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

October 2025

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 so that parameters targeted by the OLS and the two-step NARDL estimators can be estimated at the same convergence rate. 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 co-editor, Degui Li and three anonymous referees provided helpful comments for which the author is most grateful. The author is also indebted to Chun-Kyu Cho for his support and discussions with Matthew Greenwood-Nimmo, Peter Phillips, and Yongcheol Shin. Cho acknowledges the research grant provided by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF2020S1A5A2A0104-0235).

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 (2024) 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 the 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 data analysis asymptotically.

This study addresses a gap in the literature concerning NARDL model estimation, although it is not the first to address the asymptotically singular matrix problem associated with NARDL. Cho et al. (2024) 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. They show that the long-run parameter estimator in 2SNARDL is super-consistent. We provide theoretical grounds to compare these two estimators by deriving the limit distribution of OLS and showing that the parameters targeted by both estimators are estimated at the same convergence rate despite the asymptotic singular matrix problem. As a result, OLS estimates the targeted parameter with an asymptotically negligible bias, whereas 2SNARDL removes it by first estimating the asymptotic covariance matrix between the differenced regressor and cointegration error. From this feature, if the sample size is sufficiently large, OLS can more straightforwardly estimate the parameter. If the sample size is small but the covariance matrix estimator successfully removes the bias, 2SNARDL can be useful for practical purposes.

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 directions (see Phillips, 1995, and his subsequent works for early demonstrations). Similarly, the issue of singular matrices emerges in the context of maximum-likelihood (ML) and nonlinear least squares (NLS) estimations. In these cases, researchers find the limit distributions using higher-order approximations of the nonlinear model or likelihood. This slows the convergence rate of the estimator compared to 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. First, 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. Using this process, we isolate the singular matrix problem from the transformation form and show that the parameter targeted by OLS is estimated by the estimator implied by 2SNARDL at the same convergence rate as OLS.

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 estimator in 2SNARDL.

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. Moreover, our findings show that a 1% exogenous GDP tax decrease increases the log real GDP by about 3% in the long run, which aligns closely with the estimates by Romer and Romer (2010) and confirms their findings. Here, we illustrate our methodology using both OLS and 2SNARDL.

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 other primitive estimators. Section 4 discusses the limit distribution of an OLS estimator, which varies depending on parameter values and specific conditions for limit distributions. Section 5 examines the large-sample properties of the standard Wald test for the NARDL hypothesis, and discusses another Wald test for supplementary purposes. Section 6 presents Monte Carlo simulations for the Wald tests, and 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.

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 $\int \mathscr{B}$ or $\int d\mathscr{B}$ means $\int_0^1 \mathscr{B}(u)du$ or $\int_0^1 d\mathscr{B}(u)$, respectively, where $\mathscr{B}(\cdot)$ is a Brownian motion.

2 Motivation and the NARDL Model in the Literature

This section briefly summarizes NARDL and motivates the current study by relating OLS to the asymptotically singular matrix problem.

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

$$y_t = \alpha_{0*} + \xi_* t + \sum_{j=1}^{p} \phi_{j*} y_{t-j} + \sum_{j=0}^{q} (\boldsymbol{\theta}_{j*}^{+\prime} \boldsymbol{x}_{t-j}^{+} + \boldsymbol{\theta}_{j*}^{-\prime} \boldsymbol{x}_{t-j}^{-}) + e_t, \tag{1}$$

where $\boldsymbol{x}_t \in \mathbb{R}^k$ $(k \in \mathbb{N})$,

$$\boldsymbol{x}_t^+ := \sum_{j=1}^{t} \Delta \boldsymbol{x}_j^+, \quad \boldsymbol{x}_t^- := \sum_{j=1}^{t} \Delta \boldsymbol{x}_j^-, \quad \Delta \boldsymbol{x}_t^+ := \max[\boldsymbol{0}, \Delta \boldsymbol{x}_t], \quad \Delta \boldsymbol{x}_t^- := \min[\boldsymbol{0}, \Delta \boldsymbol{x}_t],$$

$\{e_t, \mathscr{F}_t\}$ is a martingale difference array (MDA), and $\mathscr{F}_t$ is the smallest σ-algebra driven by $\{y_{t-1}, \boldsymbol{x}_t^+, \boldsymbol{x}_t^-, y_{t-2}, \boldsymbol{x}_{t-1}^+, \boldsymbol{x}_{t-1}^-, \ldots\}$ such that $\Delta \boldsymbol{x}_t$ is a stationary process, and Δ denotes the differencing operator, so that $\Delta \boldsymbol{x}_t := \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$. 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 $\boldsymbol{x}_t$ is possibly different, the NARDL process generalizes the ARDL process. If $\boldsymbol{\theta}_{j*}^+ = \boldsymbol{\theta}_{j*}^-$ for each j, the NARDL process reduces to the ARDL process. In addition, the NARDL process 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. If $\xi_* = 0$, the NARDL process in (1) reduces to their definition.

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

$$\Delta y_t = \rho_* y_{t-1} + \boldsymbol{\theta}_*^{+\prime} \boldsymbol{x}_{t-1}^+ + \boldsymbol{\theta}_*^{-\prime} \boldsymbol{x}_{t-1}^- + \xi_*(t-1) + \alpha_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^{+\prime} \Delta \boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_{j*}^{-\prime} \Delta \boldsymbol{x}_{t-j}^- \right) + e_t,$$

where ρ_*, $\boldsymbol{\theta}_*^+$, $\boldsymbol{\theta}_*^-$, φ_{j*} $(j = 1, 2, \ldots, p-1)$, $\boldsymbol{\pi}_{j*}^+$, and $\boldsymbol{\pi}_{j*}^-$ $(j = 0, 1, \ldots, q-1)$ are defined by the parameters in (1), as given in Cho et al. (2023b), and $\alpha_* := \alpha_{0*} + \xi_*$. At the end of Section A.1 in the Online Supplement, we represent the parameters in the above equations using the parameters in (1). For notational simplicity, we further rewrite this as

$$\Delta y_t = \rho_* y_{t-1} + \boldsymbol{\theta}_*' \ddot{\boldsymbol{x}}_{t-1} + \xi_*(t-1) + \alpha_* + \boldsymbol{\varphi}_* \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\pi}_*' \Delta \widetilde{\boldsymbol{x}}_t + e_t \tag{2}$$

by letting

$$\begin{aligned}
\boldsymbol{\theta}_* &:= [\boldsymbol{\theta}_*^{+\prime}, \boldsymbol{\theta}_*^{-\prime}]', \quad \ddot{\boldsymbol{x}}_{t-1} := [\boldsymbol{x}_{t-1}^{+\prime}, \boldsymbol{x}_{t-1}^{-\prime}]', \\
\boldsymbol{\varphi}_* &:= [\varphi_{1*}, \varphi_{2*}, \ldots, \varphi_{p-1*}]', \quad \Delta \boldsymbol{y}_{t-1} := [\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1}]', \\
\boldsymbol{\pi}_* &:= [\boldsymbol{\pi}_*^{+\prime}, \quad \boldsymbol{\pi}_*^{-\prime}]', \quad \boldsymbol{\pi}_*^+ := [\boldsymbol{\pi}_{0*}^{+\prime}, \boldsymbol{\pi}_{1*}^{+\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{+\prime}]', \quad \boldsymbol{\pi}_*^- := [\boldsymbol{\pi}_{0*}^{-\prime}, \boldsymbol{\pi}_{1*}^{-\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{-\prime}]', \\
\Delta \widetilde{\boldsymbol{x}}_t &:= [\Delta \widetilde{\boldsymbol{x}}_t^{+\prime}, \Delta \widetilde{\boldsymbol{x}}_t^{-\prime}]', \quad \Delta \widetilde{\boldsymbol{x}}_t^+ := [\Delta \boldsymbol{x}_t^{+\prime}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{+\prime}]', \quad \text{and} \quad \Delta \widetilde{\boldsymbol{x}}_t^- := [\Delta \boldsymbol{x}_t^{-\prime}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{-\prime}]'.
\end{aligned}$$

Here, y_{t-1}, $\ddot{\boldsymbol{x}}_{t-1}$ and $(t-1)$ on the right side of (2) can be used to describe the long-run relationship. If y_{t-1} is further cointegrated with $\ddot{\boldsymbol{x}}_{t-1}$, (2) can be rewritten into the following error-correction form:

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \boldsymbol{\varphi}_*' \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\pi}_*' \Delta \widetilde{\boldsymbol{x}}_t + e_t, \tag{3}$$

such that the cointegration error is defined as $u_{t-1} := y_{t-1} - \boldsymbol{\beta}_*' \ddot{\boldsymbol{x}}_{t-1} - \zeta_*(t-1) - \nu_*$, with

$$\boldsymbol{\beta}_* := [\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime}]', \quad \boldsymbol{\beta}_*^+ := -(\boldsymbol{\theta}_*^+/\rho_*), \quad \boldsymbol{\beta}_*^- := -(\boldsymbol{\theta}_*^-/\rho_*), \quad \text{and} \quad \zeta_* := -(\xi_*/\rho_*).$$

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 $\Delta \boldsymbol{y}_{t-1}$ and $\Delta \widetilde{\boldsymbol{x}}_t$. Therefore, it follows that $\gamma_* := \alpha_* + \rho_* \nu_*$.

The NARDL process captures an asymmetric cointegrating relationship between nonstationary processes. If we let

$$\boldsymbol{\mu}_* := [\boldsymbol{\mu}_*^{+\prime}, \boldsymbol{\mu}_*^{-\prime}]', \quad \boldsymbol{\mu}_*^+ := \mathbb{E}[\Delta \boldsymbol{x}_t^+] \quad \text{and} \quad \boldsymbol{\mu}_*^- := \mathbb{E}[\Delta \boldsymbol{x}_t^-], \quad \text{then} \quad \boldsymbol{\mu}_*^+ + \boldsymbol{\mu}_*^- \equiv \mathbb{E}[\Delta \boldsymbol{x}_t],$$

by construction, because $\Delta \boldsymbol{x}_t \equiv \Delta \boldsymbol{x}_t^+ + \Delta \boldsymbol{x}_t^-$. Therefore, if we further let $\boldsymbol{s}_t^+ := \Delta \boldsymbol{x}_t^+ - \boldsymbol{\mu}_*^+$ and $\boldsymbol{s}_t^- := \Delta \boldsymbol{x}_t^- - \boldsymbol{\mu}_*^-$, it follows that

$$\ddot{\boldsymbol{x}}_t = \boldsymbol{\mu}_* t + \mathbf{m}_t \quad \text{by letting} \quad \mathbf{m}_t := [\mathbf{m}_t^{+\prime}, \mathbf{m}_t^{-\prime}]', \quad \mathbf{m}_t^+ := \sum_{j=1}^t \boldsymbol{s}_j^+ \quad \text{and} \quad \mathbf{m}_t^- := \sum_{j=1}^t \boldsymbol{s}_j^-. \tag{4}$$

From (4), $\ddot{\boldsymbol{x}}_t$ is clearly a unit-root process with nonzero time trends. Moreover, Δy_t is not necessarily distributed around zero, even when $\boldsymbol{x}_t$ is a unit-root process without a time trend. From (3) and by noting that $\mathbb{E}[u_t] = 0$ and $\mathbb{E}[e_t] = 0$, we have

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

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

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

This fact implies that y_t is a unit-root process with a deterministic time trend. This cointegrating relationship is more general than that assumed by a linear cointegration. By imposing $\boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$, the error-correction form in (3) reduces to the linear cointegrating relationship between y_t and $\boldsymbol{x}_t$.

The time trend in the dependent variable y_t has further implications in the model specification. As (5)

implies, if we suppose that $\boldsymbol{x}_t$ is an integrated series without a time trend, the NADL model can capture a cointegrating relationship between the variable without time trend $\boldsymbol{x}_t$ and the variable with time trend y_t. This means that a time trend has to be assumed by the NARDL structure for the dependent variable as given in (5). This also addresses another possibility that the NARDL model may not fully explain the time trend in y_t, and the specification in (2) is provided for such a case. By including $(t-1)$ as an additional explanatory variable, we estimate its coefficient. If the estimated coefficient of $(t-1)$ is significant, it signals that the NARDL structure cannot be regarded as the single source of the trend in y_t.

Despite the popularity in the empirical literature of estimating the unknown parameters in (2) by OLS, there is no theoretical foundation for doing so. This is mainly attributed to the asymptotically singular matrix problem. For this examination, we first let

$$\begin{aligned}\boldsymbol{z}_t &:= [\ \boldsymbol{z}_{1t}' \ \vdots\ \boldsymbol{z}_{2t}'\ ]' \\ &:= [\ y_{t-1} \quad \boldsymbol{r}_{t-1}' \ \vdots\ \boldsymbol{z}_{2t}'\ ]' \\ &:= [\ y_{t-1} \quad \ddot{\boldsymbol{x}}_{t-1}' \quad (t-1) \quad 1 \ \vdots\ \Delta \boldsymbol{y}_{t-1}' \quad \Delta \widetilde{\boldsymbol{x}}_t'\ ]',\end{aligned}$$

so that $\boldsymbol{r}_{t-1} = [\ddot{\boldsymbol{x}}_{t-1}', (t-1), 1]'$. Here, $\boldsymbol{z}_t \in \mathbb{R}^{2+p+2k(1+q)}$ is partitioned into two variables such that $\boldsymbol{z}_{1t} \in \mathbb{R}^{3+2k}$ and $\boldsymbol{z}_{2t} \in \mathbb{R}^{p+2kq-1}$ collect the variables in the long- and short-run equations, respectively. Next, we let

$$\boldsymbol{\alpha}_* := [\ \rho_* \quad \boldsymbol{\theta}_*' \quad \xi_* \quad \alpha_* \ \vdots\ \boldsymbol{\varphi}_*' \quad \boldsymbol{\pi}_*'\ ]'.$$

From this, the OLS estimator is written as

$$\widehat{\boldsymbol{\alpha}}_T := [\ \widehat{\rho}_T \quad \widehat{\boldsymbol{\theta}}_T' \quad \widehat{\xi}_T \quad \widehat{\alpha}_T \ \vdots\ \widehat{\boldsymbol{\varphi}}_T' \quad \widehat{\boldsymbol{\pi}}_T'\ ]' := \left(\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t'\right)^{-1} \sum_{t=1}^{T} \boldsymbol{z}_t \Delta y_t.$$

For later purposes, we partition $\widehat{\boldsymbol{\theta}}_T$ and $\widehat{\boldsymbol{\pi}}_T$ into $[\widehat{\boldsymbol{\theta}}_T^{+\prime}, \widehat{\boldsymbol{\theta}}_T^{-\prime}]'$ and $[\widehat{\boldsymbol{\pi}}_T^{+\prime}, \widehat{\boldsymbol{\pi}}_T^{-\prime}]'$, respectively. 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)* $\{(\Delta y_t, \Delta \boldsymbol{x}_t', u_t)' \in \mathbb{R}^{k+2} : t = \ldots, -1, 0, 1, \ldots\}$ *is a strictly stationary mixing process with* ϕ *of size* $-r/(2(r-1))$ *or* α *of size* $-r/(r-2)$ *and* $r > 2$*;*

(ii) $\mathbb{E}[|\Delta x_{ti}|^r] < \infty$ $(i = 1, 2, \ldots, k)$, $\mathbb{E}[|u_t|^r] < \infty$, $\mathbb{E}[|e_t|^2] < \infty$, *and* $\delta_* \neq 0$, *where* x_{ti} *is the* i*-th row element of* $\boldsymbol{x}_t$*;*

(iii) $\boldsymbol{\Sigma}_* := \lim_{T\to\infty} var[T^{-1/2}\sum_{t=1}^{T}\boldsymbol{w}_t]$ *is positive definite, where*

$$\boldsymbol{w}_t := [\ \boldsymbol{w}_{1t}' \ \vdots\ \boldsymbol{w}_{2t}'\ ] := [\ \boldsymbol{s}_{t-1}' \quad u_{t-1} \ \vdots\ e_t \quad e_t u_{t-1} \quad e_t \boldsymbol{z}_{2t}'\ ]'$$

and $\boldsymbol{s}_{t-1} := [\boldsymbol{s}_{t-1}^{+\prime}, \boldsymbol{s}_{t-1}^{-\prime}]'$*;*

(iv) for some $\boldsymbol{\alpha}_*$ *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) $\{e_t, \mathscr{F}_t\}$ *is an MDA.* □

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 $\ddot{\boldsymbol{x}}_t$. If no cointegrating relationship exists, so that if $\boldsymbol{\beta}_* = \mathbf{0}$ and $\zeta_* = 0$, then from (3), it follows that

$$y_t = \alpha_* + \sum_{j=1}^{p}\phi_{j*}y_{t-j} + \boldsymbol{\pi}_*'\Delta\widetilde{\boldsymbol{x}}_t + e_t.$$

Assumption 1 (*iv*) applied to this equation implies that y_t is a stationary process, which contradicts the assumption that y_t is a unit-root process. Therefore, it is necessary to assume $\rho_* = 0$ for $(\boldsymbol{\beta}_*', \zeta_*)' = \mathbf{0}$. In this study, we assume the cointegrating relationship established by (3) and examine the limit behavior of OLS.

(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 $\boldsymbol{\beta}_* = \mathbf{0}$ and $\zeta_* = 0$.

(e) Assumption 1 (*v*) is a standard condition for the error term in the ARDL and NARDL processes (e.g., Pesaran and Shin, 1998; Pesaran et al., 2001; Shin et al., 2014; Cho et al., 2024). □

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

Lemma 1. *Under Assumption 1,*

(i)

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

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

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

(iii)

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

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

(iv)

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

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

Although the OLS estimator $\widehat{\boldsymbol{\alpha}}_T:=(\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t')^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t\Delta y_t$ is popular for empirical analyses, Lemma 1 now clearly shows that its theoretical foundation cannot be established without tackling the asymptotic singular matrix problem. Note that we obtain $\mathbf{M}$ by employing different rates of convergence for each element, and it is singular because its first $(1+2k)$ columns are proportional to the $(2+2k)$-th column.

The singular matrix can be verified by conducting simple Monte Carlo simulations. By denoting the model and data-generating process (DGP) conditions for Tables 3 and 5 given below as independent and serially correlated cases, respectively, we compute the trace and determinant of $\mathbf{D}^{-1}(\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t')\mathbf{D}^{-1}$ for T =50, 100, 200, 300, 400, and 500. After iterating 10000 independent experiments, the scatter diagram between T and the average of the 10000 traces is drawn for both cases in Figure 1 (a). From this, the trace converges to a finite number for both cases as T increases, indicating that $\mathbf{D}^{-1}(\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t')\mathbf{D}^{-1}=O_{\mathbb{P}}(1)$. Meanwhile, the scatter diagram between T and the average of the 10000 determinants $\times 10^{12}$ in Figure 1 (b) shows that the determinant converges to zero. This implies that $\mathbf{D}^{-1}(\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t')\mathbf{D}^{-1}$ is asymptotically singular as Lemma 1 indicates.

We provide the proof of Lemma 1 in the Online Supplement by extending lemma 1 of Cho et al. (2024). Their lemma demonstrates 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 asymptotically singular matrix problem.

The singular matrix problem arises from the deterministic time trend in the integrated series. More specifically, (5) shows that $\mathbb{E}[\Delta y_{t-1}] \neq 0$, so that y_{t-1} is an integrated series with a time trend. Similarly, $\mathbb{E}[\Delta \widetilde{\boldsymbol{x}}_{t-1}^{+}]$ and $\mathbb{E}[\Delta \widetilde{\boldsymbol{x}}_{t-1}^{-}]$ differ from zero by construction, so that both $\boldsymbol{x}_{t-1}^{+}$ and $\boldsymbol{x}_{t-1}^{-}$ are also integrated series with time trends, meaning that all of y_{t-1}, $\ddot{\boldsymbol{x}}_{t-1}$, and $(t-1)$ on the right side of (2) are driven by the time trend $(t-1)$, and estimating their coefficients suffers from the singular matrix problem. This aspect is evident from Lemma 1, showing that the first $(2+2k)$ columns of $\mathbf{M}$ are proportional to each other, and they are involved with y_{t-1}, $\ddot{\boldsymbol{x}}_{t-1}$, and $(t-1)$. Cho et al. (2024) note this aspect and estimate the unknown long-run parameters by first removing the trends from y_{t-1} and $\ddot{\boldsymbol{x}}_{t-1}$, proposing the 2SNARDL estimation.

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 the nonlinear model or likelihood, which results in a slower convergence rate compared to the standard case (e.g., Teräsvirta, 1994; Cho and White, 2007, 2010; Cho and Phillips, 2018).

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 $(\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t')^{-1}$ is determined by k, p, and q, and they can be given arbitrarily, the task becomes challenging.

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, and we express its limit distribution through those of the primitive estimators. This method shows that the type of transformation is the source of the singularity problem.

We achieve the alternative representation in three steps. First, we estimate the long-run parameters using OLS. Second, we provide a short-run parameter estimator, allowing the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ to be expressed as a bilinear transformation of the long- and short-run primitive estimators. Finally, we express the long- and short-run primitive estimators as other bilinear transformations of further primitive estimators that do not suffer from a singular matrix problem, enabling $\widehat{\boldsymbol{\alpha}}_T$ to be rewritten as a transformation of these primitive

estimators.

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

$$\begin{aligned} y_{t-1} &= \boldsymbol{r}_{t-1}'\boldsymbol{\upsilon}_* + u_{t-1} \\ &= \boldsymbol{\beta}_*'\ddot{\boldsymbol{x}}_{t-1} + \zeta_*(t-1) + \nu_* + u_{t-1} \end{aligned} \tag{6}$$

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

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

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

Second Step We let $\boldsymbol{\omega}_* := [\rho_*, \boldsymbol{\eta}_*', \psi_*, \gamma_*, \boldsymbol{\varphi}_*', \boldsymbol{\pi}_*']'$, such that $\boldsymbol{\eta}_* := \boldsymbol{\theta}_* + \rho_*\boldsymbol{\beta}_*$ and $\psi_* := \xi_* + \rho_*\zeta_*$, and $\dot{\boldsymbol{z}}_t := [u_{t-1}, \boldsymbol{r}_{t-1}', \boldsymbol{z}_{2t}']'$ to specify the error-correction model

$$\begin{aligned} \Delta y_t &= \dot{\boldsymbol{z}}_t'\boldsymbol{\omega}_* + e_t \\ &= \rho_* u_{t-1} + \boldsymbol{\eta}_*'\ddot{\boldsymbol{x}}_{t-1} + \psi_*(t-1) + \gamma_* + \boldsymbol{\varphi}_*\Delta\boldsymbol{y}_{t-1} + \boldsymbol{\pi}_*'\Delta\widetilde{\boldsymbol{x}}_t + e_t \end{aligned} \tag{7}$$

by combining (2) and (6). Then, we estimate the parameters in (7) by

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

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

The long- and short-run parameter estimators have a regular relationship in terms of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$. From the definitions of the long-run and trend coefficients, $\boldsymbol{\beta}_*$ and ζ_*, we must have $\boldsymbol{\eta}_* = \mathbf{0}$ and $\psi_* = 0$. Even so, we estimate them through the short-run parameter estimator $\widetilde{\boldsymbol{\omega}}_T$ by including $\boldsymbol{r}_{t-1}$ as an auxiliary regressor. The purpose of this inclusion is given in the following proposition:

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

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

That is, by including the regressor in the first step, $\boldsymbol{r}_{t-1}$, as an auxiliary regressor, we can represent the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ as a bilinear transformation between the long- and short-run primitive estimators $\widetilde{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$. From Proposition 1, it follows that

$$\widehat{\boldsymbol{\alpha}}_T = [\widetilde{\rho}_T,\ -\widetilde{\rho}_T\widetilde{\boldsymbol{\beta}}_T' + \widetilde{\boldsymbol{\eta}}_T',\ -\widetilde{\rho}_T\widetilde{\zeta}_T + \widetilde{\psi}_T,\ -\widetilde{\rho}_T\widetilde{\nu}_T + \widetilde{\gamma}_T,\ \widetilde{\boldsymbol{\varphi}}_T',\ \widetilde{\boldsymbol{\pi}}_T']'.$$

Note that if both $\widetilde{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$ are consistent for $\boldsymbol{\upsilon}_*$ and $\boldsymbol{\omega}_*$, respectively,

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

by noting that $\boldsymbol{\eta}_* = \mathbf{0}$ and $\psi_* = 0$. Proposition 1 follows by applying Lemma A.1 in the Online Supplement. By letting y_t, $\boldsymbol{x}_t$, and $\boldsymbol{z}_t$ of Lemma A.1 be Δy_t, y_{t-1}, and $[\boldsymbol{r}_{t-1}', \boldsymbol{z}_{2t}']'$ of (7), respectively, Proposition 1 follows. Here, we can let $\widehat{\boldsymbol{\upsilon}}_t$ in Lemma A.1 be $\widetilde{u}_{t-1}$.

Third Step Although Proposition 1 represents $\widehat{\boldsymbol{\alpha}}_T$ as a bilinear transformation between the long- and short-run primitive estimators, both $\widetilde{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$ suffer from an asymptotically singular matrix problem; see Lemma A.3 in the Online Supplement. Again, the asymptotic singular matrix problem arises because the regressors for $\widetilde{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$ are driven by the time trend $(t-1)$. That is, $\ddot{\boldsymbol{x}}_{t-1}$ and $(t-1)$ in $\boldsymbol{r}_{t-1}$ and $\widetilde{\boldsymbol{z}}_t$ are asymptotically correlated, producing singular matrix problems as before. Therefore, we again represent both primitive estimators using other primitive estimators that do not suffer from an asymptotically singular matrix problem.

We first represent the long-run parameter estimator $\widetilde{\boldsymbol{\upsilon}}_T$ using other estimators. Note that both (4) and (6) imply that

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

This representation is obtained by collecting the deterministic time trend as a single regressor $(t-1)$ and the remaining term $\mathbf{m}_{t-1}$ of $\ddot{\boldsymbol{x}}_{t-1}$ as another regressor, so that the coefficient of $(t-1)$, ϑ_*, is now defined by those of $\ddot{\boldsymbol{x}}_{t-1}$ and $(t-1)$. If $\mathbf{m}_{t-1}$ is observed, we can estimate the coefficients in (8) by regressing y_{t-1} against $[\mathbf{m}_{t-1}', (t-1), 1]'$. However, $\mathbf{m}_{t-1}$ is not. We instead predict $\mathbf{m}_{t-1}$ first, so that we can use its predictor as a regressor. For this, we estimate the unknown parameters $\boldsymbol{\mu}_*$ in (4) by

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

We use $\ddot{\mathbf{m}}_{t-1}$ as a regressor to estimate the coefficients in (8). That is, we let

$$\ddot{\boldsymbol{r}}_{t-1} := [\ddot{\mathbf{m}}'_{t-1}, (t-1), 1]'$$

and regress y_{t-1} against $\ddot{\boldsymbol{r}}_{t-1}$, to obtain

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

Note that this primitive estimator $\ddot{\boldsymbol{\upsilon}}_T$ is obtained by first removing the time trend from $\ddot{\boldsymbol{x}}_{t-1}$ to predict $\mathbf{m}_{t-1}$ so that the asymptotic singular matrix problem can be avoided and we can estimate the long-run parameter consistently. For later purposes, we partition $\ddot{\boldsymbol{\beta}}_T$ such that $\ddot{\boldsymbol{\beta}}_T \equiv [\ddot{\boldsymbol{\beta}}^{+\prime}_T, \ddot{\boldsymbol{\beta}}^{-\prime}_T]'$ and let $\ddot{u}_{t-1} := y_{t-1} - \ddot{\boldsymbol{r}}'_{t-1}\ddot{\boldsymbol{\upsilon}}_T$.

The interrelationship between the long-run parameter and primitive estimators, $\widetilde{\boldsymbol{\upsilon}}_T$ and $\ddot{\boldsymbol{\upsilon}}_T$, is stated in the following proposition:

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

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

That is, the long-run parameter estimator $\widetilde{\boldsymbol{\upsilon}}_T$ is represented as a bilinear transformation between the primitive estimators, $\ddot{\boldsymbol{\mu}}_T$ and $\ddot{\boldsymbol{\upsilon}}_T$. This representation is particularly useful, because neither $\ddot{\boldsymbol{\mu}}_T$ nor $\ddot{\boldsymbol{\upsilon}}_T$ suffers from an asymptotically singular matrix problem, as Lemma 2 verifies below. Further, $\widetilde{u}_{t-1} = \ddot{u}_{t-1}$ using the definition of $\ddot{\mathbf{m}}_{t-1} := \ddot{\boldsymbol{x}}_{t-1} - (t-1)\ddot{\boldsymbol{\mu}}_T$, and if both $\ddot{\boldsymbol{\mu}}_T$ and $\ddot{\boldsymbol{\upsilon}}_T$ are consistent for $\boldsymbol{\mu}_*$ and $\bar{\boldsymbol{\upsilon}}_*$, respectively,

$$\widetilde{\boldsymbol{\upsilon}}_T = [\ddot{\boldsymbol{\beta}}'_T, \ddot{\vartheta}_T - \ddot{\boldsymbol{\mu}}'_T\ddot{\boldsymbol{\beta}}_T, \ddot{\nu}_T]' \xrightarrow{\mathbb{P}} [\boldsymbol{\beta}'_*, \vartheta_* - \boldsymbol{\mu}'_*\boldsymbol{\beta}_*, \nu_*]' = [\boldsymbol{\beta}'_*, \zeta_*, \nu_*]' =: \boldsymbol{\upsilon}_*$$

by noting that $\vartheta_* := \boldsymbol{\beta}'_*\boldsymbol{\mu}_* + \zeta_*$, leading to the consistency of $\widetilde{\boldsymbol{\upsilon}}_T$ for $\boldsymbol{\upsilon}_*$. Proposition 2 is established by applying Lemma A.2 in the Online Supplement. If we let $\ddot{\boldsymbol{x}}'_{t-1}$, $(t-1)$, and 1 be $\boldsymbol{x}_t$, $\boldsymbol{z}_t$, and $\boldsymbol{w}_t$, respectively, in Lemma A.2, then Proposition 2 follows.

Next, we represent the short-run parameter estimator $\widetilde{\boldsymbol{\omega}}_T$ 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 parameter estimator. Both $\ddot{\boldsymbol{x}}_{t-1}$ and $(t-1)$ on the right side of (7) are driven by the deterministic trend. We first collect the deterministic time trends in $\ddot{\boldsymbol{x}}_{t-1}$ and $(t-1)$ as a single regressor

and reparameterize the right side. Specifically, we combine (4) and (7), obtaining

$$\Delta y_t = \rho_* u_{t-1} + \boldsymbol{\eta}_*' \mathbf{m}_{t-1} + \varsigma_*(t-1) + \gamma_* + \boldsymbol{\varphi}_*' \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\pi}_*' \Delta \widetilde{\boldsymbol{x}}_t + e_t, \quad \text{where} \quad \varsigma_* := \boldsymbol{\mu}_*' \boldsymbol{\eta}_* + \psi_*. \tag{9}$$

Here, the coefficient of $(t-1)$ is obtained as ς_* by collecting the coefficients of the time trends in $\ddot{\boldsymbol{x}}_{t-1}$ and $(t-1)$, but the coefficient of $\mathbf{m}_{t-1}$ keeps the same as that of $\ddot{\boldsymbol{x}}_{t-1}$ on the right side of (7) from the fact that $\mathbf{m}_{t-1}$ is the remainder of $\ddot{\boldsymbol{x}}_{t-1}$ obtained by removing the trend. Furthermore, ς_* must be 0 from the fact that $\boldsymbol{\eta}_* = \mathbf{0}$ and $\psi_* = 0$, but $\mathbf{m}_{t-1}$ and $(t-1)$ are included as auxiliary regressors to apply Lemma A.2 in the Online Supplement again. That is, by regressing Δy_t against

$$\ddot{\boldsymbol{z}}_t := [\ \ddot{u}_{t-1} \ \vdots\ \ddot{\boldsymbol{r}}_{t-1}' \ \vdots\ \boldsymbol{z}_{2t}' \ ]',$$

we estimate the unknown parameters in (9) by OLS. For this, u_{t-1} and $\mathbf{m}_{t-1}$ in (9) are replaced with $\ddot{u}_{t-1}$ and $\ddot{\mathbf{m}}_{t-1}$, respectively as both u_{t-1} and $\mathbf{m}_{t-1}$ are unobservable. The parameter estimator is denoted as follows:

$$\ddot{\boldsymbol{\tau}}_T := [\ \ddot{\rho}_T \ \vdots\ \ddot{\boldsymbol{\tau}}_{1T}' \ \vdots\ \ddot{\boldsymbol{\tau}}_{2T}' \ ]' := [\ \ddot{\rho}_T \ \vdots\ \ddot{\boldsymbol{\eta}}_T' \quad \ddot{\varsigma}_T \quad \ddot{\gamma}_T \ \vdots\ \ddot{\boldsymbol{\tau}}_{2T}' \ ]' := \left(\sum_{t=1}^{T} \ddot{\boldsymbol{z}}_t \ddot{\boldsymbol{z}}_t' \right)^{-1} \sum_{t=1}^{T} \ddot{\boldsymbol{z}}_t \Delta y_t, \quad \text{estimating}$$

$$\boldsymbol{\tau}_* := [\ \rho_* \ \vdots\ \boldsymbol{\tau}_{1*}' \ \vdots\ \boldsymbol{\tau}_{2*}' \ ]' := [\ \rho_* \ \vdots\ \boldsymbol{\eta}_*' \quad \varsigma_* \quad \gamma_* \ \vdots\ \boldsymbol{\varphi}_*' \quad \boldsymbol{\pi}_*' \ ]'.$$

Here, for later purposes, we partitioned the parameter $\boldsymbol{\tau}_*$ and its estimator $\ddot{\boldsymbol{\tau}}_T$ into the smaller sets, and we let $\ddot{\boldsymbol{\eta}}_T \equiv [\ddot{\boldsymbol{\eta}}_T^{+\prime}, \ddot{\boldsymbol{\eta}}_T^{-\prime}]'$ and $\ddot{\boldsymbol{\pi}}_T \equiv [\ddot{\boldsymbol{\pi}}_T^{+\prime}, \ddot{\boldsymbol{\pi}}_T^{-\prime}]'$, such that $\ddot{\boldsymbol{\pi}}_T$ is the vector formed by the last $2kq$ elements of $\ddot{\boldsymbol{\tau}}_{2T}$. 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 parameter and primitive estimators, $\widetilde{\boldsymbol{\omega}}_T$ and $\ddot{\boldsymbol{\tau}}_T$, is stated in the following proposition:

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

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

That is, by including $\ddot{\mathbf{m}}_{t-1}$ and $(t-1)$ as auxiliary regressors for $\ddot{\boldsymbol{\tau}}_T$, we represent the short-run primitive estimator $\widetilde{\boldsymbol{\omega}}_T$ as a bilinear transformation between $\ddot{\boldsymbol{\mu}}_T$ and $\ddot{\boldsymbol{\tau}}_T$. From Proposition 3, it follows that

$$\widetilde{\boldsymbol{\omega}}_T = [\ddot{\rho}_T,\ \ddot{\boldsymbol{\eta}}_T',\ -\ddot{\boldsymbol{\mu}}_T' \ddot{\boldsymbol{\eta}}_T + \ddot{\varsigma}_T,\ \ddot{\gamma}_T,\ \ddot{\boldsymbol{\tau}}_{2T}']'.$$

Therefore, $\mathbf{Q}_T\ddot{\boldsymbol{\tau}}_T$ and $\ddot{\boldsymbol{\tau}}_T$ are almost identical. The only difference between them is in the coefficient of $(t-1)$. The estimated coefficient of $(t-1)$ in $\mathbf{Q}_T\ddot{\boldsymbol{\tau}}_T$ is equal to $-\ddot{\boldsymbol{\mu}}_T'\ddot{\boldsymbol{\eta}}_T + \ddot{\varsigma}_T$, whereas the corresponding estimator in $\ddot{\boldsymbol{\tau}}_T$ is $\ddot{\varsigma}_T$. If both $\ddot{\boldsymbol{\mu}}_T$ and $\ddot{\boldsymbol{\tau}}_T$ are consistent for $\boldsymbol{\mu}_*$ and $\boldsymbol{\tau}_*$, respectively, it follows that

$$\widetilde{\boldsymbol{\omega}}_T \xrightarrow{\mathbb{P}} [\rho_*,\, \boldsymbol{\eta}_*',\, -\boldsymbol{\mu}_*'\boldsymbol{\eta}_* + \varsigma_*,\, \gamma_*,\, \boldsymbol{\tau}_{2*}']' = [\rho_*,\, \boldsymbol{\eta}_*',\, \psi_*,\, \gamma_*,\, \boldsymbol{\tau}_{2*}']' = \boldsymbol{\omega}_*$$

by noting that $\varsigma_* := \boldsymbol{\mu}_*'\boldsymbol{\eta}_* + \psi_*$, and the last equality holds by the definition of $\boldsymbol{\omega}_*$. Furthermore, none of $\ddot{\boldsymbol{\mu}}_T$ and $\ddot{\boldsymbol{\tau}}_T$ suffers from an asymptotically singular matrix problem, as Lemma 2 shows below, meaning that $\widetilde{\boldsymbol{\omega}}_T$ is represented using further primitive estimators without an asymptotically singular matrix problem. Proposition 3 is again established by Lemma A.2 in the Online Supplement. By letting $\ddot{\boldsymbol{x}}_{t-1}$, $(t-1)$, and $[\ddot{u}_{t-1}, 1, \boldsymbol{z}_{2t}']'$ be $\boldsymbol{x}_t$, $\boldsymbol{z}_t$, and $\boldsymbol{w}_t$, respectively, of Lemma A.2, Proposition 3 follows.

Using these alternative forms for the long- and short-run parameter estimators, $\widetilde{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$, we now represent the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ as a transform of the further primitive estimators: $\ddot{\boldsymbol{\mu}}_T$, $\ddot{\boldsymbol{\upsilon}}_T$, and $\ddot{\boldsymbol{\tau}}_T$.

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

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

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

Proposition 4 follows simply by combining Propositions 1 and 3, in which we demonstrated that $\widehat{\boldsymbol{\alpha}}_T = \mathbf{R}_T\widetilde{\boldsymbol{\omega}}_T$ and $\widetilde{\boldsymbol{\omega}}_T = \mathbf{Q}_T\ddot{\boldsymbol{\tau}}_T$, respectively. The individual elements of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ can be rewritten as follows:

$$\widehat{\boldsymbol{\alpha}}_T = \left[\ddot{\rho}_T, -\ddot{\rho}_T\ddot{\boldsymbol{\beta}}_T' + \ddot{\boldsymbol{\eta}}_T', -\ddot{\rho}_T(\ddot{\vartheta}_T - \ddot{\boldsymbol{\mu}}_T'\ddot{\boldsymbol{\beta}}_T) - \ddot{\boldsymbol{\mu}}_T'\ddot{\boldsymbol{\eta}}_T + \ddot{\varsigma}_T, -\ddot{\rho}_T\ddot{\nu}_T + \ddot{\gamma}_T, \ddot{\boldsymbol{\tau}}_{2T}'\right]' \tag{10}$$

by using Proposition 2: $[\widetilde{\boldsymbol{\beta}}_T', \widetilde{\zeta}_T, \widetilde{\nu}_T]' = [\ddot{\boldsymbol{\beta}}_T', \ddot{\vartheta}_T - \ddot{\boldsymbol{\mu}}_T'\ddot{\boldsymbol{\beta}}_T, \ddot{\nu}_T]'$. Note that the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ is now represented using the primitive estimators $\ddot{\boldsymbol{\mu}}_T$, $\ddot{\boldsymbol{\upsilon}}_T$, and $\ddot{\boldsymbol{\tau}}_T$.

Remarks. **(a)** If $\ddot{\boldsymbol{\mu}}_T$, $\ddot{\boldsymbol{\upsilon}}_T$ and $\ddot{\boldsymbol{\tau}}_T$ are consistent for $\boldsymbol{\mu}_*$, $\boldsymbol{\upsilon}_*$ and $\boldsymbol{\tau}_*$, respectively, then

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_T &\xrightarrow{\mathbb{P}} \left[\rho_*, -\rho_*\boldsymbol{\beta}_*' + \boldsymbol{\eta}_*', -\rho_*\zeta_* - \boldsymbol{\mu}_*'\boldsymbol{\eta}_* + \varsigma_*, -\rho_*\nu_* + \gamma_*, \boldsymbol{\tau}_{2*}'\right]' \\ &= \left[\rho_*, -\rho_*\boldsymbol{\beta}_*', -\rho_*\zeta_*, -\rho_*\nu_* + \gamma_*, \boldsymbol{\tau}_{2*}'\right]', \end{aligned} \tag{11}$$

because $\zeta_* = \vartheta_* - \boldsymbol{\mu}_*'\boldsymbol{\beta}_*$, $\boldsymbol{\eta}_* = \mathbf{0}$ and $\varsigma_* = 0$. In addition, the final limit is identical to $\boldsymbol{\alpha}_* := [\rho_*, \boldsymbol{\theta}_*', \xi_*, \alpha_*, \boldsymbol{\varphi}_*', \boldsymbol{\pi}_*']'$ by the definitions of $\boldsymbol{\beta}_*$, ζ_*, γ_*, and $\boldsymbol{\tau}_{2*}$. From this, the OLS estimator is consistent for its target parameter $\boldsymbol{\alpha}_*$. This fact implies that the consistence of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ can be revealed through the consistence of $\ddot{\boldsymbol{\mu}}_T$, $\ddot{\boldsymbol{\upsilon}}_T$ and $\ddot{\boldsymbol{\tau}}_T$.

(b) By Proposition 4, the weak limit of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ can also be represented using the weak limits of the primitive estimators: $\ddot{\boldsymbol{\mu}}_T$, $\ddot{\boldsymbol{\upsilon}}_T$, and $\ddot{\boldsymbol{\tau}}_T$; see Theorem 1.

(c) The convergence rate of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ is slower than $\mathbf{D}$ in Lemma 1. For example, for the long-run OLS estimator $\widehat{\boldsymbol{\theta}}_T$, we have $\widehat{\boldsymbol{\theta}}_T = -\ddot{\rho}_T\ddot{\boldsymbol{\beta}}_T + \ddot{\boldsymbol{\eta}}_T$ from (10), and the convergence rates of $\ddot{\rho}_T$ and $(\ddot{\boldsymbol{\beta}}_T, \ddot{\boldsymbol{\eta}}_T)$ are $T^{1/2}$ and T, respectively, by Lemmas 3 and 4 given below. From this, the convergence rate of $\widehat{\boldsymbol{\theta}}_T$ is determined as $T^{1/2}$, which is slower than $T^{3/2}$ given in $\mathbf{D}$. Similar arguments apply to the other elements of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$, making its convergence rate slower than $\mathbf{D}$. 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). □

We now show that the primitive estimators $\ddot{\boldsymbol{\tau}}_T$ and $\ddot{\boldsymbol{\upsilon}}_T$ do not suffer from an asymptotically singular matrix problem. First, we let

$$\boldsymbol{\mathscr{B}}(\cdot) := [\boldsymbol{\mathscr{B}}_m(\cdot)', \mathscr{B}_u(\cdot), \mathscr{B}_e(\cdot), \mathscr{B}_{ue}(\cdot), \boldsymbol{\mathscr{B}}_{\boldsymbol{z}e}(\cdot)']' := \boldsymbol{\Sigma}_*^{1/2}\boldsymbol{\mathscr{W}}(\cdot),$$

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

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

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

The following lemma shows that neither $\ddot{\boldsymbol{\tau}}_T$ nor $\ddot{\boldsymbol{\upsilon}}_T$ suffers from an asymptotically singular matrix problem. As such, the long- and short-run parameter estimators, $\widetilde{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$, become bilinear transformations of the primitive estimators that do not suffer from this problem.

Lemma 2. *Given Assumption 1,*

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

(ii)

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

where $\bar{\boldsymbol{\mathscr{B}}}_m(\cdot):=\boldsymbol{\mathscr{B}}_m(\cdot)-3(\cdot)\int r\boldsymbol{\mathscr{B}}_m$*;*

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

(iv)

$$\ddot{\mathbf{D}}_2^{-1}\left(\sum_{t=1}^{T}\boldsymbol{z}_{2t}\ddot{\boldsymbol{r}}_{t-1}'\right)\ddot{\mathbf{D}}_1^{-1}\Rightarrow\boldsymbol{\mathscr{M}}_{21}:=\begin{bmatrix}\delta_*\boldsymbol{\iota}_{p-1}\int\bar{\boldsymbol{\mathscr{B}}}_m' & \frac{1}{2}\delta_*\boldsymbol{\iota}_{p-1} & \delta_*\boldsymbol{\iota}_{p-1}\\ \boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_*\int\bar{\boldsymbol{\mathscr{B}}}_m' & \frac{1}{2}\boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_* & \boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_*\end{bmatrix};$$

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

(vi)

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

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

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

4 Limit Distribution of the OLS Estimator

In this section, we derive the limit distribution of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ using the weak limits of the primitive estimators. In addition, we establish the consistency of the long-run parameter and primitive estimators $(\ddot{\boldsymbol{\upsilon}}_T,\ddot{\boldsymbol{\tau}}_T)$ for their target parameters $(\bar{\boldsymbol{\upsilon}}_*,\boldsymbol{\tau}_*)$, from which the consistency of $\widehat{\boldsymbol{\alpha}}_T$ follows.

We proceed in three steps. First, we derive the weak limit of the long-run parameter estimator $\ddot{\boldsymbol{\upsilon}}_T$ and show its consistency. Second, we conduct the same procedure for the primitive estimator $\ddot{\boldsymbol{\tau}}_T$. Finally, we derive the weak limit of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ using the first two results.

First Step We derive the weak limit of the long-run parameter estimators $\ddot{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\upsilon}}_T$. For this, note that $\widetilde{\boldsymbol{\upsilon}}_T=\mathbf{P}_T\ddot{\boldsymbol{\upsilon}}_T$, from Proposition 2. The weak limit of the long-run parameter estimator $\widetilde{\boldsymbol{\upsilon}}_T$ is then determined

by each element on the right side. If we let $\mathbf{P}$ be the limit of $\mathbf{P}_T$, that is,

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

its consistency follows from the consistency of $\ddot{\boldsymbol{\mu}}_T$ for $\boldsymbol{\mu}_*$. Next, we obtain the limit distribution of the primitive estimator $\ddot{\boldsymbol{\upsilon}}_T$ by noting that $\mathbf{m}_{t-1} = \ddot{\mathbf{m}}_{t-1} + (\ddot{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1)$. If we rewrite (8) as

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

with $\vartheta_{T*} := \boldsymbol{\beta}_*'(\ddot{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) + \vartheta_*$, it follows that

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

Using this arrangement, we provide the weak limit of the primitive estimator $\ddot{\boldsymbol{\upsilon}}_T$ in the following lemma:

Lemma 3. *Let* $\boldsymbol{\varrho}_{m*} := \lim_{T\to\infty} T^{-1}\sum_{t=1}^T \sum_{\tau=1}^{t-1} \mathbb{E}[\boldsymbol{s}_\tau u_t]$. *Given Assumption 1,*

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

$$\boldsymbol{\mathscr{S}} := \begin{bmatrix} \boldsymbol{\mathscr{S}}_1 \\ \mathscr{S}_2 \\ \mathscr{S}_3 \end{bmatrix} := \begin{bmatrix} \int \boldsymbol{\mathscr{B}}_m d\mathscr{B}_u - 3\int r\boldsymbol{\mathscr{B}}_m \int r d\mathscr{B}_u + \boldsymbol{\varrho}_{m*} \\ \int r d\mathscr{B}_u \\ \int d\mathscr{B}_u \end{bmatrix},$$

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

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

Remarks. **(a)** By Lemma 3, none of $(\ddot{\boldsymbol{\upsilon}}_T - \bar{\boldsymbol{\upsilon}}_{T*})$ and $(\widetilde{\boldsymbol{\upsilon}}_T - \boldsymbol{\upsilon}_*)$ follows a mixed normal distribution. Further, $\boldsymbol{\mathscr{S}}_1$ is influenced by the asymptotic bias $\boldsymbol{\varrho}_{m*}$. This asymptotic bias corresponds to that arising when estimating a cointegrating relationship by OLS (e.g., Engle and Granger, 1987; Phillips and Hansen, 1990).

(b) In the Online Supplement, we prove Lemma 3 (*i*) by noting that $\ddot{\mathbf{D}}_1(\ddot{\boldsymbol{\upsilon}}_T - \bar{\boldsymbol{\upsilon}}_{T*}) = (\ddot{\mathbf{D}}_1^{-1}\sum_{t=1}^T \ddot{\boldsymbol{r}}_{t-1}\ddot{\boldsymbol{r}}_{t-1}'\ddot{\mathbf{D}}_1^{-1})^{-1}\ddot{\mathbf{D}}_1^{-1}\sum_{t=1}^T \ddot{\boldsymbol{r}}_{t-1}u_{t-1}$. We focus on deriving the weak limit of $\ddot{\mathbf{D}}_1^{-1}\sum_{t=1}^T \ddot{\boldsymbol{r}}_{t-1}u_{t-1}$, because Lemma 2 (*ii*) already provides the weak limit of $(\ddot{\mathbf{D}}_1^{-1}\sum_{t=1}^T \ddot{\boldsymbol{r}}_{t-1}\ddot{\boldsymbol{r}}_{t-1}'\ddot{\mathbf{D}}_1^{-1})^{-1}$. Next, we note that $\widetilde{\boldsymbol{\upsilon}}_T - \boldsymbol{\upsilon}_* = (\mathbf{P}_T - \mathbf{P})(\ddot{\boldsymbol{\upsilon}}_T - \bar{\boldsymbol{\upsilon}}_{T*}) + \mathbf{P}(\ddot{\boldsymbol{\upsilon}}_T - \bar{\boldsymbol{\upsilon}}_{T*}) + (\mathbf{P}_T - \mathbf{P})(\bar{\boldsymbol{\upsilon}}_{T*} - \bar{\boldsymbol{\upsilon}}_*) + \mathbf{P}(\bar{\boldsymbol{\upsilon}}_{T*} - \bar{\boldsymbol{\upsilon}}_*) + (\mathbf{P}_T - \mathbf{P})\bar{\boldsymbol{\upsilon}}_*$, and prove Lemma 3 (*ii*) by examining the limit of each component on the right

side and showing that $\widetilde{\boldsymbol{\upsilon}}_T - \boldsymbol{\upsilon}_* = \mathbf{P}(\ddot{\boldsymbol{\upsilon}}_T - \bar{\boldsymbol{\upsilon}}_{T*}) + O_{\mathbb{P}}(T^{-3/2})$ and $\ddot{\mathbf{D}}_\dagger(\ddot{\boldsymbol{\upsilon}}_T - \bar{\boldsymbol{\upsilon}}_{T*}) = O_{\mathbb{P}}(1)$.

(c) The consistency of the estimators follows from Lemma 3: $\ddot{\boldsymbol{\upsilon}}_T - \bar{\boldsymbol{\upsilon}}_{T*} \xrightarrow{\mathbb{P}} \mathbf{0}$ and $\widetilde{\boldsymbol{\upsilon}}_T \xrightarrow{\mathbb{P}} \boldsymbol{\upsilon}_*$. We also have $\bar{\boldsymbol{\upsilon}}_{T*} \xrightarrow{\mathbb{P}} \bar{\boldsymbol{\upsilon}}_* := [\boldsymbol{\beta}_*', \vartheta_*, \nu_*]'$, because $\vartheta_{T*} := \boldsymbol{\beta}_*'(\ddot{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) + \vartheta_* \xrightarrow{\mathbb{P}} \vartheta_*$, from $\ddot{\boldsymbol{\mu}}_T = \boldsymbol{\mu}_* + O_{\mathbb{P}}(T^{-1/2})$. This implies that $\mathbf{P}_T \xrightarrow{\mathbb{P}} \mathbf{P}$, so that $\widetilde{\boldsymbol{\upsilon}}_T = \mathbf{P}_T\ddot{\boldsymbol{\upsilon}}_T \to \mathbf{P}\bar{\boldsymbol{\upsilon}}_* = \boldsymbol{\upsilon}_*$, by the definition of $\boldsymbol{\nu}_* := [\boldsymbol{\beta}_*', \zeta_*, \nu_*]'$ in (8).

(d) Lemma 3 (*ii*) implies that the weak limits of $\widetilde{\boldsymbol{\beta}}_T$ and $\widetilde{\zeta}_T$ are linearly correlated, because they are obtained as $\boldsymbol{\mathscr{L}}_1$ and $-\boldsymbol{\mu}_*'\boldsymbol{\mathscr{L}}_1$, respectively. This feature confirms how the long parameter estimator $\widetilde{\boldsymbol{\upsilon}}_T$ suffers from the asymptotically singular matrix problem, although $\ddot{\boldsymbol{\upsilon}}_T$ does not.

(e) 2SNARDL estimates the long- and short-run parameters separately, and the convergence rate of the long-run 2SNARDL estimator for $\boldsymbol{\beta}_*$ is T, which is also the convergence rate of $\ddot{\boldsymbol{\beta}}_T$ in $\ddot{\boldsymbol{\upsilon}}_T$. That is, the same convergence rate is shared by the long-run parameter estimators in the primitive and 2SNARDL estimators. As we detail below, the convergence rate of the OLS estimator for $\boldsymbol{\theta}_*$ is $\sqrt{T}$, but $\ddot{\boldsymbol{\upsilon}}_T$ estimates the long-run parameter in the long-run equation with a faster convergence rate. We use this feature for inference by defining a Wald test based on $\ddot{\boldsymbol{\upsilon}}_T$; see Section 5. □

Second Step We derive the weak limit of the primitive estimator $\ddot{\boldsymbol{\tau}}_T$. We first rewrite (9) as

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

where

$$\begin{aligned}\boldsymbol{\tau}_{T*} &:= [\rho_*, \boldsymbol{\tau}_{1T}', \boldsymbol{\tau}_{2*}']' \quad \text{and} \\ \boldsymbol{\tau}_{1T} &:= [(\boldsymbol{\eta}_* + \rho_*(\ddot{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*))', \psi_* + \boldsymbol{\eta}_*'\ddot{\boldsymbol{\mu}}_T + \rho_*(\ddot{\vartheta}_T - \vartheta_{T*}), \gamma_* + \rho_*(\ddot{\nu}_T - \nu_*)]'.\end{aligned}$$

This equality is established by noting that

$$u_{t-1} = \ddot{u}_{t-1} + (\ddot{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)'\ddot{\mathbf{m}}_{t-1} + (\ddot{\vartheta}_T - (\boldsymbol{\beta}_*'\ddot{\boldsymbol{\mu}}_T + \zeta_*))(t-1) + (\ddot{\nu}_T - \nu_*),$$

which follows from (12) because $\widetilde{u}_{t-1} := y_{t-1} - \boldsymbol{r}_{t-1}'\widetilde{\boldsymbol{\upsilon}}_T = \ddot{u}_{t-1} := y_{t-1} - \ddot{\boldsymbol{r}}_{t-1}'\ddot{\boldsymbol{\upsilon}}_T$. Using this representation, we again decompose the primitive estimator $\ddot{\boldsymbol{\tau}}_T$ into $\boldsymbol{\tau}_{T*}$ and a remainder:

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

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

Lemma 4. *Given Assumption 1,*

(i) $\ddot{\mathbf{D}}(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) \Rightarrow \boldsymbol{\mathscr{D}} := [\mathscr{D}_1, \boldsymbol{\mathscr{D}}_2', \mathscr{D}_3, \mathscr{D}_4, \boldsymbol{\mathscr{D}}_5']' := \boldsymbol{\mathscr{M}}^{-1}\boldsymbol{\mathscr{J}}$, *where* $\boldsymbol{\mathscr{J}} := [\int d\mathscr{B}_{ue}, \int \bar{\boldsymbol{\mathscr{B}}}_m' d\mathscr{B}_e, \int r d\mathscr{B}_e, \int d\mathscr{B}_e, \int d\boldsymbol{\mathscr{B}}_{ze}']'$;

(ii) $\ddot{\mathbf{D}}(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) \Rightarrow \boldsymbol{\mathscr{D}} - \rho_*[0, \boldsymbol{\mathscr{L}}', \mathbf{0}']'$. □

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

(b) In the Online Supplement, 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 $\ddot{\mathbf{D}}^{-1}(\sum_{t=1}^T \ddot{\boldsymbol{z}}_t\ddot{\boldsymbol{z}}_t')\ddot{\mathbf{D}}^{-1} \Rightarrow \boldsymbol{\mathscr{M}}$, we focus on showing that $\ddot{\mathbf{D}}^{-1}\sum_{t=1}^T \ddot{\boldsymbol{z}}_t e_t \Rightarrow \boldsymbol{\mathscr{J}}$. In addition, we note that $\ddot{\mathbf{D}}(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) = \ddot{\mathbf{D}}(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) + [0, -\rho_* T(\ddot{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)', -\rho_* T^{3/2}(\ddot{\vartheta}_T - \vartheta_{T*}), -\rho_* T^{1/2}(\ddot{\nu}_T - \nu_*), \mathbf{0}']'$, and exploit Lemmas 3 (*ii*) and 4 (*i*) to prove Lemma 4 (*ii*).

(c) Unlike the weak limit of the long-run parameter estimator $\widetilde{\boldsymbol{\upsilon}}_T$ given in Lemma 3 (*ii*), the weak limit of $\ddot{\mathbf{D}}(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*)$ cannot be written as a linear combination of other weak limits. This confirms that the primitive estimator $\ddot{\boldsymbol{\tau}}_T$ does not suffer from an asymptotic singular matrix problem.

(d) Because Lemma 4 assumes that $\rho_* < 0$, the weak limit in Lemma 4 (*ii*) cannot be used to derive the null limit distribution of the t-statistic testing $\rho_* = 0$. □

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

$$(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) = (\mathbf{T}_T - \mathbf{T})\boldsymbol{\tau}_* + \mathbf{T}(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) + (\mathbf{T}_T - \mathbf{T})(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*), \tag{15}$$

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

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

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

Before providing the weak limit, we note that different weak limits are obtained under different parameter value conditions. For example, if $\beta_{j*}^+ \neq 0$, $\beta_{j*}^- \neq 0$, and $\zeta_* \neq 0$, for each $j = 1, 2, \ldots, k$, we can rewrite

(15) as

$$\sqrt{T}(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) = \boldsymbol{c}_*\sqrt{T}(\ddot{\rho}_T - \rho_*) + \sqrt{T}\boldsymbol{d}_T + o_{\mathbb{P}}(1), \tag{16}$$

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

In the following theorem, we provide the weak limits of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ under various parameter conditions. For notational simplicity, we let $\boldsymbol{\mathscr{D}}_2 \equiv [\boldsymbol{\mathscr{D}}_2^{+\prime}, \boldsymbol{\mathscr{D}}_2^{-\prime}]'$. Recall that $\boldsymbol{\mathscr{D}}_2$ is the weak limit of $\ddot{\boldsymbol{\eta}}_T - (\boldsymbol{\eta}_* + \rho_*(\ddot{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*))$.

Theorem 1. *Given Assumption 1,*

(i) if for each $j = 1, 2, \ldots, k$, $\beta_{j*}^+ \neq 0$, $\beta_{j*}^- \neq 0$, *and* $\zeta_* \neq 0$, *then* $\sqrt{T}(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) \Rightarrow \boldsymbol{c}_*\mathscr{D}_1 + [0, \mathbf{0}_{2k\times 1}', 0, \mathscr{D}_4, \boldsymbol{\mathscr{D}}_5']'$;

(ii) if $\boldsymbol{\beta}_*^+ = \mathbf{0}$, *but for each* $j = 1, 2, \ldots, k$, $\beta_{j*}^- \neq 0$, *and* $\zeta_* \neq 0$, *then* $\ddot{\mathbf{D}}_+(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) \Rightarrow [\mathscr{D}_1, \boldsymbol{\mathscr{D}}_2^{+\prime}, -\boldsymbol{\beta}_*^{-\prime}\mathscr{D}_1, -\zeta_*\mathscr{D}_1, \mathscr{D}_4 - \nu_*\mathscr{D}_1, \boldsymbol{\mathscr{D}}_5']'$, *where* $\ddot{\mathbf{D}}_+ := diag[T^{1/2}, T\mathbf{I}_k, T^{1/2}\mathbf{I}_{k+2}, \ddot{\mathbf{D}}_2]$;

(iii) if $\boldsymbol{\beta}_*^- = \mathbf{0}$, *but for each* $j = 1, 2, \ldots, k$, $\beta_{j*}^+ \neq 0$, *and* $\zeta_* \neq 0$, *then* $\ddot{\mathbf{D}}_-(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) \Rightarrow [\mathscr{D}_1, -\boldsymbol{\beta}_*^{+\prime}\mathscr{D}_1, \boldsymbol{\mathscr{D}}_2^{-\prime}, -\zeta_*\mathscr{D}_1, \mathscr{D}_4 - \nu_*\mathscr{D}_1, \boldsymbol{\mathscr{D}}_5']'$, *where* $\ddot{\mathbf{D}}_- := diag[T^{1/2}\mathbf{I}_{k+1}, T\mathbf{I}_k, T^{1/2}\mathbf{I}_2, \ddot{\mathbf{D}}_2]$; *and*

(iv) if for each $j = 1, 2, \ldots, k$, $\beta_{j*}^+ \neq 0$, $\beta_{j*}^- \neq 0$, *but* $\zeta_* = 0$, $\ddot{\mathbf{D}}_\odot(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) \Rightarrow [\mathscr{D}_1, -\boldsymbol{\beta}_*'\mathscr{D}_1, -\boldsymbol{\mu}_*'\boldsymbol{\mathscr{D}}_2, \mathscr{D}_4 - \nu_*\mathscr{D}_1, \boldsymbol{\mathscr{D}}_5']'$, *where* $\ddot{\mathbf{D}}_\odot := diag[T^{1/2}\mathbf{I}_{2k+1}, T, T^{1/2}, \ddot{\mathbf{D}}_2]$. □

Remarks. **(a)** The limit distribution of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ is determined by $\boldsymbol{\mathscr{D}}$ 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 $\ddot{\boldsymbol{\nu}}_T$ and $\ddot{\boldsymbol{\tau}}_T$, none of which follows a mixed normal distribution asymptotically, but $\widehat{\boldsymbol{\alpha}}_T$ follows a mixed normal distribution.

(b) Although the time trend and nonstationary regressors are included as regressors, the convergence rate of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ 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) In the Online Supplement, 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 $\widehat{\boldsymbol{\theta}}_T$:

$$(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_*) = -\boldsymbol{\beta}_*(\ddot{\rho}_T - \rho_*) + (\ddot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\ddot{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*). \tag{17}$$

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

(d) By Theorem 1 (*iv*), the 2SNARDL estimator has the same convergence rate as the OLS estimator when it is properly transformed to estimate the parameter targeted by OLS. Specifically, 2SNARDL estimates $\boldsymbol{\upsilon}_*$ and $\boldsymbol{\tau}_*$ separately, and Cho, Greenwood-Nimmo, and Shin (2023a) show that the 2SNARDL estimator for $\boldsymbol{\beta}_*$ and ρ_* has convergence rate T and $\sqrt{T}$, respectively. By this feature, if we multiply the two estimators, $\boldsymbol{\theta}_*$ can be estimated by 2SNARDL and its convergence rate becomes $\sqrt{T}$. We also note that this convergence rate is identical to that of $\widehat{\boldsymbol{\theta}}_T$, meaning that the OLS and 2SNARDL estimators converges to $\boldsymbol{\theta}_*$ at the same convergence rate. In Section A.4.1 of the Online Supplement, we confirm this by simulation.

(e) By Theorem 1 (*i*), the weak limit of $(\widehat{\rho}_T, \widehat{\boldsymbol{\theta}}_T', \widehat{\xi}_T)'$ is $(1, \boldsymbol{\beta}_*', -\zeta_*)'\mathscr{D}_1$, indicating that the estimates are linearly correlated at the limit. This result highlights how the asymptotically singular matrix problem arises and affects the convergence rate of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$. Although $\mathscr{D}_1$ is not associated with an asymptotically singular matrix problem, the type of the transformation means that the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ suffers from the singularity problem.

(f) Despite the asymptotic singularity problem associated with the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$, its weak limit given in Theorem 1 (*i*) is determined by $\mathscr{D}_1$, $\mathscr{D}_4$, and $\boldsymbol{\mathscr{D}}_5$. This implies the following. First, the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ follows a mixed normal distribution as pointed out above. Therefore, if the standard t-test applies to the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$, it follows a mixed normal distribution under the null hypothesis and the condition in Theorem 1 (*i*). We verify this feature by simulation in the Online Supplement; see Section A.4.2. Second, both $\mathscr{D}_1$ and $\boldsymbol{\mathscr{D}}_5$ are the weak limits of the last two OLS estimators obtained by regressing Δy_t against $(1, u_{t-1}, \boldsymbol{z}_{2t}')'$ or $(1, \ddot{u}_{t-1}, \boldsymbol{z}_{2t}')'$. Thus, the null weak limit of the t-statistic testing $\rho_*(< 0)$ is equivalent to those of the t-statistics testing the long-run parameters $\boldsymbol{\beta}_*$ and ξ_*.

(g) If the zero coefficient conditions in Theorems 1 (*ii*, *iii*, and *iv*) hold, the limit distribution of the OLS estimator restricted by the zero condition is determined by the next-order term in (15) as mentioned above. If $\boldsymbol{\beta}_*^+ = \mathbf{0}$, (17) implies that $T(\widehat{\boldsymbol{\theta}}_T^+ - \boldsymbol{\theta}_*^+) = T\{(\ddot{\boldsymbol{\eta}}_T^+ - \boldsymbol{\eta}_*^+) - \rho_*(\ddot{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)\} \Rightarrow \boldsymbol{\mathscr{D}}_2^+$ by Lemma 4 (*i*). In parallel, if $\boldsymbol{\beta}_*^- = \mathbf{0}$, then $T(\widehat{\boldsymbol{\theta}}_T^- - \boldsymbol{\theta}_*^-) = T\{(\ddot{\boldsymbol{\eta}}_T^- - \boldsymbol{\eta}_*^-) - \rho_*(\ddot{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)\} \Rightarrow \boldsymbol{\mathscr{D}}_2^-$; and if $\zeta_* = 0$, then $T(\widehat{\xi}_T - \xi_*) = -\boldsymbol{\mu}_*'T\{(\ddot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\ddot{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)\} + O_{\mathbb{P}}(T^{-1/2}) \Rightarrow -\boldsymbol{\mu}_*'\boldsymbol{\mathscr{D}}_2$, meaning that the weak limits of $\widehat{\boldsymbol{\theta}}_T$ and $\widehat{\xi}_T$ are determined by $\boldsymbol{\mathscr{D}}_2$. Given that $\boldsymbol{\mathscr{D}}_2$ follows a mixed

normal distribution, if the standard t-test applies to $\widehat{\boldsymbol{\theta}}_T$ or $\widehat{\xi}_T$, its null weak limit also follows a mixed normal distribution.

(h) There is a caveat to Theorems 1 (*i*, *ii*, *iii*, and *iv*). If $\boldsymbol{\beta}_* = \mathbf{0}$ 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 $\boldsymbol{\beta}_*^+$, $\boldsymbol{\beta}_*^-$, 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 also implies that the limit distributions of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ are different from those under $\boldsymbol{\beta}_* = \mathbf{0}$ and $\zeta_* = 0$ under the parameter value conditions in Theorem 1.

(j) For empirical application, we recommend the following procedure. 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. Next, if ρ_* turns out to be negative, we can estimate the unknown coefficients by OLS. As the OLS estimator is asymptotically mixed normal, the null limit distribution of the t-test is also mixed normal for each parameter except for the t-statistic testing $\rho_* = 0$. This aspect further implies that the asymptotic critical values of the t-test can be obtained from the standard normal distribution table. Third, when a joint hypothesis is tested, we can apply the 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 $\rho_* = 0$. Therefore, its asymptotic critical values can be obtained from the chi-square distribution table. Note that the second and third steps are exactly the same procedure as for a standard regression analysis. Finally, when estimating the long-run parameter in the long-run equation, using the primitive estimator $\ddot{\boldsymbol{\upsilon}}_T$ or 2SNARDL is recommended. □

Before closing this section, we briefly discuss applying other estimations to the NARDL model as a future research topic. Phillips and Hansen (1990) examine estimating unknown cointegrating relationships by instrumental variable (IV) estimation. The IV or generalized method of moments (GMM) estimation can be applied to NARDL estimation, but its asymptotic distribution needs to be obtained by overcoming asymptotic singular matrix problems similarly to OLS. Alternatively, the IV or GMM estimator has to be defined to reflect the orthogonality condition properly to avoid asymptotic singular matrix problems under the NARDL framework. We leave this as a future research topic.

5 Hypotheses Testing

In this section, we develop a testing methodology by applying the Wald test principle to NARDL. In particular, we suppose that a cointegrating relationship holds between y_t and $\ddot{\boldsymbol{x}}_t$, under Assumption 1, by supposing that $\boldsymbol{\theta}_*^+ \neq \mathbf{0}$ and/or $\boldsymbol{\theta}_*^- \neq \mathbf{0}$, and develop the testing methodology. In addition to the standard Wald test, we apply the Wald test principle to the primitive estimators introduced in Sections 3 and 4.

Our main interest is in testing the symmetry conditions. The NARDL process reduces to the ARDL process if $\boldsymbol{\theta}_*^+ = \boldsymbol{\theta}_*^-$ and $\boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-$. Under the ARDL process condition, it is inefficient to estimate the parameters using NARDL, making it necessary to test the parameter symmetry conditions. We specify the following three hypothesis systems:

$$\mathscr{H}_0' : \boldsymbol{\theta}_*^+ = \boldsymbol{\theta}_*^- \quad \text{vs.} \quad \mathscr{H}_1' : \boldsymbol{\theta}_*^+ \neq \boldsymbol{\theta}_*^-;$$

$$\mathscr{H}_0'' : \boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^- \quad \text{vs.} \quad \mathscr{H}_1'' : \boldsymbol{\pi}_*^+ \neq \boldsymbol{\pi}_*^-;$$

$$\mathscr{H}_0''' : \boldsymbol{\theta}_*^+ = \boldsymbol{\theta}_*^- \ \text{ and } \ \boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^- \quad \text{vs.} \quad \mathscr{H}_1''' : \boldsymbol{\theta}_*^+ \neq \boldsymbol{\theta}_*^- \ \text{ or } \ \boldsymbol{\pi}_*^+ \neq \boldsymbol{\pi}_*^-.$$

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

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

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

$\widehat{\sigma}_{e,T}^2 := T^{-1} \sum_{t=1}^T (\Delta y_t - \boldsymbol{z}_t' \widehat{\boldsymbol{\alpha}}_T)^2$, and $\widehat{\mathbf{R}}_1 := [\mathbf{0}_{k\times 1}, \mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k\times(1+p+2kq)}]$. As we are presuming that $\boldsymbol{\theta}_*^+ \neq \mathbf{0}$ and/or $\boldsymbol{\theta}_*^- \neq \mathbf{0}$, we cannot suppose the environment assumed in Theorems 1 (*ii* or *iii*) to obtain the null limit distribution of the Wald test. In addition to this, we cannot directly apply the limit distribution of the OLS estimator in Theorems 1 (*i* and *iv*) to obtain the null limit distribution. Note that $\widehat{\mathbf{R}}_1 \widehat{\boldsymbol{\alpha}}_T = (\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-)$, so that $\sqrt{T}(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) \Rightarrow (\boldsymbol{\beta}_*^- - \boldsymbol{\beta}_*^+)\mathscr{D}_1$. Thus, the weak limit in Theorem 1 (*i* or *iv*) does not help us to obtain the null limit distribution of the Wald test, because $\boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$ under the null, meaning that $\sqrt{T}(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-)$ is asymptotically negligible. The null limit behavior has to be obtained by the next-order

term of $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-)$. Specifically, (10) implies that

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

under $\mathscr{H}_0'$, where we let $\boldsymbol{\eta}_{T*} := \boldsymbol{\eta}_* + \rho_*(\ddot{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)$ denote the vector formed by the first $2k$-row elements of $\boldsymbol{\tau}_{1T}$, and (18) is established by letting $\boldsymbol{\eta}_{T*} := [\boldsymbol{\eta}_{T*}^{+\prime}, \boldsymbol{\eta}_{T*}^{-\prime}]'$ and noting that $\widetilde{\boldsymbol{\beta}}_T = \ddot{\boldsymbol{\beta}}_T$. Therefore, the null limit distribution of the Wald test $W_T^{(1)}$ is determined not by Theorem 1 (*i* or *iv*), but by Lemma 4. All these imply $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = O_{\mathbb{P}}(T^{-1})$, so that if we let $\mathbf{R}_1 := [\mathbf{0}_{k\times 1}, \mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k\times(1+p+2kq)}]$, $T(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) \Rightarrow \mathbf{R}_1\mathscr{D}$ under $\mathscr{H}_0'$ by Lemma 4. However, its alternative behavior is different. From $\widehat{\mathbf{R}}_1\widehat{\boldsymbol{\alpha}}_T = (\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-)$, its divergence rate is determined by Theorem 1 (*i* or *iv*), meaning that $\widehat{\mathbf{R}}_1\widehat{\boldsymbol{\alpha}}_T = O_{\mathbb{P}}(T^{-1/2})$. That is, the convergence and divergence rates of the test basis are unbalanced between $\mathscr{H}_0'$ and $\mathscr{H}_1'$, respectively.

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

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

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

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

and $\ddot{\sigma}_{u,T}^2 := T^{-1}\sum_{t=1}^T \ddot{u}_t^2$. Here, $(\ddot{\boldsymbol{\beta}}_T^+ - \ddot{\boldsymbol{\beta}}_T^-)$ is not the same as $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-)$. Their difference is caused by $(\ddot{\boldsymbol{\eta}}_T^+ - \ddot{\boldsymbol{\eta}}_T^-)$ and $\ddot{\rho}_T$, as implied by (18) and the fact that $(\widetilde{\boldsymbol{\beta}}_T^+ - \widetilde{\boldsymbol{\beta}}_T^-) = (\ddot{\boldsymbol{\beta}}_T^+ - \ddot{\boldsymbol{\beta}}_T^-)$. Here, the Wald test $\mathscr{W}_T^{(1)}$ is defined only by the test basis with power. Neither $(\ddot{\boldsymbol{\eta}}_T^+ - \ddot{\boldsymbol{\eta}}_T^-)$ nor $\ddot{\rho}_T$ is capable of contributing to the power of the test. Furthermore, $(\ddot{\boldsymbol{\beta}}_T^+ - \ddot{\boldsymbol{\beta}}_T^-) = O_{\mathbb{P}}(T^{-1})$ under both hypotheses. From this, the convergence and divergence rates of the Wald test are the same under both $\mathscr{H}_0'$ and $\mathscr{H}_1'$.

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

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

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

We define another supplementary Wald test using the primitive estimator, as in the earlier case. From Proposition 4, it follows that $(\widehat{\boldsymbol{\varphi}}_T', \widehat{\boldsymbol{\pi}}_T^{+\prime}, \widehat{\boldsymbol{\pi}}_T^{-\prime})' = (\ddot{\boldsymbol{\varphi}}_T', \ddot{\boldsymbol{\pi}}_T^{+\prime}, \ddot{\boldsymbol{\pi}}_T^{-\prime})' := \ddot{\boldsymbol{\tau}}_{2T}$. That is, the short-run OLS and primitive estimators are the same, implying that we can test $\mathscr{H}_0''$ by using $\ddot{\boldsymbol{\tau}}_T$. Lemma 2 (*vi*) implies that the primitive estimator $\ddot{\boldsymbol{\tau}}_T$ is not associated with an asymptotically singular matrix problem. From this, we define the following Wald test:

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

That is, $\mathscr{W}_T^{(2)}$ is defined by applying the Wald test principle to the primitive estimator $\ddot{\boldsymbol{\tau}}_T$, estimating the parameters in (9) by noting that $\widehat{\mathbf{R}}_2 \ddot{\mathbf{D}} \ddot{\boldsymbol{\tau}}_T = \sqrt{T}(\ddot{\boldsymbol{\pi}}_T^+ - \ddot{\boldsymbol{\pi}}_T^-)$. Although the test basis of the Wald test $\mathscr{W}_T^{(2)}$ is the same as that of $W_T^{(2)}$, from $\ddot{\boldsymbol{\pi}}_T^+ - \ddot{\boldsymbol{\pi}}_T^- = \widehat{\boldsymbol{\pi}}_T^+ - \widehat{\boldsymbol{\pi}}_T^-$, their weight matrices are different. The Wald test $\mathscr{W}_T^{(2)}$ is defined using the weight matrix $\ddot{\mathbf{W}}_T^{(2)}$. This weight matrix is more relevant to the test basis, because $\ddot{\boldsymbol{\tau}}_T$ is obtained by regressing $\ddot{\boldsymbol{z}}_t$ against Δy_t.

Testing $\mathscr{H}_0'''$ versus $\mathscr{H}_1'''$ Finally, we apply the Wald test principle to test $\mathscr{H}_0'''$. As before, we cannot suppose the environment in Theorems 1 (*ii* and *iii*) under $\mathscr{H}_0'''$. We apply the Wald test principle to the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ as follows:

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

and $\widehat{\mathbf{R}}_3 := \mathrm{diag}[\widehat{\mathbf{R}}_1, \widehat{\mathbf{R}}_2]$, so that $\widehat{\mathbf{R}}_3 \widehat{\boldsymbol{\alpha}}_T = (\widehat{\boldsymbol{\theta}}_T^{+\prime} - \widehat{\boldsymbol{\theta}}_T^{-\prime}, \widehat{\boldsymbol{\pi}}_T^{+\prime} - \widehat{\boldsymbol{\pi}}_T^{-\prime})'$. Here, the null limit distribution of $\widehat{\mathbf{R}}_3 \widehat{\boldsymbol{\alpha}}_T$ is obtained by combining the results given above. That is, $T(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) \Rightarrow \mathbf{R}_1 \boldsymbol{\mathscr{D}}$ and $\sqrt{T}(\widehat{\boldsymbol{\pi}}_T^+ - \widehat{\boldsymbol{\pi}}_T^-) \Rightarrow \widehat{\mathbf{R}}_2 \boldsymbol{\mathscr{D}}$ under $\mathscr{H}_0'''$. Therefore, if we further let $\mathbf{R}_3 := \mathrm{diag}[\mathbf{R}_1, \widehat{\mathbf{R}}_2]$, it follows that $(T(\widehat{\boldsymbol{\theta}}_T^{+\prime} - \widehat{\boldsymbol{\theta}}_T^{-\prime}), \sqrt{T}(\widehat{\boldsymbol{\pi}}_T^{+\prime} - \widehat{\boldsymbol{\pi}}_T^{-\prime}))' \Rightarrow \mathbf{R}_3 \boldsymbol{\mathscr{D}}$ under $\mathscr{H}_0'''$. However, $\widehat{\mathbf{R}}_3 \widehat{\boldsymbol{\alpha}}_T = O_{\mathbb{P}}(T^{-1/2})$ under $\mathscr{H}_1'''$ because $(\widehat{\boldsymbol{\theta}}_T^{+\prime} - \widehat{\boldsymbol{\theta}}_T^{-\prime})' = O_{\mathbb{P}}(T^{-1/2})$ under $\mathscr{H}_1'''$ as discussed above.

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

$$\mathbb{H}_0''' : \boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^- \quad \text{and} \quad \boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^- \quad \text{vs.} \quad \mathbb{H}_1''' : \boldsymbol{\beta}_*^+ \neq \boldsymbol{\beta}_*^- \quad \text{or} \quad \boldsymbol{\pi}_*^+ \neq \boldsymbol{\pi}_*^-,$$

and define the following supplementary Wald test:

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

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

We now provide the limit behaviors of the Wald tests under the null and alternative using the weak limits of the primitive estimators, $\ddot{\boldsymbol{\upsilon}}_T$ and $\ddot{\boldsymbol{\tau}}_T$. For notational simplicity, we partition $\mathbb{H}_0'''$ into $\mathbb{H}_{01}''' : \boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$ and $\mathbb{H}_{02}''' : \boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-$. In parallel, we also partition $\mathbb{H}_1'''$ into $\mathbb{H}_{11}''' : \boldsymbol{\beta}_*^+ \neq \boldsymbol{\beta}_*^-$ and $\mathbb{H}_{12}''' : \boldsymbol{\pi}_*^+ \neq \boldsymbol{\pi}_*^-$. We introduce this partition because the power behavior of the Wald test $\mathscr{W}_T^{(3)}$ depends on the sub-hypotheses from the fact that the convergence rate of $(\ddot{\boldsymbol{\beta}}_T^+ - \ddot{\boldsymbol{\beta}}_T^-)$ differs from that of $(\ddot{\boldsymbol{\pi}}_T^+ - \ddot{\boldsymbol{\pi}}_T^-)$. We also let $\sigma_e^2 := \mathbb{E}[e_t^2]$. We summarize their limit behaviors in the following theorem:

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

(i) *(a)* $W_T^{(1)} \Rightarrow \boldsymbol{\mathscr{D}}' \mathbf{R}_1' (\sigma_e^2 \mathbf{R}_1 \boldsymbol{\mathscr{M}}^{-1} \mathbf{R}_1')^{-1} \mathbf{R}_1 \boldsymbol{\mathscr{D}}$ *under* $\mathscr{H}_0'$, *where* $\mathbf{R}_1 := [\mathbf{0}_{k\times 1}, \mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k\times(1+p+2kq)}]$;

(b) $W_T^{(2)} \Rightarrow \boldsymbol{\mathscr{D}}' \widehat{\mathbf{R}}_2' (\sigma_e^2 \widehat{\mathbf{R}}_2 \boldsymbol{\mathscr{M}}^{-1} \widehat{\mathbf{R}}_2')^{-1} \widehat{\mathbf{R}}_2 \boldsymbol{\mathscr{D}}$ *under* $\mathscr{H}_0''$;

(c) $W_T^{(3)} \Rightarrow \boldsymbol{\mathscr{D}}' \mathbf{R}_3' (\sigma_e^2 \mathbf{R}_3 \boldsymbol{\mathscr{M}}^{-1} \mathbf{R}_3')^{-1} \mathbf{R}_3 \boldsymbol{\mathscr{D}}$ *under* $\mathscr{H}_0'''$, *where* $\mathbf{R}_3 := diag[\mathbf{R}_1, \widehat{\mathbf{R}}_2]$;

(ii) *(a)* *for any* $c_T' = o(T)$, $\lim_{T\to\infty} \mathbb{P}(W_T^{(1)} > c_T) = 0$ *under* $\mathscr{H}_1'$;

(b) *for any* $c_T' = o(T)$, $\lim_{T\to\infty} \mathbb{P}(W_T^{(2)} > c_T') = 0$ *under* $\mathscr{H}_1''$;

(c) *for any* $c_T' = o(T)$, $\lim_{T\to\infty} \mathbb{P}(W_T^{(3)} > c_T') = 0$ *under* $\mathscr{H}_1'''$;

(iii) *(a)* $\mathscr{W}_T^{(1)} \Rightarrow \boldsymbol{\mathscr{L}}' \ddot{\mathbf{R}}_1' (\sigma_u^2 \ddot{\mathbf{R}}_1 \boldsymbol{\mathscr{M}}_{11}^{-1} \ddot{\mathbf{R}}_1')^{-1} \ddot{\mathbf{R}}_1 \boldsymbol{\mathscr{L}}$ *under* $\mathbb{H}_1'$;

(b) $\mathscr{W}_T^{(2)} \Rightarrow \boldsymbol{\mathscr{D}}' \widehat{\mathbf{R}}_2' (\sigma_e^2 \widehat{\mathbf{R}}_2 \boldsymbol{\mathscr{M}}^{-1} \widehat{\mathbf{R}}_2')^{-1} \widehat{\mathbf{R}}_2 \boldsymbol{\mathscr{D}}$ *under* $\mathscr{H}_0''$;

(c) $\mathscr{W}_T^{(3)} \Rightarrow \boldsymbol{\mathscr{L}}' \ddot{\mathbf{R}}_1' (\sigma_u^2 \ddot{\mathbf{R}}_1 \boldsymbol{\mathscr{M}}_{11}^{-1} \ddot{\mathbf{R}}_1')^{-1} \ddot{\mathbf{R}}_1 \boldsymbol{\mathscr{L}} + \boldsymbol{\mathscr{D}}' \widehat{\mathbf{R}}_2' (\sigma_e^2 \widehat{\mathbf{R}}_2 \boldsymbol{\mathscr{M}}^{-1} \widehat{\mathbf{R}}_2')^{-1} \widehat{\mathbf{R}}_2 \boldsymbol{\mathscr{D}}$ *under* $\mathbb{H}_0'''$; *and*

(iv) *(a)* *for any* $c_T = o(T^2)$, $\lim_{T\to\infty} \mathbb{P}(\mathscr{W}_T^{(1)} > c_T) = 0$ *under* $\mathbb{H}_1'$;

(b) *for any* $c_T' = o(T)$, $\lim_{T\to\infty} \mathbb{P}(\mathscr{W}_T^{(2)} > c_T') = 0$ *under* $\mathscr{H}_1''$;

(c) *for any* $c_T = o(T^2)$ *and* $c_T' = o(T)$,

(1) $\lim_{T\to\infty}\mathbb{P}(\mathscr{W}_T^{(3)} > c_T) = 0$ *under* $\mathbb{H}_{11}''' \bigcap \mathbb{H}_{02}'''$*;*

(2) $\lim_{T\to\infty}\mathbb{P}(\mathscr{W}_T^{(3)} > c_T') = 0$ *under* $\mathbb{H}_{01}''' \bigcap \mathbb{H}_{12}'''$*; and*

(3) $\lim_{T\to\infty}\mathbb{P}(\mathscr{W}_T^{(3)} > c_T) = 0$ *under* $\mathbb{H}_{11}''' \bigcap \mathbb{H}_{12}'''$. □

Remarks. **(a)** The asymptotic behaviors of the Wald tests $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ are determined by the weak limit of the primitive estimator $(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*})$. For example, it follows from (18) that $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-) - (\ddot{\rho}_T - \rho_*)(\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-) + (\ddot{\boldsymbol{\eta}}_T^+ - \boldsymbol{\eta}_{T*}^+) - (\ddot{\boldsymbol{\eta}}_T^- - \boldsymbol{\eta}_{T*}^-) + o_{\mathbb{P}}(T^{-1})$. Therefore, if we impose $\mathscr{H}_0'$, $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\ddot{\boldsymbol{\eta}}_T^+ - \boldsymbol{\eta}_{T*}^+) - (\ddot{\boldsymbol{\eta}}_T^- - \boldsymbol{\eta}_{T*}^-) + o_{\mathbb{P}}(T^{-1})$. In other words, $\widehat{\mathbf{R}}_1\widehat{\boldsymbol{\alpha}}_T = \mathbf{R}_1(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) + o_{\mathbb{P}}(T^{-1})$. This fact implies that the null weak limit of the Wald test $W_T^{(1)}$ is determined by that of $(\ddot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*})$, which is given in Lemma 4. Theorem 2 (*i.a*) reports the weak limit obtained in this way. Here, we note that the test basis has a different weak limit under $\mathscr{H}_1'$. From (18), it follows that $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-) - (\ddot{\rho}_T - \rho_*)(\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-) + o_{\mathbb{P}}(T^{-1/2})$ 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. By applying similar arguments to the other Wald tests, $W_T^{(2)}$ and $W_T^{(3)}$, we provide their limit behaviors in Theorem 2 (*i.b*, *i.c*, *ii.b*, and *ii.c*). Consequently, the Wald tests $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ asymptotically follow mixed chi-squared distributions under the null hypotheses, because all they have the null limit distributions characterized by $\mathscr{D}$ following a mixed normal distribution.

(b) If we partition $\mathscr{L}_1$ such that $\mathscr{L}_1 \equiv (\mathscr{L}_1^{+\prime}, \mathscr{L}_1^{-\prime})'$, $\ddot{\mathbf{R}}_1\mathscr{L} = \mathscr{L}_1^+ - \mathscr{L}_1^-$, so that the null weak limit in Theorem 2 (*iii.a*) is given as $(\mathscr{L}_1^+ - \mathscr{L}_1^-)'(\sigma_u^2\ddot{\mathbf{R}}_1\mathscr{M}_{11}^{-1}\ddot{\mathbf{R}}_1')^{-1}(\mathscr{L}_1^+ - \mathscr{L}_1^-)$. In parallel, if we let $\mathscr{D}_5 \equiv [\mathscr{D}_5^{\dagger\prime}, \mathscr{D}_5^{+\prime}, \mathscr{D}_5^{-\prime}]'$, such that $\mathscr{D}_5^{\dagger}$, $\mathscr{D}_5^+$, and $\mathscr{D}_5^-$ are the weak limits of $\ddot{\boldsymbol{\varphi}}_T$, $\ddot{\boldsymbol{\pi}}_T^+$, and $\ddot{\boldsymbol{\pi}}_T^-$, respectively, then $\widehat{\mathbf{R}}_2\mathscr{D} = \mathscr{D}_5^+ - \mathscr{D}_5^-$, and so the weak the limit in Theorem 2 (*iii.b*) can be rewritten as $(\mathscr{D}_5^+ - \mathscr{D}_5^-)'(\sigma_e^2\widehat{\mathbf{R}}_2\mathscr{M}^{-1}\widehat{\mathbf{R}}_2')^{-1}(\mathscr{D}_5^+ - \mathscr{D}_5^-)$.

(c) The testing methodology of using the supplementary tests depends on the hypotheses. We can rewrite the null weak limit of the Wald test $\mathscr{W}_T^{(1)}$ as $\mathscr{S}'\mathscr{M}_{11}^{-1}\ddot{\mathbf{R}}_1'(\sigma_u^2\ddot{\mathbf{R}}_1\mathscr{M}_{11}^{-1}\ddot{\mathbf{R}}_1')^{-1}\ddot{\mathbf{R}}_1\mathscr{M}_{11}^{-1}\mathscr{S}$ using the definition of $\mathscr{L} := \mathscr{M}_{11}^{-1}\mathscr{S}$, meaning that the Wald test $\mathscr{W}_T^{(1)}$ does not follow a mixed chi-squared distribution. This is mainly because the weak limit $\mathscr{S}$ does not follow a mixed normal distribution and is also influenced by the asymptotic bias $\boldsymbol{\varrho}_{m*}$. The same result holds for the Wald test $\mathscr{W}_T^{(3)}$ from the fact that $\mathscr{W}_T^{(3)} = \mathscr{W}_T^{(1)} + \mathscr{W}_T^{(2)}$. However, the null weak limit of the Wald test $\mathscr{W}_T^{(2)}$ is characterized by $\mathscr{D}$, meaning that its null weak limit follows a mixed chi-squared distribution. Therefore, tabulated asymptotic critical values are available for the Wald test $\mathscr{W}_T^{(2)}$. For both Wald tests, $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$, we can apply a resampling method; see Section 6.

(d) The divergence speed of the Wald test $\mathscr{W}_T^{(3)}$ depends on the alternative hypothesis. If $\mathbb{H}_0'''$ is negated only because of the asymmetric long-run parameters: $\boldsymbol{\beta}_*^+ \neq \boldsymbol{\beta}_*^-$, the divergence speed is T^2. In contrast, if only the short-run parameters are asymmetric: $\boldsymbol{\pi}_*^+ \neq \boldsymbol{\pi}_*^-$, its divergence speed is T. If

both the long- and short-run parameters are asymmetric, the divergence speed is determined by T^2. Theorem 2 (*iv.c*) summarizes these results. □

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 can apply. Section 6 details the residual bootstrap procedure applied to the Wald test.

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 $\Delta x_t = 1/2 + v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. 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 $\bar{\boldsymbol{\upsilon}}_{T*}$ and $\boldsymbol{\tau}_{T*}$ separately by specifying the following models:

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

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

where we set $\widehat{m}_t^+ := x_t^+ - t\widehat{\mu}_T^+$, $\widehat{m}_t^- := x_t^- - t\widehat{\mu}_T^-$, and $\widetilde{u}_t := y_t - \boldsymbol{r}_t' \widetilde{\boldsymbol{\upsilon}}_T$, with $\widehat{\mu}_T^+ := (\sum_{t=1}^{T-1} t^2)^{-1} \sum_{t=1}^{T-1} t x_t^+$, $\widehat{\mu}_T^- := (\sum_{t=1}^{T-1} t^2)^{-1} \sum_{t=1}^{T-1} t x_t^-$, $\widetilde{\boldsymbol{\upsilon}}_T := (\sum_{t=1}^{T} \boldsymbol{r}_{t-1} \boldsymbol{r}_{t-1}')^{-1} \sum_{t=1}^{T} \boldsymbol{r}_{t-1} y_{t-1}$, and $\boldsymbol{r}_{t-1} := [x_{t-1}^+, x_{t-1}^-, (t-1), 1]'$. Second, we specify the following hypothesis systems:

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

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

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

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

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

We conduct simulations under the following two DGP conditions. 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 $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$ are not mixed chi-squared. 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 $(\widehat{\rho}_T, \widehat{\theta}_T, \widehat{\xi}_T, \widehat{\alpha}_T, \widehat{\varphi}_T, \widehat{\pi}_T)$. We also let the residual be $\widehat{e}_t := \Delta y_t - \widehat{\rho}_T y_{t-1} - \widehat{\theta}_T x_{t-1} - \widehat{\xi}_T(t-1) - \widehat{\alpha}_T - \widehat{\varphi}_T \Delta y_{t-1} - \widehat{\pi}_T \Delta x_t$.

S2: We construct a resampled series as follows. First, we resample $\widehat{e}_t$ with replacement and denote it as e_t^b. Next, we let

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

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 $\mathscr{W}_T^{(1),b}$ denote the bootstrapped $\mathscr{W}_T^{(1)}$.

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 $p_T^{(1)} := B^{-1} \sum_{b=1}^B \mathbb{1}\{\mathscr{W}_T^{(1),b} > \mathscr{W}_T^{(1)}\}$ for $\mathscr{W}_T^{(1)}$. If $p_T^{(1)}$ is less than the significance level, we reject the null hypothesis. □

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

We conducted our simulation following 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 evaluated the Wald tests by mixed chi-squared distributions and reported the empirical rejection rates. The simulation results are summarized as follows:

(a) Table 2 shows the empirical Wald test 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 the Wald test $\mathscr{W}_T^{(2)}$, but both Wald tests $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$ do experience level distortions. For the current DGP, the Wald tests $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$ reject $\mathbb{H}_0'$ and $\mathbb{H}_0'''$ more often than the significance levels. This confirms that the Wald tests $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$ do not follow mixed chi-squared distributions under the null. However, the Wald test $\mathscr{W}_T^{(2)}$ does not suffer from level distortions, indicating that $\ddot{\pi}_T^+$ and $\ddot{\pi}_T^-$ follow mixed normal distributions, as Lemma 4 (*i*) establishes. Finally, the Wald tests $W_T^{(1)}$, $W_T^{(2)}$, 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 Wald test rejection rates under the alternative DGP. The summary results are as follows. First, with the bootstrap method, the empirical rejection rates of the Wald tests $\mathscr{W}_T^{(1)}$, $\mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(3)}$ tend toward 100% as T increases, implying their consistency against $\mathbb{H}_1'$, $\mathscr{H}_1''$, and $\mathbb{H}_1'''$, respectively. Second, when applying the critical values obtained from the mixed chi-squared distribution, the empirical rejection rates of the Wald tests $\mathscr{W}_T^{(1)}$, $\mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(3)}$ 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, the Wald tests $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$ are more powerful than the Wald tests $W_T^{(1)}$ and $W_T^{(3)}$, respectively, for small T. In contrast, both Wald tests $\mathscr{W}_T^{(2)}$ and $W_T^{(2)}$ show similar power patterns. □

We conduct another simulation assuming that Δx_t is serially correlated. Instead of $\Delta x_t = 1/2 + v_t$, we set $\Delta x_t = 1/4 + \kappa_* \Delta x_{t-1} + v_t$ and apply the same residual bootstrap method. Then, assuming $\kappa_* = 1/2$, we conduct simulations by setting $B = 500$ and independently iterating the experiment 5000 times for $T = 100$, 200, 300, 400, and 500. Here, both y_t and x_t become integrated series with time trends as before. The simulation results are presented in Tables 4 and 5 in the same format as in Tables 2 and 3, respectively. The simulation results are summarized as follows:

(a) Table 4 presents the empirical Wald test rejection rates obtained using the null DGP condition. For each value of T, the empirical rejection rates are very close to the nominal significance levels when using the residual bootstrap method. In contrast, the asymptotic critical values introduce level distortions for the Wald tests $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$, as observed earlier. Overall, these simulation results are similar to those shown in Table 2.

(b) Table 5 presents the empirical Wald test rejection rates obtained using the alternative DGP condition. The results are as follows. First, when the residual bootstrap method is used, the empirical rejection

rates of the Wald tests $\mathscr{W}_T^{(1)}$, $\mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(3)}$ tend toward 100% as T increases, implying their consistency against $\mathbb{H}_1'$, $\mathscr{H}_1''$, and $\mathbb{H}_1'''$, respectively. Second, when the asymptotic critical values are used, the empirical rejection rates also converge toward 100% as T increases, although the level distortions are difficult to control for the Wald tests $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$ under the null. Third, the standard Wald tests exhibit consistent power. As T increases, the Wald tests $W_T^{(1)}$, $W_T^{(2)}$, and $W_T^{(3)}$ reject the null with rejection rates tending toward 100%. Finally, the Wald tests $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(3)}$ are more powerful than the Wald tests $W_T^{(1)}$ and $W_T^{(3)}$, respectively, for small T. However, both Wald tests $\mathscr{W}_T^{(2)}$ and $W_T^{(2)}$ exhibit similar power patterns.

(c) These simulation results are qualitatively the same as those for the serially uncorrelated Δx_t. □

Before moving to the empirical section, we describe other simulation experiences on OLS and 2SNARDL. As the two estimators target the same parameter, each has relative advantages over the other. The OLS contains a finite sample bias that is driven by $\boldsymbol{\varrho}_{m*}$ 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 $\Delta \boldsymbol{x}_t$ and u_t, affecting the performance of 2SNARDL. Meanwhile, OLS is straightforward to estimate. From this comparison, if T is sufficiently large, the OLS can be preferred to the 2SNARDL for practical purposes. However, if T is not large but the asymptotic covariance matrix estimator can remove the bias, 2SNARDL can be preferred over OLS.

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, causing the OLS estimator to suffer from endogenous bias. Although not all fiscal factors are endogenous in terms of output growth, using all tax changes to analyze GDP growth can result in biased OLS estimates. Blanchard and Perotti (2002) address this by using structural vector autoregression (SVAR). They 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 these tax changes 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 models under various economic 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 outcomes from conventional analyses.

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

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

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:

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

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 can also convert (20) as follows:

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

For this estimation, we first examine the partial sum processes of the exogenous tax changes used in our empirical analysis. Figure 1 illustrates the partial sum processes due to the exogenous tax changes.[1] 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, which we discuss one by one. 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.

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

[2] We 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}.

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:

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

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

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

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. We can use the Wald tests defined in Section 5 for this inference.

7.2 Empirical Results

This section presents the estimation and inference results, divided into two parts. The first part presents the estimation results using the tax change data outlined in the previous section. The second part measures the fiscal policy effect using the tax ratio data, as used by Romer and Romer (2010). We have limited the sample period to 1947Q1–2007Q4 by excluding periods with missing observations.

7.2.1 Tax Changes Measured by Log Transformation of Tax

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, $\tau_{1,t}$, and $\tau_{2,t}$ 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 6. 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 6 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 $\mathscr{W}^{(1)}_T$ and $W^{(1)}_T$ provide p-values that make it difficult to reject the symmetry hypothesis.

(d) Short-run symmetry is confirmed by the Wald tests $\mathscr{W}^{(2)}_T$ and $W^{(2)}_T$, 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 $\mathscr{W}^{(3)}_T$ or $W^{(3)}_T$. 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. □

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 7, which is structured similarly to Table 6. This separate investigation is conducted to validate the inference results presented in Table 6. We summarize the results in Table 7.

(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 6. 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. □

The results in Tables 6 and 7 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.

7.2.2 Tax Changes Measured by Tax-to-GDP Ratio

This section extends the work of Romer and Romer (2010) to investigate the long- and short-run relationships between fiscal shocks and real GDP. 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). As with Tables 6 and

7, we estimate the models using OLS and 2SNARDL. The estimation and inference results are presented in Tables 8 and 9, respectively. We summarize the results as follows:

(a) For the exogenous tax change, the coefficient of y_{t-1} in Table 8 is statistically significant at the 25% level for the NARDL and ARDL models, by the t-test. The F-test does not reject the hypothesis of no cointegration. However, the coefficient of u_{t-1} in the t-test of Table 9 is statistically significant at the 25% and 10% levels for the NARDL and ARDL models, respectively. We give more weight to the inference results in Table 9 than to those in Table 8, because the long-run parameters can be estimated super-consistently by 2SNARDL. In addition, neither $\mathscr{W}_T^{(1)}$ nor $W_T^{(1)}$ rejects the symmetry hypothesis in the long-run parameters, indicating that the ARDL model can estimate a cointegrating relationship between y_t and r_t more efficiently than 2SNARDL can.

(b) The ARDL model estimation indicates 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.

(c) The Wald tests $\mathscr{W}_T^{(2)}$ and $W_T^{(2)}$ do not reject the hypothesis of symmetric short-run parameters. Furthermore, neither $\mathscr{W}_T^{(3)}$ nor $W_T^{(3)}$ rejects the hypothesis of long- and short-run symmetry. This confirms that ARDL is appropriate for the long- and short-run relationships between y_t and r_t.

(d) For endogenous and aggregate tax changes, there is negligible evidence of cointegration between real GDP and the tax change. None of the t- or F-tests in Table 8 rejects the hypothesis of no cointegration. Furthermore, none of the coefficients of u_{t-1} in Table 9 is statistically significant. Specifically, for the aggregate tax change, the unit-root test cannot confirm that r_t is nonstationary; see the Online Supplement. Therefore, we conclude that the long- and short-run relationships can be properly estimated only by the exogenous tax change. □

In summary, our empirical results obtained using the specification in Romer and Romer (2010) provide qualitatively the same results as those in Section 7.2.1. In particular, the long-run relationship between y_t and r_t captured by the cointegration coefficient is close to their estimate.

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 weak limits. This representation and the weak limits allowed us to derive the weak limit of the OLS and enabled us to demonstrate that the convergence rate of the OLS is identical to that of the estimator implied by the 2SNARDL estimator despite the asymptotic 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 to examine asymmetric long- and/or short-run parameters. 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, affecting 2SNARDL. 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 but the asymptotic covariance matrix estimator successfully removes the bias, 2SNARDL can be preferred over OLS.

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, using the empirical data from Romer and Romer (2010). Because all tax changes for deficit reduction are increases, and most changes for long-run growth are 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. We also find that a 1% exogenous increase in the tax-to-GDP ratio reduces the log real GDP by about 3% in the long run, confirming the estimation results of Romer and Romer (2010).

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. 7, 22, 35

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. 31, 32

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 Univresity. 21, 36, 16

——— (2023b): "Recent Developments of the Autoregressive Distributed Lag Modelling Framework," *Journal of Economic Surveys*, 37, 7–32. 4

——— (2024): "Two-Step Nonlinear ARDL Estimation: Theory and Application," Discussion paper, School of Economics, Yonsei Univresity. 1, 7, 8, 9

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, 9, 15, 20

CHO, J. S. AND H. L. WHITE (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720. 2, 9, 15, 20

——— (2010): "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157, 458–480. 2, 9, 15, 20

CLOYNE, J. (2013): "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom," *American Economic Review*, 103, 1507–1528. 32

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. 10

DEMIREL, U. D. (2021): "The Short-Term Effects of Tax Changes: The Role of State Dependence," *Journal of Monetary Economics*, 117, 918–934. 32

ENGLE, R. F. AND C. W. J. GRANGER (1987): "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276. 1, 17

GHASSIBE, M. AND F. ZANETTI (2021): "State Dependence of Fiscal Multipliers: The Source of Fluctuations Natters," *Journal of Monetary Economics*, 132, 1–23. 32

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. 32

MERTENS, K. AND J. L. M. OLEA (2018): "Marginal Tax Rates and Income: New Time Series Evidence," *Quarterly Journal of Economics*, 133, 1803–1884. 32

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. 32

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. 4, 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. 7, 22, 35, 47, 49

PHILLIPS, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306. 7

——— (1995): "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1023–1078. 2, 9

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. 7, 17, 22

PHILLIPS, P. C. B. AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346. 34, 36, 20, 25

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, 3, 31, 32, 33, 34, 36, 37, 38, 47, 48, 49, 50, 25

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, 7

TERÄSVIRTA, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208–218. 2, 9, 15, 20

WHITE, H. L. (2001): *Asymptotic Theory for Econometricians*, San Diego (CA): Academic Press. 7

Wald Test	Null Hypothesis	Symmetry	Null Limit Distribution
$W_T^{(1)} := \widehat{\boldsymbol{\alpha}}_T' \widehat{\mathbf{R}}_1' \left(\widehat{\mathbf{W}}_T^{(1)} \right)^{-1} \widehat{\mathbf{R}}_1 \widehat{\boldsymbol{\alpha}}_T$	$\mathscr{H}_0' : \boldsymbol{\theta}_*^+ = \boldsymbol{\theta}_*^-$	Long-run	Mixed Chi-squared
$W_T^{(2)} := \widehat{\boldsymbol{\alpha}}_T' \widehat{\mathbf{R}}_2' (\widehat{\mathbf{W}}_T^{(2)})^{-1} \widehat{\mathbf{R}}_2 \widehat{\boldsymbol{\alpha}}_T$	$\mathscr{H}_0'' : \boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-$	Short-run	Mixed Chi-squared
$W_T^{(3)} := \widehat{\boldsymbol{\alpha}}_T' \widehat{\mathbf{R}}_3' (\widehat{\mathbf{W}}_T^{(3)})^{-1} \widehat{\mathbf{R}}_3 \widehat{\boldsymbol{\alpha}}_T$	$\mathscr{H}_0''' : \boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$ and $\boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-$	Long- and Short-run	Mixed Chi-squared
$\mathscr{W}_T^{(1)} := \ddot{\boldsymbol{\upsilon}}_T' \ddot{\mathbf{D}}_1 \ddot{\mathbf{R}}_1' (\ddot{\mathbf{W}}_T^{(1)})^{-1} \ddot{\mathbf{R}}_1 \ddot{\mathbf{D}}_1 \ddot{\boldsymbol{\upsilon}}_T$	$\mathbb{H}_0' : \boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$	Long-run	Nonstandard
$\mathscr{W}_T^{(2)} := \ddot{\boldsymbol{\tau}}_T' \ddot{\mathbf{D}} \widehat{\mathbf{R}}_2' (\ddot{\mathbf{W}}_T^{(2)})^{-1} \widehat{\mathbf{R}}_2 \ddot{\mathbf{D}} \ddot{\boldsymbol{\tau}}_T$	$\mathscr{H}_0'' : \boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-$	Short-run	Mixed Chi-squared
$\mathscr{W}_T^{(3)} := \ddot{\boldsymbol{\delta}}_T' \bar{\mathbf{D}} \ddot{\mathbf{R}}_3' (\ddot{\mathbf{W}}_T^{(3)})^{-1} \ddot{\mathbf{R}}_3 \bar{\mathbf{D}} \ddot{\boldsymbol{\delta}}_T$	$\mathbb{H}_0''' : \boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$ and $\boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-$	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 $\widehat{\boldsymbol{\alpha}}_T$, whereas $\mathscr{W}_T^{(1)}$, $\mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(3)}$ are defined by the primitive estimators, viz., $\ddot{\boldsymbol{\upsilon}}_T$, $\ddot{\boldsymbol{\tau}}_T$, and $\ddot{\boldsymbol{\delta}}_T$, 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
$\mathscr{W}_T^{(1)}$	10%	11.00	9.84	10.56	10.44	10.78	45.62	46.14	47.56	47.86	47.24
	5%	6.06	5.00	5.08	5.00	5.20	37.30	37.92	39.50	39.52	39.14
	1%	1.24	1.00	0.86	1.02	0.86	23.86	24.62	25.40	25.70	26.20
$\mathscr{W}_T^{(2)}$	10%	10.00	10.52	10.02	9.58	9.72	11.66	11.72	10.74	10.12	10.04
	5%	5.10	5.18	5.36	4.88	4.56	6.46	5.80	5.84	5.30	5.00
	1%	1.16	1.18	1.12	0.90	0.78	1.66	1.34	1.12	1.02	1.04
$\mathscr{W}_T^{(3)}$	10%	11.16	10.34	10.74	10.22	9.84	41.52	42.14	43.04	43.18	42.54
	5%	5.90	5.26	5.16	5.18	4.94	32.92	33.60	34.10	34.46	34.52
	1%	1.04	1.10	0.98	0.92	0.96	20.14	20.88	21.36	21.12	22.28
$W_T^{(1)}$	10%	9.34	9.06	10.02	10.42	10.20	14.38	11.58	11.68	11.76	10.98
	5%	4.58	4.52	5.44	5.18	5.18	7.50	6.10	6.42	6.14	5.96
	1%	0.88	0.96	1.02	0.94	1.04	2.18	1.54	1.44	1.46	1.22
$W_T^{(2)}$	10%	9.98	9.90	10.54	9.64	10.20	11.96	10.94	11.14	10.02	10.46
	5%	5.10	5.12	5.38	4.78	5.30	6.46	6.04	5.92	5.10	5.74
	1%	0.82	1.06	1.04	1.00	1.18	1.36	1.48	1.20	1.08	1.08
$W_T^{(3)}$	10%	9.72	9.70	10.62	10.02	10.66	14.22	12.04	12.06	11.36	11.64
	5%	4.82	4.98	5.36	4.88	5.14	8.04	6.42	6.38	5.68	5.98
	1%	0.68	1.08	1.02	1.08	1.04	1.80	1.54	1.38	1.08	1.28

Table 2: EMPIRICAL REJECTION RATES OF THE WALD TESTS (IN PERCENT). This table shows the empirical rejection rates of the Wald statistics testing $\mathbb{H}_0' : \beta_* = 0$, $\mathscr{H}_0'' : \pi_*^+ = \pi_*^-$, and $\mathbb{H}_0''' : \beta_* = 0$ and $\pi_*^+ = \pi_*^-$. The total number of repetitions is 5000, and the bootstrap iteration is 500. DGP: $\Delta y_t = \rho_* u_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t$, $u_t = y_t - \beta_*^+ x_t^+ - \beta_*^- x_t^-$, $\Delta x_t = 1/2 + v_t$, and $(e_t, v_t)' \sim$ IID $N(\mathbf{0}_2, \mathbf{I}_2)$ with $(\rho_*, \pi_*^+, \pi_*^-, \beta_*^+, \beta_*^-) = (-1/2, 1/2, 1/2, 1, 1)$. Here, $\mathscr{W}_T^{(1)}$, $\mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(3)}$ 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
$\mathscr{W}_T^{(1)}$	10%	45.38	87.08	97.72	99.80	99.94	68.44	95.00	99.32	99.96	100.00
	5%	32.98	79.54	96.02	99.46	99.84	61.76	93.48	98.98	99.92	100.00
	1%	13.60	57.06	87.90	97.52	99.48	47.92	88.92	98.12	99.86	99.98
$\mathscr{W}_T^{(2)}$	10%	15.78	24.18	30.30	37.58	44.38	18.64	25.82	31.42	38.80	44.86
	5%	8.98	15.00	20.10	26.82	31.22	11.42	16.78	21.36	27.76	32.44
	1%	2.22	5.16	7.18	10.54	13.30	3.44	6.18	8.28	11.68	14.54
$\mathscr{W}_T^{(3)}$	10%	43.56	86.04	97.50	99.76	99.94	64.26	94.04	99.00	99.94	100.00
	5%	31.44	77.96	95.78	99.42	99.78	57.12	91.96	98.50	99.90	100.00
	1%	12.88	56.40	87.02	97.46	99.48	43.80	86.34	97.54	99.76	99.96
$W_T^{(1)}$	10%	42.88	83.22	97.46	99.54	99.98	53.02	86.30	97.92	99.60	99.98
	5%	30.50	75.20	95.18	99.12	99.90	41.82	79.88	96.34	99.20	99.94
	1%	12.14	53.98	87.00	97.34	99.58	22.92	63.56	90.38	98.12	99.62
$W_T^{(2)}$	10%	16.68	23.28	31.82	38.08	44.70	19.70	24.96	32.98	39.06	45.66
	5%	9.82	14.52	21.38	26.56	32.62	12.34	16.00	22.72	27.98	33.76
	1%	2.60	3.98	7.48	10.82	13.82	4.00	5.08	8.42	12.16	15.08
$W_T^{(3)}$	10%	38.04	78.18	96.12	99.22	99.94	47.26	81.52	96.64	99.30	99.96
	5%	26.50	69.44	93.50	98.74	99.80	36.14	74.16	94.54	98.94	99.88
	1%	9.80	47.46	83.48	96.32	99.30	18.66	56.34	86.96	97.16	99.46

Table 3: EMPIRICAL REJECTION RATES OF THE WALD TESTS (IN PERCENT). This table shows the empirical rejection rates of the Wald statistics testing $\mathbb{H}_0' : \beta_* = 0$, $\mathscr{H}_0'' : \pi_*^+ = \pi_*^-$, and $\mathbb{H}_0''' : \beta_* = 0$ and $\pi_*^+ = \pi_*^-$. The total number of repetitions is 5000, and the bootstrap iteration is 500. DGP: $\Delta y_t = \rho_* u_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t$, $u_t = y_t - \beta_*^+ x_t^+ - \beta_*^- x_t^-$, $\Delta x_t = 1/2 + v_t$, and $(e_t, v_t)' \sim$ IID $N(\mathbf{0}_2, \mathbf{I}_2)$ with $(\rho_*, \pi_*^+, \pi_*^-, \beta_*^+, \beta_*^-) = (-1/2, 1/8, -1/8, 1/4, -1/4)$. Here, $\mathscr{W}_T^{(1)}$, $\mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(3)}$ 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
$\mathscr{W}_T^{(1)}$	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
$\mathscr{W}_T^{(2)}$	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
$\mathscr{W}_T^{(3)}$	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 4: EMPIRICAL REJECTION RATES OF THE WALD TESTS (IN PERCENT). This table shows the empirical rejection rates of the Wald statistics testing $\mathbb{H}_0' : \beta_* = 0$, $\mathscr{H}_0'' : \pi_*^+ = \pi_*^-$, 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 $(e_t, v_t)' \sim$ 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, $\mathscr{W}_T^{(1)}$, $\mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(3)}$ 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
$\mathscr{W}_T^{(1)}$	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
$\mathscr{W}_T^{(2)}$	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
$\mathscr{W}_T^{(3)}$	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 5: EMPIRICAL REJECTION RATES OF THE WALD TESTS (IN PERCENT). This table shows the empirical rejection rates of the Wald statistics testing $\mathbb{H}_0' : \beta_* = 0$, $\mathscr{H}_0'' : \pi_*^+ = \pi_*^-$, 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 $(e_t, v_t)' \sim$ 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, $\mathscr{W}_T^{(1)}$, $\mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(3)}$ 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
$\mathscr{W}_T^{(1)}$					9.8862	0.0048	9.8910
					(0.4344)	(0.9696)	(0.5000)
$\mathscr{W}_T^{(2)}$					1.4822	3.8085	5.2908
					(0.7472)	(0.1902)	(0.6567)
$\mathscr{W}_T^{(3)}$					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 6: 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	3.7413***	4.9730**	Constant	6.0785***	4.8528***	5.8407**
		(2.5919)	(1.2787)	(2.3861)		(0.0709)	(1.3852)	(2.8526)
	τ^{+}_{t-1}	-0.4329***	0.2715***	0.2759	τ_{t-1}	-0.2328***	0.2523***	0.1564
		(0.13189)	(0.0546)	(0.1723)		(0.0709)	(0.0591)	(0.1716)
	τ^{-}_{t-1}	-0.3202*	0.3276*	0.2598				
		(0.1722)	(0.1757)	(0.1826)				
	Trend	0.8340***	0.8461***	0.8436***	Trend	0.8261***	0.8367***	0.8287***
		(0.0184)	(0.0090)	(0.0169)		(0.0127)	(0.0098)	(0.0202)
Short-Run	u_{t-1}	-0.0683*	-0.0817*	-0.0497†	u_{t-1}	-0.0708**	-0.0755*	-0.0404
		(0.0191)	(0.0223)	(0.0158)		(0.0185)	(0.0220)	(0.0148)
	Constant	0.6734***	0.6918***	0.5683***	Constant	0.5752***	0.6590***	0.6295***
		(0.1424)	(0.1458)	(0.1472)		(0.5752)	(0.1392)	(0.1404)
	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*	0.0398	$\Delta\tau_{t}$	-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)				

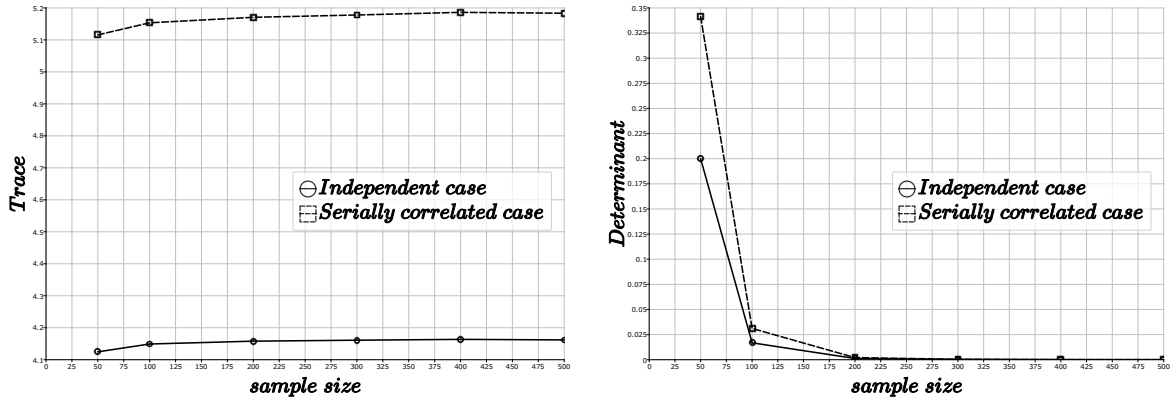
Table 7: 2SNARDL ESTIMATION OF THE NARDL AND ARDL MODELS. This table presents the 2SNARDL estimation using the quarterly data from Romer and Romer (2010). The left and right panels display estimated parameters for (22) and (20), respectively. †, *, **, and *** imply that the tests are significant at 25%, 10%, 5%, and 1%, respectively.

NARDL Model				ARDL Model			
Variables \ Tax	Exo. ratio	Endo. ratio	Sum. ratio	Variables \ Tax	Exo. ratio	Endo. ratio	Sum. ratio
y_{t-1}	-0.0609†	-0.0633	-0.0460	y_{t-1}	-0.0552†	-0.0658	-0.0400
	(0.0184)	(0.0251)	(0.0177)		(0.0163)	(0.0247)	(0.0160)
r^{+}_{t-1}	-0.1992	0.0850	0.0037	r_{t-1}	-0.1125	0.1003	-0.0351
	(0.1349)	(0.0954)	(0.0789)		(0.0743)	(0.0925)	(0.0618)
r^{-}_{t-1}	-0.0970	0.0401	-0.0566				
	(0.0801)	(0.1345)	(0.0645)				
Trend	0.0499***	0.0488**	0.0324**	Trend	0.0419***	0.0520***	0.0317***
	(0.0173)	(0.0200)	(0.0136)		(0.0127)	(0.0194)	(0.0135)
Constant	0.2683	0.3673	0.3685*	Constant	0.4054**	0.4700***	0.5261***
	(0.2363)	(0.2380)	(0.2047)		(0.1787)	(0.1663)	(0.1543)
Δy_{t-1}	0.3095***	0.3144***	0.3049***	Δy_{t-1}	0.3042***	0.3115***	0.2921***
	(0.0639)	(0.0672)	(0.0657)		(0.0636)	(0.0660)	(0.0650)
Δy_{t-2}	0.1223*	0.1265**	0.1139*	Δy_{t-2}	0.1168*	0.1242*	0.1073*
	(0.0652)	(0.0660)	(0.0651)		(0.0644)	(0.0657)	(0.0648)
Δr^{+}_{t}	0.1638	0.2465	0.1778	Δr_{t}	0.1725	0.2553	0.1467
	(0.6215)	(0.2788)	(0.2621)		(0.618)0	(0.2776)	(0.2609)
Δr^{-}_{t}	-0.2706	-0.1477	-0.2350				
	(0.2734)	(0.3051)	(0.2087)				
AIC	2.6831	2.6925	2.6895	AIC	2.6739	2.6780	2.6814
BIC	2.8133	2.8226	2.8196	BIC	2.7751	2.7792	2.7826
t-test	-3.3087†	-2.5198	-2.5917	t-test	-3.3867†	-2.6602	-2.4960
F-test	3.9276	3.5933	3.4873	F-test	5.7417	5.2684	4.8210
$\mathscr{W}^{(1)}_{T}$					52.7574	2.1358	48.9740
					(0.1294)	(0.7290)	(0.2581)
$\mathscr{W}^{(2)}_{T}$					0.0722	0.8121	0.4782
					(0.8015)	(0.4021)	(0.4981)
$\mathscr{W}^{(3)}_{T}$					52.8296	2.9480	49.4522
					(0.1340)	(0.7595)	(0.2610)
$W^{(1)}_{T}$					0.5384	0.3344	0.8601
					(0.5697)	(0.6463)	(0.5035)
$W^{(2)}_{T}$					0.4079	0.9926	1.5118
					(0.5278)	(0.3177)	(0.2279)
$W^{(3)}_{T}$					1.0869	1.2200	2.1652
					(0.6618)	(0.6150)	(0.4681)

Table 8: 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) using r_t instead of τ_t, 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. ratio	Endo. ratio	Sum. ratio	Variables \ Tax	Exo. ratio	Endo. ratio	Sum. ratio
Long-Run	Constant	-2.5602	-2.5888**	-2.9531	Constant	-0.3468	-2.2311*	-1.0492
		(2.4603)	(1.1990)	(2.5202)		(2.3897)	(1.2927)	(2.8199)
	r^+_{t-1}	-6.7407***	4.2616***	4.5455***	r_{t-1}	-3.0752**	3.3960***	2.2608*
		(2.0794)	(0.5100)	(1.2686)		(1.2428)	(0.5463)	(1.2356)
	r^-_{t-1}	-2.8360*	5.1176***	2.762**3				
		(1.5211)	(0.9583)	(1.2373)				
	Trend	0.8364***	0.8391***	0.8435***	Trend	0.8242***	0.8370***	0.8295***
		(0.0166)	(0.0081)	(0.0170)		(0.0161)	(0.0087)	(0.0190)
Short-Run	u_{t-1}	-0.0609†	-0.0633	-0.0460	u_{t-1}	-0.0558*	-0.0627	-0.0378
		(0.0184)	(0.0251)	(0.0177)		(0.0162)	(0.0248)	(0.0160)
	Constant	0.6763***	0.6693***	0.6719***	Constant	0.5730***	0.6828***	0.6517***
		(0.1484)	(0.1510)	(0.1534)		(0.1470)	(0.1468)	(0.1464)
	Δy_{t-1}	0.3095***	0.3144***	0.3049***	Δy_{t-1}	0.3038***	0.3127***	0.3007***
		(0.0639)	(0.0672)	(0.0657)		(0.0635)	(0.0669)	(0.0656)
	Δy_{t-2}	0.1223*	0.1265*	0.1139*	Δy_{t-2}	0.1149*	0.1220*	0.1070
		(0.0652)	(0.0660)	(0.0651)		(0.0644)	(0.0658)	(0.0648)
	Δr^+_t	0.1638	0.2465	0.1778	Δr_t	-0.2094	0.0801	-0.0737
		(0.6215)	(0.2788)	(0.2621)		(0.2435)	(0.2087)	(0.1596)
	Δr^-_t	-0.2706	-0.1477	-0.2350				
		(0.2734)	(0.3051)	(0.2087)				

Table 9: 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: TRACE AND DETERMINANT OF $\mathbf{D}^{-1}(\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t')\mathbf{D}^{-1}$. Figure 1 (a) shows the average of the traces of $\mathbf{D}^{-1}(\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t')\mathbf{D}^{-1}$ that are obtained by 10000 independent experiments under the DGP conditions given in Tables 3 and 5. Figure 1 (b) shows the average of the determinants $\times 10^{12}$ of the same matrices.

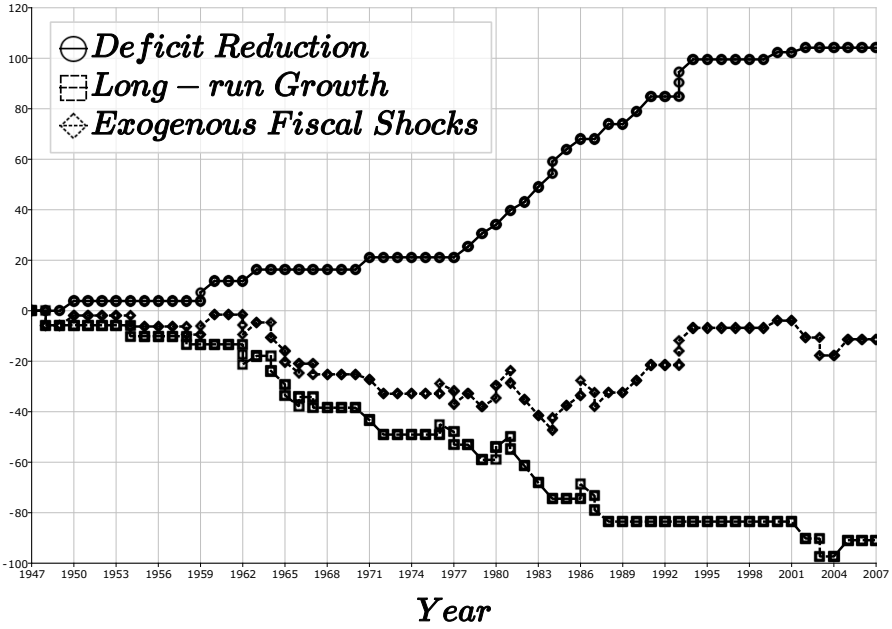


Figure 2: PARTIAL SUM PROCESS FORMED BY EXOGENOUS FISCAL SHOCKS. The solid, dashed, and dotted lines represent deficit reduction, long-run growth, and exogenous fiscal shocks, respectively.