarrow \begin{pmatrix} B_\varepsilon(r) \\ B_v(r) \end{pmatrix} = \widetilde{\Omega}^{1/2}\widetilde{W}(r),\ r \in [0,1],$$

where $\widetilde{W}(r) = (w_\varepsilon(r), w_v'(r))'$ is standard Brownian motion with

$$\widetilde{\Omega} = \lim\left(\mathrm{var}\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{\eta}_t\right)\right) = \begin{bmatrix} \sigma_\varepsilon^2 & \Omega_{\varepsilon v} \\ \Omega_{v\varepsilon} & \Omega_{vv} \end{bmatrix} > 0.$$

IMOLS estimator (Vogelsang and Wagner, 2014). By summing both sides of (1) and adding the original regressors x_t, the following integrated and modified regression can be obtained:

$$S_t^y = S_t^{f\prime}\delta + S_t^{x\prime}\beta + x_t'\gamma + S_t^u \tag{2}$$

with $S_t^y = \sum_{j=1}^{t} y_j$ and S_t^f, S_t^x, and S_t^u defined in an analogous manner. The IMOLS estimator of $\theta = (\delta', \beta', \gamma')'$ is given by

$$\widehat{\theta} = (\widehat{\delta}', \widehat{\beta}', \widehat{\gamma}')' = (S^{\widetilde{x}\prime}S^{\widetilde{x}})^{-1}S^{\widetilde{x}\prime}S^y,$$

where we let $S^{\widetilde{x}} = (S^f \vdots S^x \vdots X)$ with $S^f = (S_1^f, \ldots S_T^f)'$, $S^x = (S_1^x, \ldots S_T^x)'$, and $X = (x_1, \ldots x_T)'$.

Note that, unlike FMOLS or Dynamic OLS (DOLS, Phillips and Loretan (1991)) estimators, IMOLS estimator does not require the choice of any tuning parameter, such as the bandwidth (FMOLS) or the number of leads and lags (DOLS).

Define

$$A_{IM} = \begin{pmatrix} T^{-1/2}\tau_F^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & T^{-1}I_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_k \end{pmatrix},$$

where τ_F is a diagonal matrix with diagonal elements $1, T, T^2, ..., T^{p-1}$ satisfying

$$T^{-1}\tau_F^{-1}\sum_{t=1}^{[rT]} f_t \to \int_0^r f(s)ds,\ r\in[0,1]$$

as T grows, with $f(s) = (1, s, s^2, ..., s^{p-1})$. Under the null of cointegration, Vogelsang and Wagner (2014) show

$$A_{IM}^{-1}\begin{pmatrix}\widehat{\delta}-\delta\\ \widehat{\beta}-\beta\\ \widehat{\gamma}-\Omega_{vv}^{-1}\Omega_{vu}\end{pmatrix} = \begin{pmatrix}T^{1/2}\tau_F(\widehat{\delta}-\delta)\\ T(\widehat{\beta}-\beta)\\ (\widehat{\gamma}-\Omega_{vv}^{-1}\Omega_{vu})\end{pmatrix} \tag{3}$$
$$\xrightarrow{d} \sigma_{uv}\Pi^{-1}\left(\int g(s)g(s)'ds\right)^{-1}\int g(s)w_{u\cdot v}(s)ds,$$

where $\Pi = \mathrm{diag}(I_p, \Omega_{vv}^{1/2}, \Omega_{vv}^{1/2})$, and $g(r) = (\int_0^r f(s)'ds, \int_0^r w_v(s)'ds, w_v(r)')'$.

3. IMOLS-BASED COINTEGRATION TESTS

Denote the residuals in (2) by $\widetilde{S}_t^u$ and consider its difference

$$\triangle\widetilde{S}_t^u = \widetilde{S}_t^u - \widetilde{S}_{t-1}^u = u_t - f_t'(\widehat{\delta}-\delta) - x_t'(\widehat{\beta}-\beta) - \triangle x_t'\widehat{\gamma} = \widehat{u}_t - \triangle x_t'\widehat{\gamma},$$

where $\widetilde{S}_t^u = S_t^y - S_t^{f\prime}\widehat{\delta} - S_t^{x\prime}\widehat{\beta} - x_t'\widehat{\gamma}$. With $\theta_1 = (\delta',\ \beta')'$, $\theta_0 = (\delta', \beta', \mathbf{0})'$, and $\theta_* = (\delta', \beta', \Omega_{vu}'\Omega_{vv}^{-1})'$,

$$\triangle\widetilde{S}_t^u = u_t - (f_t', x_t')(\widehat{\theta}_1 - \theta_1) - \triangle x_t'\widehat{\gamma} = u_t - (f_t', x_t', \triangle x_t')(\widehat{\theta} - \theta_0)$$
$$= u_t - (f_t', x_t', \triangle x_t')(\widehat{\theta} - \theta_*) - \triangle x_t'\Omega_{vv}^{-1}\Omega_{vu}.$$

The proposed KPSS-type test statistic is given by

$$KPSS^{Fb} = \frac{T^{-2}\sum_{t=2}^{T}\left(\sum_{j=2}^{t}\triangle\widetilde{S}_j^u\right)^2}{\widetilde{\sigma}_{uv}^2} = \frac{T^{-2}\sum_{t=2}^{T}\left(\widetilde{S}_t^u - \widetilde{S}_1^u\right)^2}{\widetilde{\sigma}_{uv}^2},$$

where $\widetilde{\sigma}_{uv}^2$ is a HAC estimator constructed using $\triangle\widetilde{S}_t^u$:

$$\widetilde{\sigma}_{uv}^2 = \frac{1}{T}\sum_{i=2}^{T}\sum_{j=2}^{T} k\left(\frac{|i-j|}{M}\right)\triangle\widetilde{S}_i^u\triangle\widetilde{S}_j^u.$$

The Bartlett kernel[3] $k(z) = (1-|z|)\mathbf{1}(|z|\le 1)$ is considered in this paper, but other kernels may be used.

[3] Thus, $\widetilde{\sigma}_{uv}^2$ is a Newey-West estimator (Newey and West, 1987).

Theorem 1. *Under Assumption 1 (H_0), it holds that as $T \to \infty$,*

(a) $T^{-1/2}\left(\widetilde{S}^u_{[rT]} - \widetilde{S}^u_1\right) = T^{-1/2}\sum_{t=2}^{[rT]} \triangle \widetilde{S}^u_t \Rightarrow \sigma_{uv}\widetilde{P}(r),\ r \in [0,1]$, *where*

$$\widetilde{P}(r) \equiv \int_0^r dw_{uv}(s) - g(r)'\left[\int_0^1 g(s)g(s)'ds\right]^{-1}\int_0^1 (G(1)-G(s))dw_{uv}(s);$$

(b) under small-b asymptotics, $\widetilde{\sigma}^2_{uv} \xrightarrow{d} \sigma^2_{uv}\left(1+d'_\gamma d_\gamma\right)$, *where d_γ denotes the last k components of* $(\int_0^1 g(s)g(s)'ds)^{-1}\int_0^1 (G(1)-G(s))dw_{uv}(s)$ *with* $G(s) = \int_0^s g(u)du$;

(c) under small-b asymptotics,

$$KPSS^{Fb} \xrightarrow{d} \frac{\int_0^1 \widetilde{P}(r)^2 dr}{1+d'_\gamma d_\gamma}. \qquad (4)$$

Under Assumption 2 (H_1),

(a′) $T^{-3/2}\sum_{t=2}^{[rT]} \triangle \widetilde{S}^u_t \Rightarrow \sigma_\varepsilon \widetilde{H}(r)$, *where*

$$\widetilde{H}(r) \equiv \int_0^r w_\varepsilon(s)ds - g(r)'\left[\int_0^1 g(s)g(s)'ds\right]^{-1}\int_0^1 g(s)\int_0^s w_\varepsilon(u)duds;$$

(b′) under small-b asymptotics, $\widetilde{\sigma}^2_{uv} = O_p(MT)$;

(c′) under small-b asymptotics, $KPSS^{Fb} = O_p\left(\frac{T}{M}\right)$, *which implies* $KPSS^{Fb} \to \infty$ *as* $T \to \infty$.

Proof. Proofs for parts (a) and (b) are provided in Vogelsang and Wagner (2014). See Lemma 2 and Theorem 3, respectively. For parts (c) through (c'), see the Appendix.

Remark 1. Theorem 1 shows that the $KPSS^{Fb}$ test is consistent against the alternative hypothesis of non-cointegration under the traditional small-*b* asymptotic framework. However, the limit in part (c) does not capture the impact of bandwidth. Fixed-*b* limit distributions are derived under the assumption that the bandwidth (M) is proportional to the sample size (T) (i.e., $M = bT$ with $b \in (0,1]$). Typically, fixed-*b* limits depend on the bandwidth ratio (b) and the kernel being used. The next Corollary provides the fixed-*b* limits for $KPSS^{Fb}$.

Corollary 1. *Assume $M = bT$ for a fixed value of $b \in (0,1]$. Then, under Assumption 1 (H_0), as $T \to \infty$,*

$$KPSS^{Fb} \xrightarrow{d} \frac{\int_0^1 \widetilde{P}(u)^2 du}{\int_0^1\int_0^1 k\left(\frac{|r-s|}{b}\right)d\widetilde{P}(r)d\widetilde{P}(s)} \equiv \frac{\int_0^1 \widetilde{P}(u)^2 du}{\mathbf{P}(b,\widetilde{P})}, \qquad (5)$$

where $\widetilde{P}(r)$ is defined in Theorem 1.[4] *Under Assumption 2 (H_1), as $T \to \infty$,*

$$KPSS^{Fb} \xrightarrow{d} \frac{\int_0^1 \widetilde{H}(u)^2 du}{\int_0^1 \int_0^1 k\left(\frac{|r-s|}{b}\right) d\widetilde{H}(r) d\widetilde{H}(s)} \equiv \frac{\int_0^1 \widetilde{H}(u)^2 du}{\mathbf{P}(b,\widetilde{H})}, \tag{6}$$

where $\widetilde{H}(r)$ is defined in Theorem 1.

Proof. Theorem 1 (parts (a) and (a')) and the continuous mapping theorem immediately yield the desired result.

Remark 2. The fixed-b percentiles for the random variable in (5) can be simulated by using independent and identically distributed (IID) $N(0,1)$ pseudo random numbers. Researchers can pick a fixed-b critical value for a given specific value of $b = M/T$.

In the simulation study (Section 4), the AR(1) plug-in data-dependent bandwidth rule proposed by Andrews (1991) will be used. Below the lmits of the AR(1) coefficient calculated using $\left\{\triangle \widetilde{S}_t^u\right\}$ are derived under H_0 and H_1, respectively. This result can explain the empirical findings reported in Section 4.

Consider the AR(1) plug-in data-dependent rule (Andrews, 1991) with the Bartlett kernel:

$$M = 1.1447\left(\widehat{\alpha}(1)T\right)^{\frac{1}{3}} \quad \text{with} \quad \widehat{\alpha}(1) = \frac{4\widehat{\phi}^2}{(1-\widehat{\phi})^2(1+\widehat{\phi})^2}, \tag{7}$$

where $\widehat{\phi}$ is the estimated coefficient in the AR(1) regression with $\triangle \widetilde{S}_t^u$, $t = 3,...,T$:

$$\widehat{\phi} = \frac{\sum_{t=3}^T \triangle \widetilde{S}_t^u \triangle \widetilde{S}_{t-1}^u}{\sum_{t=3}^T \left(\triangle \widetilde{S}_{t-1}^u\right)^2}.$$

Define

$$\Sigma_{\eta\eta} \equiv p\lim \frac{1}{T}\sum \eta_t \eta_t' = \begin{pmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{uv}' & \Sigma_{vv} \end{pmatrix} \quad \text{and}$$

$$\Sigma_{\eta,\eta-1} \equiv p\lim \frac{1}{T}\sum \eta_t \eta_{t-1}' = \begin{pmatrix} \Sigma_{u,u-1} & \Sigma_{u,v-1} \\ \Sigma_{v,u-1} & \Sigma_{v,v-1} \end{pmatrix},$$

where $\eta_t = (u_t, v_t')'$. Theorem 2 provides the limits of $\widehat{\phi}$ under the null and alternative hypotheses, respectively.

[4] An equivalent expression of $\mathbf{P}(b,\widetilde{P})$ is available in Cho and Vogelsang (2017).

Theorem 2. *Under H_0, as $T \to \infty$,*

$$\widehat{\phi} \xrightarrow{d} \frac{\Sigma_{u,u-1} - \Sigma_{u,v-1}\gamma_0^\infty - \gamma_0^{\infty\prime}\Sigma_{v,u-1} + \gamma_0^{\infty\prime}\Sigma_{v,v-1}\gamma_0^\infty}{\Sigma_{uu} - 2\Sigma_{uv}\gamma_0^\infty + \gamma_0^{\infty\prime}\Sigma_{vv}\gamma_0^\infty}$$
$$= \frac{(1,\ -\gamma_0^{\infty\prime})\Sigma_{\eta,\eta-1}\begin{pmatrix} 1 \\ -\gamma_0^\infty \end{pmatrix}}{(1,\ -\gamma_0^{\infty\prime})\Sigma_{\eta\eta}\begin{pmatrix} 1 \\ -\gamma_0^\infty \end{pmatrix}},$$

where γ_0^∞ denotes the weak limit of the IMOLS estimator $\widehat{\gamma}$ under H_0. Under H_1, as $T \to \infty$,

$$\widehat{\phi} \xrightarrow{d} \frac{\gamma_1^{\infty\prime}\Sigma_{v,v-1}\gamma_1^\infty}{\gamma_1^{\infty\prime}\Sigma_{vv}\gamma_1^\infty}, \tag{8}$$

where γ_1^∞ denotes the weak limit of $\frac{\widehat{\gamma}}{T}$ under H_1. Here, γ_0^∞ and γ_1^∞ are given by (see the proof for Lemma 1 in the Appendix)

$$\gamma_0^\infty \equiv \Omega_{vv}^{-1}\Omega_{vu} + \sigma_{uv}\Omega_{vv}^{-1/2}d_\gamma \quad \textit{and} \quad \gamma_1^\infty \equiv \Omega_{vv}^{-1/2}\widetilde{d}_\gamma,$$

where d_γ denotes the last k components of $(\int_0^1 g(s)\, g(s)'\, ds)^{-1} \int_0^1 (G(1) - G(s))\, dw_{uv}(s)$, and $\widetilde{d}_\gamma$ denotes the last k components of $(\int_0^1 g(s) g(s)' ds)^{-1} (\int_0^1 g(s) \int_0^s \sigma_\varepsilon w_\varepsilon(u)\, du ds)$.

Proof. See the Appendix.

Theorem 2 demonstrates that under H_0, the limit of $\widehat{\phi}$ depends on the autocovariances of $\eta_t = (u_t, v_t')'$. However, under H_1, the limit of $\widehat{\phi}$ primarily depends on the persistence of $\triangle x_t \equiv v_t$, with no dependence on the persistence of u_t.[5] Given this result, it follows that $\widehat{\alpha}(1)$ in (7) has a nondegenerate limit, and the *DD* bandwidth is $O_p\left(T^{1/3}\right)$.[6] Therefore, the *DD* bandwidth may not be too large under H_1, and the power of the test is not necessarily low. This property is unique because other residuals, such as the OLS residuals, lead to a bandwidth of $O_p(T)$ under H_1, as documented in Xiao and Phillips (2002). Furthermore, Theorem 2 explains why the power of the $KPSS^{Fb}$ test may be substantially affected by the persistence in $\triangle x_t$.

[5]With $k = 1$, the limit of $\widehat{\phi}$ under H_1 is reduced to the first order autocorrelation of $\triangle x_t$ (equals θ in the simulation setting). Therefore, for an alternative DGP, the resulting *DD* bandwidth tends to be large when θ is large, leading to reduced power compared to the case where θ is small.

[6]This is because $\widehat{\gamma} = O_p(1)$ under H_0 (cointegration) and $\widehat{\gamma} = O_p(T)$ under H_1 (non cointegration) as presented in Lemma 1 (a). Hence the terms involving $\widehat{\gamma}$ dominate other terms in both the denominator and numerator of $\widehat{\phi}$, under H_1.

4. SIMULATION STUDY

4.1. DGP SPECIFICATIONS

To evaluate the performance of the test in finite samples, the following data generation process is considered.

$$y_t = 1 + x_{1t} + 2x_{2t} + u_t = 1 + (1,\ 2)x_t + u_t,$$

where $u_t = \alpha u_{t-1} + \widetilde{\varepsilon}_t$, $\triangle x_t = v_t = (v_{1t}, v_{2t})'$, and $v_t = \mathrm{diag}(\theta, \theta)v_{t-1} + (\widetilde{\xi}_{1t}, \widetilde{\xi}_{2t})'$ with[7]

$$\begin{pmatrix} \widetilde{\varepsilon}_t \\ \widetilde{\xi}_{1t} \\ \widetilde{\xi}_{2t} \end{pmatrix} \sim \mathbf{NID} \begin{pmatrix} 1 & \rho & \frac{\rho}{2} \\ \rho & 1 & 0 \\ \frac{\rho}{2} & 0 & 1 \end{pmatrix}.$$

The key parameter is α. If $\alpha = 1$, the regression is a spurious regression with no cointegrating relationship. If $|\alpha| < 1$, then there exists a cointegrating relationship. The parameters α and θ primarily control the persistence of the regression error term, u_t, and $v_t = \triangle x_t$, respectively. The strength of the endogeneity is controlled by ρ.

For comparisons, the results for FMOLS-based KPSS (S_2^+; Harris and Inder, 1994) and CUSUM (*CS*; Xiao and Phillips, 2002) tests are reported. A variety of bandwidth rules are considered: conventional non-data-dependent rules, the fixed-b rule ($M = bT$), and the AR(1) plug-in data-dependent (*DD*) rule of Andrews (1991). In this section, the results with conventional non-data-dependent rules are reported for $M1 = [4(T/100)^{1/4}]$ and $M2 = [2T^{1/3}]$. The number of simulation replications is 5,000. The 95% critical values for the case of $f_t = 1$ and $k = 2$ are 0.2210 for S_2^+ and 1.0413 for *CS*. The traditional critical value for $KPSS^{Fb}$ associated with the limit in (4) is 0.049, and fixed-b critical values for $KPSS^{Fb}$ are presented in Table 2 for the selected values of b.

---

[7]For nonzero values of ρ, it is required that $-\frac{2}{\sqrt{5}} < \rho < \frac{2}{\sqrt{5}} \risingdotseq 0.8944$ to ensure the covariance matrix

$$\mathbf{V} = \begin{pmatrix} 1 & \rho & \frac{\rho}{2} \\ \rho & 1 & 0 \\ \frac{\rho}{2} & 0 & 1 \end{pmatrix}$$

is positive definite since the eigenvalues of this matrix are 1 and $1 \pm \frac{\sqrt{5}}{2}\rho$.

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7
ρ	0	0.5	0.5	0	0.8	0.8	0.8
α	0.5	0.5	0.9	0.8	0.9	0.9	0.9
θ	0.5	0.5	0.5	0.8	0.2	0.5	0.8
	DGP1A		DGP3A	DGP4A	DGP5A	DGP6A	DGP7A
ρ	0		0.5	0	0.8	0.8	0.8
α	1		1	1	1	1	1
θ	0.5		0.5	0.8	0.2	0.5	0.8

Table 1: DGP SPECIFICATIONS FOR MONTE CARLO SIMULATIONS. There exists a cointegrating relationship in DGP1 through DGP7, while there is no cointegration in DGP1A through DGP7A. The parameter θ controls the persistence of $\triangle x_t$, and the strength of the endogeneity is controlled by ρ.

b	95% percentile
0.02	0.0499
0.04	0.0516
0.06	0.0541
0.08	0.0577
0.1	0.0627
0.2	0.1147
0.3	0.1850
0.4	0.2491
0.5	0.3001
1	0.5081

Table 2: FIXED-b CRITICAL VALUES (NOMINAL SIZE OF 5%, INTERCEPT ONLY AND $k = 2$). The critical values were simulated using IID standard normal pseudo random numbers with $T = 1,000$ and $50,000$ replications.

4.2. EMPIRICAL SIZE AND POWER

Tables 3 and 4 provide the simulation results for the CS, S_2^+ and $KPSS^{Fb}$ tests. For the FMOLS-based tests (CS and S_2^+), the bandwidth $M2$ delivers mild size distortion and reasonable power. The Andrews' DD bandwidth also provides a good size property. However, its power is very low. This result corroborates a well-known property of the DD bandwidth that is documented in Xiao and

Phillips (2002): The *DD* bandwidth is $O_p(T)$ under H_1, which renders the test inconsistent.

		FM-CUSUM (CS)			FM-KPSS (S_2^+)			IM-KPSS($KPSS^{Fb}$)				
(ρ,α,θ)	T	DD	$M1$	$M2$	DD	$M1$	$M2$	DD	$0.02T$	$0.08T$	$0.1T$	$0.2T$
	100	.031	.057	.024	.044	.073	.033	.142	.416	.124	.121	.135
DGP1	200	.043	.067	.030	.055	.082	.040	.082	.170	.058	.060	.077
(0,.5,.5)	500	.048	.077	.040	.056	.081	.046	.072	.076	.044	.039	.040
	1000	.054	.080	.046	.056	.075	.048	.065	.056	.040	.034	.033
	100	.024	.050	.017	.040	.065	.029	.142	.408	.124	.124	.128
DGP2	200	.037	.062	.024	.051	.072	.036	.087	.177	.063	.065	.073
(.5,.5,.5)	500	.047	.075	.039	.055	.081	.047	.075	.075	.047	.044	.037
	1000	.049	.078	.041	.056	.075	.048	.068	.058	.042	.034	.031
	100	.025	.337	.069	.046	.410	.119	.177	.771	.194	.138	.061
DGP3	200	.021	.434	.100	.039	.483	.138	.170	.565	.113	.077	.036
(.5,.9,.5)	500	.034	.538	.129	.047	.514	.144	.157	.282	.061	.044	.032
	1000	.037	.536	.124	.050	.473	.128	.138	.143	.046	.037	.031
	100	.012	.129	.027	.025	.158	.047	.062	.788	.100	.062	.067
DGP4	200	.021	.186	.042	.039	.197	.068	.052	.534	.051	.032	.050
(0,.8,.8)	500	.038	.223	.062	.049	.209	.074	.068	.198	.042	.032	.035
	1000	.045	.225	.065	.057	.195	.073	.069	.100	.041	.033	.032
	100	.035	.488	.104	.040	.575	.159	.327	.518	.249	.215	.112
DGP5	200	.019	.596	.128	.031	.654	.169	.176	.209	.123	.126	.082
(.8,.9,.2)	500	.019	.691	.126	.027	.682	.144	.042	.035	.056	.069	.068
	1000	.014	.673	.107	.024	.634	.115	.015	.013	.040	.045	.046
	100	.023	.428	.079	.042	.522	.137	.172	.658	.167	.132	.074
DGP6	200	.017	.555	.108	.033	.620	.152	.112	.295	.077	.070	.046
(.8,.9,.5)	500	.018	.667	.116	.029	.660	.141	.044	.053	.040	.041	.039
	1000	.016	.657	.105	.027	.621	.118	.023	.021	.033	.034	.029
	100	.040	.076	.006	.036	.116	.024	.053	.862	.132	.071	.055
DGP7	200	.004	.157	.024	.011	.194	.047	.057	.648	.072	.044	.036
(.8,.9,.8)	500	.011	.315	.060	.022	.306	.078	.069	.234	.049	.035	.034
	1000	.021	.542	.102	.036	.491	.115	.063	.099	.041	.036	.029

Table 3: EMPIRICAL SIZE, $\alpha \neq 1$. For the FMOLS-based tests (CS and S_2^+), the non data-dependent bandwidths $M1 = [4(T/100)^{1/4}], M2 = [2T^{1/3}]$ and the data-dependent (DD) bandwidth are used. The DD bandwidth is calculated using the OLS residuals in (1). The 95% critical values are 0.2210 for S_2^+ and 1.0413 for CS. For the $KPSS^{Fb}$ test, DD bandwidth is calculated using the IMOLS residuals, and the fixed-b critical value corresponding to the specific value of $b = DD/T$ is used in each replication. The number of simulation replications are $5,000$.

		FM-CUSUM (CS)			FM-KPSS (S_2^+)			IM-KPSS ($KPSS^{Fb}$)				
(ρ,α,θ)	T	DD	$M1$	$M2$	DD	$M1$	$M2$	DD	$0.02T$	$0.08T$	$0.1T$	$0.2T$
	100	.133	.577	.187	.136	.675	.289	.228	.820	.277	.195	.062
DGP1A	200	.155	.813	.370	.142	.863	.493	.308	.711	.224	.152	.027
$(0,1,.5)$	500	.176	.978	.722	.148	.985	.792	.511	.625	.208	.134	.012
	1000	.188	.998	.900	.144	.999	.930	.626	.610	.212	.136	.011
	100	.114	.555	.171	.117	.659	.276	.232	.826	.279	.197	.058
DGP3A	200	.136	.830	.373	.128	.869	.488	.319	.716	.229	.153	.024
$(.5,1,.5)$	500	.154	.977	.730	.131	.983	.806	.517	.630	.221	.149	.013
	1000	.166	.998	.903	.131	.999	.932	.626	.605	.214	.133	.015
	100	.121	.483	.130	.135	.586	.236	.068	.921	.227	.136	.048
DGP4A	200	.149	.776	.333	.152	.833	.455	.087	.852	.183	.107	.024
$(0,1,.8)$	500	.181	.975	.702	.170	.982	.783	.237	.718	.185	.117	.013
	1000	.193	.998	.898	.169	.998	.927	.401	.648	.201	.127	.011
	100	.102	.580	.189	.106	.678	.285	.435	.751	.352	.264	.080
DGP5A	200	.116	.830	.388	.117	.874	.500	.530	.673	.267	.190	.034
$(.8,1,.2)$	500	.133	.980	.742	.113	.988	.809	.660	.628	.239	.162	.017
	1000	.144	.999	.905	.119	.998	.929	.687	.604	.220	.142	.015
	100	.063	.546	.170	.098	.647	.266	.234	.832	.287	.197	.061
DGP6A	200	.080	.817	.373	.112	.868	.489	.317	.714	.334	.150	.023
$(.8,1,.5)$	500	.088	.980	.737	.114	.987	.805	.527	.638	.225	.147	.014
	1000	.103	.998	.903	.118	.998	.929	.631	.608	.214	.137	.014

Table 4: EMPIRICAL POWER, $\alpha = 1$. For the FMOLS-based tests (CS and S_2^+), the non data-dependent bandwidths $M1 = [4(T/100)^{1/4}], M2 = [2T^{1/3}]$ and the data-dependent (DD) bandwidth are used. The DD bandwidth is calculated using the OLS residuals in (1). The 95% critical values are 0.2210 for S_2^+ and 1.0413 for CS. For the $KPSS^{Fb}$ test, DD bandwidth is calculated using the IMOLS residuals, and the fixed-b critical value corresponding to the specific value of $b = DD/T$ is used in each replication. The number of simulation replications are $5,000$.

The $KPSS^{Fb}$ test is conducted using fixed-b critical value corresponding to specific values of b. With the fixed-b rule $(M = bT)$, fixed-b inference exhibits good size properties (see Table 3 and the solid lines in Figure 1[8]). As shown in Table 3, using large values of b helps mitigates size distortion. However, the test power deteriorates as b increases (Table 4), and a large bandwidth can lead to very low power, which may be smaller than empirical size. Figure 1 illustrates this point. In Figure 1, for $b = 0.2$, the distribution of the test statistic under H_1 (dashed line) is located left to the distribution under H_0 (solid line). A similar

[8]The solid lines in Figure 1 are the finite sample distributions of the statistic under DGP1 for $b \in \{0.02, 0.08, 0.2\}$ with $T = 500$. The fixed-b critical values are marked on the horizontal axis by $c1$ (0.0499), $c2$ (0.0577), and $c3$ (0.1147) for $b = 0.02$, 0.08, and 0.2, respectively.

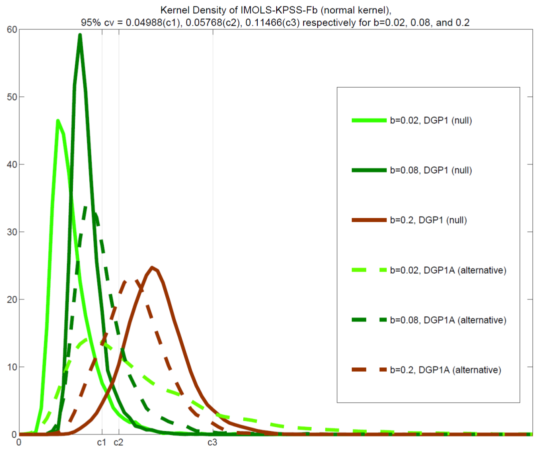

Figure 1: KERNEL DENSITY OF $KPSS^{Fb}$, DGP1 AND DGP1A, $T = 500$. The solid lines are the kernel density of the test statistic for $b = 0.02, 0.08$, and 0.2 under DGP1. The dashed lines are the kernel density of the test statistic for $b = 0.02, 0.08$, and 0.2 under DGP1A.

pattern may be presumably found in the distributions of the fixed-b limits given in (5) and (6).

As shown in Table 3, the $KPSS^{Fb}$ test controls the size reasonably well when the *DD* bandwidth is employed. Notably, simulation results reveal an interesting characteristic of the IMOLS residual: the *DD* bandwidth does not yield low power for the $KPSS^{Fb}$ test, unlike for the FMOLS-based tests. Theorem 2 indicates that the magnitude of the *DD* bandwidth can account for this outcome: the *DD* bandwidth calculated with $\triangle\widetilde{S}_t^u$ is $O_p\left(T^{1/3}\right)$ under H_1, rather than $O_p\left(T\right)$.[9] Consequently, the value of b given by the *DD* scheme shrinks to zero as the sample size grows, thereby ensuring substantial test power. However, as shown in (8) the limit of the AR(1) coefficient estimate heavily depends on the first order autocorrelation of $\triangle x_t$ under H_1. This implies that test power may be crucially affected by the temporal dependence of $\triangle x_t$. The empirical power reported in Table 4 also makes this point. For DGP7A, which only differs in the value of θ from DGP5A and 6A, the rejection rate is only 24.9% (for T=500). For DGP5A and DGP6A, the rejection rates are 52.7% and 66.0%, respectively.

[9]See Xiao and Phillips (2002).

4.3. MODIFIED RESIDUALS

In this subsection some modified residuals ($\triangle\breve{S}_t^{m_1}$ and $\triangle\breve{S}_t^{m_2}$) are examined whether they can improve the power of the test. Recall that $\triangle\widetilde{S}_t^u = \widehat{u}_t - \triangle x_t'\widehat{\gamma}$. Consider the following two modified residuals:

$$\triangle\breve{S}_t^{m_1} = \widehat{u}_t - T^{-c}\triangle x'\widehat{\gamma} \text{ with } c \in (0,1/2), \text{ and}$$

$$\triangle\breve{S}_t^{m_2} = \widehat{u}_t - T^{-c}v_t^{s\prime}\widehat{\gamma}, \text{ with } c \in [0,1/2)$$

where $\{v_t^s\}$ is a sequence of IID pseudo random numbers with zero mean and unit variance. Denote the coefficient estimate in the AR(1) regression with $\left\{\triangle\breve{S}_t^{m_1}\right\}$ and $\left\{\triangle\breve{S}_t^{m_2}\right\}$ by $\widehat{\phi}^{m_1}(c)$ and $\widehat{\phi}^{m_2}(c)$, respectively, and the corresponding Andrews' *DD* bandwidths by $DD^{m_1}(c)$ and $DD^{m_2}(c)$. Note that with $c = 0$, $\triangle\breve{S}_t^{m_1}$ is identical to $\triangle\widetilde{S}_t^u$. As c increases, both $\triangle\breve{S}_t^{m_1}$ and $\triangle\breve{S}_t^{m_2}$ come close to $\widehat{u}_t$, and $\widehat{\phi}^{m_1}(c)$ and $\widehat{\phi}^{m_2}(c)$ is mainly determined by the temporal dependence in $\widehat{u}_t$, which may lead to smaller size distortions for null DGPs with α being close to 1, but lower the test power under H_1. Hence there is a trade-off between size and power: the higher c, less size distortion but lower power.

By applying Lemma 1 in the Appendix, one can show that for $0 < c < \frac{1}{2}$,

$$\widehat{\phi}^{m_1}(c) = \frac{\frac{1}{T}\sum_{t=3}^{T} u_t u_{t-1} + o_p(1)}{\frac{1}{T}\sum_{t=2}^{T-1} u_t^2 + o_p(1)} \xrightarrow{p} \frac{\Sigma_{u,u-1}}{\Sigma_{uu}} = corr(u_t, u_{t-1}) \text{ under } H_0,$$

and

$$\widehat{\phi}^{m_1}(c) = \frac{\frac{1}{T^{3-2c}}\widehat{\gamma}'\sum_{t=3}^{T} T^{-2c}\triangle x_t \triangle x_{t-1}'\widehat{\gamma} + o_p(1)}{\frac{1}{T^{3-2c}}\widehat{\gamma}'\sum_{t=2}^{T-1} T^{-2c}\triangle x_t \triangle x_t'\widehat{\gamma} + o_p(1)} \xrightarrow{d} \frac{\gamma_1^{\infty\prime}\Sigma_{v,v-1}\gamma_1^{\infty}}{\gamma_1^{\infty\prime}\Sigma_{vv}\gamma_1^{\infty}} \text{ under } H_1.$$

Thus, under null DGPs, $\widehat{\phi}^{m_1}(c)$ only reflects the persistence in u_t in the limit. For the concerned case, where α is large, it can be expected that plugging in $\widehat{\phi}^{m_1}(c)$ in (7) may produce a larger bandwidth than using $\widehat{\phi}$. Using large bandwidths can lead to better size accuracy in the fixed-b inference as demonstrated in Table 1. Under H_1, the limit of $\widehat{\phi}^{m_1}(c)$ stays the same as that of $\widehat{\phi}$. Next, denote $\Sigma_{v,v-1}^s = E\left(v_t^s v_{t-1}^{s\prime}\right)$, and $\Sigma_{vv}^s = \mathrm{var}(v_t^s)$ and note that $\Sigma_{v,v-1}^s = \mathbf{0}$ by construction. It can be shown that

$$\widehat{\phi}^{m_2}(c) \xrightarrow{d} \begin{cases} \frac{\Sigma_{u,u-1}}{\Sigma_{uu}} = corr(u_t, u_{t-1}) & \text{for } 0 < c < \frac{1}{2} \\ \frac{\Sigma_{u,u-1}}{\Sigma_{uu} + \gamma_0^{\infty\prime}\gamma_0^{\infty}} & \text{for } c = 0 \end{cases} \text{ under } H_0,$$

and

$$\widehat{\phi}^{m_2}(c) \xrightarrow{d} \frac{\gamma_1^{\infty\prime}\Sigma_{v,v-1}^{s}\gamma_1^{\infty}}{\gamma_1^{\infty\prime}\Sigma_{vv}^{s}\gamma_1^{\infty}} = 0 \text{ for } 0 \le c < \frac{1}{2} \text{ under } H_1.$$

Note that the limit of $\widehat{\phi}^{m_2}(c)$ is given by zero under H_1, regardless of the value of θ. Hence the limit of $\widehat{\alpha}(1)$ is zero, and $DD^{m_2}(c) = o_p\left(T^{1/3}\right)$. Therefore, it can be expected that the test power may be further improved, and the power may be less sensitive to the persistence in $\triangle x_t$.

Table 5 reports the simulation results when employing the two modified residuals to obtain data driven bandwidths. As predicted by the above analysis, $DD^{m_1}(c)$ leads to less size distortion but lower power, compared to the original *DD* bandwidth. Unreported simulation results indicate that the power decreases with c and varies depending on the value of θ. Notably, the $DD^{m_2}(c)$ delivers substantially improved power, with a significant reduction in the dependence of test power on θ, compared to both the original *DD* and $DD^{m_1}(c)$. In particular, Table 5 demonstrates that the $DD^{m_2}(c)$ bandwidth with $c = 0.05$ delivers less size distortion and more balanced test performance for the $KPSS^{Fb}$ test if the sample size is greater than 200.

Table 6 presents empirical size for additional null DGPs with greater persistence in the regression error ($\alpha = 0.95$) along with the empirical power ($\alpha = 1$). The test power of *CS* is much higher, compared to $KPSS^{Fb}$. For the persistent null DGPs, FMOLS-based tests experience severe size distortions, where the over-rejection problem do not improve with the sample size. The $KPSS^{Fb}$ test may also suffer from the size distortion, but the size accuracy improves with the sample size. The degree of over-rejection with $KPSS^{Fb}$ depends on the DGP parameter values of ρ and θ as well as on the choice of data-dependent bandwidth rule. Overall, for the DGPs with $\rho = 0.5$, the FMOLS-based *CS* test with *M*2 bandwidth delivers more balanced performance between size and power although the test is subject to size distortion. The $KPSS^{Fb}$ test may display rejection rates close to 5% when the $DD^{m_2}(c)$ bandwidth is used with a large value of c (i.e. c =0.2). However, this size improvement comes at the cost of substantial power loss. When the degree of endogeneity is relatively strong ($\rho = 0.8$), the $KPSS^{Fb}$ test with $DD^{m_2}(0)$ and $DD^{m_2}(0.05)$ exhibits a substantially improved size accuracy.[10] The power is lower than the *CS* test, but the size-power tradeoff does not seem too costly.

[10]Table 6 shows that the performance of the $KPSS^{Fb}$ test under persistent null DGPs ($\alpha = 0.95$) gets worse for $\rho = 0.5$, i.e. more over-rejections. Note that the IMOLS residuals $\triangle\widetilde{S}_t^u$ equals $\widehat{u}_t$ minus $\triangle x_t'\widehat{\gamma}$. Hence the poor size performance may be because the persistence in the IMOLS residuals and the modified residuals gets higher as ρ gets smaller.

		DD	$DD^{m_1}(0.05)$				$DD^{m_2}(0)$			
(ρ,α,θ)	DGP \ T	1000	100	200	500	1000	100	200	500	1000
$(0,.5,.5)$	DGP1	.065	.141	.082	.071	.064	.154	.092	.077	.071
$(0,1,.5)$	DGP1A	.626	.186	.257	.460	.594	.352	.505	.768	.858
$(.5,.5,.5)$	DGP2	.068	.139	.086	.074	.068	.171	.112	.090	.078
$(.5,.9,.5)$	DGP3	.138	.145	.133	.125	.110	.279	.302	.286	.243
$(.5,1,.5)$	DGP3A	.626	.192	.262	.469	.595	.354	.516	.763	.858
$(0,.8,.8)$	DGP4	.069	.056	.050	.067	.069	.077	.069	.083	.081
$(0,1,.8)$	DGP4A	.401	.057	.064	.187	.342	.132	.248	.552	.770
$(.8,.9,.2)$	DGP5	.015	.267	.143	.035	.013	.370	.206	.050	.016
$(.8,1,.2)$	DGP5A	.687	.372	.464	.623	.675	.480	.612	.750	.782
$(.8,.9,.5)$	DGP6	.023	.135	.087	.038	.019	.310	.209	.074	.030
$(.8,1,.5)$	DGP6A	.631	.195	.261	.478	.595	.386	.526	.764	.859
$(.8,.9,.8)$	DGP7	.063	.039	.036	.053	.047	.127	.164	.140	.112
$(.8,1,.8)$	DGP7A	.405	.061	.066	.193	.348	.178	.294	.568	.765

			$DD^{m_2}(0.05)$				$DD^{m_2}(0.2)$			
(ρ,α,θ)	DGP \ T		100	200	500	1000	100	200	500	1000
$(0,.5,.5)$	DGP1		.145	.085	.074	.068	.137	.079	.072	.064
$(0,1,.5)$	DGP1A		.263	.388	.645	.770	.115	.144	.308	.468
$(.5,.5,.5)$	DGP2		.156	.099	.082	.072	.137	.086	.075	.068
$(.5,.9,.5)$	DGP3		.211	.216	.201	.173	.107	.092	.098	.082
$(.5,1,.5)$	DGP3A		.270	.398	.647	.772	.118	.152	.320	.473
$(0,.8,.8)$	DGP4		.065	.058	.074	.074	.045	.045	.067	.068
$(0,1,.8)$	DGP4A		.086	.149	.386	.612	.052	.032	.094	.207
$(.8,.9,.2)$	DGP5		.313	.163	.038	.014	.216	.108	.028	.012
$(.8,1,.2)$	DGP5A		.411	.525	.694	.743	.236	.286	.471	.573
$(.8,.9,.5)$	DGP6		.228	.147	.052	.023	.127	.074	.037	.019
$(.8,1,.5)$	DGP6A		.300	.415	.652	.771	.148	.176	.341	.481
$(.8,.9,.8)$	DGP7		.090	.111	.101	.080	.051	.043	.060	.052
$(.8,1,.8)$	DGP7A		.119	.203	.433	.618	.053	.056	.148	.257

Table 5: EMPIRICAL REJECTION RATES, $KPSS^{Fb}$ WITH $DD^{m_1}(c)$ AND $DD^{m_2}(c)$. This table reports the empirical size ($\alpha < 1$) and power ($\alpha = 1$) for the $KPSS^{Fb}$ test. The sample sizes are $100, 200, 500$, and 1000. The DD, $DD^{m1}(\cdot)$, and $DD^{m2}(\cdot)$ bandwidths are calculated using the IMOLS residuals and the modified IMOLS residuals (See Section 4.3). The number of simulation replications are $5,000$.

	CS		*KPSS^{Fb}*							
	M2		*DD*		$DD^{m_2}(0)$		$DD^{m_2}(0.05)$		$DD^{m_2}(0.2)$	
$(\rho,\alpha,\theta)\setminus T$	500	1000	500	1000	500	1000	500	1000	500	1000
(.5, .95, .2)	.296	.300	.443	.404	.548	.520	.506	.411	.199	.168
(.5, 1, .2)	.731	.905	.653	.681	.751	.784	.689	.750	.457	.575
(.5, .95, .5)	.287	.295	.275	.245	.490	.466	.360	.202	.130	.080
(.5, 1, .5)	.730	.903	.517	.626	.763	.858	.647	.772	.320	.473
(.5, .95, .8)	.262	.275	.106	.113	.281	.280	.180	.183	.063	.072
(.5, 1, .8)	.715	.899	.249	.405	.558	.773	.403	.625	.108	.212
(.8, .95, .2)	.386	.364	.137	.038	.187	.054	.137	.038	.071	.020
(.8, 1, .2)	.742	.905	.660	.687	.750	.782	.694	.743	.471	.573
(.8, .95, .5)	.369	.351	.100	.033	.213	.071	.140	.045	.061	.021
(.8, 1, .5)	.737	.903	.527	.631	.764	.859	.652	.771	.341	.481
(.8, .95, .8)	.310	.314	.075	.045	.444	.333	.124	.068	.053	.036
(.8, 1, .8)	.717	.895	.249	.405	.568	.765	.433	.618	.148	.257

Table 6: EMPIRICAL SIZE AND POWER, $\alpha = 0.95$ AND 1. This table reports the empirical size ($\alpha = 0.95$) and power ($\alpha = 1$) for the CS and $KPSS^{Fb}$ test. The sample sizes are 500 and 1000. The non data-dependent bandwidth M2 for the CS test is given by $[2T^{1/3}]$. For the $KPSS^{Fb}$ test, *DD*, $DD^{m1}(\cdot)$, and $DD^{m2}(\cdot)$ bandwidths are calculated using the IMOLS residuals and the two modified IMOLS residuals, respectively (See Section 4.3), and the inference is conducted with fixed-*b* critical values. The number of simulation replications are $5,000$.

5. CONCLUSION

This study focuses on IMOLS residual-based cointegration tests. The proposed $KPSS^{Fb}$ test, constructed using the IMOLS residual $\triangle\widetilde{S}_t^u$, is shown to be consistent under the traditional small-*b* asymptotic framework. Additionally, its fixed-*b* null limit distribution is pivotal, enabling fixed-*b* inference. Simulation results indicate that the $KPSS^{Fb}$ test works reasonably well with a fixed value of *b* being used, if *b* is not large. However, the power of the test with a large value of *b* (e.g. $b = 0.2$) is very low. In contrast, using the Andrews' AR(1) plug-in data-driven bandwidth and fixed-*b* critical values, the $KPSS^{Fb}$ test performs reasonably well, in terms of both size and power. A key finding is that the *DD* bandwidth scheme, when applied to IMOLS residual $\triangle\widetilde{S}_t^u$, results in a bandwidth that can produce a consistent test, whereas FMOLS-based tests are no longer consistent with the *DD* scheme. However, it is shown that the *DD* bandwidth and test power are overly sensitive to the degree of temporal dependence in $\triangle x_t$: the test power may be low (high) for the DGPs with a large (small) value of θ. To address this issue, two modified residuals are proposed to obtain alternative data-driven bandwidths. Simulation results demonstrate that the bandwidth based on the modified residual $\triangle \breve{S}_t^{m_2}$ can improve test power, and the power is less affected by the persistence in $\triangle x_t$.

From a practitioner's perspective, the FMOLS-based tests (CS and S_2^+) with a relatively large non data-dependent bandwidth such as $M2$ may be a reasonable choice as long as the persistence in the regression error is not high (Table 3). However, one should note that the data-dependent bandwidth should be avoided due to the low power (Table 4). In Table 5, the $KPSS^{Fb}$ test with the DD or modified DD bandwidths exhibit better size property compared to the FMOLS-based tests, as long as the sample size is not small ($T \geq 200$). However, the power is much lower than those of the FMOLS-based tests. A reasonable bandwidth choice for the $KPSS^{Fb}$ test may be the $DD^{m_2}(c)$ bandwidth with $c = 0.05$, which which the test shows moderate size distortion and more balanced test performance, if the sample size is not small .

When the persistence of the regression error is high, for example $\alpha = 0.95$, the FMOLS-based tests also suffer from severe size distortions and over-rejection is not alleviated even in large samples (Table 6). The simulation results show that the performance of the $KPSS^{Fb}$ test may depend on the degree of endogeneity (ρ): (1) when ρ is not large (i.e. ρ =0.5), the $KPSS^{Fb}$ test shows smaller size distortion in the simulation if the modified bandwidth $DD^{m_2}(c)$ is applied with a relatively large value of c such as 0.2 (Table 6). But it should be noted that this bandwidth choice leads to low power; (2) when ρ is large (ρ =0.8), the $KPSS^{Fb}$ test shows more balanced performance compared to the CS test, particularly if the $DD^{m_2}(c)$ bandwidth is used with $c = 0.05$.[11]

REFERENCES

Andrews, D. W. K. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica* 59, 817-854.

Caner, M. and Kilian, L. (2001). "Size distortions of tests of the null hypothesis of stationarity: Evidence and implications for the PPP debate," *Journal of International Money and Finance* 20, 639-657.

[11] The bandwidth ratio (b) is given by $DD^{m_2}(0.05)/T$ and the fixed-b critical values should be used.

Cho, C. K. and Vogelsang, T. J. (2017). “Fixed-b inference for testing structural change in a time series regression,” *Econometrics* 5, 2.

Choi, I. and Ahn, B. C. (1995). “Testing for cointegration in a system of equations,” *Econometric Theory* 11, 952-983.

Gabriel, V. J. (2003). “Tests for the null hypothesis of cointegration: A Monte Carlo comparison,” *Econometric Reviews* 22, 411-435.

Hansen, B. E. (1992). “Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends,” *Journal of Econometrics* 53, 87-121.

Harris, D. and Inder, B. (1994). “A test of the null hypothesis of cointegration,” *Non-stationary time series analysis and cointegration*, 133-152.

Kiefer, N. M. and Vogelsang, T. J. (2005). “A new asymptotic theory for heteroskedasticity-autocorrelation robust tests,” *Econometric Theory* 21, 1130-1164.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root,” *Journal of Econometrics* 54, 154-178.

Müller, U. K. (2005). “Size and power of tests of stationarity in highly autocorrelated time series,” *Journal of Econometrics* 128, 195-213.

Müller, U. K. (2014). “HAC corrections for strongly autocorrelated time series,” *Journal of Business & Economic Statistics* 32, 311-322.

Newey, W. K. and West, K. D. (1987). “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica* 55, 703-708.

Phillips, P. C. B. and Loretan, M. (1991). “Estimating long run economic equilibria,” *Review of Economic Studies* 58, 407-436.

Phillips, P. C. B. (1988). “Spectral regression for cointegrated time series,” Cowles Foundation for Research in Economics at Yale University.

Shin, Y. (1994). “A residual-based test of the null of cointegration against the alternative of no cointegration,” *Econometric Theory* 10, 91-115.

Vogelsang, T. J. and Wagner, M. (2014). "Integrated modified OLS estimation and fixed-b inference for cointegrating regressions," *Journal of Econometrics* 178, 741-760.

Xiao, Z. and Phillips, P. C. B. (2002). "A CUSUM test for cointegration using regression residuals," *Journal of Econometrics* 108, 43-61.

A. MATHEMATICAL PROOFS

Proof of Theorem 1. (*a*) See Vogelsang and Wagner (2014, Lemma 2).

(*b*) See Vogelsang and Wagner (2014, Theorem 3).

(*c*) Parts (*a*) and (*b*) immediately yield the desired result.

(a') To derive the limit of the IMOLS estimator $\widehat{\theta} = (\widehat{\delta}', \widehat{\beta}', \widehat{\gamma}')'$ under the alternative of non-cointegration, denote $\theta_0 = (\delta', \beta', \mathbf{0})'$ and note that

$$\widehat{\theta} = (S^{\widetilde{x}\prime} S^{\widetilde{x}})^{-1},\ S^{\widetilde{x}\prime} S^{y} = (S^{\widetilde{x}\prime} S^{\widetilde{x}})^{-1} S^{\widetilde{x}\prime} (S^{f\prime}\delta + S^{x\prime}\beta + X'\cdot\mathbf{0} + S^{u}).$$

Then it can be shown

$$T^{-1}A_{IM}^{-1}(\widehat{\theta}-\theta_0) = T^{-1}A_{IM}^{-1}(S^{\widetilde{x}\prime} S^{\widetilde{x}})^{-1} S^{\widetilde{x}\prime} S^{u} = (T^{-2}A_{IM}S^{\widetilde{x}\prime} S^{\widetilde{x}} A_{IM})^{-1} T^{-3}A_{IM}S^{\widetilde{x}\prime} S^{u}$$
$$\xrightarrow{d} \sigma_{\varepsilon}\Pi^{-1}\left(\int_0^1 g(s)g(s)'ds\right)^{-1}\int_0^1 g(s)\int_0^s w_{\varepsilon}(u)duds.$$

Next, consider the residuals $\widetilde{S}_t^u = -S_t^{\widetilde{x}\prime}(\widehat{\theta}-\theta_0) + S_t^u$ with $S_t^{\widetilde{x}\prime} = (S_t^{f\prime}, S_t^{x\prime}, x_t')$. Then for the differenced residuals $\triangle\widetilde{S}_t^u = -\widetilde{x}_t'(\widehat{\theta}-\theta_0) + u_t$ with $\widetilde{x}_t' = (f_t', x_t', \triangle x_t')$, it can be shown that

$$T^{-3/2}\sum_{t=2}^{[rT]}\triangle\widetilde{S}_t^u = T^{-3/2}\sum_{t=2}^{[rT]}u_t - T^{-3/2}\sum_{t=2}^{[rT]}\widetilde{x}_t'(\widehat{\theta}-\theta_0)$$
$$= T^{-3/2}\sum_{t=2}^{[rT]}u_t - T^{-1/2}\sum_{t=2}^{[rT]}\widetilde{x}_t'A_{IM}\left[T^{-1}A_{IM}^{-1}(\widehat{\theta}-\theta_0)\right]$$
$$\Rightarrow \sigma_{\varepsilon}\int_0^r w_{\varepsilon}(s)ds - \sigma_{\varepsilon}g(r)'\left(\int_0^1 g(s)g(s)'ds\right)^{-1}\int_0^1 g(s)\int_0^s w_{\varepsilon}(u)duds$$
$$\equiv \sigma_{\varepsilon}\widetilde{H}(r).$$

(b') Consider the HAC estimator $\widetilde{\sigma}_{uv}^2 = \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) \frac{1}{T} \sum_t \triangle \widetilde{S}_t^u \triangle \widetilde{S}_{t+h}^u$ with $2 \leq t, t+h \leq T$. Substituting $\triangle \widetilde{S}_t^u = -\widetilde{x}_t'(\widehat{\theta} - \theta_0) + u_t$ yields

$$\widetilde{\sigma}_{uv}^2 = \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) \left(\frac{1}{T} \sum_t u_t u_{t+h} - \frac{1}{T} \sum_t u_t \widetilde{x}_{t+h}'(\widehat{\theta} - \theta_0) \right. \tag{9}$$
$$\left. - \frac{1}{T} \sum_t u_{t+h} \widetilde{x}_t'(\widehat{\theta} - \theta_0) + \frac{1}{T} \sum_t (\widehat{\theta} - \theta_0)' \widetilde{x}_t \widetilde{x}_{t+h}'(\widehat{\theta} - \theta_0) \right).$$

To prove that $\frac{1}{MT} \widetilde{\sigma}_{uv}^2$ is $O_p(1)$, consider the first term in (9):

$$\frac{1}{MT} \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) \frac{1}{T} \sum_t u_t u_{t+h} = \frac{1}{M} \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) \frac{1}{T} \sum_t \frac{u_t}{T^{1/2}} \frac{u_{t+h}}{T^{1/2}}. \tag{10}$$

By the same arguments in Phillips (1988, page 432) or Xiao and Phillips (2002, page 59), the expression in (10) converges in distribution to $2\pi K(0) \int_0^1 \sigma_\varepsilon^2 w_\varepsilon(r)^2 dr$. The second term in (9) can be shown to converge as

$$\begin{aligned}
&\frac{1}{MT} \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) \frac{1}{T} \sum_t u_t \widetilde{x}_{t+h}'(\widehat{\theta} - \theta_0) \\
&= \frac{1}{M} \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) \frac{1}{T} \sum_t u_t \widetilde{x}_{t+h}' A_{IM} \frac{1}{T} A_{IM}^{-1}(\widehat{\theta} - \theta_0) \\
&= \frac{1}{M} \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) \left(\frac{1}{T} \sum_t \frac{u_t}{\sqrt{T}} f_{t+h}' \tau_F^{-1}, \ \frac{1}{T} \sum_t \frac{u_t}{\sqrt{T}} \frac{x_{t+h}'}{\sqrt{T}}, \ \sum_t \frac{u_t}{\sqrt{T}} \frac{\triangle x_{t+h}'}{\sqrt{T}} \right) \\
&\quad \times \frac{1}{T} A_{IM}^{-1}(\widehat{\theta} - \theta_0) \\
&\xrightarrow{d} \left(2\pi K(0) \int_0^1 \sigma_\varepsilon w_\varepsilon(r) f(r)' dr, \ 2\pi K(0) \int_0^1 \sigma_\varepsilon w_\varepsilon(r) B_v(r)' dr, \right. \\
&\quad \left. 2\pi K(0) \int_0^1 \sigma_\varepsilon w_\varepsilon(r) dB_v(r)' + \Omega_{\varepsilon v}' \right) \\
&\quad \times \Pi^{-1} \left(\int_0^1 g_1(s) g_1(s)' ds \right)^{-1} \int_0^1 g(s) \int_0^s \sigma_\varepsilon w_\varepsilon(u) du ds.
\end{aligned}$$

Finally, the last term in (9) equals

$$
\begin{aligned}
&\frac{1}{MT}\sum_{h=-M}^{M} k\left(\frac{h}{M}\right)\frac{1}{T}\sum_{t}(\widehat{\theta}-\theta_0)'\widetilde{x}_t\widetilde{x}_{t+h}'(\widehat{\theta}-\theta_0) \qquad (11)\\
&=\frac{1}{M}\sum_{h=-M}^{M} k\left(\frac{h}{M}\right)\frac{1}{T}\sum_{t}(\widehat{\theta}-\theta_0)'\frac{1}{T}A_{IM}^{-1}(\sqrt{T}A_{IM}\widetilde{x}_t\widetilde{x}_{t+h}'A_{IM}\sqrt{T})\frac{1}{T}A_{IM}^{-1}(\widehat{\theta}-\theta_0)\\
&=(\widehat{\theta}-\theta_0)'\frac{1}{T}A_{IM}^{-1}\left[\frac{1}{M}\sum_{h=-M}^{M} k\left(\frac{h}{M}\right)\frac{1}{T}\sum_{t}\sqrt{T}A_{IM}\widetilde{x}_t\widetilde{x}_{t+h}'A_{IM}\sqrt{T}\right]\frac{1}{T}A_{IM}^{-1}(\widehat{\theta}-\theta_0).
\end{aligned}
$$

Here, the expression in the bracket of the last line equals

$$
\frac{1}{M}\sum_{h=-M}^{M} k\left(\frac{h}{M}\right)\times\begin{pmatrix}
\frac{1}{T}\sum_t \tau_F^{-1}f_t f_{t+h}'\tau_F^{-1} & \frac{1}{T}\sum_t \tau_F^{-1}f_t T^{-1/2}x_{t+h}' & \sum_t \tau_F^{-1}f_t T^{-1/2}\triangle x_{t+h}'\\
\frac{1}{T}\sum_t T^{-1/2}x_t f_{t+h}'\tau_F^{-1} & \frac{1}{T}\sum_t T^{-1/2}x_t T^{-1/2}x_{t+h}' & \sum_t T^{-1/2}x_t T^{-1/2}\triangle x_{t+h}'\\
\sum_t T^{-1/2}\triangle x_t f_{t+h}'\tau_F^{-1} & \sum_t T^{-1/2}\triangle x_t T^{-1/2}x_{t+h}' & \sum_t T^{-1/2}\triangle x_t T^{-1/2}\triangle x_{t+h}'
\end{pmatrix},
$$

which converges in distribution to

$$
\begin{pmatrix}
2\pi K(0)\int_0^1 f(r)f(r)'dr & 2\pi K(0)\int_0^1 f(r)B_v(r)'dr & 2\pi K(0)\int_0^1 f(r)dB_v(r)'\\
2\pi K(0)\int_0^1 B_v(r)f(r)'dr & 2\pi K(0)\int_0^1 B_v(r)B_v(r)'dr & 2\pi K(0)\int_0^1 B_v(r)dB_v(r)'+\Omega_{vv}\\
2\pi K(0)\int_0^1 dB_v(r)f(r)' & 2\pi K(0)\int_0^1 dB_v(r)B_v(r)'+\Omega_{vv} & \mathbf{0}
\end{pmatrix}.
$$

Note that the last diagonal block in above expression is zero because

$$
\frac{1}{M}\sum_{h=-M}^{M} k\left(\frac{h}{M}\right)\sum_{t}T^{-1/2}\triangle x_t T^{-1/2}\triangle x_{t+h}' = \frac{1}{M}\sum_{h=-M}^{M} k\left(\frac{h}{M}\right)\frac{1}{T}\sum_{t}\triangle x_t\triangle x_{t+h}'
$$

and this is nothing more than $\frac{1}{M}$ times a HAC estimator of the long-run variance of $\{\triangle x_t = v_t\}$ so that the limit is zero as T and M grow. Therefore, $\frac{1}{MT}\widetilde{\sigma}_{uv}^2 = O_p(1)$ under H_1.

(c') Immediately from parts (a') and (b'),

$$
\begin{aligned}
KPSS^{Fb} &= \frac{T^{-2}\sum_{t=2}^{T}\left(\widehat{S}_t^u-\widehat{S}_1^u\right)^2}{\widetilde{\sigma}_{uv}^2} = \frac{T^{-2}\sum_{t=2}^{T}\left(\sum_{j=2}^{t}\triangle\widetilde{S}_j^u\right)^2}{\widetilde{\sigma}_{uv}^2}\\
&= \frac{T^2T^{-1}\sum_{t=2}^{T}\left(T^{-3/2}\sum_{j=2}^{t}\triangle\widetilde{S}_j^u\right)^2}{\widetilde{\sigma}_{uv}^2} = \frac{T^2O_p(1)}{O_p(MT)} = O_p\left(\frac{T}{M}\right).
\end{aligned}
$$

Hence $KPSS^{Fb}$ diverges under H_1 because $\frac{M}{T}$ shrinks to zero as T grows in the framework of the traditional asymptotics.

Lemma 1. *The stochastic orders of the terms appearing in the expression of* $\widehat{\phi}$ *are given as below.*

(a) $\widehat{\gamma} = O_p(1)$ *under Assumption 1 (null of cointegration); and* $O_p(T)$ *under Assumption 2 (alternative of non-cointegration).*

(b) $\sum_{t=1}^{T} \widehat{u}_t^2 = O_p(T)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(c) $\sum_{t=1}^{T} u_t \triangle x_t' \widehat{\gamma} = O_p(T)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(d) $\sum_{t=1}^{T} (f_t' x_t') (\widehat{\theta}_1 - \theta_1) \triangle x_t' \widehat{\gamma} = O_p(1)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(e) $\widehat{\gamma}' \sum_{t=1}^{T} \triangle x_t \triangle x_t' \widehat{\gamma} = O_p(T)$ *under Assumption 1; and* $O_p(T^3)$ *under Assumption 2.*

(f) $\sum_{t=1}^{T} u_t u_{t-1} = O_p(T)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(g) $\sum_{t=1}^{T} u_t (f_{t-1}' \; x_{t-1}') (\widehat{\theta}_1 - \theta_1) = O_p(1)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(h) $\sum_{t=1}^{T} (f_t' \; x_t') (\widehat{\theta}_1 - \theta_1) u_{t-1} = O_p(1)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(i) $(\widehat{\theta}_1 - \theta_1)' \sum_{t=1}^{T} \begin{pmatrix} f_t \\ x_t \end{pmatrix} (f_{t-1}' \; x_{t-1}') (\widehat{\theta}_1 - \theta_1) = O_p(1)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(j) $\sum_{t=1}^{T} u_t \triangle x_{t-1}' \widehat{\gamma} = O_p(T)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(k) $\sum_{t=1}^{T} (f_t' \; x_t') (\widehat{\theta}_1 - \theta_1) \triangle x_{t-1}' \widehat{\gamma} = O_p(1)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(l) $\sum_{t=1}^{T} \triangle x_t' \widehat{\gamma} u_{t-1} = O_p(T)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(m) $\widehat{\gamma}' \sum_{t=1}^{T} \triangle x_t (f_{t-1}' \; x_{t-1}') (\widehat{\theta}_1 - \theta_1) = O_p(1)$ *under Assumption 1; and* $O_p(T^2)$ *under Assumption 2.*

(n) $\widehat{\gamma}' \sum_{t=1}^{T} \triangle x_t \triangle x_{t-1}' \widehat{\gamma} = O_p(T)$ *under Assumption 1; and* $O_p(T^3)$ *under Assumption 2.*

Proof of Lemma 1. Define $\overline{A}_{IM} = \text{diag}(T^{-1/2}\tau_F^{-1}, T^{-1}I_k)$. Consider $\widehat{\theta}_1 = (\widehat{\delta}', \widehat{\beta}')'$. By the Frisch-Waugh-Lovell Theorem,

$$\widehat{\theta}_1 = \left(W'(I - P_X)W\right)^{-1} W'(I - P_X)S^y \text{ with } W = (S^f \vdots S^x),$$

and

$$\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1) = \left(\overline{A}_{IM}\left(W'W - W'P_X W\right)\overline{A}_{IM}\right)^{-1} \times \overline{A}_{IM}\left(W'S^u - W'P_X S^u\right).$$

One can show that $T^{-2}\overline{A}_{IM}W'W\overline{A}_{IM}$, $T^{-2}\overline{A}_{IM}W'P_XW\overline{A}_{IM}$, and $T^{-2}\overline{A}_{IM}(W'S^u - W'P_XS^u)$ are $O_p(1)$, from which it can be deduced that $\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1)$ is $O_p(1)$ under Assumption 1.

Next, suppose that Assumption 2 holds (non-cointegration). Note that

$$\begin{aligned} &T^{-1}\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1) \\ &= \left(T^{-2}\overline{A}_{IM}\left(W'W - W'P_XW\right)\overline{A}_{IM}\right)^{-1} T^{-3}\overline{A}_{IM}\left(W'S^u - W'P_XS^u\right). \end{aligned} \tag{12}$$

The limit of the expression inside the inverse is given by

$$T^{-2}\overline{A}_{IM}(W'W - W'P_XW)\overline{A}_{IM} \xrightarrow{d} \overline{\Pi}\Psi_1\overline{\Pi},$$

where

$$\overline{\Pi} = \begin{pmatrix} I_p & \mathbf{0} \\ \mathbf{0} & \Omega_{vv}^{1/2} \end{pmatrix}$$

and

$$\Psi_1 = \int_0^1 g_1(r)g_1'(r)dr - \int_0^1 g_1(r)w_v'(r)dr\left[\int_0^1 w_v(r)w_v'(r)dr\right]^{-1}\int_0^1 w_v(r)g_1'(r)dr$$

with $g_1(r) \equiv (\int_0^r f(s)'ds, \int_0^r w_v(s)'\, ds)'$. The term outside the inverse in (12) also converges since

$$\begin{aligned} &T^{-3}\overline{A}_{IM}(W'S^u - W'P_XS^u) \\ &\xrightarrow{d} \sigma_\varepsilon\overline{\Pi}\left[\int_0^1 g_1(r)\left(\int_0^r w_\varepsilon(s)ds\right)dr\right. \\ &\quad \left. - \int_0^1 g_1(s)w_v'(s)ds\left(\int_0^1 w_v(s)w_v'(s)ds\right)^{-1}\int_0^1 w_v(r)\left(\int_0^r w_\varepsilon(s)ds\right)dr\right] \\ &\equiv \sigma_\varepsilon\overline{\Pi}\Psi_2. \end{aligned}$$

Therefore, $T^{-1}\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1) \xrightarrow{d} \sigma_\varepsilon\overline{\Pi}^{-1}\Psi_1^{-1}\Psi_2$.

(*a*) Under Assumption 1, $\widehat{\gamma} \xrightarrow{d} \Omega_{vv}^{-1}\Omega_{vu} + \sigma_{uv}\Omega_{vv}^{-1/2}d_\gamma \equiv \gamma_0^\infty$ as shown in Theorem 2 of VW (2014), where d_γ is defined in part (*b*) of Theorem 1. Next, under Assumption 2, by denoting $\theta = (\delta', \beta', 0')'$, one can show

$$\begin{aligned} T^{-1}A_{IM}^{-1}(\widehat{\theta} - \theta) &= \left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1} T^{-3}A_{IM}S^{\widetilde{x}\prime}S^u \\ &\xrightarrow{d} (\Pi')^{-1}\left(\int_0^1 g(s)g(s)'ds\right)^{-1}\Pi^{-1}\left(\Pi\int_0^1 g(s)\int_0^s \sigma_\varepsilon w_\varepsilon(u)duds\right), \end{aligned}$$

from which it is deduced that $T^{-1}\widehat{\gamma} \xrightarrow{d} \Omega_{vv}^{-1/2}\widetilde{d_\gamma} \equiv \gamma_1^\infty$, where $\widetilde{d_\gamma}$ is the last k components of $(\int_0^1 g(s)g(s)'\,ds)^{-1}(\int_0^1 g(s)\int_0^s \sigma_\varepsilon w_\varepsilon(u)\,du ds)$.

(b) Suppose Assumption 1 holds. Recall that $\widehat{u_t} = u_t - (f_t' x_t')(\widehat{\theta}_1 - \theta_1)$ and

$$\widehat{u}_t^2 = u_t^2 + (\widehat{\theta}_1 - \theta_1)' \begin{pmatrix} f_t \\ x_t \end{pmatrix} \times \left(f_t'\ x_t'\right)(\widehat{\theta}_1 - \theta_1) - 2(\widehat{\theta}_1 - \theta_1)' \begin{pmatrix} f_t u_t \\ x_t u_t \end{pmatrix}.$$

The sum over t of the second term, $(\widehat{\theta}_1 - \theta_1)'\overline{A}_{IM}^{-1}\overline{A}_{IM}\sum_{t=1}^T \begin{pmatrix} f_t \\ x_t \end{pmatrix}(f_t'\ x_t')\overline{A}_{IM}\overline{A}_{IM}^{-1}$ $(\widehat{\theta}_1 - \theta_1)$ is $O_p(1)$ since

$$\overline{A}_{IM}\sum_{t=1}^T \begin{pmatrix} f_t \\ x_t \end{pmatrix}\left(f_t'\ x_t'\right)\overline{A}_{IM} \xrightarrow{d} \int_0^1 \begin{pmatrix} f(r) \\ B_v(r) \end{pmatrix}(f(r)B_v(r))dr$$

and $\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1) = O_p(1)$. The sum of the third term, $2(\widehat{\theta}_1 - \theta_1)'\sum_{t=1}^T \begin{pmatrix} f_t u_t \\ x_t u_t \end{pmatrix} =$ $2(\widehat{\theta}_1 - \theta_1)'\overline{A}_{IM}^{-1}\overline{A}_{IM}\sum_{t=1}^T \begin{pmatrix} f_t u_t \\ x_t u_t \end{pmatrix}$ is also $O_p(1)$ because

$$\overline{A}_{IM}\sum_{t=1}^T \begin{pmatrix} f_t u_t \\ x_t u_t \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \int_0^1 f(r)\,dB_u(r) \\ \int_0^1 B_v(r)\,dB_u(r) + \sum_{k=0}^\infty E(v_t u_{t+k}) \end{pmatrix}.$$

So it holds that $\frac{1}{T}\sum_{t=1}^T \widehat{u}_t^2 = \frac{1}{T}\sum_{t=1}^T u_t^2 + o_p(1)$.

Next, under Assumption 2, note that $\sum_{t=1}^T u_t^2$ is $O_p\left(T^2\right)$. Now by noting $T^{-1}\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1) = O_p(1)$, one can show both

$$\frac{1}{T^2}(\widehat{\theta}_1 - \theta_1)'\sum_{t=1}^T \begin{pmatrix} f_t \\ x_t \end{pmatrix}(f_t'\ x_t')(\widehat{\theta}_1 - \theta_1) \quad \text{and} \quad \frac{2}{T^2}(\widehat{\theta}_1 - \theta_1)'\sum_{t=1}^T \begin{pmatrix} f_t u_t \\ x_t u_t \end{pmatrix}$$

are $O_p(1)$. Hence, $\sum_{t=1}^T \widehat{u}_t^2 = O_p\left(T^2\right)$.

(c) Given the result (a) of this Lemma, $\frac{1}{T}\sum_{t=1}^T u_t \triangle x_t'\widehat{\gamma} \xrightarrow{d} \Sigma_{uv}\gamma_0^\infty$ under Assumption 1. Under Assumption 2,

$$\frac{1}{T^2}\sum_{t=1}^T u_t \triangle x_t'\widehat{\gamma} = \frac{1}{T}\sum_{t=1}^T u_t \triangle x_t'\frac{\widehat{\gamma}}{T} \xrightarrow{d} \left(\int_0^1 \sigma_\varepsilon w_\varepsilon(r)\,dB_v'(r) + \sum_{k=0}^\infty E\left(\varepsilon_0 v_k'\right)\right)\gamma_1^\infty.$$

(d) Under Assumption 1, $\sum_{t=1}^T (f_t'\ x_t')(\widehat{\theta}_1 - \theta_1)\triangle x_t'\widehat{\gamma} = (\widehat{\theta}_1 - \theta_1)'\overline{A}_{IM}^{-1}\overline{A}_{IM}\sum_{t=1}^T$ $(f_t'\ x_t')'\triangle x_t'\widehat{\gamma}$ can be shown to converge by using $\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1) = O_p(1)$, part ($a$)

of this Lemma, and by noting

$$\overline{A}_{IM}\sum_{t=1}^{T}\left(f_t'\; x_t'\right)'\triangle x_t' \xrightarrow{d} \int_0^1 f(r)dB_v'(r), \int_0^1 B_v(r)dB_v'(r) + \sum_{k=0}^{\infty} E\left(v_t v_{t+k}'\right).$$

Under Assumption 2,

$$\frac{1}{T^2}\sum_{t=1}^{T}\left(f_t'\; x_t'\right)(\widehat{\theta}_1-\theta_1)\triangle x_t'\widehat{\gamma} = (\widehat{\theta}_1-\theta_1)'\overline{A}_{IM}^{-1}T^{-1}\overline{A}_{IM}^{-1}\sum_{t=1}^{T}(f_t'\; x_t')\triangle x_t'\frac{\widehat{\gamma}}{T}$$

can be shown to converge by noting $T^{-1}\overline{A}_{IM}^{-1}(\widehat{\theta}_1-\theta_1) = O_p(1)$ and by applying part (a) of this Lemma.

(e) With $\frac{1}{T}\sum_{t=1}^{T}\triangle x_t\triangle x_t' \xrightarrow{p} \Sigma_{vv}$, $\frac{1}{T}\widehat{\gamma}'\sum_{t=1}^{T}\triangle x_t\triangle x_t'\widehat{\gamma} \xrightarrow{d} \gamma_0^{\infty\prime}\Sigma_{vv}\gamma_0^{\infty}$ under Assumption 1, and $\frac{1}{T^3}\widehat{\gamma}'\sum_{t=1}^{T}\triangle x_t\triangle x_t'\widehat{\gamma} \xrightarrow{d} \gamma_1^{\infty\prime}\Sigma_{vv}\gamma_1^{\infty}$ under Assumption 2 by part (a).

(f) Under Assumption 1, $\frac{1}{T}\sum_{t=1}^{T}u_t u_{t-1} \xrightarrow{p} \Sigma_{u,u-1}$. Under Assumption 2,

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T}u_t u_{t-1} &= \frac{1}{T^2}\sum_{t=1}^{T}u_t\left(u_t-\varepsilon_t\right) = \frac{1}{T^2}\sum_{t=1}^{T}u_t^2 - \frac{1}{T^2}\sum_{t=1}^{T}u_t\varepsilon_t \\ &= \frac{1}{T^2}\sum_{t=1}^{T}u_t^2 - \frac{1}{T}O_p(1) \xrightarrow{d} \int_0^1 B_\varepsilon(r)^2\,dr.\end{aligned}$$

Thus, $\sum_{t=1}^{T}u_t u_{t-1} = O_p\left(T^2\right)$.

(g) $\sum_{t=1}^{T}u_t\left(f_{t-1}'\; x_{t-1}'\right)(\widehat{\theta}_1-\theta_1) = \sum_{t=1}^{T}u_t\left(f_{t-1}'\; x_{t-1}'\right)\overline{A}_{IM}\overline{A}_{IM}^{-1}(\widehat{\theta}_1-\theta_1)$. Under Assumption 1, it holds that $\overline{A}_{IM}^{-1}(\widehat{\theta}_1-\theta_1) = O_p(1)$, $\sum_{t=1}^{T}u_t f_{t-1}'T^{-\frac{1}{2}}\tau_F^{-1} = \sum_{t=1}^{T}T^{-\frac{1}{2}}u_t f_t'\tau_F^{-1} + o_p(1)$ and $\sum_{t=1}^{T}u_t x_{t-1}'T^{-1} = T^{-1}\sum_{t=1}^{T}u_t x_t' - T^{-1}\sum_{t=1}^{T}u_t\, v_t = O_p(1)$. Therefore, $\sum_{t=1}^{T}u_t\left(f_{t-1}'\; x_{t-1}'\right)(\widehat{\theta}_1-\theta_1) = O_p(1)$. Suppose Assumption 2 holds. Write the expression as

$$\frac{1}{T^2}\sum_{t=1}^{T}u_t(f_{t-1}'\; x_{t-1}')(\widehat{\theta}_1-\theta_1) = \sum_{t=1}^{T}T^{-1}u_t\left(f_{t-1}'\; x_{t-1}'\right)\overline{A}_{IM}T^{-1}\overline{A}_{IM}^{-1}(\widehat{\theta}_1-\theta_1).$$

But, $T^{-1}\overline{A}_{IM}^{-1}(\widehat{\theta}_1-\theta_1) = O_p(1)$,

$$\sum_{t=1}^{T}T^{-1}u_t f_{t-1}'T^{-\frac{1}{2}}\tau_F^{-1} = \frac{1}{T}\sum_{t=1}^{T}T^{-\frac{1}{2}}u_t f_t'\tau_F^{-1} + o_p(1) \xrightarrow{d} \int_0^1 B_\varepsilon(r)f(r)dr$$

and $T^{-2}\sum_{t=1}^{T}u_t x_{t-1}' = \frac{1}{T}\sum_{t=1}^{T}T^{-1/2}u_t T^{-1/2}x_t' + o_p(1) \xrightarrow{d} \int_0^1 B_\varepsilon(r)\,B_v'(r)\,dr.$

(h) Similar arguments in the proof of (g) can be applied.

(i) Similar arguments in the proof of (b) can be applied.

(j) Proof is analogous to (c). Under Assumption 1,

$$\frac{1}{T}\sum_{t=1}^{T} u_t \triangle x'_{t-1}\widehat{\gamma} = \frac{1}{T}\sum_{t=1}^{T} u_t v'_{t-1}\widehat{\gamma} \xrightarrow{d} \Sigma_{u,v-1}\gamma_0^{\infty}.$$

Under Assumption 2,

$$\frac{1}{T^2}\sum_{t=1}^{T} u_t \triangle x'_{t-1}\widehat{\gamma} = \frac{1}{T}\sum_{t=1}^{T} u_t v'_{t-1}\frac{\widehat{\gamma}}{T} = \left(\frac{1}{T}\sum_{t=2}^{T} u_t v'_t + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t v'_{t-1}\right)\frac{\widehat{\gamma}}{T}$$

$$\xrightarrow{d} \left(\int_0^1 B_\varepsilon(r)\,dB'_v(r) + \sum_{k=0}^{\infty} E(\varepsilon_t v'_{t+k}) + \Sigma_{\varepsilon,v-1}\right)\gamma_1^{\infty}$$

with $\Sigma_{\varepsilon,v-1} = p\lim\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t v'_{t-1}\right)$, which yields the desired result.

(k) Suppose Assumption 1 holds. Rewrite the expression as

$$(\widehat{\theta}_1 - \theta_1)\overline{A}_{IM}^{-1}\overline{A}_{IM}\sum_{t=1}^{T}\begin{pmatrix} f_t v'_{t-1} \\ x_t v'_{t-1}\end{pmatrix}\widehat{\gamma} = (\widehat{\theta}_1 - \theta_1)\overline{A}_{IM}^{-1}\begin{pmatrix} \sum_{t=1}^{T} T^{-1/2}\tau_F^{-1} f_t v'_{t-1} \\ \sum_{t=1}^{T} T^{-1} x_t v'_{t-1}\end{pmatrix}\widehat{\gamma}.$$

Note that $\frac{1}{T}\sum_{t=1}^{T} x_t v'_{t-1} = \frac{1}{T}\sum_{t=2}^{T-1} x_t v'_t + \frac{1}{T}\sum_{t=1}^{T} v_t v'_{t-1} \xrightarrow{d} \int_0^1 B_v(r)\,dB'_v(r) + \sum_{k=0}^{\infty} E\left(v_t v'_{t+k}\right) + \Sigma_{v,v-1}$, which, being combined with $\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1) = O_p(1)$ and part (a) of this Lemma, yields the desired result. Under Assumption 2,

$$\frac{1}{T^2}(\widehat{\theta}_1 - \theta_1)\overline{A}_{IM}^{-1}\overline{A}_{IM}\sum_{t=1}^{T}\begin{pmatrix} f_t v'_{t-1} \\ x_t v'_{t-1}\end{pmatrix}\widehat{\gamma} = \frac{1}{T}(\widehat{\theta}_1 - \theta_1)\overline{A}_{IM}^{-1}\overline{A}_{IM}\sum_{t=1}^{T}\begin{pmatrix} f_t v'_{t-1} \\ x_t v'_{t-1}\end{pmatrix}\frac{\widehat{\gamma}}{T},$$

which can be shown to be $O_p(1)$ by noting $T^{-1}\overline{A}_{IM}^{-1}(\widehat{\theta}_1 - \theta_1) = O_p(1)$ and using part (a) of this Lemma.

(l) Under Assumption 1, $\frac{1}{T}\sum_{t=1}^{T}\triangle x'_t\widehat{\gamma}u_{t-1} = \widehat{\gamma}'\frac{1}{T}\sum_{t=1}^{T}\triangle x_t u_{t-1}$ converges to $\gamma_0^{\infty\prime}\Sigma_{v,u-1}$. Under Assumption 2,

$$\frac{1}{T^2}\sum_{t=1}^{T}\triangle x'_t\widehat{\gamma}u_{t-1} = \frac{\widehat{\gamma}'}{T}\frac{1}{T}\sum_{t=1}^{T} v_t(u_t - \varepsilon_t)$$

$$\xrightarrow{d} \gamma_1^{\infty\prime}\left(\int_0^1 B_u(r)dB_v(r) + \sum_{k=0}^{\infty} E(u_t v_{t+k}) - \Sigma_{v\varepsilon}\right),$$

yielding $\sum_{t=1}^{T}\triangle x'_t\widehat{\gamma}u_{t-1} = O_p\left(T^2\right)$.

(m) Proof is analogous to (d).

(n) Under Assumption 1, $\frac{1}{T}\widehat{\gamma}'\sum_{t=1}^{T}\triangle x_t\triangle x_{t-1}'\widehat{\gamma}$ converges to $\gamma_0^{\infty\prime}\Sigma_{v,v-1}\gamma_0^{\infty}$. Under Assumption 2, $\frac{1}{T^2}\widehat{\gamma}'\sum_{t=1}^{T}\triangle x_t\triangle x_{t-1}'\widehat{\gamma}$ converges to $\gamma_1^{\infty\prime}\Sigma_{v,v-1}\gamma_1^{\infty}$

Proof of Theorem 2. Rewrite $\widehat{\phi}$ by using $\triangle\widetilde{S}_t^u = u_t - (f_t'\ x_t')\left(\widehat{\theta}_1 - \theta_1\right) - \triangle x_t'\widehat{\gamma} = \widehat{u}_t - \triangle x_t'\widehat{\gamma}$ as

$$\widehat{\phi} = \frac{\sum_{t=3}^{T}\triangle\widetilde{S}_t^u\triangle\widetilde{S}_{t-1}^u}{\sum_{t=3}^{T}\left(\triangle\widetilde{S}_{t-1}^u\right)^2} = \frac{\sum_{t=3}^{T}\left(\Xi_t^{B1}+\Xi_t^{B2}\right)}{\sum_{t=2}^{T-1}\left(\Xi_t^{A1}+\Xi_t^{A2}\right)},$$

where $\Xi_t^{A1} = \widehat{u}_t^2$,

$$\Xi_t^{A2} = -\left[2\left(u_t - \left(f_t'\ x_t'\right)\left(\widehat{\theta}_1 - \theta_1\right)\right) - \triangle x_t'\widehat{\gamma}\right]\times\triangle x_t'\widehat{\gamma},$$

$$\Xi_t^{B1} = \Big[u_t u_{t-1} - u_t\left(f_{t-1}'\ x_{t-1}'\right)\left(\widehat{\theta}_1 - \theta_1\right) - \left(f_t'\ x_t'\right)\left(\widehat{\theta}_1 - \theta_1\right)u_{t-1} + (\widehat{\theta}_1 - \theta_1)'\left(f_t'\ x_t'\right)'\left(f_{t-1}'\ x_{t-1}'\right)(\widehat{\theta}_1 - \theta_1)\Big], \quad \text{and}$$

$$\Xi_t^{B2} = \Big[-u_t\triangle x_{t-1}'\widehat{\gamma} + \left(f_t'\ x_t'\right)\left(\widehat{\theta}_1 - \theta_1\right)\triangle x_{t-1}'\widehat{\gamma} - \triangle x_t'\widehat{\gamma}u_{t-1} + \triangle x_t'\widehat{\gamma}(f_{t-1}'\ x_{t-1}')(\widehat{\theta}_1 - \theta_1) + \widehat{\gamma}'\triangle x_t\triangle x_{t-1}'\widehat{\gamma}\Big].$$

First, under the null of cointegration, by the results in Lemma 1 in the Appendix, the denominator and numerator of $\widehat{\phi}$, if divided by T, jointly converge in distribution to

$$\frac{1}{T}\sum_{t=2}^{T-1}\left(\Xi_t^{A1}+\Xi_t^{A2}\right) \xrightarrow{d} \Sigma_{uu} - 2\Sigma_{uv}\gamma_0^{\infty} + \gamma_0^{\infty\prime}\Sigma_{vv}\gamma_0^{\infty},$$

and

$$\frac{1}{T}\sum_{t=3}^{T}\left(\Xi_t^{B1}+\Xi_t^{B2}\right) \xrightarrow{d} \Sigma_{u,u-1} - \Sigma_{u,v-1}\gamma_0^{\infty} + \gamma_0^{\infty\prime}\Sigma_{v,v-1}\gamma_0^{\infty} - \gamma_0^{\infty\prime}\Sigma_{v,u-1},$$

respectively. Thus, by the continuous mapping principle,

$$\widehat{\phi} \xrightarrow{d} \frac{\Sigma_{u,u-1} - \Sigma_{u,v-1}\gamma_0^{\infty} + \gamma_0^{\infty\prime}\Sigma_{v,v-1}\gamma_0^{\infty} - \gamma_0^{\infty\prime}\Sigma_{v,u-1}}{\Sigma_{uu} - 2\Sigma_{uv}\gamma_0^{\infty} + \gamma_0^{\infty\prime}\Sigma_{vv}\gamma_0^{\infty}}.$$

Next, under the alternative of non-cointegration, again by using Lemma 1, one can show the denominator and numerator of $\widehat{\phi}$, if divided by T^3, jointly converge in distribution to

$$\frac{1}{T^3}\sum_{t=2}^{T}\left(\Xi_t^{A1}+\Xi_t^{A2}\right)\xrightarrow{d}\gamma_1^{\infty\prime}\Sigma_{vv}\gamma_1^{\infty}\quad\text{and}\quad\frac{1}{T^3}\sum_{t=3}^{T}\left(\Xi_t^{B1}+\Xi_t^{B2}\right)\xrightarrow{d}\gamma_1^{\infty\prime}\Sigma_{v,v-1}\gamma_1^{\infty},$$

respectively, which yields

$$\widehat{\phi}\xrightarrow{d}\frac{\gamma_1^{\infty\prime}\Sigma_{v,v-1}\gamma_1^{\infty}}{\gamma_1^{\infty\prime}\Sigma_{vv}\gamma_1^{\infty}}.$$

B. THE $KPSS^{FB}$ TEST WITH TRENDING STOCHASTIC REGRESSORS, WITH DETRENDING

In the main paper, the $I(1)$ regressors have been assumed to contain no drift (i.e., no deterministic linear trend). This assumption is relaxed in this section. Consider the model

$$y_t = f_t'\delta + x_t'\beta + u_t,\quad x_t = \alpha_x + \delta_x t + x_t^0,$$

with $\triangle x_t^0 = v_t$ and $f_t = (1,t)'$ and maintain Assumption 1. Without a drift ($\delta_x = 0$), it holds that

$$T^{-1/2}A_{IM}S_{[rT]}^{\widetilde{x}} = T^{-1/2}A_{IM}\begin{pmatrix}\sum_{t=1}^{[rT]} f_t\\ \sum_{t=1}^{[rT]} x_t\\ x_{[rT]}\end{pmatrix}\Rightarrow \Pi g(r),$$

from which (3) in the main paper can be derived. However, with a nonzero drift in x_t, this result would not hold because the linear trend in x_t would dominate the other stochastic components.

Now consider a normalizing matrix

$$A_{IM}^{D} = \begin{pmatrix} T^{-1/2}\tau_F^{-1} & \mathbf{0} & \mathbf{0}\\ -T^{-1}D_x & T^{-1}I_k & \mathbf{0}\\ -D_x^F & \mathbf{0} & I_k\end{pmatrix}\ \text{with}\ \left(A_{IM}^{D\prime}\right)^{-1} = \begin{pmatrix} T^{1/2}\tau_F & T^{1/2}\tau_F D_x' & T^{1/2}\tau_F D_x^{F\prime}\\ \mathbf{0} & TI_k & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & I_k\end{pmatrix},$$

where $D_x = \left(\mathbf{0} \vdots \delta_x\right)$ and $D_x^F = \left(\delta_x \vdots \mathbf{0}\right)$.

Theorem 3. *With the linear trends in stochastic regressors, under Assumption 1, it holds that, as* $T \to \infty$, $T^{-1/2} A_{IM}^{D} S_{[rT]}^{\widetilde{x}} \Rightarrow \Pi g(r), r \in [0,1]$,

$$\left(A_{IM}^{D\prime}\right)^{-1}(\widehat{\theta}-\theta_{*})+\begin{pmatrix} T^{1/2}\tau_F D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

$$\xrightarrow{d} \sigma_{uv}\Pi^{-1}\left(\int_0^1 g(s)'g(s)ds\right)^{-1}\int_0^1 g(s)w_{uv}(s)ds,$$

$$T^{-1/2}\sum_{t=2}^{[rT]}\triangle\widetilde{S}_t^u \Rightarrow \sigma_{uv}\widetilde{P}(r),\ r\in[0,1],$$

and $KPSS^{Fb}$ *weakly converges to the limit in (5) with* $f(r) = (1,r)'$.

Proof of Theorem 3.

$$T^{-1/2}A_{IM}^{D}S_{[rT]}^{\widetilde{x}} = T^{-1/2}A_{IM}^{D}\begin{pmatrix}\sum_{t=1}^{[rT]} f_t \\ \sum_{t=1}^{[rT]} x_t \\ x_{[rT]}\end{pmatrix} = \begin{pmatrix} T^{-1}\tau_F^{-1}\sum_{t=1}^{[rT]} f_t \\ -T^{-3/2}D_x\sum_{t=1}^{[rT]} f_t + T^{-3/2}\sum_{t=1}^{[rT]} x_t \\ -T^{-1/2}D_x^F\sum_{t=1}^{[rT]} f_t + T^{-1/2}x_{[rT]}\end{pmatrix}$$

$$= \begin{pmatrix} T^{-1}\tau_F^{-1}\sum_{t=1}^{[rT]} f_t \\ -T^{-3/2}\left(\frac{[rT]([rT]+1)}{2}\right)\delta_x + [rT]T^{-3/2}\alpha_x + T^{-3/2}\left(\frac{[rT]([rT]+1)}{2}\right)\delta_x + T^{-3/2}\sum_{t=1}^{[rT]} x_t^0 \\ -[rT]T^{-1/2}\delta_x + T^{-1}\alpha_x + [rT]T^{-1/2}\delta_x + T^{-1/2}x_{[rT]}^0\end{pmatrix}$$

$$= \begin{pmatrix} T^{-1}\tau_F^{-1}\sum_{t=1}^{[rT]} f_t \\ [rT]T^{-3/2}\alpha_x + T^{-3/2}\sum_{t=1}^{[rT]} x_t^0 \\ T^{-1}\alpha_x + T^{-1/2}x_{[rT]}^0\end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \int_0^r f(s)ds \\ \Omega_{vv}^{-1/2}\int_0^r W_v(s)ds \\ \Omega_{vv}^{-1/2}W_v(r)\end{pmatrix} = \begin{pmatrix} I_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{vv}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Omega_{vv}^{-1/2}\end{pmatrix}\begin{pmatrix}\int_0^r f(s)ds \\ \int_0^r W_v(s)ds \\ W_v(r)\end{pmatrix} = \Pi g(r).$$

Note that premultiplication by A_{IM}^{D} eliminates the nonzero drift (linear trend) in the second and third segment of $A_{IM}^{D}S_{[rT]}^{\widetilde{x}}$. Next, rewrite the IMOLS estimator as

$$\begin{aligned}&\left(A_{IM}^{D\prime}\right)^{-1}(\widehat{\theta}-\theta_{*}) \\ &= \left(T^{-2}A_{IM}^{D}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}^{D\prime}\right)^{-1}\left(T^{-2}A_{IM}^{D}S^{\widetilde{x}\prime}S^{u}\right) - \left(A_{IM}^{D\prime}\right)^{-1}\left(\mathbf{0},\mathbf{0},\Omega_{vu}'\Omega_{vv}^{-1}\right)' \\ &= \left(T^{-2}A_{IM}^{D}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}^{D\prime}\right)^{-1}\left(T^{-2}A_{IM}^{D}S^{\widetilde{x}\prime}S^{u}\right) - \begin{pmatrix} T^{1/2}\tau_F D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \Omega_{vv}^{-1}\Omega_{vu}\end{pmatrix},\end{aligned}$$

and with a further rearrangement,

$$(A_{IM}^{D\prime})^{-1}(\widehat{\theta}-\theta_*)+\begin{pmatrix} T^{1/2}\tau_F D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$
$$=\left(T^{-2}A_{IM}^{D}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}^{D\prime}\right)^{-1}\left(T^{-2}A_{IM}^{D}S^{\widetilde{x}\prime}S^{u}\right)-\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \Omega_{vv}^{-1}\Omega_{vu} \end{pmatrix}.$$

Now, it is straightforward to prove the second result by noting that the limit of the right hand side is $\sigma_{uv}\Pi^{-1}\left(\int_0^1 g(s)'g(s)ds\right)^{-1}\int_0^1 g(s)w_{uv}(s)ds$, as shown by Vogelsang and Wagner (2014). This limit is identical to the null limiting distribution of $A_{IM}^{-1}(\widehat{\theta}-\theta_*)$ in the case of $\delta_x=\mathbf{0}$.

Thereafter, consider the scaled partial sum of $\triangle\widetilde{S}_t^u=u_t-\triangle x_t'\Omega_{vv}^{-1}\Omega_{vu}-\widetilde{x}_t'(\widehat{\theta}-\theta_*)$, where $\triangle x_t'=\delta_x'+\triangle x_t^{0\prime}$, and $\widetilde{x}_t'=(f_t',x_t',\triangle x_t')$.

$$\begin{aligned} &T^{-1/2}\sum_{t=2}^{[rT]}\triangle\widetilde{S}_t^u \\ &=T^{-1/2}\sum_{t=2}^{[rT]}u_t-T^{-1/2}\left(x_{[rT]}'-x_1'\right)\Omega_{vv}^{-1}\Omega_{vu}-T^{-1/2}\sum_{t=2}^{[rT]}\widetilde{x}_t'(\widehat{\theta}-\theta_*) \\ &=T^{-1/2}\sum_{t=2}^{[rT]}u_t-T^{-1/2}\left([rT]\delta_x'-\delta_x'+x_{[rT]}^{0\prime}\right)\Omega_{vv}^{-1}\Omega_{vu} \qquad (13)\\ &-T^{-1/2}\sum_{t=2}^{[rT]}\left(f_t',x_t',\triangle x_t'\right)A_{IM}^{D\prime}\left(A_{IM}^{D\prime}\right)^{-1}\left(\widehat{\theta}-\theta_*\right)+o_p(1). \end{aligned}$$

However, one can rewrite the second summation as

$$\begin{aligned} &\sum_{t=2}^{[rT]}\left(f_t',x_t',\triangle x_t'\right)A_{IM}^{D\prime}\left(A_{IM}^{D\prime}\right)^{-1}(\widehat{\theta}-\theta_*) \\ &=\sum_{t=2}^{[rT]}\left(f_t',x_t',\triangle x_t'\right)A_{IM}^{D\prime}\left[\left(A_{IM}^{D\prime}\right)^{-1}(\widehat{\theta}-\theta_*)+\begin{pmatrix} T^{1/2}\tau_F D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}\right] \\ &-\sum_{t=2}^{[rT]}\left(f_t',x_t',\triangle x_t'\right)A_{IM}^{D\prime}\begin{pmatrix} T^{1/2}\tau_F D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \end{aligned}$$

By noting that $A_{IM}^{D\prime}\begin{pmatrix} T^{1/2}\tau_F D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$, one can show

$$\sum_{t=2}^{[rT]} \left(f_t', x_t', \triangle x_t'\right) A_{IM}^{D\prime} \begin{pmatrix} T^{1/2}\tau_F D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$
$$= \sum_{t=2}^{[rT]} \delta_x'\Omega_{vv}^{-1}\Omega_{vu} = \left([rT]\delta_x' - \delta_x'\right)\Omega_{vv}^{-1}\Omega_{vu}.$$

Hence, the δ_x'-terms in (13) are canceled out, yielding

$$T^{-1/2}\sum_{t=2}^{[rT]} \triangle \widetilde{S}_t^u = T^{-1/2}\sum_{t=2}^{[rT]} u_t - T^{-1/2} x_{[rT]}^{0\prime}\Omega_{vv}^{-1}\Omega_{vu}$$
$$-T^{-1/2}\sum_{t=2}^{[rT]} \left(f_t', x_t', \triangle x_t'\right) A_{IM}^{D\prime}\left[\left(A_{IM}^{D\prime}\right)^{-1}(\widehat{\theta} - \theta_*) + \begin{pmatrix} T^{1/2}\tau_F D_x^{F\prime}\Omega_{vv}^{-1}\Omega_{vu} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}\right]$$
$$+ o_p(1).$$

This is an analogous expression to that in Vogelsang and Wagner (2014, p.759). Apply the first two results in this Theorem to get $T^{-1/2}\sum_{t=2}^{[rT]} \triangle \widetilde{S}_t^u \Rightarrow \sigma_{uv}\widetilde{P}(r)$, $r \in [0,1]$. Upon this result, the limit of $KPSS^{Fb}$ can be immediately derived.