

Estimation and Inference of Nonlinear Autoregressive Distributed Lag Models with Time Trend by Ordinary Least Squares

JIN SEO CHO

School of Economics, Yonsei University, Seodaemun-gu, Seoul 03722, Korea

jinseocho@yonsei.ac.kr

July 2026

Abstract

This study examines the large-sample behavior of an ordinary least squares (OLS) estimator within a correctly specified nonlinear autoregressive distributed lag (NARDL) model for nonstationary data. Although the OLS estimator suffers from an asymptotically singular matrix problem, it remains consistent for unknown model parameters and asymptotically follows a mixed normal distribution. Additionally, we examine the large-sample behavior of the standard Wald test, as defined by the OLS estimator, for asymmetries in long- and short-run NARDL parameters, and enhance this analysis with a super-consistent long-run parameter estimator. We then confirm the theory on the Wald test using Monte Carlo simulations. Finally, using U.S. GDP and exogenous fiscal shock data, we demonstrate use of the OLS estimator and show statistical evidence of long- and short-run symmetries between the effects of tax increases and decreases on U.S. GDP.

Key Words: Nonlinear autoregressive distributed lag model; OLS estimation; Singular matrix; Limit distribution; Wald test; Exogenous fiscal shocks; GDP.

JEL Classifications: C12, C13, C22, E62.

Acknowledgments: The editors-in-chief, Peter Phillips and Yixiao Sun, the co-editor, Degui Li, and four anonymous referees provided helpful comments for which the author is most grateful. The author is also indebted to Chun-Kyu Cho for his unwavering support and discussions with Yingjie Feng, Matthew Greenwood-Nimmo, Ruqing Jiang, Sainan Jin, Xun Lu, Yongcheol Shin, Liangjun Su, and seminar participants at Tsinghua University. Cho acknowledges Chang-am fellowship and the research grant provided by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF2020S1A5A2A01040235).

1 Introduction

The nonlinear autoregressive distributed lag (NARDL) model is widely used to estimate the asymmetric cointegrating relationship between nonstationary variables. Since [Shin, Yu, and Greenwood-Nimmo \(2014\)](#) introduced the NARDL model, researchers have revisited many long-run relationships, and modified linear ones using different slope coefficients based on the signs of the variables. For example, [Borenstein, Cameron, and Gilbert \(1997\)](#) identified the so-called rockets and feathers in gasoline prices, showing that upward cost shocks in crude oil prices pass through faster than downward shocks, thus affecting other economic variables asymmetrically. [Chesnes \(2016\)](#) empirically confirms this feature using the NARDL model.

Despite its popularity, estimating the NARDL model using ordinary least squares (OLS) lacks theoretical foundation. [Cho, Greenwood-Nimmo, and Shin \(2026\)](#) noted that OLS suffers from an asymptotically singular matrix problem, and since then, no theoretical justification has emerged for estimating the NARDL parameter using OLS. Currently, many empirical studies continue to estimate the unknown parameter using OLS, without an established theory.

In this study, we revisit OLS and provide its large-sample theory. Although OLS suffers from the problem of an asymptotically singular matrix, we find that it consistently estimates the unknown parameter. Furthermore, we establish a theoretical basis for applying Wald test principle to OLS by deriving its asymptotic distribution, which allows us to test for the NARDL hypothesis. OLS asymptotically follows a mixed normal distribution under some mild regularity conditions, validating the use of OLS for empirical analysis.

This study addresses a gap in the literature concerning NARDL estimation, although it is not the first to address the asymptotically singular matrix problem associated with NARDL. [Cho et al. \(2026\)](#) explained and resolved this problem by using a two-step procedure for estimating the NARDL parameter, termed two-step NARDL (2SNARDL). They separately estimate the long- and short-run parameters, following the method used by [Engle and Granger \(1987\)](#), which avoids the asymptotically singular matrix problem. We provide theoretical grounds to compare these two estimators by deriving the limit distribution of OLS and showing that the parameter targeted by OLS is estimated at the same convergence rate as 2SNARDL. From this, if the sample size is sufficiently large, OLS can estimate the parameter more straightforwardly than 2SNARDL. Meanwhile, 2SNARDL is more useful than OLS if the sample size is small and the asymptotic bias ignored by OLS can be efficiently removed.

Singular matrices are common in econometric asymptotics, particularly in multivariate regressions, and do not necessarily indicate the absence of a limit theory; rather, they lead to significant changes in that theory. In time-series analysis involving deterministic trends and nonstationary cointegrated regressors, researchers typically resolve singularity issues by applying multiple rates of convergence in various direc-

tions (see [Phillips, 1995](#), and his subsequent works for early demonstrations). Similarly, singular matrices emerge in the context of maximum-likelihood (ML) and nonlinear least squares (NLS) estimations, letting researchers find the limit distribution using higher-order approximation of the nonlinear model. From this the convergence rate of the estimator becomes slower than the standard case (e.g., [Teräsvirta, 1994](#); [Cho and White, 2007, 2010](#); [Cho and Phillips, 2018](#)).

In the current study's OLS analysis, applying a higher-order approximation to its limit distribution is necessary, but challenging. Unlike ML and NLS, the model is linear, which means the higher-order expansion should directly apply to OLS, rather than the model itself. This task is particularly challenging, because it involves applying the higher-order expansion to the inverse matrix within OLS. Because the dimension of the matrix is arbitrary, approximating it using a higher-order expansion is not straightforward.

We address this issue by indirectly obtaining the asymptotic distribution of OLS. For this, we represent OLS as a transformation of other primitive estimators that do not suffer from a singular matrix problem, and then we derive their weak limits to achieve the desired limit distribution. This process elaborates on the Frisch-Waugh-Lovell (FWL) projection method in [Park and Phillips \(1988, 1989\)](#) so that OLS can be represented by the primitive estimators without omitting any of its elements, identifying a channel through which the unknown parameters and the primitive estimators affect the limit distribution. Using this we isolate the singular matrix problem and show that the convergence rate of OLS depends on the unknown parameter values.

The indirectly obtained limit distribution is useful for inferring the NARDL parameter. We can develop a test methodology using this limit distribution. In addition, using the limit distributions of the primitive estimators, we can establish an additional Wald test to compensate for the slower convergence rate of OLS. The primitive estimators prove to be super-consistent for the long-run parameter, which is also the same convergence rate as the long-run 2SNARDL estimator.

The limit theory of OLS guides proper empirical data analysis. In this demonstration, we analyze the empirical data provided by [Romer and Romer \(2010\)](#) to examine the long- and short-run relationships between U.S. GDP and fiscal exogenous shocks. [Romer and Romer \(2010\)](#) measure legislated exogenous tax changes related to the U.S. GDP using narrative records such as presidential speeches and *Congressional reports*. By applying the NARDL model, we determine the long- and short-run relationships between tax increases and decreases in relation to the U.S. GDP. This empirical project also serves to illustrate our methodology in a standard setting.

We conduct the empirical project not only for demonstration, but also to identify how the long- and short-run relationships between U.S. GDP and tax decreases differ from those with tax increases. According to [Romer and Romer \(2010\)](#), exogenous tax changes are classified into those aimed at deficit reduction and

those intended for long-run growth. All tax changes for deficit reduction relate to tax increases, while most changes for long-run growth involve tax decreases. The NARDL model, which we use, separates shocks into negative and positive ones and estimates their possibly different effects on the dependent variable. By applying the NARDL model, we analyze the long- and short-run relationships between the tax changes for deficit reduction and those for long-run growth in terms of U.S. GDP. Our investigation shows that the long- and short-run parameters are indeed symmetric, meaning that the tax changes for deficit reduction and those for long-run growth affect the U.S. GDP symmetrically.

The remainder of this paper is structured as follows. Section 2 provides an overview of the NARDL model and discusses the asymptotically singular matrix problem associated with OLS. Section 3 defines primitive estimators and presents OLS as a bilinear transform of the primitive estimators. Section 4 discusses the limit distribution of OLS. Section 5 examines the large-sample properties of the standard Wald test for the NARDL hypothesis, and Section 6 presents Monte Carlo simulations for the Wald tests. Section 7 presents the empirical illustration. Finally, Section 8 concludes the paper. All mathematical proofs are available in the Online Supplement, in which we also provide other simulation evidence for OLS and empirical supplements.

Before moving to the next section, we present the notation used throughout this paper. We provide the weak limit of an estimator by a stochastic integral. Denoting the weak limit by $\int \mathcal{B}$ or $\int d\mathcal{B}$ means $\int_0^1 \mathcal{B}(u)du$ or $\int_0^1 d\mathcal{B}(u)$, respectively, where $\mathcal{B}(\cdot)$ is a Brownian motion.

2 Motivation and the NARDL Model in the Literature

This section briefly summarizes NARDL and motivates the current study.

2.1 NARDL and Error-Correction Models

We consider a NARDL(p, q) process augmented by a time trend:

$$y_t = \alpha_{0*} + \xi_* t + \sum_{j=1}^p \phi_{j*} y_{t-j} + \sum_{j=0}^q (\theta_{j*}^+ x_{t-j}^+ + \theta_{j*}^- x_{t-j}^-) + e_t, \quad (1)$$

where $\mathbf{x}_t \in \mathbb{R}^k$ ($k \in \mathbb{N}$), $\mathbf{x}_t^+ := \sum_{j=1}^t \Delta \mathbf{x}_j^+$, $\mathbf{x}_t^- := \sum_{j=1}^t \Delta \mathbf{x}_j^-$, $\Delta \mathbf{x}_t^+ := \max[\mathbf{0}, \Delta \mathbf{x}_t]$, $\Delta \mathbf{x}_t^- := \min[\mathbf{0}, \Delta \mathbf{x}_t]$, $\{e_t, \mathcal{F}_t\}$ is a martingale difference array (MDA), and $\mathcal{F}_t$ is the smallest σ -algebra driven by $\{y_{t-1}, \mathbf{x}_t^+, \mathbf{x}_t^-, y_{t-2}, \mathbf{x}_{t-1}^+, \mathbf{x}_{t-1}^-, \dots\}$ such that $\Delta \mathbf{x}_t$ is a stationary process, and Δ denotes the differencing operator, so that $\Delta \mathbf{x}_t := \mathbf{x}_t - \mathbf{x}_{t-1}$. Here, the max and min operators applied to a vector operate element-wise. The NARDL process in (1) is defined by introducing the asymmetric effect to the autoregressive distributed lag (ARDL) process proposed by Pesaran and Shin (1998). By supposing that the response of

dependent variable y_t to the positive and negative parts of x_t is possibly different, the NARDL process generalizes the ARDL process. If $\theta_{j*}^+ = \theta_{j*}^-$ for each j , the NARDL process reduces to the ARDL process. In addition, the NARDL process in (1) is more general than those defined by Shin et al. (2014) and Cho, Greenwood-Nimmo, and Shin (2023b), because the latter do not allow for a time trend on the right side.

The NARDL process is closely related to an error-correction representation. We first rewrite (1) as follows:

$$\Delta y_t = \rho_* y_{t-1} + \theta_*^{+'} x_{t-1}^+ + \theta_*^{-'} x_{t-1}^- + \xi_*(t-1) + \alpha_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\pi_{j*}^{+'} \Delta x_{t-j}^+ + \pi_{j*}^{-'} \Delta x_{t-j}^- \right) + e_t,$$

where ρ_* , θ_*^+ , θ_*^- , φ_{j*} ($j = 1, 2, \dots, p-1$), π_{j*}^+ , and π_{j*}^- ($j = 0, 1, \dots, q-1$) are defined by the parameters in (1), as given in Cho et al. (2023b). Specifically, $\rho_* := \sum_{j=1}^p \phi_{j*} - 1$, $\theta_*^+ := \sum_{j=0}^q \theta_{j*}^+$, $\theta_*^- := \sum_{j=0}^q \theta_{j*}^-$, $\pi_{0*}^+ := \theta_{0*}^+$, $\pi_{0*}^- := \theta_{0*}^-$, and for $\ell = 1, 2, \dots, p-1$ and $j = 1, 2, \dots, q-1$, $\varphi_{\ell*} := -\sum_{i=\ell+1}^p \phi_{i*}$, $\pi_{j*}^+ := -\sum_{i=j+1}^q \theta_{i*}^+$, $\pi_{j*}^- := -\sum_{i=j+1}^q \theta_{i*}^-$, and $\alpha_* = \alpha_{0*} + \xi_*$. By plugging these parameters into (2), the NARDL(p, q) process in (1) is obtained. For notational simplicity, we further rewrite this as

$$\Delta y_t = \rho_* y_{t-1} + \theta_*' \ddot{x}_{t-1} + \xi_*(t-1) + \alpha_* + \kappa_*' z_{2t} + e_t \quad (2)$$

by letting $\theta_* := [\theta_*^{+'}, \theta_*^{-'}]'$, $\ddot{x}_{t-1} := [x_{t-1}^{+'}, x_{t-1}^{-'}]'$, $\kappa_* := [\varphi_*', \pi_*']'$, and $z_{2t} := [\Delta y_{t-1}', \Delta \tilde{x}_t']'$, such that

$$\begin{aligned} \varphi_* &:= [\varphi_{1*}, \dots, \varphi_{p-1*}]', \quad \Delta y_{t-1} := [\Delta y_{t-1}, \dots, \Delta y_{t-p+1}]', \\ \pi_* &:= [\pi_*^{+'}, \pi_*^{-'}]', \quad \pi_*^+ := [\pi_{0*}^{+'}, \dots, \pi_{q-1*}^{+'}]', \quad \pi_*^- := [\pi_{0*}^{-'}, \dots, \pi_{q-1*}^{-'}]', \\ \Delta \tilde{x}_t &:= [\Delta \tilde{x}_t^{+'}, \Delta \tilde{x}_t^{-'}]', \quad \Delta \tilde{x}_t^+ := [\Delta x_t^{+'}, \dots, \Delta x_{t-q+1}^{+'}]', \quad \text{and} \quad \Delta \tilde{x}_t^- := [\Delta x_t^{-'}, \dots, \Delta x_{t-q+1}^{-'}]'. \end{aligned}$$

Here, if y_{t-1} is cointegrated with $\ddot{x}_{t-1}$, (2) can be rewritten into the following error-correction form:

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \kappa_*' z_{2t} + e_t, \quad (3)$$

such that the cointegration error is defined as $u_{t-1} := y_{t-1} - \beta_*' \ddot{x}_{t-1} - \zeta_*(t-1) - \nu_*$ with $\beta_* := [\beta_*^{+'}, \beta_*^{-'}]'$, $\beta_*^+ := -(\theta_*^+/\rho_*)$, $\beta_*^- := -(\theta_*^-/\rho_*)$, and $\zeta_* := -(\xi_*/\rho_*)$. Here, the intercept ν_* is introduced so that $\mathbb{E}[u_t] = 0$, and u_t is assumed to be stationary and possibly correlated with Δy_{t-1} and $\Delta \tilde{x}_t$. From this, it follows that $\gamma_* = \alpha_* + \rho_* \nu_*$.

The NARDL process captures an asymmetric cointegrating relationship between nonstationary processes. If we let $\mu_* := [\mu_*^{+'}, \mu_*^{-'}]'$, $\mu_*^+ := \mathbb{E}[\Delta x_t^+]$ and $\mu_*^- := \mathbb{E}[\Delta x_t^-]$, then $\mu_*^+ + \mu_*^- \equiv \mathbb{E}[\Delta x_t]$

by construction. Therefore, if we further let $s_t^+ := \Delta x_t^+ - \mu_*^+$ and $s_t^- := \Delta x_t^- - \mu_*^-$, it follows that

$$\ddot{x}_t = \mu_* t + \mathbf{m}_t \quad \text{by letting} \quad \mathbf{m}_t := [\mathbf{m}_t^{+'}, \mathbf{m}_t^{-'}]', \quad \mathbf{m}_t^+ := \sum_{j=1}^t s_j^+ \quad \text{and} \quad \mathbf{m}_t^- := \sum_{j=1}^t s_j^-. \quad (4)$$

From (4), $\ddot{x}_t$ is clearly a unit-root process with nonzero time trends. Moreover, Δy_t is not necessarily distributed around zero, even when x_t is a unit-root process without a time trend. From (3), we have

$$\delta_* := \mathbb{E}[\Delta y_t] = \frac{1}{\varrho_*} \left[\gamma_* + \sum_{j=0}^{q-1} \pi_{j*} \mu_* \right], \quad \text{where} \quad \pi_{j*} := [\pi_{j*}^{+'}, \pi_{j*}^{-'}]' \quad \text{and} \quad \varrho_* := 1 - \sum_{j=1}^{p-1} \varphi_{j*},$$

so that if $d_t := \Delta y_t - \delta_*$, then

$$y_t = \delta_* t + \sum_{j=1}^t d_j. \quad (5)$$

This fact implies that y_t is a unit-root process with a deterministic time trend, which has further implications. First, even when $\mathbb{E}[x_t] = \mathbf{0}$, the NARDL model captures a cointegrating relationship between the variables without and with time trend, viz., x_t and y_t . Second, the NARDL model includes the time trend in (2) as an explanatory variable by noting that the trend in y_t could also be influenced by exogenous factors other than the NARDL effect. If the estimated coefficient of the time trend is insignificant, it signals that the trend in y_t is solely sourced from the NARDL effect.

2.2 OLS for NARDL and Asymptotic Singular Matrix

Shin et al. (2014) estimate (2) by OLS without distinguishing the long- and short-run equation, which is now very popular in the empirical literature. For this examination, we first let

$$\mathbf{z}_t := \begin{bmatrix} \mathbf{z}_{1t}' & \mathbf{z}_{2t}' \end{bmatrix}' := \begin{bmatrix} y_{t-1} & \mathbf{r}_{t-1}' & \mathbf{z}_{2t}' \end{bmatrix}' := \begin{bmatrix} y_{t-1} & \ddot{x}_{t-1}' & (t-1) & 1 & \mathbf{z}_{2t}' \end{bmatrix}',$$

so that $\mathbf{r}_{t-1} = [\ddot{x}_{t-1}', (t-1), 1]'$. Here, $\mathbf{z}_t \in \mathbb{R}^{2+p+2k(1+q)}$ is partitioned into two variables such that $\mathbf{z}_{1t} \in \mathbb{R}^{3+2k}$ and $\mathbf{z}_{2t} \in \mathbb{R}^{p+2kq-1}$. Next, we let $\alpha_* := \begin{bmatrix} \rho_* & \theta_*' & \xi_* & \alpha_* & \kappa_*' \end{bmatrix}'$. From this, the OLS estimator is written as

$$\hat{\alpha}_T := \begin{bmatrix} \hat{\rho}_T & \hat{\theta}_T' & \hat{\xi}_T & \hat{\alpha}_T & \hat{\kappa}_T' \end{bmatrix}' := \left(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t' \right)^{-1} \sum_{t=1}^T \mathbf{z}_t \Delta y_t.$$

For later purposes, we partition $\hat{\theta}_T$ and $\hat{\kappa}_T$ into $[\hat{\theta}_T^{+'}, \hat{\theta}_T^{-'}]'$ and $[\hat{\varphi}_T', \hat{\pi}_T']' \equiv [\hat{\varphi}_T', \hat{\pi}_T^{+'}, \hat{\pi}_T^{-'}]'$, respectively.

Despite its popularity, there is no theoretical foundation for the OLS. This is mainly attributed to the

asymptotically singular matrix problem. Even under mild regularity conditions, the OLS estimator suffers from an asymptotically singular matrix problem. For this discussion, we impose the following conditions, which are standard in the literature:

- Assumption 1.** (i) $\{(\Delta y_t, \Delta \mathbf{x}'_t, u_t)' \in \mathbb{R}^{k+2} : t = \dots, -1, 0, 1, \dots\}$ is a strictly stationary ϕ -mixing process of size $-r/(2(r-1))$ or α -mixing process of size $-r/(r-2)$ and $r > 2$;
- (ii) $\mathbb{E}[|\Delta x_{ti}|^r] < \infty$ ($i = 1, 2, \dots, k$), $\mathbb{E}[|u_t|^r] < \infty$, $\mathbb{E}[|e_t|^2] < \infty$, and $\delta_* \neq 0$, where x_{ti} is the i -th row element of $\mathbf{x}_t$;
- (iii) $\Sigma_* := \lim_{T \rightarrow \infty} \text{var}[T^{-1/2} \sum_{t=1}^T \mathbf{w}_t]$ is positive definite, where

$$\mathbf{w}_t := \begin{bmatrix} \mathbf{w}'_{1t} & \mathbf{w}'_{2t} \end{bmatrix} := \begin{bmatrix} \mathbf{s}'_{t-1} & u_{t-1} & e_t & e_t u_{t-1} & e_t \mathbf{z}'_{2t} \end{bmatrix}'$$

$$\text{and } \mathbf{s}_{t-1} := [\mathbf{s}_{t-1}^{+'}, \mathbf{s}_{t-1}^{-'}]';$$

- (iv) for some α_* with $\rho_* < 0$, Δy_t is generated by (2) such that $|L_*| > 1$, where $1 - \sum_{j=1}^p \phi_{j*} L_*^j \equiv 0$;
- (v) $\{e_t, \mathcal{F}_t\}$ is an MDA such that $\mathbb{E}[e_t^2 | \mathcal{F}_t] = \sigma_e^2$. $\square$

- Remarks.** (a) Under the same assumption, we investigate the limit behavior of the OLS estimator below.
- (b) Assumptions 1 (i and ii) assume mixing and moment conditions to apply the functional central limit theorem (FCLT) to a partial-sum data process. The FCLT is popular for deriving the limit distribution of the OLS applied to estimate a cointegrating relationship (e.g., Phillips and Hansen, 1990; Phillips, 1991; White, 2001, chapter 7).
- (c) Assumption 1 (iv) assumes $\rho_* < 0$ for a cointegrating relationship between y_t and $\tilde{\mathbf{x}}_t$. If no cointegrating relationship exists, so that if $\beta_* = \mathbf{0}$ and $\zeta_* = 0$, then it follows from (3) that

$$y_t = \alpha_* + \sum_{j=1}^p \phi_{j*} y_{t-j} + \pi_*' \Delta \tilde{\mathbf{x}}_t + e_t.$$

Assumption 1 (iv), however, implies that y_t is a stationary process, contradicting the assumption that y_t is a unit-root process. From this, it is necessary to assume $\rho_* = 0$ for $(\beta_*', \zeta_*)' = \mathbf{0}$. In this study, we assume the cointegrating relationship established by (3).

- (d) Although it is not our main interest, Pesaran, Shin, and Smith (2001) and Banerjee, Dolado, and Mestre (1998) have previously examined estimating the unknown parameter by OLS under $\rho_* = 0$. They assumed a linear model and developed a testing methodology for $\rho_* = 0$. Their methodology is applicable even in the context of NARDL, because assuming no cointegration under the condition of linearity is equivalent to assuming $\beta_* = \mathbf{0}$ and $\zeta_* = 0$.
- (e) Assumption 1 (v) is a standard condition for the error term in the ARDL and NARDL processes (e.g.,

We now show how the asymptotically singular matrix problem is associated with the OLS. For this, we provide the following lemma.

Lemma 1. *Under Assumption 1,*

(i)

$$\mathbf{D}_1^{-1} \left(\sum_{t=1}^T \mathbf{z}_{1t} \mathbf{z}'_{1t} \right) \mathbf{D}_1^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_{11} := \begin{bmatrix} \begin{bmatrix} \frac{1}{3} \delta_* \\ \frac{1}{3} \boldsymbol{\mu}_* \\ \frac{1}{3} \end{bmatrix} & \vdots & \begin{bmatrix} \frac{1}{2} \delta_* \\ \frac{1}{2} \boldsymbol{\mu}_* \\ \frac{1}{2} \end{bmatrix} \\ [\delta_*, \boldsymbol{\mu}'_*, 1] & & \end{bmatrix},$$

where $\mathbf{D}_1 := \text{diag}[T^{3/2} \mathbf{I}_{2+2k}, T^{1/2}]$;

(ii) $\mathbf{D}_2^{-1} \left(\sum_{t=1}^T \mathbf{z}_{2t} \mathbf{z}'_{2t} \right) \mathbf{D}_2^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_{22} := \mathbb{E}[\mathbf{z}_{2t} \mathbf{z}'_{2t}]$, where $\mathbf{D}_2 := T^{1/2} \mathbf{I}_{p+2kq-1}$;

(iii)

$$\mathbf{D}_2^{-1} \left(\sum_{t=1}^T \mathbf{z}_{2t} \mathbf{z}'_{1t} \right) \mathbf{D}_1^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_{21} := \begin{bmatrix} \begin{bmatrix} \frac{1}{2} \delta_* \boldsymbol{\iota}_{p-1} \\ \frac{1}{2} \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* \end{bmatrix} & \vdots & \begin{bmatrix} \frac{1}{2} \delta_* \boldsymbol{\iota}_{p-1} \\ \frac{1}{2} \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* \end{bmatrix} \\ [\delta_*, \boldsymbol{\mu}'_*, 1] & & \end{bmatrix},$$

where for each $\ell \in \mathbb{N}$, $\boldsymbol{\iota}_\ell := [1, 1, \dots, 1]' \in \mathbb{R}^\ell$;

(iv)

$$\mathbf{D}^{-1} \left(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}'_t \right) \mathbf{D}^{-1} \xrightarrow{\mathbb{P}} \mathbf{M} := \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix},$$

which is singular, where $\mathbf{D} := \text{diag}[\mathbf{D}_1, \mathbf{D}_2]$ and $\mathbf{M}_{12} := \mathbf{M}'_{21}$. $\square$

Here, $\mathbf{M}$ is singular because its first $(1 + 2k)$ columns are proportional to the $(2 + 2k)$ -th column.

Remarks. (a) Lemma 1 clearly implies that the theoretical foundation of the OLS cannot be established without tackling the asymptotically singular matrix problem. The asymptotically singular matrix problem arises from the deterministic time trend in the regressors. Specifically, $\mathbb{E}[\Delta y_{t-1}] \neq 0$ by (5), so that y_{t-1} is an integrated series with a deterministic time trend. Likewise, both $\mathbf{x}_{t-1}^+$ and $\mathbf{x}_{t-1}^-$ are integrated series with deterministic time trends. Therefore, all of y_{t-1} , $\ddot{\mathbf{x}}_{t-1}$, and $(t - 1)$ are driven by the time trend, leading to the asymptotically singular matrix problem. This aspect is also evident from Lemma 1. The first $(2 + 2k)$ columns of $\mathbf{M}$ are proportional to each other, and they are the terms involved with y_{t-1} , $\ddot{\mathbf{x}}_{t-1}$, and $(t - 1)$.

(b) Cho et al. (2026) note the asymptotically singular matrix and estimate the unknown long-run parameters by first removing the trends from y_{t-1} and $\ddot{\mathbf{x}}_{t-1}$, proposing 2SNARDL estimation. We prove Lemma 1 in the Online Supplement by extending lemma 1 of Cho et al. (2026). Their lemma demon-

- strates another singular matrix problem by supposing no time trend in (1). Lemma 1 verifies that augmenting the time trend on the right side is not helpful in eliminating the singular matrix problem.
- (c) We can easily verify the asymptotically singular matrix problem by conducting simple Monte Carlo simulations. By denoting the model and data-generating process (DGP) conditions for Table 3, we compute the minimum diagonal element and determinant of $\mathbf{D}^{-1}(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t') \mathbf{D}^{-1}$ for $T = 25, 50, 75, 100, 125, \dots, 475, \text{ and } 500$. After iterating 10000 independent experiments, Figure 1 (a) draws a scatter diagram between T and the average of the 10000 minimum diagonal elements. From this, we can see that the minimum diagonal element converges to a finite number, indicating that $\mathbf{D}^{-1}(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t') \mathbf{D}^{-1} = O_{\mathbb{P}}(1)$. Meanwhile, Figure 1 (b) draws a scatter diagram between T and the average of the 10000 determinants $\times 10^{12}$. This second figure shows that the determinant converges to zero, meaning that $\mathbf{D}^{-1}(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t') \mathbf{D}^{-1}$ is asymptotically singular. This corroborates Lemma 1.
- (d) Singular matrices often appear in econometric asymptotics, especially in multivariate regressions. These matrices do not necessarily indicate the absence of a limit theory; rather, they typically lead to significant changes in that theory. In time-series analysis, when dealing with deterministic trends and nonstationary cointegrated regressors, researchers use multiple rates of convergence in different directions to address this issue (e.g., Phillips, 1995, and his subsequent works). In the cases of ML and NLS estimations, researchers obtain the limit distribution by applying higher-order approximations to nonlinear models, resulting in a slower convergence rate compared to the standard case (e.g., Teräsvirta, 1994; Cho and White, 2007, 2010; Cho and Phillips, 2018).
- (e) In the current context, it is not straightforward to apply a higher-order expansion to OLS. The model is already linear, unlike ML and NLS. This means that it is necessary to apply the higher-order expansion to OLS itself. However, given that the dimension of $(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t')^{-1}$ is determined by k, p , and q , and they can be given arbitrarily, the task becomes challenging. $\square$

3 An Alternative Representation of the OLS Estimator

Owing to the complexity of the higher-order expansion, we use a different approach to obtain the desired limit distribution. We represent the OLS estimator as a transformation of other primitive estimators that do not suffer from an asymptotic singularity problem.

3.1 Alternative Representation

We achieve the alternative representation in two steps. First, we estimate the long- and short-run parameters using OLS and represent the OLS estimator $\hat{\alpha}_T$ as a bilinear transformation of the long- and short-run OLS

estimators. Next, we express the long- and short-run OLS estimators as other bilinear transformations of further primitive estimators that do not suffer from an asymptotically singular matrix problem.

3.1.1 First Step: Representing OLS Estimator by Long- and Short-Run Estimators

We estimate the long-run parameter using OLS. That is, after specifying the long-run equation

$$y_{t-1} = \mathbf{r}'_{t-1} \mathbf{v}_* + u_{t-1} = \boldsymbol{\beta}'_* \ddot{\mathbf{x}}_{t-1} + \zeta_*(t-1) + \nu_* + u_{t-1} \quad (6)$$

by letting $\mathbf{v}_* := [\boldsymbol{\beta}'_*, \zeta_*, \nu_*]'$, we let the long-run OLS estimator be

$$\tilde{\mathbf{v}}_T := [\tilde{\boldsymbol{\beta}}'_T, \tilde{\zeta}_T, \tilde{\nu}_T]' := \left(\sum_{t=1}^T \mathbf{r}_{t-1} \mathbf{r}'_{t-1} \right)^{-1} \sum_{t=1}^T \mathbf{r}_{t-1} y_{t-1}.$$

For later purposes, we let $\tilde{u}_{t-1} := y_{t-1} - \mathbf{r}'_{t-1} \tilde{\mathbf{v}}_T$ and partition $\tilde{\boldsymbol{\beta}}_T$ such that $\tilde{\boldsymbol{\beta}}_T \equiv [\tilde{\boldsymbol{\beta}}_T^{+'}, \tilde{\boldsymbol{\beta}}_T^{-'}]'$.

We next estimate the short-run parameter by letting $\boldsymbol{\omega}_* := [\rho_*, \boldsymbol{\eta}'_*, \psi_*, \gamma_*, \boldsymbol{\kappa}'_*]'$, such that $\boldsymbol{\eta}_* := \boldsymbol{\theta}_* + \rho_* \boldsymbol{\beta}_*$ and $\psi_* := \xi_* + \rho_* \zeta_*$, and $\dot{\mathbf{z}}_t := [u_{t-1}, \mathbf{r}'_{t-1}, \mathbf{z}'_{2t}]'$ to specify the error-correction model

$$\Delta y_t = \dot{\mathbf{z}}'_t \boldsymbol{\omega}_* + e_t = \rho_* u_{t-1} + \boldsymbol{\eta}'_* \ddot{\mathbf{x}}_{t-1} + \psi_*(t-1) + \gamma_* + \boldsymbol{\kappa}'_* \mathbf{z}_{2t} + e_t. \quad (7)$$

We obtain this equation by combining (2) and (6) and estimate the parameters in (7) by

$$\tilde{\boldsymbol{\omega}}_T := [\tilde{\rho}_T, \tilde{\boldsymbol{\eta}}'_T, \tilde{\psi}_T, \tilde{\gamma}_T, \tilde{\boldsymbol{\kappa}}'_T]' := \left(\sum_{t=1}^T \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}'_t \right)^{-1} \sum_{t=1}^T \tilde{\mathbf{z}}_t \Delta y_t, \text{ where } \tilde{\mathbf{z}}_t := [\tilde{u}_{t-1}, \mathbf{r}'_{t-1}, \mathbf{z}'_{2t}]',$$

estimating the short-run parameter $\boldsymbol{\omega}_*$. Note that $\tilde{\boldsymbol{\omega}}_T$ is obtained by replacing u_{t-1} with $\tilde{u}_{t-1}$. For later purposes, we partition $\boldsymbol{\eta}_*$ into $[\boldsymbol{\eta}_*^{+'}, \boldsymbol{\eta}_*^{-'}]'$.

The long- and short-run OLS estimators have a regular relationship in terms of the OLS estimator $\hat{\boldsymbol{\alpha}}_T$. The following proposition reveals the relationship:

Proposition 1. *Under Assumption 1, $\hat{\boldsymbol{\alpha}}_T = \mathbf{R}_T \tilde{\boldsymbol{\omega}}_T$, where*

$$\mathbf{R}_T := \begin{bmatrix} \mathbf{R}_T^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p+2kq-1} \end{bmatrix}, \quad \text{and} \quad \mathbf{R}_T^{11} := \begin{bmatrix} 1 & \mathbf{0}_{1 \times (2+2k)} \\ -\tilde{\mathbf{v}}_T & \mathbf{I}_{2+2k} \end{bmatrix}. \quad \square$$

Remarks. (a) By Proposition 1, the OLS estimator $\hat{\boldsymbol{\alpha}}_T$ is now represented as a bilinear transformation

between the long- and short-run OLS estimators $\tilde{\mathbf{v}}_T$ and $\tilde{\boldsymbol{\omega}}_T$. That is,

$$\hat{\boldsymbol{\alpha}}_T = [\tilde{\rho}_T, -\tilde{\rho}_T \tilde{\boldsymbol{\beta}}_T' + \tilde{\boldsymbol{\eta}}_T', -\tilde{\rho}_T \tilde{\zeta}_T + \tilde{\psi}_T, -\tilde{\rho}_T \tilde{\nu}_T + \tilde{\gamma}_T, \tilde{\boldsymbol{\kappa}}_T']'.$$

If both $\tilde{\mathbf{v}}_T$ and $\tilde{\boldsymbol{\omega}}_T$ are consistent for $\mathbf{v}_*$ and $\boldsymbol{\omega}_*$, respectively, it follows that

$$\hat{\boldsymbol{\alpha}}_T \xrightarrow{\mathbb{P}} [\rho_*, -\rho_* \boldsymbol{\beta}_*' + \boldsymbol{\eta}_*', -\rho_* \zeta_* + \psi_*, -\rho_* \nu_* + \gamma_*, \boldsymbol{\kappa}_*']' = [\rho_*, \boldsymbol{\theta}_*, \xi_*, \alpha_*, \boldsymbol{\kappa}_*']'.$$

by noting that $\boldsymbol{\eta}_* = \mathbf{0}$ and $\psi_* = 0$.

- (b) Proposition 1 follows by applying Lemma A.2 in the Online Supplement. By letting y_t , $\mathbf{x}_t$, $\mathbf{z}_t$, and $\mathbf{w}_t$ of Lemma A.2 be Δy_t , y_{t-1} , $\mathbf{r}_{t-1}$, and $\mathbf{z}_{2t}$ of (7), respectively, Proposition 1 follows. Here, $\hat{\mathbf{v}}_t$ in Lemma A.2 corresponds to $\tilde{u}_{t-1}$. Lemma A.2 is established by elaborating on the FWL projection method in Park and Phillips (1988, 1989). It complements the FWL method by augmenting a representation for the estimator ignored by FWL, so that the OLS estimator $\hat{\boldsymbol{\alpha}}_T$ could be written without omitting any of its element as given in Proposition 1. By this feature, we call Lemma A.2 the augmented FWL (AFWL). For the goal to derive the joint limit distribution of $\hat{\boldsymbol{\alpha}}_T$, it is important to derive the joint limit distributions of the long- and short-run OLS estimators without omitting any of its element, but FWL omits the representation of the coefficient, if its variable is concentrated out. For this reason, we complement FWL using AFWL and provide the alternative representation for $\hat{\boldsymbol{\alpha}}_T$. $\square$

3.1.2 Second Step: Representing Long- and Short-Run Estimators by Primitive Estimators

Although Proposition 1 represents $\hat{\boldsymbol{\alpha}}_T$ as a bilinear transformation between the long- and short-run OLS estimators, they both suffer from asymptotically singular matrix problems. It is mainly because $\tilde{\mathbf{v}}_T$ and $\tilde{\boldsymbol{\omega}}_T$ are defined by regressors including $\tilde{\mathbf{x}}_{t-1}$ and $(t-1)$ that are driven by the time trend; see Lemma A.3 in the Online Supplement. Therefore, we again represent both estimators using other primitive estimators that do not suffer from an asymptotically singular matrix problem.

Long-Run Estimator We first represent the long-run OLS estimator $\tilde{\mathbf{v}}_T$ using other estimators. Note that both (4) and (6) imply that

$$y_{t-1} = \mathbf{r}_{t-1}' \mathbf{v}_* + u_{t-1} = \boldsymbol{\beta}_*' \mathbf{m}_{t-1} + \vartheta_*(t-1) + \nu_* + u_{t-1}, \quad \text{where } \vartheta_* := \boldsymbol{\beta}_*' \boldsymbol{\mu}_* + \zeta_*. \quad (8)$$

This representation is obtained by collecting the deterministic time trend $(t-1)$ and the remainder $\mathbf{m}_{t-1}$ in (4) as regressors. However, $\mathbf{m}_{t-1}$ is unobservable. So, we first predict it by estimating the unknown

parameters μ_* in (4):

$$\ddot{\mu}_T := \left(\sum_{t=1}^T (t-1)^2 \right)^{-1} \sum_{t=1}^T (t-1) \ddot{x}_{t-1}, \quad \text{obtaining} \quad \ddot{m}_{t-1} := \ddot{x}_{t-1} - (t-1) \ddot{\mu}_T,$$

and regressing y_{t-1} against $\ddot{r}_{t-1} := [\ddot{m}'_{t-1}, (t-1), 1]'$, to obtain

$$\ddot{v}_T := [\ddot{\beta}'_T, \ddot{\vartheta}_T, \ddot{\nu}_T]' := \left(\sum_{t=1}^T \ddot{r}_{t-1} \ddot{r}'_{t-1} \right)^{-1} \sum_{t=1}^T \ddot{r}_{t-1} y_{t-1}, \quad \text{estimating} \quad \bar{v}_* := [\beta'_*, \vartheta_*, \nu_*]'$$

For later purposes, we partition $\ddot{\beta}_T$ such that $\ddot{\beta}_T \equiv [\ddot{\beta}_T^+, \ddot{\beta}_T^-]'$ and let $\ddot{u}_{t-1} := y_{t-1} - \ddot{r}'_{t-1} \ddot{v}_T$.

The interrelationship between the long-run OLS and primitive estimators is stated in the following proposition:

Proposition 2. *Given Assumption 1, $\tilde{v}_T := [\tilde{\beta}'_T, \tilde{\zeta}_T, \tilde{\nu}_T]' = \mathbf{P}_T \ddot{v}_T = [\ddot{\beta}'_T, \ddot{\vartheta}_T - \ddot{\mu}'_T \ddot{\beta}_T, \ddot{\nu}_T]'$, where*

$$\mathbf{P}_T := \begin{bmatrix} \mathbf{I}_{2k} & \mathbf{0} \\ \mathbf{P}_T^{21} & \mathbf{I}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{P}_T^{21} := \begin{bmatrix} -\ddot{\mu}'_T \\ \mathbf{0}_{1 \times 2k} \end{bmatrix}. \quad \square$$

Remarks. (a) By Proposition 2, the long-run OLS estimator $\tilde{v}_T$ is represented as a bilinear transformation between the primitive estimators, $\ddot{\mu}_T$ and $\ddot{v}_T$. This representation is particularly useful, because neither $\ddot{\mu}_T$ nor $\ddot{v}_T$ suffers from an asymptotically singular matrix problem; see Lemma 2 below.

(b) If both $\ddot{\mu}_T$ and $\ddot{v}_T$ are consistent for μ_* and $\bar{v}_*$, respectively,

$$\tilde{v}_T = [\ddot{\beta}'_T, \ddot{\vartheta}_T - \ddot{\mu}'_T \ddot{\beta}_T, \ddot{\nu}_T]' \xrightarrow{\mathbb{P}} [\beta'_*, \vartheta_* - \mu'_* \beta_*, \nu_*]' = [\beta'_*, \zeta_*, \nu_*]' =: \mathbf{v}_*$$

by noting that $\vartheta_* := \beta'_* \mu_* + \zeta_*$, leading to the consistency of $\tilde{v}_T$ for $\mathbf{v}_*$. Furthermore, $\tilde{u}_{t-1} = \ddot{u}_{t-1}$ using the definition of $\ddot{m}_{t-1} := \ddot{x}_{t-1} - (t-1) \ddot{\mu}_T$.

(c) Proposition 2 is also established by AFWL. If we let y_{t-1} , $\ddot{x}'_{t-1}$, $(t-1)$, and 1 be y_t , $\mathbf{x}_t$, $\mathbf{z}_t$, and $\mathbf{w}_t$, respectively, in Lemma A.2, then Proposition 2 follows. $\square$

Short-Run Estimator Next, we represent the short-run OLS estimator $\tilde{\omega}_T$ as a bilinear transformation between two other estimators that do not suffer from an asymptotically singular matrix problem. This representation is parallel to the one for the long-run OLS estimator. As both $\ddot{x}_{t-1}$ and $(t-1)$ on the right side of (7) are driven by the deterministic trend, we reparameterize (7) using (4):

$$\Delta y_t = \rho_* u_{t-1} + \eta'_* \mathbf{m}_{t-1} + \varsigma_*(t-1) + \gamma_* + \kappa'_* \mathbf{z}_{2t} + e_t, \quad \text{where} \quad \varsigma_* := \mu'_* \eta_* + \psi_*. \quad (9)$$

Here, the coefficient of $(t-1)$ is obtained as ς_* by collecting the coefficients of the time trends in $\ddot{x}_{t-1}$ and $(t-1)$, but the coefficient of $\mathbf{m}_{t-1}$ keeps the same as that of $\ddot{x}_{t-1}$ on the right side of (7).

We estimate the unknown parameters in (9) by OLS. For this, we replace u_{t-1} and $\mathbf{m}_{t-1}$ in (9) with $\ddot{u}_{t-1}$ and $\ddot{\mathbf{m}}_{t-1}$, respectively as they are unobservable. By letting $\ddot{\mathbf{z}}_t := [\ddot{u}_{t-1} \mid \ddot{\mathbf{r}}'_{t-1} \mid \mathbf{z}'_{2t}]'$, the parameter estimator is obtained as follows:

$$\ddot{\boldsymbol{\tau}}_T := [\ddot{\rho}_T \mid \ddot{\boldsymbol{\tau}}'_{1T} \mid \ddot{\boldsymbol{\kappa}}'_T]' := [\ddot{\rho}_T \mid \ddot{\boldsymbol{\eta}}'_T \mid \ddot{\varsigma}_T \mid \ddot{\gamma}_T \mid \ddot{\boldsymbol{\kappa}}'_T]' := \left(\sum_{t=1}^T \ddot{\mathbf{z}}_t \ddot{\mathbf{z}}'_t \right)^{-1} \sum_{t=1}^T \ddot{\mathbf{z}}_t \Delta y_t, \quad \text{estimating}$$

$$\boldsymbol{\tau}_* := [\rho_* \mid \boldsymbol{\tau}'_{1*} \mid \boldsymbol{\kappa}'_*]' := [\rho_* \mid \boldsymbol{\eta}'_* \mid \varsigma_* \mid \gamma_* \mid \boldsymbol{\kappa}'_*]'$$

Here, for later purposes, we let $\ddot{\boldsymbol{\eta}}_T \equiv [\ddot{\boldsymbol{\eta}}_T^{+'}, \ddot{\boldsymbol{\eta}}_T^{-'}]'$ and $\ddot{\boldsymbol{\kappa}}_T \equiv [\ddot{\boldsymbol{\varphi}}'_T, \ddot{\boldsymbol{\pi}}'_T]' \equiv [\ddot{\boldsymbol{\varphi}}'_T, \ddot{\boldsymbol{\pi}}_T^{+'}, \ddot{\boldsymbol{\pi}}_T^{-'}]'$. In addition, we separately estimate ψ_* by $\ddot{\psi}_T := -\ddot{\boldsymbol{\mu}}'_T \ddot{\boldsymbol{\eta}}_T + \ddot{\varsigma}_T$ by applying the plug-in principle to $\varsigma_* := \boldsymbol{\mu}'_* \boldsymbol{\eta}_* + \psi_*$.

The interrelationship between the short-run OLS and primitive estimators, $\tilde{\boldsymbol{\omega}}_T$ and $\ddot{\boldsymbol{\tau}}_T$, is stated in the following proposition:

Proposition 3. *Given Assumption 1, $\tilde{\boldsymbol{\omega}}_T = \mathbf{Q}_T \ddot{\boldsymbol{\tau}}_T$, where*

$$\mathbf{Q}_T := \begin{bmatrix} \mathbf{I}_{1+2k} & \mathbf{0} \\ \mathbf{Q}_T^{21} & \mathbf{I}_{p+2kq+1} \end{bmatrix} \quad \text{and} \quad \mathbf{Q}_T^{21} := \begin{bmatrix} 0 & -\ddot{\boldsymbol{\mu}}'_T \\ \mathbf{0}_{(p+2kq) \times 1} & \mathbf{0}_{(p+2kq) \times 2k} \end{bmatrix}. \quad \square$$

Remarks. (a) By Proposition 3, the short-run OLS estimator $\tilde{\boldsymbol{\omega}}_T$ is represented as a bilinear transformation between $\ddot{\boldsymbol{\mu}}_T$ and $\ddot{\boldsymbol{\tau}}_T$, which do not suffer from asymptotically singular matrix problems; see Lemma 2 below.

(b) If both $\ddot{\boldsymbol{\mu}}_T$ and $\ddot{\boldsymbol{\tau}}_T$ are consistent for $\boldsymbol{\mu}_*$ and $\boldsymbol{\tau}_*$, respectively, it follows that

$$\begin{aligned} \tilde{\boldsymbol{\omega}}_T &= [\ddot{\rho}_T, \ddot{\boldsymbol{\eta}}'_T, -\ddot{\boldsymbol{\mu}}'_T \ddot{\boldsymbol{\eta}}_T + \ddot{\varsigma}_T, \ddot{\gamma}_T, \ddot{\boldsymbol{\kappa}}'_T]' \\ &\xrightarrow{\mathbb{P}} [\rho_*, \boldsymbol{\eta}'_*, -\boldsymbol{\mu}'_* \boldsymbol{\eta}_* + \varsigma_*, \gamma_*, \boldsymbol{\kappa}'_*]' = [\rho_*, \boldsymbol{\eta}'_*, \psi_*, \gamma_*, \boldsymbol{\kappa}'_*]' =: \boldsymbol{\omega}_* \end{aligned}$$

by noting that $\varsigma_* := \boldsymbol{\mu}'_* \boldsymbol{\eta}_* + \psi_*$.

(c) As before, Proposition 3 is established by AFWL. By letting Δy_t , $\ddot{x}_{t-1}$, $(t-1)$, and $[\ddot{u}_{t-1}, 1, \mathbf{z}'_{2t}]'$ be y_t , $\mathbf{x}_t$, $\mathbf{z}_t$, and $\mathbf{w}_t$, respectively, in Lemma A.2, Proposition 3 follows. $\square$

OLS Estimator Using the alternative representations for the long- and short-run OLS estimators, we now represent the OLS estimator $\hat{\boldsymbol{\alpha}}_T$ as a transform of the primitive estimators: $\ddot{\boldsymbol{\mu}}_T$, $\ddot{\mathbf{v}}_T$, and $\ddot{\boldsymbol{\tau}}_T$. The following proposition reveals it:

Proposition 4. Given Assumption 1, $\hat{\alpha}_T = \mathbf{T}_T \ddot{\tau}_T$, where $\mathbf{T}_T := \mathbf{R}_T \mathbf{Q}_T$ such that

$$\mathbf{R}_T \mathbf{Q}_T = \mathbf{T}_T := \begin{bmatrix} \mathbf{T}_T^{11} & \mathbf{0} \\ \mathbf{T}_T^{21} & \mathbf{I}_{p+2kq+1} \end{bmatrix}, \quad \mathbf{T}_T^{11} := \begin{bmatrix} 1 & \mathbf{0}_{1 \times 2k} \\ -\tilde{\beta}_T & \mathbf{I}_{2k} \end{bmatrix} \quad \text{and}$$

$$\mathbf{T}_T^{21} := \begin{bmatrix} -\tilde{\zeta}_T & -\ddot{\mu}'_T \\ -\tilde{\nu}_T & \mathbf{0}_{1 \times 2k} \\ \mathbf{0}_{(p+2kq-1) \times 1} & \mathbf{0}_{(p+2kq-1) \times 2k} \end{bmatrix}. \quad \square$$

Remarks. (a) Proposition 4 follows by combining Propositions 1 and 3, in which we demonstrated that $\hat{\alpha}_T = \mathbf{R}_T \tilde{\omega}_T$ and $\tilde{\omega}_T = \mathbf{Q}_T \ddot{\tau}_T$, respectively, so that

$$\hat{\alpha}_T = \left[\ddot{\rho}_T, -\ddot{\rho}_T \ddot{\beta}'_T + \ddot{\eta}'_T, -\ddot{\rho}_T (\ddot{\vartheta}_T - \ddot{\mu}'_T \ddot{\beta}_T) - \ddot{\mu}'_T \ddot{\eta}_T + \ddot{\zeta}_T, -\ddot{\rho}_T \ddot{\nu}_T + \ddot{\gamma}_T, \ddot{\kappa}'_T \right]' \quad (10)$$

by further noting Proposition 2: $[\tilde{\beta}'_T, \tilde{\zeta}_T, \tilde{\nu}_T]' = [\ddot{\beta}'_T, \ddot{\vartheta}_T - \ddot{\mu}'_T \ddot{\beta}_T, \ddot{\nu}_T]'$. An alternative representation of $\hat{\alpha}_T$ is now provided without omitting any of its elements.

(b) If $\ddot{\mu}_T, \ddot{\nu}_T$ and $\ddot{\tau}_T$ are consistent for μ_*, ν_* and τ_* , respectively, then

$$\begin{aligned} \hat{\alpha}_T &\xrightarrow{\mathbb{P}} [\rho_*, -\rho_* \beta'_* + \eta'_*, -\rho_* \zeta_* - \mu'_* \eta_* + \varsigma_*, -\rho_* \nu_* + \gamma_*, \kappa'_*]' \\ &= [\rho_*, -\rho_* \beta'_*, -\rho_* \zeta_*, -\rho_* \nu_* + \gamma_*, \kappa'_*]' = [\rho_*, \theta'_*, \xi_*, \alpha_*, \kappa'_*]' =: \alpha_*, \end{aligned} \quad (11)$$

where the first equality holds because $\zeta_* = \vartheta_* - \mu'_* \beta_*$, $\eta_* = \mathbf{0}$ and $\varsigma_* = 0$. In addition, the second equality holds by the definitions of β_*, ζ_* , and γ_* . From this, the consistence of $\hat{\alpha}_T$ can be established.

(c) By Proposition 4, we can obtain the weak limit of $\hat{\alpha}_T$ using the weak limits of the primitive estimators: $\ddot{\mu}_T, \ddot{\nu}_T$, and $\ddot{\tau}_T$; see Theorem 1.

(d) The convergence rate of the OLS estimator $\hat{\alpha}_T$ is slower than $\mathbf{D}$ in Lemma 1. For example, for the long-run OLS estimator $\hat{\theta}_T$, we have $\hat{\theta}_T = -\ddot{\rho}_T \ddot{\beta}_T + \ddot{\eta}_T$ from (10), and the convergence rates of $\ddot{\rho}_T$ and $(\ddot{\beta}_T, \ddot{\eta}_T)$ are $T^{1/2}$ and T , respectively, by Lemmas 3 and 4 given below. From this, if for $j = 1, 2, \dots, k$, $\theta_{j*}^\pm \neq 0$, then the convergence rate of $\hat{\theta}_T$ is determined as $T^{1/2}$, which is slower than $T^{3/2}$ given in $\mathbf{D}$. Similar arguments apply to the other elements in $\hat{\alpha}_T$. This slower convergence rate is the same effect as expected when the asymptotic distribution of an estimator is obtained by a higher-order expansion (e.g., Teräsvirta, 1994; Cho and White, 2007, 2010; Cho and Phillips, 2018). Theorem 1 below elaborates on the convergence rates under various parameter conditions. $\square$

3.2 Nonsingular Primitive Estimators

We now show that the primitive estimators $\tilde{\tau}_T$ and $\tilde{v}_T$ do not suffer from an asymptotically singular matrix problem by deriving their limit distributions. By the definition of $\ddot{\mu}_T$, it does not suffer from an asymptotic singularity. We therefore focus on the limit distributions of $\tilde{\tau}_T$ and $\tilde{v}_T$. For this we let

$$\mathcal{B}(\cdot) := [\mathcal{B}_s(\cdot)', \mathcal{B}_u(\cdot), \mathcal{B}_e(\cdot), \mathcal{B}_{ue}(\cdot), \mathcal{B}_{ze}(\cdot)']' := \Sigma_*^{1/2} \mathcal{W}(\cdot),$$

where $\mathcal{W}(\cdot)$ is a vector of $(2 + p + 2k(1 + q))$ independent standard Wiener processes, and Σ_* is the global covariance matrix given in Assumption 1. Here, $\mathcal{B}(\cdot)$ is the Brownian motion obtained by applying FCLT to $\mathbf{B}_T(\cdot) := T^{-1/2} \sum_{t=1}^{[\cdot]T} \mathbf{w}_t$; see Lemma B.1 in the Online Supplement. For later purposes, we also partition $\mathbf{B}_T(\cdot)$ in parallel to $\mathcal{B}(\cdot)$:

$$\mathbf{B}_T(\cdot) := [\mathbf{B}_{sT}(\cdot)', B_{uT}(\cdot), B_{eT}(\cdot), B_{ueT}(\cdot), \mathbf{B}_{zeT}(\cdot)']' := \frac{1}{\sqrt{T}} \sum_{t=1}^{[\cdot]T} [s'_{t-1}, u_{t-1}, e_t, u_{t-1}e_t, \mathbf{z}'_{2t}e_t]'$$

by noting that $\mathbf{w}_t := [s'_{t-1}, u_{t-1}, e_t, u_{t-1}e_t, \mathbf{z}'_{2t}e_t]'$.

The following lemma shows that neither $\tilde{\tau}_T$ nor $\tilde{v}_T$ suffers from an asymptotically singular matrix problem.

Lemma 2. *Given Assumption 1,*

(i) $T^{-1/2} \left(\sum_{t=1}^T \ddot{u}_{t-1} \ddot{r}_{t-1} \right) \ddot{\mathbf{D}}_1^{-1} \Rightarrow \mathbf{M}_{1u} := \mathbf{0}_{(2k+2) \times 1}$, where $\ddot{\mathbf{D}}_1 := \text{diag}[T\mathbf{I}_{2k}, T^{3/2}, T^{1/2}]$;

(ii)

$$\ddot{\mathbf{D}}_1^{-1} \left(\sum_{t=1}^T \ddot{r}_{t-1} \ddot{r}'_{t-1} \right) \ddot{\mathbf{D}}_1^{-1} \Rightarrow \mathcal{M}_{11} := \begin{bmatrix} \int \bar{\mathcal{B}}_s \bar{\mathcal{B}}_s' & \mathbf{0}_{2k \times 1} & \int \widetilde{\mathcal{B}}_s \\ \mathbf{0}_{1 \times 2k} & \frac{1}{3} & \frac{1}{2} \\ \int \widetilde{\mathcal{B}}_s' & \frac{1}{2} & 1 \end{bmatrix},$$

where $\bar{\mathcal{B}}_s(\cdot) := \mathcal{B}_s(\cdot) - 3(\cdot) \int r \mathcal{B}_s$ and $\widetilde{\mathcal{B}}_s(\cdot) := \mathcal{B}_s(\cdot) - \frac{3}{2}(\cdot) \int r \mathcal{B}_s$;

(iii) $\ddot{\mathbf{D}}_2^{-1} \left(\sum_{t=1}^T \ddot{u}_{t-1} \mathbf{z}_{2t} \right) \ddot{\mathbf{D}}_2^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_{2u} := \mathbb{E}[u_{t-1} \mathbf{z}_{2t}]$, where $\ddot{\mathbf{D}}_2 := T^{1/2} \mathbf{I}_{p+2kq-1}$;

(iv)

$$\ddot{\mathbf{D}}_2^{-1} \left(\sum_{t=1}^T \mathbf{z}_{2t} \ddot{r}'_{t-1} \right) \ddot{\mathbf{D}}_2^{-1} \Rightarrow \mathcal{M}_{21} := \begin{bmatrix} \delta_* \boldsymbol{\iota}_{p-1} \int \widetilde{\mathcal{B}}_s' & \frac{1}{2} \delta_* \boldsymbol{\iota}_{p-1} & \delta_* \boldsymbol{\iota}_{p-1} \\ \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* \int \widetilde{\mathcal{B}}_s' & \frac{1}{2} \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* & \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* \end{bmatrix};$$

(v) $\ddot{\mathbf{D}}_2^{-1} \left(\sum_{t=1}^T \mathbf{z}_{2t} \mathbf{z}'_{2t} \right) \ddot{\mathbf{D}}_2^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_{22}$; and

(vi)

$$\ddot{\mathbf{D}}^{-1} \left(\sum_{t=1}^T \ddot{\mathbf{z}}_t \ddot{\mathbf{z}}_t' \right) \ddot{\mathbf{D}}^{-1} \Rightarrow \mathcal{M} := \begin{bmatrix} \sigma_u^2 & \mathbf{M}_{u1} & \mathbf{M}_{u2} \\ \mathbf{M}_{1u} & \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathbf{M}_{2u} & \mathcal{M}_{21} & \mathcal{M}_{22} \end{bmatrix},$$

where $\ddot{\mathbf{D}} := \text{diag}[T^{1/2}, \ddot{\mathbf{D}}_1, \ddot{\mathbf{D}}_2]$, $\mathcal{M}_{12} := \mathcal{M}'_{21}$, $\mathbf{M}_{u1} := \mathbf{M}'_{1u}$, $\mathbf{M}_{u2} := \mathbf{M}'_{2u}$, and $\sigma_u^2 := \mathbb{E}[u_t^2]$. $\square$

Both weak limits $\mathcal{M}_{11}$ and $\mathcal{M}$ given in Lemmas 2 (ii and vi) are nonsingular almost surely, meaning that the weak limits of $(\ddot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} \ddot{\mathbf{r}}_{t-1}' \ddot{\mathbf{D}}_1^{-1})^{-1}$ and $(\ddot{\mathbf{D}}^{-1} \sum_{t=1}^T \ddot{\mathbf{z}}_t \ddot{\mathbf{z}}_t' \ddot{\mathbf{D}}^{-1})^{-1}$ are obtained as $\mathcal{M}_{11}^{-1}$ and $\mathcal{M}^{-1}$, respectively, where the inverse matrices are those within the long- and short-run OLS estimators, $\ddot{\mathbf{v}}_T := (\sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} \ddot{\mathbf{r}}_{t-1}')^{-1} \sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} y_{t-1}$ and $\ddot{\mathbf{r}}_T := (\sum_{t=1}^T \ddot{\mathbf{z}}_t \ddot{\mathbf{z}}_t')^{-1} \sum_{t=1}^T \ddot{\mathbf{z}}_t \Delta y_t$. From this, it follows that neither $\ddot{\mathbf{v}}_T$ nor $\ddot{\mathbf{r}}_T$ suffers from an asymptotically singular matrix problem.

4 Limit Distribution of the OLS Estimator

In this section, we derive the limit distribution of $\hat{\alpha}_T$ using the weak limits of the primitive estimators. We proceed in three steps. First, we derive the weak limits of $\tilde{\mathbf{v}}_T$ and $\ddot{\mathbf{v}}_T$. Second, we conduct the same procedure for the primitive estimator $\ddot{\mathbf{r}}_T$. Finally, we derive the weak limit of $\hat{\alpha}_T$.

4.1 Weak Limit of the Long-Run OLS Estimator

We derive the weak limits of $\ddot{\mathbf{v}}_T$ and $\tilde{\mathbf{v}}_T$ using Proposition 2: $\tilde{\mathbf{v}}_T = \mathbf{P}_T \ddot{\mathbf{v}}_T$. Here,

$$\mathbf{P}_T \xrightarrow{\mathbb{P}} \mathbf{P} := \begin{bmatrix} \mathbf{I}_{2k} & \mathbf{0} \\ \mathbf{P}^{21} & \mathbf{I}_2 \end{bmatrix} \quad \text{with} \quad \mathbf{P}^{21} := \begin{bmatrix} -\boldsymbol{\mu}'_* \\ \mathbf{0}_{1 \times 2k} \end{bmatrix},$$

as $\ddot{\boldsymbol{\mu}}_T$ is consistent for $\boldsymbol{\mu}_*$. Next, we obtain the limit distribution of the primitive estimator $\ddot{\mathbf{v}}_T$. For this, we note that $\mathbf{m}_{t-1} = \ddot{\mathbf{m}}_{t-1} + (\ddot{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1)$ and rewrite (8) as

$$y_{t-1} = \ddot{\mathbf{r}}_{t-1}' \bar{\mathbf{v}}_{T*} + u_{t-1} = \boldsymbol{\beta}'_* \ddot{\mathbf{m}}_{t-1} + \vartheta_{T*}(t-1) + \nu_* + u_{t-1}, \quad \text{where } \bar{\mathbf{v}}_{T*} := [\boldsymbol{\beta}'_*, \vartheta_{T*}, \nu_*]' \quad (12)$$

with $\vartheta_{T*} := \boldsymbol{\beta}'_*(\ddot{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) + \vartheta_*$. This rearranges (8) so that the right-hand side variables in the long-run equation are the same as the regressors employed for $\ddot{\mathbf{v}}_T$. From this, it follows that

$$\ddot{\mathbf{v}}_T := \left(\sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} \ddot{\mathbf{r}}_{t-1}' \right)^{-1} \sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} y_{t-1} = \bar{\mathbf{v}}_{T*} + \left(\sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} \ddot{\mathbf{r}}_{t-1}' \right)^{-1} \sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} u_{t-1}. \quad (13)$$

Using this, we derive the weak limits of $\tilde{\mathbf{v}}_T$ and $\ddot{\mathbf{v}}_T$ as follows:

Lemma 3. Let $\varrho_{m*} := \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \sum_{\tau=1}^{t-1} \mathbb{E}[s_\tau u_t]$. Given Assumption 1,

(i) $\ddot{\mathbf{D}}_1(\ddot{\mathbf{v}}_T - \bar{\mathbf{v}}_{T*}) \Rightarrow \mathcal{L} := [\mathcal{L}'_1, \mathcal{L}_2, \mathcal{L}_3]' := \mathcal{M}_{11}^{-1} \mathcal{S}$, where

$$\mathcal{S} := \begin{bmatrix} \mathcal{S}_1 \\ \mathcal{S}_2 \\ \mathcal{S}_3 \end{bmatrix} := \begin{bmatrix} \int \mathcal{B}_s d\mathcal{B}_u - 3 \int r \mathcal{B}_s \int r d\mathcal{B}_u + \varrho_{m*} \\ \int r d\mathcal{B}_u \\ \int d\mathcal{B}_u \end{bmatrix},$$

such that $\mathcal{L}_1$ and $\mathcal{S}_1 \in \mathbb{R}^{2k}$; and

(ii) $\ddot{\mathbf{D}}_\dagger(\tilde{\mathbf{v}}_T - \mathbf{v}_*) \Rightarrow [\mathcal{L}'_1, -\boldsymbol{\mu}'_* \mathcal{L}_1, \mathcal{L}_3]'$, where $\ddot{\mathbf{D}}_\dagger := \text{diag}[T\mathbf{I}_{2k}, T, T^{1/2}]$. $\square$

Remarks. (a) By Lemma 3, none of $(\ddot{\mathbf{v}}_T - \bar{\mathbf{v}}_{T*})$ and $(\tilde{\mathbf{v}}_T - \mathbf{v}_*)$ follows a mixed normal distribution, and $\mathcal{S}_1$ is influenced by the asymptotic bias ϱ_{m*} . This bias corresponds to that arising when estimating a cointegrating relationship by OLS (e.g., [Engle and Granger, 1987](#); [Phillips and Hansen, 1990](#)).

(b) We prove Lemma 3 (i) by noting that $\ddot{\mathbf{D}}_1(\ddot{\mathbf{v}}_T - \bar{\mathbf{v}}_{T*}) = (\ddot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} \ddot{\mathbf{r}}'_{t-1} \ddot{\mathbf{D}}_1^{-1})^{-1} \ddot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} u_{t-1}$ and focusing on the weak limit of $\ddot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} u_{t-1}$. Lemma 2 (ii) already derives the weak limit of $(\ddot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \ddot{\mathbf{r}}_{t-1} \ddot{\mathbf{r}}'_{t-1} \ddot{\mathbf{D}}_1^{-1})^{-1}$. Next, we note that $\tilde{\mathbf{v}}_T - \mathbf{v}_* = (\mathbf{P}_T - \mathbf{P})(\ddot{\mathbf{v}}_T - \bar{\mathbf{v}}_{T*}) + \mathbf{P}(\ddot{\mathbf{v}}_T - \bar{\mathbf{v}}_{T*}) + (\mathbf{P}_T - \mathbf{P})(\bar{\mathbf{v}}_{T*} - \bar{\mathbf{v}}_*) + \mathbf{P}(\bar{\mathbf{v}}_{T*} - \bar{\mathbf{v}}_*) + (\mathbf{P}_T - \mathbf{P})\bar{\mathbf{v}}_*$ and prove Lemma 3 (ii) by examining each component on the right side. As a result, it follows that $\tilde{\mathbf{v}}_T - \mathbf{v}_* = \mathbf{P}(\ddot{\mathbf{v}}_T - \bar{\mathbf{v}}_{T*}) + O_{\mathbb{P}}(T^{-3/2})$ and $\ddot{\mathbf{D}}_\dagger(\ddot{\mathbf{v}}_T - \bar{\mathbf{v}}_{T*}) = O_{\mathbb{P}}(1)$.

(c) The consistency of the estimators follows from Lemma 3: $\ddot{\mathbf{v}}_T - \bar{\mathbf{v}}_{T*} \xrightarrow{\mathbb{P}} \mathbf{0}$ and $\tilde{\mathbf{v}}_T \xrightarrow{\mathbb{P}} \mathbf{v}_*$. We also have $\bar{\mathbf{v}}_{T*} \xrightarrow{\mathbb{P}} \bar{\mathbf{v}}_* := [\boldsymbol{\beta}'_*, \vartheta_*, \nu_*]'$, because $\vartheta_{T*} := \boldsymbol{\beta}'_*(\ddot{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) + \vartheta_* \xrightarrow{\mathbb{P}} \vartheta_*$. This further implies that $\mathbf{P}_T \xrightarrow{\mathbb{P}} \mathbf{P}$, so that $\tilde{\mathbf{v}}_T = \mathbf{P}_T \ddot{\mathbf{v}}_T \rightarrow \mathbf{P} \bar{\mathbf{v}}_* = [\boldsymbol{\beta}'_*, \zeta_*, \nu_*]' =: \mathbf{v}_*$.

(d) By Lemma 3 (ii), the weak limits of $\tilde{\boldsymbol{\beta}}_T$ and $\tilde{\zeta}_T$ are linearly correlated. This feature confirms that the long-run OLS estimator $\tilde{\mathbf{v}}_T$ suffers from the asymptotically singularity, although $\ddot{\mathbf{v}}_T$ does not.

(e) 2SNARDL estimates the long- and short-run parameters separately, and the convergence rate of the long-run 2SNARDL estimator is T , which is also the same as the convergence rate of $\ddot{\boldsymbol{\beta}}_T$ in $\ddot{\mathbf{v}}_T$. Meanwhile, the convergence rate of the OLS estimator for $\boldsymbol{\theta}_*$ depends on the unknown parameters. $\square$

4.2 Weak Limit of the Short-Run OLS Estimator

We derive the weak limit of the primitive estimator $\ddot{\mathbf{r}}_T$. For this purpose, we first rewrite (9) as

$$\begin{aligned} \Delta y_t = \boldsymbol{\tau}'_{T*} \ddot{\mathbf{z}}_t + e_t = \rho_* \ddot{u}_{t-1} + (\boldsymbol{\eta}_* + \rho_*(\ddot{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*))' \ddot{\mathbf{m}}_{t-1} + (\psi_* + \boldsymbol{\eta}'_* \ddot{\boldsymbol{\mu}}_T + \rho_*(\ddot{\vartheta}_T - \vartheta_{T*}))(t-1) \\ + (\gamma_* + \rho_*(\ddot{\nu}_T - \nu_*)) + \boldsymbol{\kappa}'_* \mathbf{z}_{2t} + e_t, \end{aligned}$$

where

$$\begin{aligned}\tau_{T*} &:= [\rho_*, \tau'_{1T}, \kappa_*']' \quad \text{and} \\ \tau_{1T} &:= [(\eta_* + \rho_*(\ddot{\beta}_T - \beta_*))', \psi_* + \eta'_* \ddot{\mu}_T + \rho_*(\ddot{\vartheta}_T - \vartheta_{T*}), \gamma_* + \rho_*(\ddot{\nu}_T - \nu_*)']'.\end{aligned}$$

This equality is established by noting that

$$u_{t-1} = \ddot{u}_{t-1} + (\ddot{\beta}_T - \beta_*)' \ddot{m}_{t-1} + (\ddot{\vartheta}_T - \vartheta_{T*})(t-1) + (\ddot{\nu}_T - \nu_*),$$

which follows from (12) and the definition of $\ddot{u}_{t-1} := y_{t-1} - \ddot{r}'_{t-1} \ddot{v}_T$. This representation is obtained so that the right-hand side variables in the short-run equation are the same as the regressors employed for $\ddot{\tau}_T$. That is, we can decompose the primitive estimator $\ddot{\tau}_T$ into τ_{T*} and a remainder using this representation:

$$\ddot{\tau}_T := \left(\sum_{t=1}^T \ddot{z}_t \ddot{z}_t' \right)^{-1} \sum_{t=1}^T \ddot{z}_t \Delta y_t = \tau_{T*} + \left(\sum_{t=1}^T \ddot{z}_t \ddot{z}_t' \right)^{-1} \sum_{t=1}^T \ddot{z}_t e_t. \quad (14)$$

We now provide the weak limit of the primitive estimator $\ddot{\tau}_T$ in the following lemma:

Lemma 4. *Given Assumption 1,*

- (i) $\ddot{D}(\ddot{\tau}_T - \tau_{T*}) \Rightarrow \mathcal{D} := [\mathcal{D}_1, \mathcal{D}_2', \mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5']' := \mathcal{M}^{-1} \mathcal{J}$, where $\mathcal{J} := [\int d\mathcal{B}_{ue}, \int \bar{\mathcal{B}}_s' d\mathcal{B}_e, \int r d\mathcal{B}_e, \int d\mathcal{B}_e, \int d\mathcal{B}_{ze}']'$;
- (ii) $\ddot{D}(\ddot{\tau}_T - \tau_*) \Rightarrow \mathcal{D} - \rho_*[0, \mathcal{L}', \mathbf{0}']'$. □

Remarks. (a) From the definition of $\mathcal{D}$, the primitive estimator $\ddot{D}(\ddot{\tau}_T - \tau_{T*})$ asymptotically follows a mixed normal distribution, but $\ddot{D}(\ddot{\tau}_T - \tau_*)$ does not. Lemma 2 (i) already shows that the weak limit $\mathcal{L}$ does not follow a mixed normal distribution, and the limit distribution of $\ddot{D}(\ddot{\tau}_T - \tau_*)$ is formed by both $\mathcal{D}$ and $\mathcal{L}$.

- (b) We prove Lemma 4 by examining the limit behavior of each component on the right side of (14). Because Lemma 2 (vi) already shows that $\ddot{D}^{-1}(\sum_{t=1}^T \ddot{z}_t \ddot{z}_t') \ddot{D}^{-1} \Rightarrow \mathcal{M}$, we focus on showing that $\ddot{D}^{-1} \sum_{t=1}^T \ddot{z}_t e_t \Rightarrow \mathcal{J}$. In addition, we note that $\ddot{D}(\ddot{\tau}_T - \tau_{T*}) = \ddot{D}(\ddot{\tau}_T - \tau_*) + [0, -\rho_* T(\ddot{\beta}_T - \beta_*)', -\rho_* T^{3/2}(\ddot{\vartheta}_T - \vartheta_{T*}), -\rho_* T^{1/2}(\ddot{\nu}_T - \nu_*), \mathbf{0}']'$, and exploit Lemmas 3 (ii) and 4 (i) to prove Lemma 4 (ii).
- (c) Unlike the weak limit of the long-run OLS estimator $\tilde{v}_T$, the weak limit of $\ddot{D}(\ddot{\tau}_T - \tau_*)$ cannot be written as a linear combination of other weak limits. This corroborates that the primitive estimator $\ddot{\tau}_T$ does not suffer from an asymptotically singular matrix problem.
- (d) Because Lemma 4 assumes that $\rho_* < 0$, we cannot use the weak limit in Lemma 4 (ii) to derive the null limit distribution of the t -statistic testing $\rho_* = 0$. □

4.3 Weak Limit of the OLS Estimator

We finally derive the weak limit of the OLS estimator $\hat{\alpha}_T$. For this derivation, we first note that

$$(\hat{\alpha}_T - \alpha_*) = (\mathbf{T}_T - \mathbf{T})\tau_* + \mathbf{T}(\ddot{\tau}_T - \tau_*) + (\mathbf{T}_T - \mathbf{T})(\ddot{\tau}_T - \tau_*), \quad (15)$$

by (10) and (11), where $\mathbf{T}$ is the probability limit of $\mathbf{T}_T$. That is,

$$\mathbf{T} := \begin{bmatrix} \mathbf{T}^{11} & \mathbf{0} \\ \mathbf{T}^{21} & \mathbf{I}_{p+2kq+1} \end{bmatrix}, \quad \mathbf{T}^{11} := \begin{bmatrix} 1 & \mathbf{0}_{1 \times 2k} \\ -\beta_* & \mathbf{I}_{2k} \end{bmatrix}, \quad \mathbf{T}^{21} := \begin{bmatrix} -\zeta_* & -\mu'_* \\ -\nu_* & \mathbf{0}_{1 \times 2k} \\ \mathbf{0}_{(p+2kq-1) \times 1} & \mathbf{0}_{(p+2kq-1) \times 2k} \end{bmatrix}.$$

Here, the weak limits of $\mathbf{T}_T$ and $\ddot{\tau}_T$ are already provided in Lemmas 3 and 4, respectively. Using these weak limits, we derive the weak limit of the OLS estimator $\hat{\alpha}_T$.

Before stating the weak limit, we note that (15) identifies the channel through which the unknown parameters and primitive estimators affect the weak limit. It is difficult to identify *a priori* how they affect the limit distribution of $\hat{\alpha}_T$, but (15) represents $\hat{\alpha}_T$ as a transformation of the unknown parameters and the primitive estimators for the long- and short-run OLS estimators, enabling us to identify how they affect the weak limit. For example, if for each $j = 1, 2, \dots, k$, $\beta_{j*}^\pm \neq 0$ and $\zeta_* \neq 0$, we can rewrite (15) as

$$\sqrt{T}(\hat{\alpha}_T - \alpha_*) = \mathbf{c}_* \sqrt{T}(\ddot{\rho}_T - \rho_*) + \sqrt{T} \mathbf{d}_T + o_{\mathbb{P}}(1), \quad (16)$$

where $\mathbf{c}_* := [1, -\beta'_*, -\zeta_*, -\nu_*, \mathbf{0}']'$ and $\mathbf{d}_T := [0, \mathbf{0}'_{2k \times 1}, 0, \{(\ddot{\gamma}_T - \gamma_*) - \rho_*(\ddot{\nu}_T - \nu_*)\}, (\ddot{\kappa}_T - \kappa_*)']'$. This representation implies that the weak limit is determined by the weak limits of $\ddot{\rho}_T$ and $\mathbf{d}_T$. However, if β_* or ζ_* is zero, we cannot use (16) to find the weak limit, because the estimator corresponding to β_* or ζ_* is asymptotically negligible. That is, $\sqrt{T}(\hat{\theta}_T - \theta_*) = o_{\mathbb{P}}(1)$ or $\sqrt{T}(\hat{\xi}_T - \xi_*) = o_{\mathbb{P}}(1)$, respectively. For such a case, we have to exploit the next-order terms to derive the weak limit of $(\hat{\theta}_T - \theta_*)$, demonstrating that the weak limit depends on the parameter value condition on β_* and ζ_* .

In the following theorem, we collect the weak limits obtained under various parameter conditions. For notational simplicity, we let $\mathcal{D}_2 \equiv [\mathcal{D}_2^{+'}, \mathcal{D}_2^{-'}]'$. Recall that $\mathcal{D}_2$ is the weak limit of $\ddot{\eta}_T - (\eta_* + \rho_*(\ddot{\beta}_T - \beta_*))$.

Theorem 1. *Given Assumption 1,*

- (i) *if for each $j = 1, 2, \dots, k$, $\beta_{j*}^\pm \neq 0$ and $\zeta_* \neq 0$, then $\sqrt{T}(\hat{\alpha}_T - \alpha_*) \Rightarrow \mathbf{c}_* \mathcal{D}_1 + [0, \mathbf{0}'_{2k \times 1}, 0, \mathcal{D}_4, \mathcal{D}_5']'$;*
- (ii) *if $\beta_*^+ = \mathbf{0}$, but for each $j = 1, 2, \dots, k$, $\beta_{j*}^- \neq 0$ and $\zeta_* \neq 0$, then $\ddot{\mathbf{D}}_+(\hat{\alpha}_T - \alpha_*) \Rightarrow [\mathcal{D}_1, \mathcal{D}_2^{+'}, -\beta_*^{-'} \mathcal{D}_1, -\zeta_* \mathcal{D}_1, \mathcal{D}_4 - \nu_* \mathcal{D}_1, \mathcal{D}_5']'$, where $\ddot{\mathbf{D}}_+ := \text{diag}[T^{1/2}, T \mathbf{I}_k, T^{1/2} \mathbf{I}_{k+2}, \ddot{\mathbf{D}}_2]$;*

- (iii) if $\beta_*^- = \mathbf{0}$, but for each $j = 1, 2, \dots, k$, $\beta_{j*}^+ \neq 0$ and $\zeta_* \neq 0$, then $\ddot{\mathbf{D}}_-(\hat{\alpha}_T - \alpha_*) \Rightarrow [\mathcal{D}_1, -\beta_*^{+'} \mathcal{D}_1, \mathcal{D}_2^{-'}, -\zeta_* \mathcal{D}_1, \mathcal{D}_4 - \nu_* \mathcal{D}_1, \mathcal{D}_5']'$, where $\ddot{\mathbf{D}}_- := \text{diag}[T^{1/2} \mathbf{I}_{k+1}, T \mathbf{I}_k, T^{1/2} \mathbf{I}_2, \ddot{\mathbf{D}}_2]$; and
- (iv) if for each $j = 1, 2, \dots, k$, $\beta_{j*}^\pm \neq 0$, but $\zeta_* = 0$, then $\ddot{\mathbf{D}}_\odot(\hat{\alpha}_T - \alpha_*) \Rightarrow [\mathcal{D}_1, -\beta_*' \mathcal{D}_1, -\mu_*' \mathcal{D}_2, \mathcal{D}_4 - \nu_* \mathcal{D}_1, \mathcal{D}_5']'$, where $\ddot{\mathbf{D}}_\odot := \text{diag}[T^{1/2} \mathbf{I}_{2k+1}, T, T^{1/2}, \ddot{\mathbf{D}}_2]$. $\square$

Remarks. (a) The limit distribution of $\hat{\alpha}_T$ is determined by $\mathcal{D}$ under all the parameter value conditions in Theorem 1, so that it asymptotically follows a mixed normal distribution by Lemma 4. The OLS estimator is formed by the primitive estimators $\ddot{v}_T$ and $\ddot{\tau}_T$, none of which follows a mixed normal distribution asymptotically, but $\hat{\alpha}_T$ follows a mixed normal distribution.

- (b) Although the time trend and nonstationary regressors are included as regressors, the convergence rate of $\hat{\alpha}_T$ in Theorem 1 (i) is slower than $\mathbf{D}$ in Lemma 1. Again, this slower convergence rate is the same effect as expected when the asymptotic distribution of an estimator is obtained by a higher-order expansion (e.g., Teräsvirta, 1994; Cho and White, 2007, 2010; Cho and Phillips, 2018).
- (c) We prove Theorem 1 (i) by deriving the weak limit of each component on the right side of (15). For example, we first derive the following for the long-run OLS estimator $\hat{\theta}_T$:

$$(\hat{\theta}_T - \theta_*) = -\beta_*(\ddot{\rho}_T - \rho_*) + (\ddot{\eta}_T - \eta_*) - \rho_*(\ddot{\beta}_T - \beta_*). \quad (17)$$

Lemmas 3 and 4 verify that $(\ddot{\rho}_T - \rho_*) = O_{\mathbb{P}}(T^{-1/2})$ and $(\ddot{\eta}_T - \eta_*) - \rho_*(\ddot{\beta}_T - \beta_*) = O_{\mathbb{P}}(T^{-1})$, respectively. Using these, we derive that $\sqrt{T}(\hat{\theta}_T - \theta_*) \Rightarrow -\beta_* \mathcal{D}_1$.

- (d) 2SNARDL has the same convergence rate as the OLS estimator when it is transformed to estimate the parameter targeted by OLS. 2SNARDL estimates ν_* and τ_* separately, and Cho, Greenwood-Nimmo, and Shin (2023a) show that the 2SNARDL for β_* and ρ_* has convergence rate T and $\sqrt{T}$, respectively. Thus, their product estimates θ_* with convergence rate $\sqrt{T}$, which is identical to that of $\hat{\theta}_T$ under Theorem 1 (i). Section A.4.1 in the Online Supplement corroborates this by simulation.
- (e) By Theorem 1 (i), the weak limit of $(\hat{\rho}_T, \hat{\theta}_T', \hat{\xi}_T)'$ is $(1, \beta_*', -\zeta_*)' \mathcal{D}_1$, indicating that the estimates are linearly correlated at the limit. This result confirms the asymptotically singular matrix problem and demonstrates how the convergence rate of $\hat{\alpha}_T$ is determined.
- (f) Despite the asymptotic singularity problem associated with $\hat{\alpha}_T$, its weak limit given in Theorem 1 (i) is determined by $\mathcal{D}_1$, $\mathcal{D}_4$, and $\mathcal{D}_5$. This implies the following. First, the OLS estimator follows a mixed normal distribution, so that if the standard t -test applies to $\hat{\alpha}_T$, it follows a mixed normal distribution under the null hypothesis and the condition in Theorem 1 (i). We verify this by simulation in the Online Supplement; see Section A.4.2. Second, $\mathcal{D}_1$ is the weak limit of the second coefficient obtained by regressing Δy_t against $(1, \ddot{u}_{t-1}, z_{2t}')'$. Thus, the null weak limit of the t -statistic testing

ρ_* (< 0) is equivalent to those of the t -statistics testing the long-run parameters β_* and ξ_* .

- (g) If the zero coefficient conditions in Theorems 1 (ii, iii, and iv) are imposed, the limit distribution of the OLS estimator has to be determined by the next-order term in (15) as mentioned above. Specifically, if $\beta_*^+ = \mathbf{0}$, (17) implies that $T(\hat{\theta}_T^+ - \theta_*^+) = T\{(\ddot{\eta}_T^+ - \eta_*^+) - \rho_*(\ddot{\beta}_T^+ - \beta_*^+)\} \Rightarrow \mathcal{D}_2^+$ by Lemma 4 (i). In parallel, if $\beta_*^- = \mathbf{0}$, then $T(\hat{\theta}_T^- - \theta_*^-) = T\{(\ddot{\eta}_T^- - \eta_*^-) - \rho_*(\ddot{\beta}_T^- - \beta_*^-)\} \Rightarrow \mathcal{D}_2^-$; and if $\zeta_* = 0$, then $T(\hat{\xi}_T - \xi_*) = -\mu_*' T\{(\ddot{\eta}_T - \eta_*) - \rho_*(\ddot{\beta}_T - \beta_*)\} + O_{\mathbb{P}}(T^{-1/2}) \Rightarrow -\mu_*' \mathcal{D}_2$, meaning that the weak limits of $\hat{\theta}_T$ and $\hat{\xi}_T$ are determined by $\mathcal{D}_2$. Given that $\mathcal{D}_2$ follows a mixed normal distribution, if the standard t -test applies to $\hat{\theta}_T$ or $\hat{\xi}_T$, its null weak limit also follows a mixed normal distribution.
- (h) There is a caveat to Theorems 1 (i, ii, iii, and iv). If $\beta_* = \mathbf{0}$ and $\zeta_* = 0$ simultaneously, then $\rho_* = 0$, by the remark below Assumption 1, which contradicts the assumption that $\rho_* < 0$. Theorems 1 (i, ii, iii, and iv) assume an environment in which at least one of β_*^+ , β_*^- , and ζ_* is nonzero.
- (i) As a remark relevant to (h), Pesaran et al. (2001) and Banerjee et al. (1998) provide the asymptotic critical values of the F - and t -statistics testing $\rho_* = 0$, showing that their null limit distributions cannot be approximated by a mixed normal distribution. This implies that the limit distributions of $\hat{\alpha}_T$ under the conditions in Theorem 1 are different from that under $\beta_* = \mathbf{0}$ and $\zeta_* = 0$.
- (j) For empirical application, the following procedure is applicable. First, it is necessary to test $\rho_* = 0$ using Pesaran et al.'s (2001) F -test or Banerjee et al.'s (1998) t -test as a preliminary procedure. Second, if ρ_* turns out to be negative, we can estimate the unknown coefficients by OLS. The null limit distribution of the t -test is mixed normal for each parameter except for the t -statistic testing $\rho_* = 0$. Third, when a joint hypothesis is tested, we can apply standard Wald test principle to the OLS estimator. By Theorem 1, the standard Wald test defined by OLS is asymptotically mixed chi-squared unless the joint hypothesis is involved with testing $\rho_* = 0$. The second and third steps are the same as for a standard regression analysis. Finally, when estimating the long-run parameter, we recommend using the primitive estimator $\ddot{v}_T$ or 2SNARDL. $\square$

Before closing this section, we note that NARDL model can also be estimated by instrumental variable (IV) estimation by overcoming singular matrix problems similarly to OLS (e.g., Phillips and Hansen, 1990). We leave this as a future research topic.

5 Hypotheses Testing

In this section, we develop a testing methodology by Wald test principle. In particular, we suppose that a cointegrating relationship holds between y_t and $\ddot{x}_t$ by supposing that $\theta_*^+ \neq \mathbf{0}$ and/or $\theta_*^- \neq \mathbf{0}$.

Our main interest is in testing the symmetry conditions. The NARDL process reduces to the ARDL

process if $\theta_*^+ = \theta_*^-$ and $\pi_*^+ = \pi_*^-$. Under the ARDL process condition, it is inefficient to estimate the parameters in NARDL, making it necessary to test the symmetry conditions. We specify the following three hypothesis systems:

$$\mathcal{H}'_0 : \theta_*^+ = \theta_*^- \quad \text{vs.} \quad \mathcal{H}'_1 : \theta_*^+ \neq \theta_*^-;$$

$$\mathcal{H}''_0 : \pi_*^+ = \pi_*^- \quad \text{vs.} \quad \mathcal{H}''_1 : \pi_*^+ \neq \pi_*^-;$$

$$\mathcal{H}'''_0 : \theta_*^+ = \theta_*^- \text{ and } \pi_*^+ = \pi_*^- \quad \text{vs.} \quad \mathcal{H}'''_1 : \theta_*^+ \neq \theta_*^- \text{ or } \pi_*^+ \neq \pi_*^-.$$

Here, $\mathcal{H}'_0$ and $\mathcal{H}''_0$ are provided to test the long- and short-run symmetries, respectively, and $\mathcal{H}'''_0$ hypothesizes both symmetries jointly.

Testing $\mathcal{H}'_0$ versus $\mathcal{H}'_1$ We first apply Wald test principle to test $\mathcal{H}'_0$. The standard Wald test applied to OLS is defined as follows:

$$W_T^{(1)} := \hat{\alpha}'_T \hat{\mathbf{R}}'_1 \left(\widehat{\mathbf{W}}_T^{(1)} \right)^{-1} \hat{\mathbf{R}}_1 \hat{\alpha}_T, \quad \text{where} \quad \widehat{\mathbf{W}}_T^{(1)} := \hat{\sigma}_{e,T}^2 \hat{\mathbf{R}}_1 \left(\sum_{t=1}^T z_t z'_t \right)^{-1} \hat{\mathbf{R}}'_1,$$

$\hat{\sigma}_{e,T}^2 := T^{-1} \sum_{t=1}^T (\Delta y_t - z'_t \hat{\alpha}_T)^2$, and $\hat{\mathbf{R}}_1 := [\mathbf{0}_{k \times 1}, \mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k \times (1+p+2kq)}]$. As we assume that $\theta_*^+ \neq \mathbf{0}$ and/or $\theta_*^- \neq \mathbf{0}$, we cannot suppose the environment assumed in Theorems 1 (ii or iii) under the null. Furthermore, we cannot directly apply the limit distribution of the OLS estimator in Theorems 1 (i and iv) to obtain the null limit distribution of the test. Note that $\hat{\mathbf{R}}_1 \hat{\alpha}_T = (\hat{\theta}_T^+ - \hat{\theta}_T^-)$, so that $\sqrt{T}(\hat{\theta}_T^+ - \hat{\theta}_T^-) \Rightarrow (\beta_*^- - \beta_*^+) \mathcal{D}_1$. Thus, $\sqrt{T}(\hat{\theta}_T^+ - \hat{\theta}_T^-)$ is asymptotically negligible under the null. The null limit behavior has to be obtained by the next-order term of $(\hat{\theta}_T^+ - \hat{\theta}_T^-)$. Specifically, (10) implies that

$$\begin{aligned} (\hat{\theta}_T^+ - \hat{\theta}_T^-) &= (\ddot{\eta}_T^+ - \ddot{\eta}_T^-) - \ddot{\rho}_T(\tilde{\beta}_T^+ - \tilde{\beta}_T^-) \\ &= (\ddot{\eta}_T^+ - \ddot{\eta}_T^-) - (\eta_*^+ - \eta_*^-) - \rho_*(\tilde{\beta}_T^+ - \tilde{\beta}_T^-) - (\ddot{\rho}_T - \rho_*)(\tilde{\beta}_T^+ - \tilde{\beta}_T^-) \\ &= (\ddot{\eta}_T^+ - \ddot{\eta}_T^-) - (\eta_{T*}^+ - \eta_{T*}^-) + o_{\mathbb{P}}(T^{-1}) \end{aligned} \tag{18}$$

under $\mathcal{H}'_0$, where $\eta_{T*} := \eta_* + \rho_*(\tilde{\beta}_T - \beta_*)$, and (18) is established by letting $\eta_{T*} := [\eta_{T*}^+, \eta_{T*}^-]'$ and noting that $\tilde{\beta}_T = \beta_T$. Therefore, the null limit distribution of $W_T^{(1)}$ is determined not by Theorem 1 (i or iv), but by Lemma 4. All these imply $(\hat{\theta}_T^+ - \hat{\theta}_T^-) = O_{\mathbb{P}}(T^{-1})$, so that if we let $\mathbf{R}_1 := [\mathbf{0}_{k \times 1}, \mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k \times (1+p+2kq)}]$, $T(\hat{\theta}_T^+ - \hat{\theta}_T^-) \Rightarrow \mathbf{R}_1 \mathcal{D}$ under $\mathcal{H}'_0$ by Lemma 4. However, its alternative behavior is differently determined. From $\hat{\mathbf{R}}_1 \hat{\alpha}_T = (\hat{\theta}_T^+ - \hat{\theta}_T^-)$, its divergence rate is determined by Theorem 1 (i or iv), meaning that $\hat{\mathbf{R}}_1 \hat{\alpha}_T = O_{\mathbb{P}}(T^{-1/2})$. Therefore, the convergence and divergence rates of the test basis are unbalanced between $\mathcal{H}'_0$ and $\mathcal{H}'_1$, respectively.

We supplement the unbalanced rates by applying Wald test principle to the primitive estimator $\ddot{v}_T$. This is mainly because the primitive estimator $\ddot{v}_T$ enables us to define another Wald test using the next-order terms. Because testing $\mathcal{H}'_0$ is equivalent to testing

$$\mathbb{H}'_0 : \beta_*^+ - \beta_*^- = \mathbf{0},$$

we apply Wald test principle to the primitive estimator $\ddot{\beta}_T$. If we let $\ddot{\mathbf{R}}_1 := [\mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k \times 2}]$, $\ddot{\mathbf{R}}_1 \ddot{v}_{T*} = \beta_*^+ - \beta_*^-$, which we can estimate by $\ddot{\beta}_T^+ - \ddot{\beta}_T^- = \ddot{\mathbf{R}}_1 \ddot{v}_T$. Using this, we define another Wald test:

$$\mathcal{W}_T^{(1)} := \ddot{v}_T' \ddot{\mathbf{D}}_1 \ddot{\mathbf{R}}_1' (\ddot{\mathbf{W}}_T^{(1)})^{-1} \ddot{\mathbf{R}}_1 \ddot{\mathbf{D}}_1 \ddot{v}_T, \quad \text{where} \quad \ddot{\mathbf{W}}_T^{(1)} := \ddot{\sigma}_{u,T}^2 \ddot{\mathbf{R}}_1 \ddot{\mathbf{D}}_1 \left(\sum_{t=1}^T \ddot{r}_{t-1} \ddot{r}_{t-1}' \right)^{-1} \ddot{\mathbf{D}}_1 \ddot{\mathbf{R}}_1'$$

and $\ddot{\sigma}_{u,T}^2 := T^{-1} \sum_{t=1}^T \ddot{u}_t^2$. Here, $(\ddot{\beta}_T^+ - \ddot{\beta}_T^-)$ and $(\widehat{\theta}_T^+ - \widehat{\theta}_T^-)$ are different by the presence of $(\ddot{\eta}_T^+ - \ddot{\eta}_T^-)$ and $\ddot{\rho}_T$ as given in (18) such that $(\ddot{\beta}_T^+ - \ddot{\beta}_T^-) = (\widehat{\beta}_T^+ - \widehat{\beta}_T^-)$. Furthermore, $(\ddot{\beta}_T^+ - \ddot{\beta}_T^-) = O_{\mathbb{P}}(T^{-1})$ under both hypotheses, letting the convergence and divergence rates of $\mathcal{W}_T^{(1)}$ be identical under both $\mathcal{H}'_0$ and $\mathcal{H}'_1$.

Testing $\mathcal{H}''_0$ versus $\mathcal{H}''_1$ We next apply Wald test principle to test $\mathcal{H}''_0$. The Wald test principle applied to $\widehat{\alpha}_T$ delivers the following test:

$$W_T^{(2)} := \widehat{\alpha}_T' \widehat{\mathbf{R}}_2' (\widehat{\mathbf{W}}_T^{(2)})^{-1} \widehat{\mathbf{R}}_2 \widehat{\alpha}_T, \quad \text{where} \quad \widehat{\mathbf{W}}_T^{(2)} := \widehat{\sigma}_{e,T}^2 \widehat{\mathbf{R}}_2 \left(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t' \right)^{-1} \widehat{\mathbf{R}}_2'$$

and $\widehat{\mathbf{R}}_2 := [\mathbf{0}_{kq \times (2+p+2k)}, \mathbf{I}_{kq}, -\mathbf{I}_{kq}]$, so that it follows that $\widehat{\mathbf{R}}_2 \widehat{\alpha}_T = \widehat{\pi}_T^+ - \widehat{\pi}_T^-$ and $\sqrt{T} \widehat{\mathbf{R}}_2 \widehat{\alpha}_T \Rightarrow \widehat{\mathbf{R}}_2 \mathcal{D}$ using the limit distribution in Theorem 1.

We define another supplementary Wald test using the primitive estimator. From Proposition 4, it follows that $\widehat{\kappa}_T = \ddot{\kappa}_T$. That is, the short-run OLS and its primitive estimator are identical, so that we can test $\mathcal{H}''_0$ by using $\ddot{\tau}_T$. By Lemma 2 (vi), $\ddot{\tau}_T$ is not associated with an asymptotically singular matrix problem. We define the following Wald test:

$$\mathcal{W}_T^{(2)} := \ddot{\tau}_T' \ddot{\mathbf{D}} \widehat{\mathbf{R}}_2' (\ddot{\mathbf{W}}_T^{(2)})^{-1} \widehat{\mathbf{R}}_2 \ddot{\mathbf{D}} \ddot{\tau}_T, \quad \text{where} \quad \ddot{\mathbf{W}}_T^{(2)} := \widehat{\sigma}_{e,T}^2 \widehat{\mathbf{R}}_2 \ddot{\mathbf{D}} \left(\sum_{t=1}^T \ddot{z}_t \ddot{z}_t' \right)^{-1} \ddot{\mathbf{D}} \widehat{\mathbf{R}}_2'.$$

Although the test basis of the Wald test $\mathcal{W}_T^{(2)}$ is the same as that of $W_T^{(2)}$, from the fact that $\widehat{\pi}_T^+ - \widehat{\pi}_T^- = \widehat{\pi}_T^+ - \widehat{\pi}_T^-$, the weight matrix of $\mathcal{W}_T^{(2)}$ is different from $W_T^{(2)}$. The weight matrix $\ddot{\mathbf{W}}_T^{(2)}$ is more relevant to the test basis, because $\ddot{\tau}_T$ is obtained by regressing Δy_t on $\ddot{z}_t$.

Testing $\mathcal{H}_0'''$ versus $\mathcal{H}_1'''$ Finally, we apply Wald test principle to $\hat{\alpha}_T$ as follows:

$$W_T^{(3)} := \hat{\alpha}_T' \hat{\mathbf{R}}_3' (\hat{\mathbf{W}}_T^{(3)})^{-1} \hat{\mathbf{R}}_3 \hat{\alpha}_T, \quad \text{where} \quad \hat{\mathbf{W}}_T^{(3)} := \hat{\sigma}_{e,T}^2 \hat{\mathbf{R}}_3 \left(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t' \right)^{-1} \hat{\mathbf{R}}_3'$$

and $\hat{\mathbf{R}}_3 := \text{diag}[\hat{\mathbf{R}}_1, \hat{\mathbf{R}}_2]$, so that $\hat{\mathbf{R}}_3 \hat{\alpha}_T = (\hat{\theta}_T^{+'} - \hat{\theta}_T^{-'}, \hat{\pi}_T^{+'} - \hat{\pi}_T^{-'})'$. Here, the null weak limit of $\hat{\mathbf{R}}_3 \hat{\alpha}_T$ is obtained by combining the fact that $T(\hat{\theta}_T^+ - \hat{\theta}_T^-) \Rightarrow \mathbf{R}_1 \mathcal{D}$ and $\sqrt{T}(\hat{\pi}_T^+ - \hat{\pi}_T^-) \Rightarrow \hat{\mathbf{R}}_2 \mathcal{D}$ under $\mathcal{H}_0'''$. Therefore, if we let $\mathbf{R}_3 := \text{diag}[\mathbf{R}_1, \hat{\mathbf{R}}_2]$, $(T(\hat{\theta}_T^{+'} - \hat{\theta}_T^{-'}), \sqrt{T}(\hat{\pi}_T^{+'} - \hat{\pi}_T^{-'}))' \Rightarrow \mathbf{R}_3 \mathcal{D}$ under $\mathcal{H}_0'''$. However, $\hat{\mathbf{R}}_3 \hat{\alpha}_T = O_{\mathbb{P}}(T^{-1/2})$ under $\mathcal{H}_1'''$ because $(\hat{\theta}_T^{+'} - \hat{\theta}_T^{-'})' = O_{\mathbb{P}}(T^{-1/2})$ as discussed above.

We define another supplementary Wald test and overcome the unbalanced convergence and divergence rates of $W_T^{(3)}$ between $\mathcal{H}_0'''$ and $\mathcal{H}_1'''$. For this, we first reformulate $\mathcal{H}_0'''$ against $\mathcal{H}_1'''$ into

$$\mathbb{H}_0''' : \beta_*^+ = \beta_*^- \quad \text{and} \quad \pi_*^+ = \pi_*^- \quad \text{vs.} \quad \mathbb{H}_1''' : \beta_*^+ \neq \beta_*^- \quad \text{or} \quad \pi_*^+ \neq \pi_*^-,$$

and define the following supplementary Wald test:

$$\mathcal{W}_T^{(3)} := \ddot{\delta}_T' \bar{\mathbf{D}} \ddot{\mathbf{R}}_3' (\ddot{\mathbf{W}}_T^{(3)})^{-1} \ddot{\mathbf{R}}_3 \bar{\mathbf{D}} \ddot{\delta}_T,$$

where $\ddot{\delta}_T := (\ddot{v}_T', \ddot{\tau}_T')'$, $\bar{\mathbf{D}} := \text{diag}[\ddot{\mathbf{D}}_1, \ddot{\mathbf{D}}]$, $\ddot{\mathbf{R}}_3 := \text{diag}[\ddot{\mathbf{R}}_1, \hat{\mathbf{R}}_2]$, and $\ddot{\mathbf{W}}_T^{(3)} := \text{diag}[\ddot{\mathbf{W}}_T^{(1)}, \ddot{\mathbf{W}}_T^{(2)}]$. The Wald test $\mathcal{W}_T^{(3)}$ is defined by applying Wald test principle to both primitive estimators $\ddot{v}_T$ and $\ddot{\tau}_T$ and by noting that $\ddot{\mathbf{R}}_3 \bar{\mathbf{D}} \ddot{\delta}_T = (T(\ddot{\beta}_T^+ - \ddot{\beta}_T^-)', \sqrt{T}(\ddot{\pi}_T^+ - \ddot{\pi}_T^-)')'$. It straightforwardly follows that $\mathcal{W}_T^{(3)} = \mathcal{W}_T^{(1)} + \mathcal{W}_T^{(2)}$ from the fact that $\ddot{\mathbf{W}}_T^{(3)}$ is block-diagonal. From this, its null and alternative limit behaviors are determined by those of $\mathcal{W}_T^{(1)}$ and $\mathcal{W}_T^{(2)}$.

We now provide the limit behaviors of the Wald tests. For notational simplicity, we partition $\mathbb{H}_0'''$ into $\mathbb{H}_{01}''' : \beta_*^+ = \beta_*^-$ and $\mathbb{H}_{02}''' : \pi_*^+ = \pi_*^-$. In parallel, we also partition $\mathbb{H}_1'''$ into $\mathbb{H}_{11}''' : \beta_*^+ \neq \beta_*^-$ and $\mathbb{H}_{12}''' : \pi_*^+ \neq \pi_*^-$. These partitions are provided because the power of $\mathcal{W}_T^{(3)}$ depends on the sub-hypotheses from the fact that the divergence rate of $(\ddot{\beta}_T^+ - \ddot{\beta}_T^-)$ differs from that of $(\ddot{\pi}_T^+ - \ddot{\pi}_T^-)$. We also let $\sigma_e^2 := \mathbb{E}[e_t^2]$. The following theorem summarizes their limit behaviors:

Theorem 2. *Given Assumption 1, if $\theta_*^+ \neq \mathbf{0}$ and/or $\theta_*^- \neq \mathbf{0}$,*

- (i) (a) $W_T^{(1)} \Rightarrow \mathcal{D}' \mathbf{R}_1' (\sigma_e^2 \mathbf{R}_1 \mathcal{M}^{-1} \mathbf{R}_1')^{-1} \mathbf{R}_1 \mathcal{D}$ under $\mathcal{H}_0'$, where $\mathbf{R}_1 := [\mathbf{0}_{k \times 1}, \mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k \times (1+p+2kq)}]$;
- (b) $W_T^{(2)} \Rightarrow \mathcal{D}' \hat{\mathbf{R}}_2' (\sigma_e^2 \hat{\mathbf{R}}_2 \mathcal{M}^{-1} \hat{\mathbf{R}}_2')^{-1} \hat{\mathbf{R}}_2 \mathcal{D}$ under $\mathcal{H}_0''$;
- (c) $W_T^{(3)} \Rightarrow \mathcal{D}' \mathbf{R}_3' (\sigma_e^2 \mathbf{R}_3 \mathcal{M}^{-1} \mathbf{R}_3')^{-1} \mathbf{R}_3 \mathcal{D}$ under $\mathcal{H}_0'''$, where $\mathbf{R}_3 := \text{diag}[\mathbf{R}_1, \hat{\mathbf{R}}_2]$;
- (ii) (a) for any $c_T' = o(T)$, $\lim_{T \rightarrow \infty} \mathbb{P}(W_T^{(1)} > c_T') = 1$ under $\mathcal{H}_1'$;
- (b) for any $c_T' = o(T)$, $\lim_{T \rightarrow \infty} \mathbb{P}(W_T^{(2)} > c_T') = 1$ under $\mathcal{H}_1''$;

- (c) for any $c'_T = o(T)$, $\lim_{T \rightarrow \infty} \mathbb{P}(W_T^{(3)} > c'_T) = 1$ under $\mathcal{H}_1'''$;
- (iii) (a) $\mathcal{W}_T^{(1)} \Rightarrow \mathcal{L}' \ddot{\mathbf{R}}_1' (\sigma_u^2 \ddot{\mathbf{R}}_1 \mathcal{M}_{11}^{-1} \ddot{\mathbf{R}}_1')^{-1} \ddot{\mathbf{R}}_1 \mathcal{L}$ under $\mathbb{H}_0'$;
- (b) $\mathcal{W}_T^{(2)} \Rightarrow \mathcal{D}' \hat{\mathbf{R}}_2' (\sigma_e^2 \hat{\mathbf{R}}_2 \mathcal{M}^{-1} \hat{\mathbf{R}}_2')^{-1} \hat{\mathbf{R}}_2 \mathcal{D}$ under $\mathcal{H}_0''$;
- (c) $\mathcal{W}_T^{(3)} \Rightarrow \mathcal{L}' \ddot{\mathbf{R}}_1' (\sigma_u^2 \ddot{\mathbf{R}}_1 \mathcal{M}_{11}^{-1} \ddot{\mathbf{R}}_1')^{-1} \ddot{\mathbf{R}}_1 \mathcal{L} + \mathcal{D}' \hat{\mathbf{R}}_2' (\sigma_e^2 \hat{\mathbf{R}}_2 \mathcal{M}^{-1} \hat{\mathbf{R}}_2')^{-1} \hat{\mathbf{R}}_2 \mathcal{D}$ under $\mathbb{H}_0'''$; and
- (iv) (a) for any $c_T = o(T^2)$, $\lim_{T \rightarrow \infty} \mathbb{P}(\mathcal{W}_T^{(1)} > c_T) = 1$ under $\mathbb{H}_1'$;
- (b) for any $c'_T = o(T)$, $\lim_{T \rightarrow \infty} \mathbb{P}(\mathcal{W}_T^{(2)} > c'_T) = 1$ under $\mathcal{H}_1''$;
- (c) for any $c_T = o(T^2)$ and $c'_T = o(T)$,
- (1) $\lim_{T \rightarrow \infty} \mathbb{P}(\mathcal{W}_T^{(3)} > c_T) = 1$ under $\mathbb{H}_{11}''' \cap \mathbb{H}_{02}'''$;
- (2) $\lim_{T \rightarrow \infty} \mathbb{P}(\mathcal{W}_T^{(3)} > c'_T) = 1$ under $\mathbb{H}_{01}''' \cap \mathbb{H}_{12}'''$; and
- (3) $\lim_{T \rightarrow \infty} \mathbb{P}(\mathcal{W}_T^{(3)} > c_T) = 1$ under $\mathbb{H}_{11}''' \cap \mathbb{H}_{12}'''$. $\square$

Remarks. (a) The asymptotic behaviors of $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ are determined by the weak limit of the primitive estimator $(\ddot{\tau}_T - \tau_{T*})$. For example, it follows from (18) that $(\hat{\theta}_T^+ - \hat{\theta}_T^-) = (\theta_*^+ - \theta_*^-) - (\ddot{\rho}_T - \rho_*)(\beta_*^+ - \beta_*^-) + (\ddot{\eta}_T^+ - \eta_{T*}^+) - (\ddot{\eta}_T^- - \eta_{T*}^-) + o_{\mathbb{P}}(T^{-1})$, so that $(\hat{\theta}_T^+ - \hat{\theta}_T^-) = (\ddot{\eta}_T^+ - \eta_{T*}^+) - (\ddot{\eta}_T^- - \eta_{T*}^-) + o_{\mathbb{P}}(T^{-1})$ under $\mathcal{H}_0'$. In other words, $\hat{\mathbf{R}}_1 \hat{\alpha}_T = \mathbf{R}_1(\ddot{\tau}_T - \tau_{T*}) + o_{\mathbb{P}}(T^{-1})$. This fact implies that the null weak limit of $W_T^{(1)}$ is determined by that of $(\ddot{\tau}_T - \tau_{T*})$. Using Lemma 4, Theorem 2 (i.a) reports its weak limit under $\mathcal{H}_0'$. Meanwhile, the alternative weak limit of the test basis is obtained from $(\hat{\theta}_T^+ - \hat{\theta}_T^-) = (\theta_*^+ - \theta_*^-) - (\ddot{\rho}_T - \rho_*)(\beta_*^+ - \beta_*^-) + o_{\mathbb{P}}(T^{-1/2})$ and $(\ddot{\rho}_T - \rho_*) = O_{\mathbb{P}}(T^{-1/2})$. Theorem 2 (ii.a) reports the power behavior of the test implied by this representation. We apply similar arguments to the other Wald tests, $W_T^{(2)}$ and $W_T^{(3)}$ and provide their limit behaviors in Theorem 2 (i.b, i.c, ii.b, and ii.c). Consequently, $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ asymptotically follow mixed chi-squared distributions under the respective null hypotheses.

- (b) If we partition $\mathcal{L}_1$ such that $\mathcal{L}_1 \equiv (\mathcal{L}_1^{+'}, \mathcal{L}_1^{-'})'$, $\ddot{\mathbf{R}}_1 \mathcal{L} = \mathcal{L}_1^+ - \mathcal{L}_1^-$, so that the null weak limit in Theorem 2 (iii.a) is given as $(\mathcal{L}_1^+ - \mathcal{L}_1^-)' (\sigma_u^2 \ddot{\mathbf{R}}_1 \mathcal{M}_{11}^{-1} \ddot{\mathbf{R}}_1')^{-1} (\mathcal{L}_1^+ - \mathcal{L}_1^-)$. In parallel, if we let $\mathcal{D}_5 \equiv [\mathcal{D}_5^{+'}, \mathcal{D}_5^{+'}, \mathcal{D}_5^{-'}]'$, such that $\mathcal{D}_5^+$, $\mathcal{D}_5^+$, and $\mathcal{D}_5^-$ are the weak limits of $\ddot{\varphi}_T$, $\ddot{\pi}_T^+$, and $\ddot{\pi}_T^-$, respectively, then $\hat{\mathbf{R}}_2 \mathcal{D} = \mathcal{D}_5^+ - \mathcal{D}_5^-$, and so the weak the limit in Theorem 2 (iii.b) can be rewritten as $(\mathcal{D}_5^+ - \mathcal{D}_5^-)' (\sigma_e^2 \hat{\mathbf{R}}_2 \mathcal{M}^{-1} \hat{\mathbf{R}}_2')^{-1} (\mathcal{D}_5^+ - \mathcal{D}_5^-)$.
- (c) The test methodology of using the supplementary tests depends on the hypothesis of consideration. We can rewrite the null weak limit of $\mathcal{W}_T^{(1)}$ as $\mathcal{S}' \mathcal{M}_{11}^{-1} \ddot{\mathbf{R}}_1' (\sigma_u^2 \ddot{\mathbf{R}}_1 \mathcal{M}_{11}^{-1} \ddot{\mathbf{R}}_1')^{-1} \ddot{\mathbf{R}}_1 \mathcal{M}_{11}^{-1} \mathcal{S}$ using the definition of $\mathcal{L} := \mathcal{M}_{11}^{-1} \mathcal{S}$, meaning that $\mathcal{W}_T^{(1)}$ does not follow a mixed chi-squared distribution. This follows from the fact that $\mathcal{S}$ does not follow a mixed normal distribution and is also influenced by the asymptotic bias ϱ_{m*} . The same result also holds for $\mathcal{W}_T^{(3)}$. However, the null weak limit of $\mathcal{W}_T^{(2)}$ is characterized by $\mathcal{D}$, meaning that its null weak limit follows a mixed chi-squared distribution. For both $\mathcal{W}_T^{(1)}$ and $\mathcal{W}_T^{(3)}$, we can test the hypotheses by applying a resampling method; see Section 6.

- (d) The divergence speed of $\mathcal{W}_T^{(3)}$ depends on the alternative sub-hypothesis. If $\mathbb{H}_0'''$ is negated only because of the asymmetric long-run parameters: $\beta_*^+ \neq \beta_*^-$, the divergence speed of $\mathcal{W}_T^{(3)}$ is T^2 . In contrast, if only the short-run parameters are asymmetric: $\pi_*^+ \neq \pi_*^-$, its divergence speed is T . If both the long- and short-run parameters are asymmetric, the divergence speed is determined as T^2 . Theorem 2 (iv.c) summarizes these results. $\square$

Before moving to the next section, we summarize the use of the Wald tests introduced in this section. Table 1 provides their summary. We contain the formulas of the Wald tests along with their null hypotheses and null limit distributions. In case the null limit distribution is mixed chi-squared, the researcher can obtain the asymptotic critical values from the chi-square distribution table. Otherwise, the residual bootstrap is applicable. Section 6 details the residual bootstrap procedure.

6 Monte Carlo Simulations

In this section, we conduct simulations and examine the finite sample properties of the Wald tests.

For our simulation, we assume the following DGP condition:

$$y_{t-1} = \nu_* + \beta_*^+ x_{t-1}^+ + \beta_*^- x_{t-1}^- + \zeta_*(t-1) + u_{t-1} \quad \text{and}$$

$$\Delta y_t = \alpha_* + \rho_* u_{t-1} + \varphi_* \Delta y_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t,$$

where $\Delta x_t = 1/4 + \kappa_* \Delta x_{t-1} + v_t$, and $(e_t, v_t)' \sim \text{IIDN}(\mathbf{0}_2, \mathbf{I}_2)$. By this DGP condition, both y_t and x_t are integrated series with time trends. We also set $(\nu_*, \zeta_*, \alpha_*, \rho_*, \varphi_*) = (0, 0, 0, -1/2, 0)$ throughout the simulation, but adjust the value of $(\beta_*^+, \beta_*^-, \pi_*^+, \pi_*^-)$, depending on the hypotheses of interest. According to the NARDL condition, it must hold that $\theta_*^+ = -\rho_* \beta_*^+$ and $\theta_*^- = -\rho_* \beta_*^-$.

The next procedure applies to define the Wald tests. First, we estimate the unknown parameters using the primitive parameter estimators. Specifically, we estimate $\bar{\nu}_{T*}$ and τ_{T*} separately by specifying the following models:

$$y_t = \beta^+ \ddot{m}_t^+ + \beta^- \ddot{m}_t^- + \vartheta t + \nu + u_t \quad \text{and}$$

$$\Delta y_t = \rho \tilde{u}_{t-1} + \eta^+ \ddot{m}_{t-1}^+ + \eta^- \ddot{m}_{t-1}^- + \varsigma(t-1) + \gamma + \varphi \Delta y_{t-1} + \pi^+ \Delta x_t^+ + \pi^- \Delta x_t^- + e_t,$$

where we set $\ddot{m}_t^+ := x_t^+ - t\ddot{\mu}_T^+$, $\ddot{m}_t^- := x_t^- - t\ddot{\mu}_T^-$, and $\tilde{u}_t := y_t - \mathbf{r}_t' \tilde{\mathbf{v}}_T$, with $\ddot{\mu}_T^+ := (\sum_{t=1}^{T-1} t^2)^{-1} \sum_{t=1}^{T-1} t x_t^+$, $\ddot{\mu}_T^- := (\sum_{t=1}^{T-1} t^2)^{-1} \sum_{t=1}^{T-1} t x_t^-$, $\tilde{\mathbf{v}}_T := (\sum_{t=1}^T \mathbf{r}_{t-1} \mathbf{r}_{t-1}')^{-1} \sum_{t=1}^T \mathbf{r}_{t-1} y_{t-1}$, and $\mathbf{r}_{t-1} := [x_{t-1}^+, x_{t-1}^-, (t -$

1), 1]'. Second, we assume the following hypothesis systems:

$$\mathcal{H}'_0 : \theta_*^+ = \theta_*^- \text{ vs. } \mathcal{H}'_1 : \theta_*^+ \neq \theta_*^-;$$

$$\mathcal{H}''_0 : \pi_*^+ = \pi_*^- \text{ vs. } \mathcal{H}''_1 : \pi_*^+ \neq \pi_*^-;$$

$$\mathcal{H}'''_0 : \theta_*^+ = \theta_*^- \text{ and } \pi_*^+ = \pi_*^- \text{ vs. } \mathcal{H}'''_1 : \theta_*^+ \neq \theta_*^- \text{ or } \pi_*^+ \neq \pi_*^-;$$

$$\mathbb{H}'_0 : \beta_* = 0 \text{ vs. } \mathbb{H}'_1 : \beta_* \neq 0; \text{ and } \mathbb{H}'''_0 : \beta_* = 0 \text{ and } \pi_*^+ = \pi_*^- \text{ vs. } \mathbb{H}'''_1 : \beta_* \neq 0 \text{ or } \pi_*^+ \neq \pi_*^-.$$

Note that these hypotheses correspond to those in Section 5. Finally, we compute the Wald tests $\mathcal{W}_T^{(1)}$, $\mathcal{W}_T^{(2)}$, $\mathcal{W}_T^{(3)}$, $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$, as stated in Section 5.

We conduct simulations under the following two DGP conditions assuming $\kappa_* = 1/2$. First, we set $\beta_*^+ = \beta_*^- = 1$ and $\pi_*^+ = \pi_*^- = 1/2$ to generate data. This parameter condition satisfies the ARDL condition. From this, we examine the finite sample properties of the Wald tests under the null. Second, we set $\beta_*^+ = 1/4$, $\beta_*^- = -1/4$, $\pi_*^+ = 1/8$, and $\pi_*^- = -1/8$, and use this to examine the power of the tests.

A bootstrap method is used for the testing. Theorem 2 shows that the null limit distributions of the Wald tests $\mathcal{W}_T^{(1)}$ and $\mathcal{W}_T^{(3)}$ are not mixed chi-squared. So, we apply the following residual bootstrap procedure:

- S1:** After computing the Wald tests, we estimate the ARDL model by regressing Δy_t against y_{t-1} , x_{t-1} , $(t-1)$, 1, Δy_{t-1} , and Δx_t . We let the estimated linear coefficient be $(\hat{\rho}_T, \hat{\theta}_T, \hat{\xi}_T, \hat{\alpha}_T, \hat{\varphi}_T, \hat{\pi}_T)$. We also let the residual be $\hat{e}_t := \Delta y_t - \hat{\rho}_T y_{t-1} - \hat{\theta}_T x_{t-1} - \hat{\xi}_T(t-1) - \hat{\alpha}_T - \hat{\varphi}_T \Delta y_{t-1} - \hat{\pi}_T \Delta x_t$.
- S2:** We construct a resampled series as follows. First, we resample $\hat{e}_t$ with replacement and denote it as e_t^b . Next, we let

$$\Delta y_t^b := \hat{\rho}_T y_{t-1}^b + \hat{\theta}_T x_{t-1} + \hat{\xi}_T(t-1) + \hat{\alpha}_T + \hat{\varphi}_T \Delta y_{t-1}^b + \hat{\pi}_T \Delta x_t + e_t^b,$$

where y_t^b is the cumulative sum of Δy_t^b . Note that we do not resample Δx_t here. Using the resampled series, we compute the Wald tests and denote them using the superscript b . For example, we let $\mathcal{W}_T^{(1),b}$ denote the bootstrapped $\mathcal{W}_T^{(1)}$.

- S3:** We iterate the second step B times in total and compute the empirical p -value of the test. For example, we let the empirical p -value be $p_T^{(1)} := B^{-1} \sum_{b=1}^B \mathbb{1}\{\mathcal{W}_T^{(1),b} > \mathcal{W}_T^{(1)}\}$ for $\mathcal{W}_T^{(1)}$. If $p_T^{(1)}$ is less than the significance level, we reject the null hypothesis. $\square$

This bootstrap procedure can apply even when the null weak limit of the Wald test follows a mixed chi-squared distribution.

We conduct our simulation by the above residual bootstrap procedure. Assuming that $B = 500$, we

independently iterate the above experiment 5000 times for $T = 100, 200, 300, 400$, and 500 . The simulation results are reported in Tables 2 and 3. We also evaluate the Wald tests by mixed chi-squared distributions and report their empirical rejection rates. The simulation results are summarized as follows:

- (a) Table 2 shows the empirical rejection rates using the null DGP condition allowing $\beta_*^+ = \beta_*^- = 1$ and $\pi_*^+ = \pi_*^- = 1/2$. The left side panel shows the empirical rejection rates obtained by the residual bootstrap method. The rejection rates on the right side are based on the mixed chi-squared distributions. In summary, first, when using the bootstrap method, for each T , the empirical rejection rates are very close to the nominal significance levels. This implies that the bootstrap method effectively controls the test levels. Second, when considering the asymptotic critical values from the mixed chi-squared distribution, we find no significant level distortion for $\mathcal{W}_T^{(2)}$, but both $\mathcal{W}_T^{(1)}$ and $\mathcal{W}_T^{(3)}$ do experience level distortions. For the current DGP, $\mathcal{W}_T^{(1)}$ and $\mathcal{W}_T^{(3)}$ reject $\mathbb{H}_0'$ and $\mathbb{H}_0'''$ more often than the significance levels. This confirms that $\mathcal{W}_T^{(1)}$ and $\mathcal{W}_T^{(3)}$ do not follow mixed chi-squared distributions under the null. However, $\mathcal{W}_T^{(2)}$ does not suffer from level distortions, indicating that $\tilde{\pi}_T^+$ and $\tilde{\pi}_T^-$ follow mixed normal distributions, as Lemma 4 (i) establishes. Finally, $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ suffer no level distortions when using the critical values from the mixed chi-squared distributions.
- (b) Table 3 shows the empirical rejection rates under the alternative DGP. The summary results are as follows. First, with the bootstrap method, the empirical rejection rates of $\mathcal{W}_T^{(1)}$, $\mathcal{W}_T^{(2)}$, and $\mathcal{W}_T^{(3)}$ tend toward 100% as T increases, implying their consistency against $\mathbb{H}_1'$, $\mathcal{H}_1''$, and $\mathbb{H}_1'''$, respectively. Second, when applying the critical values obtained from the mixed chi-squared distribution, the empirical rejection rates of $\mathcal{W}_T^{(1)}$, $\mathcal{W}_T^{(2)}$, and $\mathcal{W}_T^{(3)}$ also converge toward 100% as T increases. However, it is difficult to control their sizes, as shown in Table 2. Third, the standard Wald tests are consistently powerful. Fourth, as in the case of the bootstrap method, $\mathcal{W}_T^{(1)}$ and $\mathcal{W}_T^{(3)}$ are more powerful than $W_T^{(1)}$ and $W_T^{(3)}$, respectively, for small T . In contrast, both $\mathcal{W}_T^{(2)}$ and $W_T^{(2)}$ show similar power patterns. $\square$

Before moving to the empirical section, we describe other simulation experiences on OLS and 2SNARDL. As the two estimators can estimate the same parameter, each has relative advantages over the other. The OLS contains a finite sample bias that is driven by ϱ_{m*} in Lemma 3 and vanishes as T increases, so that the negligible bias affects its finite sample performance. In contrast, 2SNARDL estimates the parameter by removing the bias. On the other hand, 2SNARDL is more involved than OLS. The bias is removed using an asymptotic covariance matrix estimator between Δx_t and u_t , affecting the performance of 2SNARDL. Meanwhile, OLS is straightforward to compute. From this comparison, if T is sufficiently large, OLS can be preferred to 2SNARDL for practical purposes. However, if T is not large 2SNARDL can be preferred over OLS when the bias is successfully removed. Cho et al. (2026) provide simulation results to compare the finite performance of OLS with 2SNARDL.

7 Empirical Application

This section examines the empirical data provided by [Romer and Romer \(2010\)](#) to measure the fiscal policy impact on the GDP of the United States of America. We review the literature and apply the NARDL model to examine the long- and short-run symmetries in the data.

7.1 Literature Review and Empirical Motivation

Estimating the effects of fiscal policy on output is challenging, because many fiscal factors that lead to tax changes are correlated with output, and this lets the OLS estimator suffer from endogenous bias. [Blanchard and Perotti \(2002\)](#) address this by using structural vector autoregression (SVAR). They attempt to solve the bias problem by assuming that policymakers do not respond to shocks contemporaneously, and by using information on the elasticity of revenue to create cyclically adjusted revenues. From this, they estimate that the effect of a tax cut on U.S. GDP is around 1%. However, [Romer and Romer \(2010\)](#) and [Cloyne \(2013\)](#) argue that the structural assumptions used in the SVAR model may be unrealistic in estimating the effect of fiscal policy on output.

[Romer and Romer \(2010\)](#) analyze the effects of tax changes correlated with GDP differently by performing a narrative analysis. They examine the motivations behind each tax change from 1945 to 2007 using sources such as *the Economic Report of the President* and *the Congressional Record*. They categorize legislated tax changes that altered tax liabilities from one quarter to the next into four categories: (i) tax changes to counteract changes in government spending, (ii) tax changes to offset other factors affecting near-term output, (iii) tax changes to address inherited budget deficits, and (iv) tax changes to promote long-term growth. The first two categories, considered countercyclical, are motivated by restoring output growth reduced by other factors, making it difficult to classify them as purely exogenous. In contrast, the last two categories, based on policymakers' perceptions of prudent fiscal policies or focus on increasing long-term growth, may be classified as exogenous. Consequently, they identified 54 exogenous tax changes through narrative analysis during the same period. Using these exogenous fiscal shocks, they develop a time-series model and estimate that GDP would increase by approximately 3% over three years following a tax cut of 1% of GDP. This estimate differs significantly from that of [Blanchard and Perotti \(2002\)](#).

Narrative analysis is widely used to examine the effects of fiscal shocks on GDP. For example, [Cloyne \(2013\)](#) applies this approach to UK legislation and estimates that a 1% tax cut, as a percentage of GDP, increases output by nearly 2.5% over the following three years, similar to findings in the United States of America. [Mertens and Olea \(2018\)](#) use narrative analysis to determine that the short-run tax elasticity of income is approximately 1.2% in the United States of America by analyzing exogenous variations in

the marginal tax rate. Additionally, [Gunter, Riera-Crichton, Vegh, and Vuletin \(2021\)](#) expand the use of narrative analysis to estimate the value-added-tax multipliers for 51 countries, showing that the effect of tax changes on output is highly nonlinear.

Narrative analysis is also used to specialize the time-series model in [Romer and Romer \(2010\)](#) for specific economic models under various environments. For example, [Mertens and Ravn \(2012\)](#) differentiate between surprise and anticipated tax changes to analyze the dynamic effects of tax changes on GDP, finding that anticipated tax cuts lead to a contraction in GDP. [Demirel \(2021\)](#) and [Ghassibe and Zanetti \(2021\)](#) examine the state-dependent effects of exogenous tax changes on GDP, allowing for varying tax multiplier estimations during recessions and expansions. Researchers widely apply narrative analysis to examine the effect of tax changes on GDP in other fields, enabling comparisons with conventional analysis outcomes.

The current study extends its focus beyond short-run relationships to estimate the long-run relationship between GDP and fiscal shocks. Previous studies have primarily concentrated on the short-term effects of fiscal policy. For example, one benchmark model specified by [Romer and Romer \(2010\)](#) is given as

$$\Delta y_t = \gamma_* + \sum_{j=0}^{q-1} \pi_{j*} \Delta \tau_{t-i} + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + e_t \quad (19)$$

to examine how GDP responds to exogenous tax changes, where y_t is the logarithm of real GDP and $\Delta \tau_t$ is the logarithm of an exogenous tax change. Note that all variables in (19) are differenced. Thus, it characterizes the short-run relationship between Δy_t and $\Delta \tau_t$.

Because y_t and τ_t are both observable, we can estimate their long-run relationship by applying cointegration analysis. To this end, we augment the cointegration error on the right side of (19) as follows:

$$\Delta y_t = \gamma_* + \rho_* u_{t-1} + \sum_{j=0}^{q-1} \pi_{j*} \Delta \tau_{t-i} + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + e_t, \quad (20)$$

where $u_t := y_t - \beta_* \tau_t - \zeta_* t - \nu_*$. Note that the long-run relationship between y_t and τ_t can be found by estimating the long-run coefficient β_* , and the short-run relationship can be found by estimating the coefficients of π_{j*} and φ_{j*} . To estimate the unknown parameters, we also convert (20) as follows:

$$\Delta y_t = \alpha_* + \rho_* y_{t-1} + \theta_* \tau_{t-1} + \zeta_* t + \sum_{j=0}^{q-1} \pi_{j*} \Delta \tau_{t-i} + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + e_t. \quad (21)$$

For this estimation, we first examine the partial sum processes of the exogenous tax changes used in our empirical analysis. Figure 2 illustrates the partial sum processes due to the exogenous tax changes.¹

¹Data are obtained from <https://eml.berkeley.edu/~cromer/> (Accessed: Feb. 10, 2023).

The solid and dashed lines represent the partial sum processes of tax changes for deficit reduction (τ_{1t}) and long-run growth (τ_{2t}), respectively, and the dotted line represents the partial sum process of their sum (τ_t).²

The exogenous tax changes exhibit characteristics suitable for NARDL analysis. First, tax changes for budget deficits always result in tax increases, so $\Delta\tau_{1t}$ is always positive. Second, most tax changes aimed at long-run economic growth involve tax decreases; out of 31 legislated tax changes for long-run growth, only six of them result in tax increases. Consequently, overall, the partial sum processes for deficit reduction and long-run growth remain in positive and negative regions, respectively. Although the NARDL model assumptions do not perfectly align with the characteristics of the exogenous fiscal shocks, we use the approximation of $\Delta\tau_t^+ := \max[0, \Delta\tau_t]$ for tax changes due to budget deficit and $\Delta\tau_t^- := \min[0, \Delta\tau_t]$ for tax changes due to long-run growth.

Next, we specify the following NARDL model and estimate the long- and short-run parameters:

$$\Delta y_t = \gamma_* + \rho_* u_{t-1} + \sum_{j=0}^{q-1} \left(\pi_{j*}^+ \Delta\tau_{t-i}^+ + \pi_{j*}^- \Delta\tau_{t-i}^- \right) + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + e_t, \quad (22)$$

where $u_t := y_t - \beta_*^+ \tau_t^+ - \beta_*^- \tau_t^- - \zeta_* t - \nu_*$. This can be rewritten as

$$\Delta y_t = \alpha_* + \rho_* y_{t-1} + \theta_*^+ \tau_{t-1}^+ + \theta_*^- \tau_{t-1}^- + \xi_* t + \sum_{j=0}^{q-1} \left(\pi_{j*}^+ \Delta\tau_{t-i}^+ + \pi_{j*}^- \Delta\tau_{t-i}^- \right) + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + e_t, \quad (23)$$

which we estimate by OLS. If $\beta_*^+ = \beta_*^-$ (or $\theta_*^+ = \theta_*^-$) and $\pi_{j*}^+ = \pi_{j*}^-$, (22) reduces to (20). Economically, this implies that the relationship between tax changes aimed at deficit reduction and long-run growth in real GDP is roughly symmetric in both the long and short run.

7.2 Empirical Results

This section presents the estimation and inference results using the tax change data outlined in the previous section. We have limited the sample period to 1947Q1–2007Q4 by excluding periods with missing observations.

Before presenting the estimation and inference results, we provide the basic statistical characteristics of the data. The logarithm of the GDP quantity index, multiplied by 100, is represented by y_t , and τ_t , τ_{1t} , and

²We obtain the partial sum processes by first converting the nominal tax changes into consistent values over the period 1947Q1 to 2007Q4. To this end, we first discount the nominal values using the price index implied by the nominal GDP and the quantity index for GDP in the data set, and then apply a log transformation. We find that $\Delta\tau_t = \Delta\tau_{1t} + \Delta\tau_{2t}$, $\Delta\tau_{1t} := \text{sgn}(\Delta T_{1t}) \log(|\Delta T_{1t}|/p_t)$, $\Delta\tau_{2t} := \text{sgn}(\Delta T_{2t}) \log(|\Delta T_{2t}|/p_t)$, and $p_t := NY_t/Y_t$, where ΔT_{1t} and ΔT_{2t} represent the nominal tax changes for budget deficit and long-run growth, respectively, and NY_t and Y_t represent the nominal GDP and quantity GDP index, respectively. If $\Delta T_{1t} = 0$ or $\Delta T_{2t} = 0$, we let $\Delta\tau_{1t} = 0$ or $\Delta\tau_{2t} = 0$, respectively. The partial sum processes in Figure 1 represent τ_t , τ_{1t} , and τ_{2t} .

$\tau_{2,t}$ are defined as in footnote 2. The descriptive statistics of Δy_t , $\Delta \tau_{1t}$, $\Delta \tau_{2t}$, and $\Delta \tau_t$ can be found in Table A.9 in the Online Supplement. Furthermore, our unit-root test on y_t , τ_{1t} , τ_{2t} , and τ_t follows the method of Phillips and Perron (1988), including or excluding the time trend. The test results indicate that we cannot reject the unit-root hypothesis for the series.

We report the estimation results in Table 4. The columns marked “Exo” give the parameter estimates obtained by OLS for the NARDL and ARDL models. That is, the unknown parameters in (23) and (21) are estimated by OLS. Orders for the NARDL model are based on the Akaike information criterion (AIC), with $p = 3$ and $q = 1$ for both the NARDL and the ARDL models. Standard errors are listed in parentheses below the parameter estimates. Except for the coefficient of y_{t-1} , we use the asymptotic critical values from the mixed normal distribution. For the t -test on the coefficient of y_{t-1} , we use the asymptotic critical values provided by Banerjee et al. (1998). Furthermore, we test whether all coefficients of y_{t-1} , τ_{t-1}^+ , and τ_{t-1}^- are equal to zero by applying the F -test of Pesaran et al. (2001). Finally, we test the hypotheses of symmetry between long-run parameters, short-run parameters, or both by using the Wald tests given in Section 5. The results are presented in the two bottom panels. We summarize the results in Table 4 as follows:

- (a) The coefficient of y_{t-1} is significant at the 10% and 5% levels for the NARDL and ARDL models, respectively by the t -test. Although we cannot reject the hypothesis of no cointegration by the F -test for the NARDL model, it is significant at the 10% level for the ARDL model. Our analysis using the 2SNARDL model also suggests a cointegrating relationship between the log real GDP and the log of an exogenous log tax shock.
- (b) The NARDL model estimation indicates that an increase in an exogenous tax shock measured by τ_{t-1}^+ reduces the log real GDP. In contrast, a decrease in an exogenous tax shock measured by τ_{t-1}^- increases the log of real GDP. The ARDL model shows the same relationship between the log of an exogenous tax shock and the log real GDP, aligning with standard economic theory.
- (c) The estimated coefficients of τ_{t-1}^+ and τ_{t-1}^- are almost equal in magnitude, suggesting no long-run asymmetry between the log real GDP and the log of an exogenous tax shock. We confirm this using the Wald tests. Both Wald tests $\mathcal{W}_T^{(1)}$ and $W_T^{(1)}$ provide p -values that make it difficult to reject the symmetry hypothesis.
- (d) Short-run symmetry is confirmed by the Wald tests $\mathcal{W}_T^{(2)}$ and $W_T^{(2)}$, because they do not reject the symmetry hypothesis. Moreover, neither the long-run nor the short-run symmetry hypothesis is challenged by the Wald test $\mathcal{W}_T^{(3)}$ or $W_T^{(3)}$. As such, we conclude that the ARDL model is appropriate for studying the relationship between the log real GDP and the log of an exogenous tax shock.
- (e) We present the estimation results for the log of an endogenous tax shock, calculated in the same way as for the log of an exogenous tax shock. Columns labeled “Endo” give the estimation and inference

results obtained using the log of the endogenous tax shock data. Similarly, the columns labeled “Sum” show the estimation and inference results obtained using the logs of both exogenous and endogenous tax shocks. The estimated signs of τ_{t-1}^+ , τ_{t-1}^- , and τ_{t-1} are inconsistent with the parameter values posited by standard economic theory. When using the logs of both exogenous and endogenous tax shocks jointly, there is little evidence of cointegration. These estimation results indicate that only the log of an exogenous tax shock can be used to estimate the coefficients having signs consistent with standard economic theory. $\square$

Next, we estimate the NARDL and ARDL models using the 2SNARDL method proposed by [Cho et al. \(2023a\)](#). For the ARDL model, we apply 2SNARDL estimation by imposing the short- and long-run parameter symmetry conditions. The results are presented in Table 5, which is structured similarly to Table 4. This separate investigation is conducted to corroborate the inference results presented in Table 4. We summarize the results in Table 5.

- (a) The NARDL and ARDL models estimated by 2SNARDL show that the coefficient of u_{t-1} for the exogenous tax shock is statistically significant. The significance levels are 10% and 5% for the NARDL and ARDL models, respectively. Moreover, we apply the unit-root test of [Phillips and Perron \(1988\)](#) to the cointegration residuals obtained from both models. The results reject the unit-root hypothesis. The p -values are 10.19% and 2.58% for the NARDL and ARDL models, respectively. This confirms the cointegrating relationship between the log real GDP and the log of an exogenous tax shock.
- (b) For the log of an exogenous tax shock, the NARDL model indicates that the long-run coefficient of the log tax increase is -0.4329, whereas that of the log tax decrease is -0.3202. These signs are consistent with standard economic theory, and are statistically significant at the 1% and 10% levels, respectively. For the ARDL model, the long-run coefficient of the log tax change is -0.2328. This sign is also consistent with standard economic theory, and the estimated coefficient is statistically significant at the 1% level.
- (c) Our findings for the endogenous and aggregate log tax shocks align with the results presented in Table 4. The coefficients of τ_{t-1}^+ , τ_{t-1}^- , and τ_{t-1} for the log of endogenous tax shocks are statistically significant, but their signs are inconsistent with standard economic theory. However, for the log of the aggregate tax shock, none of these coefficients are statistically significant. Moreover, the coefficient of u_{t-1} is insignificant. We believe these inconsistent results are due to the correlation between an endogenous shock and a structural error. $\square$

The results in Tables 4 and 5 suggest that by using the log of exogenous tax shocks, we can properly identify the relationship between the log real GDP and a fiscal shock. The findings indicate limited statistical support for the asymmetry between tax shocks for deficit reduction and those for long-run growth. Moreover,

the OLS and 2SNARDL estimations produce qualitatively similar results.

In the Online Supplement, we provide additional empirical results. Rather than using the tax change logarithm τ_t , we use $\Delta r_t := (\Delta T_{1t} + \Delta T_{2t})/NY_t$, which represents the tax change-to-nominal GDP ratio, to measure the effect of fiscal policy and to specify the models corresponding to (20), (21), (22), and (23). From this, we obtain empirical results similar to those using log tax data and further show that an increase in an exogenous tax shock measured by r_{t-1} reduces the long-run log real GDP by about 3%. This is close to the result of [Romer and Romer \(2010\)](#), estimating that GDP will increase by approximately 3% over three years following a tax cut of 1% of GDP; see Section [A.5](#).

8 Conclusion

OLS has an asymptotically singular matrix when used with the NARDL model. However, despite the absence of established limit theory for OLS, it remains popular in empirical literature.

This study investigates the large-sample behavior of OLS by addressing the problem of an asymptotically singular matrix. Specifically, we find that OLS is consistent for the unknown NARDL parameters and follows a mixed normal distribution asymptotically under some mild regularity conditions. This implies that the standard principles of the t - and Wald tests apply, despite the asymptotically singular matrix problem. To derive the asymptotic distribution, we first represent the OLS as a transformation of other primitive estimators that are not affected by an asymptotically singular matrix, and derive their joint weak limits. The representation is obtained by elaborating on the FWL projection method in [Park and Phillips \(1988, 1989\)](#), and the weak limits of the primitive estimators allow us to derive the joint weak limit of the OLS. They further enable us to demonstrate that the convergence rate of the OLS is identical to that of the estimator implied by 2SNARDL, despite the asymptotically singular matrix problem.

In addition, we examine the large-sample behavior of the Wald tests for the NARDL hypothesis. Beyond the standard Wald tests defined by the OLS, we develop supplementary Wald tests using primitive estimators. The null limit distributions of the standard Wald tests are mixed chi-squared, whereas those of the supplementary Wald tests differ when testing for long-run asymmetry. By applying the residual bootstrap method, Monte Carlo simulations show that the supplementary Wald tests generally perform better than the standard Wald tests.

OLS and 2SNARDL target the same parameter, and each has relative advantages over the other. The OLS contains a finite sample bias that is asymptotically negligible, affecting its finite sample performance. In contrast, 2SNARDL estimates the parameter by removing the bias. However, 2SNARDL first estimates the asymptotic covariance matrix between the differenced regressor and cointegration error. Meanwhile,

OLS is not involved in such a preliminary procedure. From this, if the sample size is sufficiently large, the OLS can be preferred to the 2SNARDL for practical purposes; if the sample size is not large, 2SNARDL can be preferred over OLS when it successfully removes the bias.

Lastly, we illustrate the proper use of the NARDL model by estimating the long- and short-run relationships between GDP and exogenous fiscal shocks due to deficit reduction and long-run growth. For this we use the empirical data from [Romer and Romer \(2010\)](#). Because all tax changes for deficit reduction are tax increases, and most changes for long-run growth are tax decreases, the NARDL model approximates the relationship between GDP and the exogenous fiscal shocks. We estimate the model and examine whether the relationships between tax increases and decreases are symmetric in both the long and short run. We find no evidence of asymmetric relationships between them.

References

- BANERJEE, A., J. DOLADO, AND R. MESTRE (1998): “Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework,” *Journal of Time Series Analysis*, 19, 267–283. [6](#), [20](#), [31](#)
- BLANCHARD, O. AND R. PEROTTI (2002): “An Empirical Characterization of the Dynamic of Changes in Government Spending and Taxes on Output,” *Quarterly Journal of Economics*, 117, 1329–1368. [28](#)
- BORENSTEIN, S., C. A. CAMERON, AND R. GILBERT (1997): “Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?” *Quarterly Journal of Economics*, 112, 305–339. [1](#)
- CHESNES, M. (2016): “Asymmetric Pass-Through in U.S. Gasoline Prices,” *Energy Journal*, 37, 153–180. [1](#)
- CHO, J. S., M. J. GREENWOOD-NIMMO, AND Y. SHIN (2023a): “Nonlinear Autoregressive Distributed Lag Model Estimation Using Data with Drifts Applied to the Relationship between Income Inequality and Economic Openness,” Discussion paper, School of Economics, Yonsei University. [19](#), [32](#)
- (2023b): “Recent Developments of the Autoregressive Distributed Lag Modelling Framework,” *Journal of Economic Surveys*, 37, 7–32. [4](#)
- (2026): “Two-Step Nonlinear ARDL Estimation: Theory and Application,” Discussion paper, School of Economics, Yonsei University. [1](#), [7](#), [27](#)
- CHO, J. S. AND P. C. B. PHILLIPS (2018): “Sequentially Testing Polynomial Model Hypotheses Using Power Transforms of Regressors,” *Journal of Applied Econometrics*, 33, 141–159. [2](#), [8](#), [13](#), [19](#)

- CHO, J. S. AND H. L. WHITE (2007): “Testing for Regime Switching,” *Econometrica*, 75, 1671–1720. 2, 8, 13, 19
- (2010): “Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models,” *Journal of Econometrics*, 157, 458–480. 2, 8, 13, 19
- CLOYNE, J. (2013): “Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom,” *American Economic Review*, 103, 1507–1528. 28
- DE JONG, R. M. AND J. DAVIDSON (2000): “The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes,” *Econometric Theory*, 16, 621–642.
- DEMIREL, U. D. (2021): “The Short-Term Effects of Tax Changes: The Role of State Dependence,” *Journal of Monetary Economics*, 117, 918–934. 29
- ENGLE, R. F. AND C. W. J. GRANGER (1987): “Co-integration and Error Correction: Representation, Estimation and Testing,” *Econometrica*, 55, 251–276. 1, 16
- GHASSIBE, M. AND F. ZANETTI (2021): “State Dependence of Fiscal Multipliers: The Source of Fluctuations Matters,” *Journal of Monetary Economics*, 132, 1–23. 29
- GUNTER, S., D. RIERA-CRICHTON, C. A. VEGH, AND G. VULETIN (2021): “Non-Linear Effects of Tax Changes on Output: The Role of the Initial Level of Taxation,” *Journal of International Economics*, 131, 103450. 29
- MERTENS, K. AND J. L. M. OLEA (2018): “Marginal Tax Rates and Income: New Time Series Evidence,” *Quarterly Journal of Economics*, 133, 1803–1884. 28
- MERTENS, K. AND M. O. RAVN (2012): “Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks,” *American Economic Journal: Economic Policy*, 4, 145–181. 29
- PARK, J. Y. AND P. C. B. PHILLIPS (1988): “Statistical Inference in Regression with Integrated Processes: Part 1,” *Econometric Theory*, 4, 468–498. 2, 10, 33
- (1989): “Statistical Inference in Regression with Integrated Processes: Part 2,” *Econometric Theory*, 5, 95–131. 2, 10, 33
- PESARAN, M. H. AND Y. SHIN (1998): “An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis,” in *Econometrics and Economic Theory: The Ragnar Frisch Centennial Symposium*,

- ed. by S. Strom, Cambridge: Cambridge University Press, Econometric Society Monographs, 371–413. 3, 7
- PESARAN, M. H., Y. SHIN, AND R. J. SMITH (2001): “Bounds Testing Approaches to the Analysis of Level Relationships,” *Journal of Applied Econometrics*, 16, 289–326. 6, 7, 20, 31, 40
- PHILLIPS, P. C. B. (1991): “Optimal Inference in Cointegrated Systems,” *Econometrica*, 59, 283–306. 6
- (1995): “Fully Modified Least Squares and Vector Autoregression,” *Econometrica*, 63, 1023–1078. 2, 8
- PHILLIPS, P. C. B. AND B. E. HANSEN (1990): “Statistical Inference in Instrumental Variable Regression with I(1) Processes,” *Review of Economic Studies*, 57, 99–125. 6, 16, 20
- PHILLIPS, P. C. B. AND P. PERRON (1988): “Testing for a Unit Root in Time Series Regression,” *Biometrika*, 75, 335–346. 31, 32
- ROMER, C. D. AND D. H. ROMER (2010): “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks,” *American Economic Review*, 100, 763–801. 2, 28, 29, 33, 34, 40, 41
- SHIN, Y., B. YU, AND M. J. GREENWOOD-NIMMO (2014): “Modelling Asymmetric Cointegration and Dynamic Multipliers in a Nonlinear ARDL Framework,” in *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*, ed. by W. Horrace and R. Sickles, New York (NY): Springer Science & Business Media, 281–314. 1, 4, 5, 7
- TERÄSVIRTA, T. (1994): “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models,” *Journal of the American Statistical Association*, 89, 208–218. 2, 8, 13, 19
- WHITE, H. L. (2001): *Asymptotic Theory for Econometricians*, San Diego (CA): Academic Press. 6

Wald Test	Null Hypothesis	Symmetry	Null Limit Distribution
$W_T^{(1)} := \hat{\alpha}'_T \hat{\mathbf{R}}'_1 (\hat{\mathbf{W}}_T^{(1)})^{-1} \hat{\mathbf{R}}_1 \hat{\alpha}_T$	$\mathcal{H}_0' : \theta_*^+ = \theta_*^-$	Long-run	Mixed Chi-squared
$W_T^{(2)} := \hat{\alpha}'_T \hat{\mathbf{R}}'_2 (\hat{\mathbf{W}}_T^{(2)})^{-1} \hat{\mathbf{R}}_2 \hat{\alpha}_T$	$\mathcal{H}_0'' : \pi_*^+ = \pi_*^-$	Short-run	Mixed Chi-squared
$W_T^{(3)} := \hat{\alpha}'_T \hat{\mathbf{R}}'_3 (\hat{\mathbf{W}}_T^{(3)})^{-1} \hat{\mathbf{R}}_3 \hat{\alpha}_T$	$\mathcal{H}_0''' : \beta_*^+ = \beta_*^-$ and $\pi_*^+ = \pi_*^-$	Long- and Short-run	Mixed Chi-squared
$\mathcal{W}_T^{(1)} := \ddot{v}'_T \ddot{\mathbf{D}}_1 \ddot{\mathbf{R}}'_1 (\ddot{\mathbf{W}}_T^{(1)})^{-1} \ddot{\mathbf{R}}_1 \ddot{\mathbf{D}}_1 \ddot{v}_T$	$\mathbb{H}_0' : \beta_*^+ = \beta_*^-$	Long-run	Nonstandard
$\mathcal{W}_T^{(2)} := \ddot{\tau}'_T \ddot{\mathbf{D}} \ddot{\mathbf{R}}'_2 (\ddot{\mathbf{W}}_T^{(2)})^{-1} \ddot{\mathbf{R}}_2 \ddot{\mathbf{D}} \ddot{\tau}_T$	$\mathcal{H}_0'' : \pi_*^+ = \pi_*^-$	Short-run	Mixed Chi-squared
$\mathcal{W}_T^{(3)} := \ddot{\delta}'_T \ddot{\mathbf{D}} \ddot{\mathbf{R}}'_3 (\ddot{\mathbf{W}}_T^{(3)})^{-1} \ddot{\mathbf{R}}_3 \ddot{\mathbf{D}} \ddot{\delta}_T$	$\mathbb{H}_0''' : \beta_*^+ = \beta_*^-$ and $\pi_*^+ = \pi_*^-$	Long- and Short-run	Nonstandard

Table 1: SUMMARY OF THE WALD TESTS. This table summarizes the hypotheses and the null limit distributions of the Wald tests defined in Section 5. $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ are defined by the OLS estimator $\hat{\alpha}_T$, whereas $\mathcal{W}_T^{(1)}$, $\mathcal{W}_T^{(2)}$, and $\mathcal{W}_T^{(3)}$ are defined by the primitive estimators, viz., $\ddot{v}_T$, $\ddot{\tau}_T$, and $\ddot{\delta}_T$, respectively. When the null limit distribution of the Wald test is mixed chi-squared, the asymptotic critical values of the test can be selected from the chi-square distribution. However, for the Wald test whose null limit distribution is nonstandard, the residual bootstrap can be applied to the test. Section 6 illustrates the application of the residual bootstrap to the Wald test.

Wald Test	Method	Bootstrap Method					Mixed Chi-squared Distribution				
	$\alpha \setminus T$	100	200	300	400	500	100	200	300	400	500
$\mathcal{W}_T^{(1)}$	10%	11.34	10.78	10.18	10.26	10.32	39.68	40.84	41.20	42.78	41.72
	5%	5.98	5.46	4.90	5.46	5.32	32.04	32.80	33.02	34.62	33.42
	1%	1.00	1.02	0.94	1.38	0.90	19.34	19.62	19.82	20.96	21.40
$\mathcal{W}_T^{(2)}$	10%	9.58	8.96	9.70	9.56	9.96	12.10	10.20	10.52	10.20	10.10
	5%	4.46	4.28	5.14	4.90	4.80	6.30	4.94	5.78	5.00	5.38
	1%	1.04	0.86	1.02	0.84	1.02	1.56	1.06	1.28	0.90	1.06
$\mathcal{W}_T^{(3)}$	10%	11.46	10.50	10.20	9.98	10.62	36.38	36.46	37.28	38.14	37.12
	5%	5.54	5.18	4.94	5.28	5.20	28.10	27.56	28.44	29.44	28.86
	1%	1.02	0.84	0.80	1.28	0.96	16.62	16.18	16.12	17.58	17.50
$W_T^{(1)}$	10%	10.24	10.12	10.62	9.34	9.90	14.44	12.78	12.10	10.90	10.96
	5%	5.38	5.52	5.14	4.34	4.86	8.56	7.22	6.46	5.26	5.56
	1%	1.24	1.14	1.08	0.92	1.08	2.84	1.92	1.32	1.14	1.34
$W_T^{(2)}$	10%	9.82	10.90	9.38	10.32	10.44	11.98	11.94	10.24	10.86	10.82
	5%	4.90	5.56	4.56	5.42	5.68	6.52	6.52	5.20	5.70	6.00
	1%	1.04	0.96	1.00	1.36	1.12	1.74	1.10	1.18	1.38	1.46
$W_T^{(3)}$	10%	9.96	11.32	10.06	9.90	10.84	14.98	13.62	11.72	10.98	11.66
	5%	5.06	5.64	4.82	4.96	5.86	8.42	7.54	6.00	5.74	6.60
	1%	1.14	1.04	1.16	1.10	1.14	2.62	1.82	1.50	1.50	1.42

Table 2: EMPIRICAL REJECTION RATES OF THE WALD TESTS (IN PERCENT). This table shows the empirical rejection rates of the Wald statistics testing $\mathbb{H}'_0 : \beta_* = 0$, $\mathcal{H}''_0 : \pi_*^+ = \pi_*^-$, and $\mathbb{H}'''_0 : \beta_* = 0$ and $\pi_*^+ = \pi_*^-$. The total number of repetitions is 5000, and the bootstrap iteration is 500. DGP: $\Delta y_t = \rho_* u_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t$, $u_t = y_t - \beta_*^+ x_t^+ - \beta_*^- x_t^-$, $\Delta x_t = 1/4 + \kappa_* \Delta x_{t-1} + v_t$, and $(e_t, v_t)' \sim \text{IID } N(\mathbf{0}_2, \mathbf{I}_2)$ with $(\kappa_*, \rho_*, \pi_*^+, \pi_*^-, \beta_*^+, \beta_*^-) = (1/2, -1/2, 1/2, 1/2, 1, 1)$. Here, $\mathcal{W}_T^{(1)}$, $\mathcal{W}_T^{(2)}$, and $\mathcal{W}_T^{(3)}$ denote the Wald tests in Section 5, and $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ are the standard Wald tests.

Wald Test	Method	Bootstrap Method					Mixed Chi-squared Distribution				
	$\alpha \setminus T$	100	200	300	400	500	100	200	300	400	500
$\mathcal{W}_T^{(1)}$	10%	52.00	91.58	99.30	99.92	100.0	73.96	97.34	99.92	99.98	100.0
	5%	39.20	86.62	98.54	99.88	100.0	67.62	96.32	99.82	99.98	100.0
	1%	17.32	70.56	94.76	99.44	99.94	55.08	93.26	99.48	99.96	100.0
$\mathcal{W}_T^{(2)}$	10%	16.48	23.52	30.24	37.72	44.20	19.12	25.62	31.06	38.96	45.46
	5%	9.16	14.64	19.44	26.76	32.58	11.80	16.78	21.04	27.98	33.92
	1%	2.32	4.30	6.52	10.92	14.18	3.94	5.82	7.46	11.64	15.32
$\mathcal{W}_T^{(3)}$	10%	50.70	90.78	99.18	99.92	100.0	69.84	96.36	99.80	99.98	100.0
	5%	36.98	85.84	98.32	99.90	99.98	63.90	95.14	99.66	99.98	100.0
	1%	16.44	69.22	94.32	99.36	99.92	51.42	91.68	99.20	99.94	100.0
$W_T^{(1)}$	10%	47.92	91.16	99.16	99.92	100.0	58.56	93.16	99.24	99.94	100.0
	5%	35.04	85.24	98.22	99.90	100.0	47.72	88.78	98.74	99.92	100.0
	1%	15.22	67.52	94.24	99.12	99.94	28.80	76.20	95.82	99.52	99.98
$W_T^{(2)}$	10%	15.42	22.08	29.92	37.12	45.18	18.28	24.42	31.34	38.48	46.34
	5%	8.74	14.10	19.78	26.18	31.78	11.28	15.62	21.00	27.80	33.48
	1%	2.10	4.76	6.88	10.38	13.88	3.30	5.98	7.80	11.42	14.92
$W_T^{(3)}$	10%	41.94	87.02	98.54	99.92	100.0	52.28	89.68	98.86	99.92	100.0
	5%	30.68	79.86	97.32	99.76	100.0	41.42	84.08	97.84	99.82	100.0
	1%	12.40	61.20	91.60	98.86	99.88	23.66	69.66	93.92	99.16	99.96

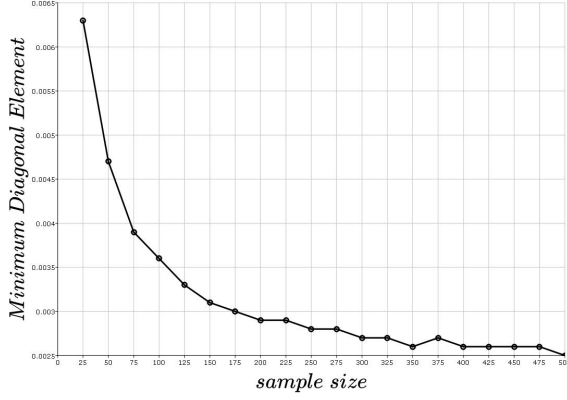
Table 3: EMPIRICAL REJECTION RATES OF THE WALD TESTS (IN PERCENT). This table shows the empirical rejection rates of the Wald statistics testing $\mathbb{H}'_0 : \beta_* = 0$, $\mathcal{H}''_0 : \pi_*^+ = \pi_*^-$, and $\mathbb{H}'''_0 : \beta_* = 0$ and $\pi_*^+ = \pi_*^-$. The total number of repetitions is 5000, and the bootstrap iteration is 500. DGP: $\Delta y_t = \rho_* u_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t$, $u_t = y_t - \beta_*^+ x_t^+ - \beta_*^- x_t^-$, $\Delta x_t = 1/4 + \kappa_* \Delta x_{t-1} + v_t$, and $(e_t, v_t)' \sim \text{IID } N(\mathbf{0}_2, \mathbf{I}_2)$ with $(\kappa_*, \rho_*, \pi_*^+, \pi_*^-, \beta_*^+, \beta_*^-) = (1/2, -1/2, 1/8, -1/8, 1/4, -1/4)$. Here, $\mathcal{W}_T^{(1)}$, $\mathcal{W}_T^{(2)}$, and $\mathcal{W}_T^{(3)}$ denote the Wald tests in Section 5, and $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ are the standard Wald tests.

NARDL Model				ARDL Model			
Variables \ Tax	Exo.	Endo.	Sum.	Variables \ Tax	Exo.	Endo.	Sum.
y_{t-1}	-0.0683*	-0.0817	-0.0497	y_{t-1}	-0.070**	-0.0755	-0.0404
	(0.0191)	(0.0223)	(0.0158)		(0.0185)	(0.0220)	(0.0148)
τ_{t-1}^+	-0.0123 [†]	0.0158**	0.0016	τ_{t-1}	-0.0139**	0.0142*	-0.0053
	(0.0078)	(0.0078)	(0.0093)		(0.0060)	(0.0077)	(0.0074)
τ_{t-1}^-	-0.0123 [†]	0.0149	-0.0093				
	(0.0083)	(0.0176)	(0.0090)				
Trend	0.0549***	0.0631***	0.0324**	Trend	0.0570***	0.0584***	0.0333***
	(0.0177)	(0.0176)	(0.0127)		(0.0151)	(0.0169)	(0.0120)
Constant	0.7588***	0.6974***	0.8357***	Constant	0.7826***	0.7542***	0.8543***
	(0.1874)	(0.1644)	(0.1744)		(0.1539)	(0.1604)	(0.1561)
Δy_{t-1}	0.3129***	0.3118***	0.3135***	Δy_{t-1}	0.3091***	0.3035***	0.3031***
	(0.0633)	(0.0658)	(0.0646)		(0.0630)	(0.0657)	(0.0648)
Δy_{t-2}	0.1265*	0.1446**	0.1190*	Δy_{t-2}	0.1304**	0.1331**	0.1091*
	(0.0646)	(0.0658)	(0.0645)		(0.0643)	(0.0653)	(0.0647)
$\Delta \tau_t^+$	-0.0029	0.0890*	0.03984 [†]	$\Delta \tau_t$	-0.0444 [†]	0.0434 [†]	-0.0122
	(0.0435)	(0.0465)	(0.0341)		(0.0272)	(0.0362)	(0.0226)
$\Delta \tau_t^-$	-0.0780**	-0.0302	-0.0683*				
	(0.0391)	(0.0598)	(0.0347)				
AIC	-6.5459	-6.5372	-6.5332	AIC	-6.5561	-6.5434	-6.5284
BIC	-6.4225	-6.4099	-6.5332	BIC	-6.4602	-6.4475	-6.4378
t -test	-3.5680*	-3.6519*	-3.1308 [†]	t -test	-3.8220**	-3.4319*	-2.7229
F -test	4.5387	4.9164	3.8509	F -test	7.3222*	6.5583*	4.7046
$\mathcal{W}_T^{(1)}$					9.8862	0.0048	9.8910
					(0.4344)	(0.9696)	(0.5000)
$\mathcal{W}_T^{(2)}$					1.4822	3.8085	5.2908
					(0.7472)	(0.1902)	(0.6567)
$\mathcal{W}_T^{(3)}$					30.9528	1.4114	32.3642
					(0.2737)	(0.2659)	(0.2718)
$W_T^{(1)}$					0.0000	0.0039	1.0285
					(0.9982)	(0.9604)	(0.4485)
$W_T^{(2)}$					1.5426	2.5180	4.4950
					(0.2265)	(0.1259)	(0.0398)
$W_T^{(3)}$					1.5527	2.5326	5.2070
					(0.5474)	(0.3706)	(0.1525)

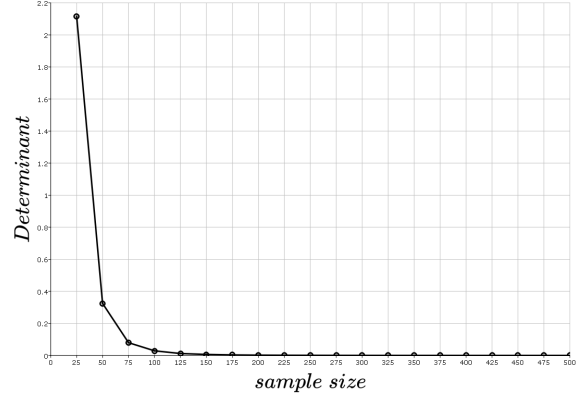
Table 4: OLS ESTIMATION OF THE NARDL AND ARDL MODELS. This table presents the OLS estimation using quarterly data from [Romer and Romer \(2010\)](#). The left and right panels display estimated parameters for (23) and (21), respectively. Figures in parentheses indicate standard errors of the OLS estimates. At the bottom of the top panels, AIC, BIC, t -test, and [Pesaran et al.'s \(2001\)](#) F -test are reported. [†], *, **, and *** indicate significance at 25%, 10%, 5%, and 1% levels, respectively. Wald tests in the last two bottom panels show the Wald tests in Section 5 and the standard Wald tests. Figures in parentheses below the Wald tests show p -values. They are obtained from 100000 bootstrap iterations.

	NARDL Model				ARDL Model			
	Variables \ Tax	Exo.	Endo.	Sum.	Variables \ Tax	Exo.	Endo.	Sum.
Long-Run	Constant	4.2360 [†] (2.5919)	3.7413*** (1.2787)	4.9730** (2.3861)	Constant	6.0785*** (1.7968)	4.8528*** (1.3852)	5.8407** (2.8526)
	τ_{t-1}^+	-0.4329*** (0.1318)	0.2715*** (0.0546)	0.2759 [†] (0.1723)	τ_{t-1}	-0.2328*** (0.0709)	0.2523*** (0.0591)	0.1564 (0.1716)
	τ_{t-1}^-	-0.3202* (0.1722)	0.3276* (0.1757)	0.2598 [†] (0.1826)				
	Trend	0.8340*** (0.0184)	0.8461*** (0.0090)	0.8436*** (0.0169)	Trend	0.8261*** (0.0127)	0.8367*** (0.0098)	0.8287*** (0.0202)
Short-Run	u_{t-1}	-0.0683* (0.0191)	-0.0817* (0.0223)	-0.0497 [†] (0.0158)	u_{t-1}	-0.0708** (0.0185)	-0.0755* (0.0220)	-0.0404 (0.0148)
	Constant	0.6734*** (0.1424)	0.6918*** (0.1458)	0.5683*** (0.1472)	Constant	0.5752*** (0.1391)	0.6590*** (0.1392)	0.6295*** (0.1404)
	Δy_{t-1}	0.3129*** (0.0633)	0.3118*** (0.0658)	0.3135*** (0.0646)	Δy_{t-1}	0.3091*** (0.0630)	0.3035*** (0.0657)	0.3031*** (0.0648)
	Δy_{t-2}	0.1265* (0.0646)	0.1446** (0.0658)	0.1190* (0.0645)	Δy_{t-2}	0.1304** (0.0643)	0.1331** (0.0653)	0.1091* (0.0647)
	$\Delta \tau_t^+$	-0.0029 (0.0435)	0.0890* (0.0465)	0.0398 [†] (0.0341)	$\Delta \tau_t$	-0.0444 [†] (0.0272)	0.0434 [†] (0.0362)	-0.0122 (0.0226)
	$\Delta \tau_t^-$	-0.0780** (0.0391)	-0.0302 (0.0598)	-0.0683* (0.0347)				

Table 5: 2SNARDL ESTIMATION OF THE NARDL AND ARDL MODELS. This table presents the 2SNARDL estimation using the quarterly data from [Romer and Romer \(2010\)](#). The left and right panels display estimated parameters for (22) and (20), respectively. [†], *, **, and *** imply that the tests are significant at 25%, 10%, 5%, and 1%, respectively.



(a) Average of Traces



(b) Average of Determinants $\times 10^{12}$

Figure 1: ASYMPTOTIC SINGULARITY OF THE MATRIX $\mathbf{D}^{-1}(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t') \mathbf{D}^{-1}$. Figure 1 (a) shows the scatter diagram between the sample size and the averaged minimum diagonal element of $\mathbf{D}^{-1}(\sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t') \mathbf{D}^{-1}$ obtained by imposing the DGP condition given in Table 3. Figure 1 (b) shows the scatter diagram between the sample size and the averaged determinant $\times 10^{12}$ of the same matrix. For the scatter diagrams, 10000 independent experiments are repeated to compute the averages from the 10000 independent matrices.

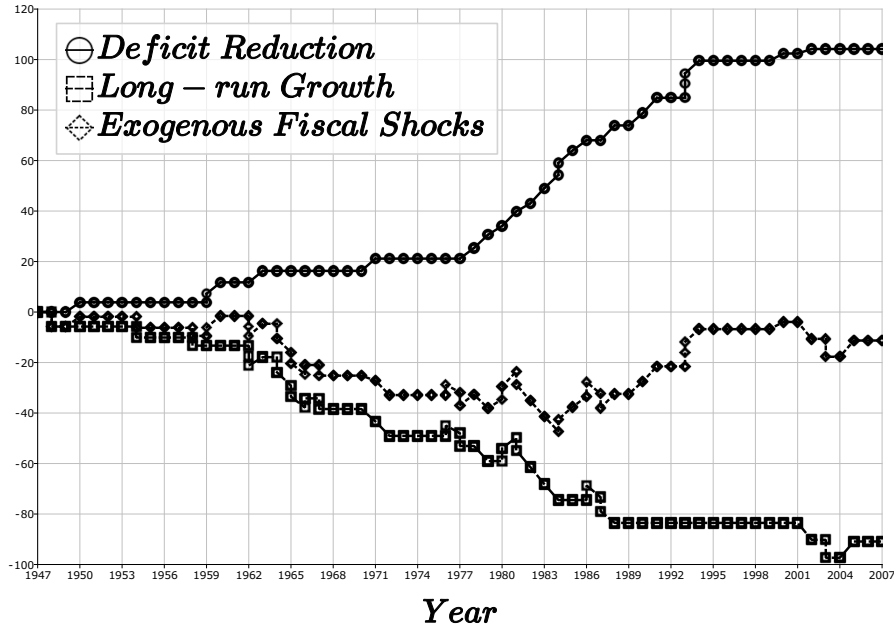


Figure 2: PARTIAL SUM PROCESS FORMED BY EXOGENOUS FISCAL SHOCKS. The solid, dashed, and dotted lines represent deficit reduction, long-run growth, and exogenous fiscal shocks, respectively.