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

JIN SEO CHO

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

jinseocho@yonsei.ac.kr

December 2024

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. We also examine the large-sample behavior of the standard Wald test, as defined by the OLS estimator, for asymmetries in long- and short-run NARDL parameters, and enhance this analysis with a super-consistent long-run parameter estimator. We then confirm the theory on the Wald test using Monte Carlo simulations. Finally, using U.S. GDP and exogenous fiscal shock data, we 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.

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. Although the long-run parameter estimator is not super-consistent, as it is in the two-step NARDL (2SNARDL) estimation proposed by Cho et al. (2024), this feature validates 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) explain and resolve this problem by using a two-step procedure for estimating the NARDL parameter, termed 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, unlike OLS. Here, we provide theoretical grounds to compare these two estimators and validate the use of OLS asymptotically.

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, which 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 convergence rate of OLS is slower than that of 2SNARDL.

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,

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 can 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.

This 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 4 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.

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 drift:

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

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

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

$\{e_t, \mathscr{F}_t\}$ 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. The NARDL process in (1) is more general than those defined by Shin et al. (2014) and Cho, Greenwood-Nimmo, and Shin (2023b), because the latter do not allow for a time trend on the right side. 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 $\rho_*, \boldsymbol{\theta}_*^+, \boldsymbol{\theta}_*^-, \varphi_{j*}$ $(j = 1, 2, \ldots, p-1)$, $\boldsymbol{\pi}_{j*}^+$, and $\boldsymbol{\pi}_{j*}^-$ $(j = 0, 1, \ldots, q-1)$ are defined by the parameters in (1), as given in Cho et al. (2023b). For notational simplicity, we further rewrite this as

$$\Delta y_t = \rho_* y_{t-1} + \boldsymbol{\theta}_*^{\prime} \ddot{\boldsymbol{x}}_{t-1} + \xi_*(t-1) + \alpha_* + \boldsymbol{\varphi}_* \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\pi}_*^{\prime} \Delta \widetilde{\boldsymbol{x}}_t + e_t \tag{2}$$

by letting

$$\boldsymbol{\theta}_* := [\boldsymbol{\theta}_*^{+\prime}, \boldsymbol{\theta}_*^{-\prime}]^{\prime}, \quad \ddot{\boldsymbol{x}}_{t-1} := [\boldsymbol{x}_{t-1}^{+\prime}, \boldsymbol{x}_{t-1}^{-\prime}]^{\prime},$$

$$\boldsymbol{\varphi}_* := [\varphi_{1*}, \varphi_{2*}, \ldots, \varphi_{p-1*}]^{\prime}, \quad \Delta \boldsymbol{y}_{t-1} := [\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1}]^{\prime},$$

$$\boldsymbol{\pi}_* := [\boldsymbol{\pi}_*^{+\prime}, \quad \boldsymbol{\pi}_*^{-\prime}]^{\prime}, \quad \boldsymbol{\pi}_*^+ := [\boldsymbol{\pi}_{0*}^{+\prime}, \boldsymbol{\pi}_{1*}^{+\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{+\prime}]^{\prime}, \quad \boldsymbol{\pi}_*^- := [\boldsymbol{\pi}_{0*}^{-\prime}, \boldsymbol{\pi}_{1*}^{-\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{-\prime}]^{\prime},$$

$$\Delta \widetilde{\boldsymbol{x}}_t := [\Delta \widetilde{\boldsymbol{x}}_t^{+\prime}, \Delta \widetilde{\boldsymbol{x}}_t^{-\prime}]^{\prime}, \quad \Delta \widetilde{\boldsymbol{x}}_t^+ := [\Delta \boldsymbol{x}_t^{+\prime}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{+\prime}]^{\prime}, \quad \text{and} \quad \Delta \widetilde{\boldsymbol{x}}_t^- := [\Delta \boldsymbol{x}_t^{-\prime}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{-\prime}]^{\prime}.$$

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

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \boldsymbol{\varphi}_* \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\pi}_*^{\prime} \Delta \widetilde{\boldsymbol{x}}_t + e_t, \tag{3}$$

where $\gamma_* := \alpha_* + \xi_* + \rho_* \nu_*$, such that the cointegration error is defined as $u_{t-1} := y_{t-1} - \boldsymbol{\beta}_*^{\prime} \ddot{\boldsymbol{x}}_{t-1} - \zeta_*(t-1) - \nu_*$, with

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

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 $\Delta \widetilde{\boldsymbol{x}}_t$.

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

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

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

$s_t^- := \Delta x_t^- - \mu_*^-$, it follows that

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

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

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

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 drift. This cointegrating relationship is more general than that assumed by a linear cointegration. By imposing $\beta_*^+ = \beta_*^-$, the error-correction form in (3) reduces to the linear cointegrating relationship between y_t and x_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 because it suffers from an asymptotically singular matrix problem. For this examination, we first let

$$z_t := [\ z_{1t}' \ \vdots \ z_{2t}' \ ]' := [\ y_{t-1} \ \ \ddot{x}_{t-1}' \ \ (t-1) \ \ 1 \ \vdots \ \Delta y_{t-1}' \ \ \Delta \tilde{x}_t' \ ]'.$$

Here, $z_t \in \mathbb{R}^{2+p+2k(1+q)}$ is partitioned into two variables such that $z_{1t} \in \mathbb{R}^{3+2k}$ and $z_{2t} \in \mathbb{R}^{p+2kq-1}$ collect the variables in the long- and short-run equations, respectively. Next, we let

$$\alpha_* := [\ \alpha_{1*}' \ \vdots \ \alpha_{2*}' \ ]' := [\ \rho_* \ \ \theta_*' \ \ \xi_* \ \ \alpha_* \ \vdots \ \varphi_*' \ \ \pi_*' \ ]',$$

From this, the OLS estimator is written as

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

For later purposes, we partition $\widehat{\theta}_T$ and $\widehat{\pi}_T$ into $[\widehat{\theta}_T^{+\prime}, \widehat{\theta}_T^{-\prime}]'$ and $[\widehat{\pi}_T^{+\prime}, \widehat{\pi}_T^{-\prime}]'$, respectively, such that $\widehat{\theta}_T^+$ and $\widehat{\theta}_T^- \in \mathbb{R}^k$, and $\widehat{\pi}_T^+$ and $\widehat{\pi}_T^- \in \mathbb{R}^{kq}$. Even under mild regularity conditions, the OLS estimator suffers from an asymptotically singular matrix problem. For this discussion, we impose the following conditions, which

5

are standard in the literature:

Assumption 1. *(i)* $\{(\Delta \boldsymbol{x}_t', u_t, e_t)' \in \mathbb{R}^{k+2} : t = \ldots, -1, 0, 1, \ldots\}$ *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*

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

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.* $\square$

Remarks. **(a)** Under the same assumption, we below investigate the limit behavior of the OLS estimator.

(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 $\boldsymbol{\beta}_* = \boldsymbol{0}$ 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 $(\boldsymbol{\beta}_*', \zeta_*)' = \boldsymbol{0}$. 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 to the context of NARDL, because assuming no cointegration under the condition of linearity is equivalent to assuming $\boldsymbol{\beta}_* = \boldsymbol{0}$ and $\zeta_* = 0$.

(e) Assumption 1 (*v*) is a standard condition for the error term in the autoregressive distributed lag (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) *if* $\mathbf{D}_1 := diag[T^{3/2}\mathbf{I}_{2+2k}, T^{1/2}]$, $\mathbf{D}_1^{-1}\left(\sum_{t=1}^{T}\boldsymbol{z}_{1t}\boldsymbol{z}_{1t}'\right)\mathbf{D}_1^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_{11}$, *where*

$$\mathbf{M}_{11} := \begin{bmatrix} \frac{1}{3}\delta_*^2 & \frac{1}{3}\delta_*\boldsymbol{\mu}_*' & \frac{1}{3}\delta_* & \frac{1}{2}\delta_* \\ \frac{1}{3}\delta_*\boldsymbol{\mu}_* & \frac{1}{3}\boldsymbol{\mu}_*\boldsymbol{\mu}_*' & \frac{1}{3}\boldsymbol{\mu}_* & \frac{1}{2}\boldsymbol{\mu}_* \\ \frac{1}{3}\delta_* & \frac{1}{3}\boldsymbol{\mu}_*' & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2}\delta_* & \frac{1}{2}\boldsymbol{\mu}_*' & \frac{1}{2} & 1 \end{bmatrix};$$

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

(iii) $\mathbf{D}_2^{-1}\left(\sum_{t=1}^{T}\boldsymbol{z}_{2t}\boldsymbol{z}_{1t}'\right)\mathbf{D}_1^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_{21}$, *where*

$$\mathbf{M}_{21} := \begin{bmatrix} \frac{1}{2}\delta_*^2\boldsymbol{\iota}_{p-1} & \frac{1}{2}\delta_*\boldsymbol{\iota}_{p-1}\boldsymbol{\mu}_*' & \frac{1}{2}\delta_*\boldsymbol{\iota}_{p-1} & \delta_*\boldsymbol{\iota}_{p-1} \\ \frac{1}{2}\delta_*\boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_* & \frac{1}{2}\boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_*\boldsymbol{\mu}_*' & \frac{1}{2}\boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_* & \boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_* \end{bmatrix};$$

(iv) *if* $\mathbf{D} := diag[\mathbf{D}_1, \mathbf{D}_2]$ *and* $\mathbf{M}_{12} := \mathbf{M}_{21}'$, $\mathbf{D}^{-1}\left(\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t'\right)\mathbf{D}^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}$, *where*

$$\mathbf{M} := \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix},$$

which is singular. □

Although the OLS estimator $\widehat{\boldsymbol{\alpha}}_T := (\sum_{t=1}^{T}\boldsymbol{z}_{1t}\boldsymbol{z}_{1t}')^{-1}\sum_{t=1}^{T}\boldsymbol{z}_{1t}\Delta y_t$ is popular for empirical analyses, Lemma 1 and its definition now clearly show 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. 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.

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. Thus, it is necessary to apply the higher-order expansion to OLS itself. However, given that the dimension of $(\sum_{t=1}^{T} z_t z_t')^{-1}$ 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 $\widehat{\alpha}_T$ to be expressed as a bilinear transformation of the long- and short-run parameter estimators. Finally, we express the long- and short-run parameter estimators as other bilinear transformations of primitive estimators that do not suffer from a singular matrix problem, enabling $\widehat{\alpha}_T$ 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

$$y_{t-1} = \beta_*' \ddot{x}_{t-1} + \zeta_*(t-1) + \nu_* + u_{t-1}, \tag{6}$$

we let the long-run parameter estimator be

$$\widehat{v}_T := [\widehat{\beta}_T', \widehat{\zeta}_T, \widehat{\nu}_T]' := \left(\sum_{t=1}^{T} \widetilde{r}_{t-1} \widetilde{r}_{t-1}' \right)^{-1} \sum_{t=1}^{T} \widetilde{r}_{t-1} y_{t-1}, \;\; \text{where } \widetilde{r}_{t-1} := [\ddot{x}_{t-1}', t-1, 1]',$$

estimating the long-run parameter $\boldsymbol{\upsilon}_* := [\boldsymbol{\beta}'_*, \zeta_*, \nu_*]'$. For later purposes, we let $\widehat{u}_{t-1} := y_{t-1} - \widetilde{\boldsymbol{r}}'_{t-1}\widehat{\boldsymbol{\upsilon}}_T$ and $\widehat{\boldsymbol{\beta}}_T \equiv [\widehat{\boldsymbol{\beta}}^{+\prime}_T, \widehat{\boldsymbol{\beta}}^{-\prime}_T]'$, such that $\widehat{\boldsymbol{\beta}}^+_T$ and $\widehat{\boldsymbol{\beta}}^-_T \in \mathbb{R}^k$.

Second Step   We specify the error-correction model

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

by combining (2) and (6), where $\boldsymbol{\eta}_* := \boldsymbol{\theta}_* + \rho_*\boldsymbol{\beta}_*$ and $\psi_* := \xi_* + \rho_*\zeta_*$. Then, we estimate the parameters in (7) by

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

estimating the short-run parameter $\boldsymbol{\omega}_* := [\ \rho_* \mid \boldsymbol{\eta}'_* \quad \psi_* \quad \gamma_* \mid \boldsymbol{\varphi}'_* \quad \boldsymbol{\pi}'_*\ ]'$. Note that $\widetilde{\boldsymbol{\omega}}_T$ is obtained by replacing u_{t-1} in (7) with $\widehat{u}_{t-1}$. For later purposes, we also partition $\boldsymbol{\eta}_*$ into $[\boldsymbol{\eta}^{+\prime}_*, \boldsymbol{\eta}^{-\prime}_*]'$, such that $\boldsymbol{\eta}^{+\prime}_*$ and $\boldsymbol{\eta}^{-\prime}_* \in \mathbb{R}^k$.

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 $\boldsymbol{\eta}_* = \boldsymbol{0}$ and $\psi_* = 0$. Even so, we estimate them through the short-run parameter estimator $\widetilde{\boldsymbol{\omega}}_T$ by including $\ddot{\boldsymbol{x}}_{t-1}$ and $t-1$ as auxiliary regressors. The purpose of this inclusion is evident in the following proposition:

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

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

That is, we can represent the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ as a bilinear transformation between the long- and short-run parameter estimators, $\widehat{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$. We obtain Proposition 1 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 $[\widetilde{\boldsymbol{r}}'_{t-1}, \boldsymbol{z}'_{2t}]'$ of (7), respectively, Proposition 1 follows. Here, we can let $\widehat{\boldsymbol{\upsilon}}_t$ in Lemma A.1 be $\widehat{u}_{t-1}$.

Third Step   Although Proposition 1 represents the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ as a bilinear transformation between the long- and short-run parameter estimators, both $\widehat{\boldsymbol{\upsilon}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$ suffer from an asymptotically singular matrix problem; see Lemma A.3 in the online supplement. Therefore, we again represent both estimators using other estimators that do not suffer from an asymptotically singular matrix problem.

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

9

(6) imply that

$$y_t = \boldsymbol{\beta}'_* \mathbf{m}_t + \vartheta_* t + \nu_* + u_t, \quad \text{where} \quad \vartheta_* := \boldsymbol{\beta}'_* \boldsymbol{\mu}_* + \zeta_*. \tag{8}$$

Therefore, if $\mathbf{m}_t$ is observed, we can estimate the coefficients in (8) by regressing y_t against $[\mathbf{m}'_t, t, 1]$. However, $\mathbf{m}_t$ is not. We instead predict $\mathbf{m}_t$ first, so that we can use its predictor as a regressor. For this, we estimate the unknown parameters $\boldsymbol{\mu}_*$ in (4) by

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

We use $\widehat{\mathbf{m}}_t$ as a regressor to estimate the coefficients in (8). That is, we let $\dot{\boldsymbol{r}}_{t-1} := [\widehat{\mathbf{m}}'_{t-1}, t - 1, 1]'$ and regress y_{t-1} against $\dot{\boldsymbol{r}}_{t-1}$, to obtain

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

For later purposes, we let $\widetilde{\boldsymbol{\beta}}_T \equiv [\widetilde{\boldsymbol{\beta}}_T^{+\prime}, \widetilde{\boldsymbol{\beta}}_T^{-\prime}]'$, such that $\widetilde{\boldsymbol{\beta}}_T^+$ and $\widetilde{\boldsymbol{\beta}}_T^- \in \mathbb{R}^k$.

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

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

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

That is, the long-run parameter estimator $\widehat{\boldsymbol{v}}_T$ is represented as a bilinear transformation between the primitive estimators, $\widehat{\boldsymbol{\mu}}_T$ and $\widetilde{\boldsymbol{v}}_T$. This representation is particularly useful, because neither $\widehat{\boldsymbol{\mu}}_T$ nor $\widetilde{\boldsymbol{v}}_T$ suffers from an asymptotically singular matrix problem, as Lemma 2 verifies below. 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. Specifically, we combine (4) and (7), obtaining

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

Then, we regress Δy_t against $\dot{\boldsymbol{z}}_t := [\; \widehat{u}_{t-1} \;\vdots\; \dot{\boldsymbol{r}}'_{t-1} \;\vdots\; \boldsymbol{z}'_{2t} \;]'$ to estimate the unknown parameters in (9) by

OLS. For this, u_{t-1} and $\mathbf{m}_{t-1}$ in (9) need to be replaced with $\widehat{u}_{t-1}$ and $\widehat{\mathbf{m}}_{t-1}$, respectively. The parameter estimator is as follows:

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

estimating $\boldsymbol{\tau}_* := [\ \rho_* \ \vdots \ \boldsymbol{\tau}_{1*}' \ \vdots \ \boldsymbol{\tau}_{2*}' \ ]' := [\ \rho_* \ \vdots \ \boldsymbol{\eta}_*' \ \ \varsigma_* \ \ \gamma_* \ \vdots \ \boldsymbol{\varphi}_*' \ \ \boldsymbol{\pi}_*' \ ]'$. Here, for later purposes, we partitioned the parameter $\boldsymbol{\tau}_*$ and its estimator $\dot{\boldsymbol{\tau}}_T$ into the smaller sets, and we let $\dot{\boldsymbol{\eta}}_T \equiv [\dot{\boldsymbol{\eta}}_T^{+\prime}, \dot{\boldsymbol{\eta}}_T^{-\prime}]'$, such that $\dot{\boldsymbol{\eta}}_T^+$ and $\dot{\boldsymbol{\eta}}_T^- \in \mathbb{R}^k$. We also let $\dot{\boldsymbol{\pi}}_T \equiv [\dot{\boldsymbol{\pi}}_T^{+\prime}, \dot{\boldsymbol{\pi}}_T^{-\prime}]'$, such that $\dot{\boldsymbol{\pi}}_T$ is the vector formed by the last $2kq$ elements of $\dot{\boldsymbol{\tau}}_{2T}$, and $\dot{\boldsymbol{\pi}}_T^{+\prime}$ and $\dot{\boldsymbol{\pi}}_T^{-\prime} \in \mathbb{R}^{kq}$. In addition, we separately estimate ψ_* by $\dot{\psi}_T := -\widehat{\boldsymbol{\mu}}_T' \dot{\boldsymbol{\eta}}_T + \dot{\varsigma}_T$ using the definition of ς_*.

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

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

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

Therefore, the short-run parameter estimator $\widetilde{\boldsymbol{\omega}}_T$ becomes a bilinear transformation between the primitive estimators, $\widehat{\boldsymbol{\mu}}_T$ and $\dot{\boldsymbol{\tau}}_T$, neither of which suffers from an asymptotically singular matrix problem, as Lemma 2 shows. Proposition 3 is again established by Lemma A.2 in the online supplement. By letting $\ddot{\boldsymbol{x}}_{t-1}$, $(t-1)$, and $[\widehat{u}_{t-1}, 1, \boldsymbol{z}_{2t}']'$ be $\boldsymbol{x}_t$, $\boldsymbol{z}_t$, and $\boldsymbol{w}_t$, respectively, of Lemma A.2, Proposition 3 follows. Here, $\mathbf{Q}_T \dot{\boldsymbol{\tau}}_T$ and $\dot{\boldsymbol{\tau}}_T$ are almost identical, from the definition of $\mathbf{Q}_T$. The only difference is in their $(2+2k)$-th rows. The $(2+2k)$-th row element of $\mathbf{Q}_T \dot{\boldsymbol{\tau}}_T$ is equal to $\dot{\psi}_T := -\widehat{\boldsymbol{\mu}}_T' \dot{\boldsymbol{\eta}}_T + \dot{\varsigma}_T$, whereas the corresponding element of $\dot{\boldsymbol{\tau}}_T$ is $\dot{\varsigma}_T$. Furthermore, if we let $\widetilde{\psi}_T$ denote the $(2+2k)$-th row of the short-run parameter estimator $\widetilde{\boldsymbol{\omega}}_T$, it follows from Proposition 3 that $\widetilde{\psi}_T = \dot{\psi}_T$.

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

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

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

11

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

Proposition 4 follows simply by combining Propositions 1 and 3. The individual elements of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ can be rewritten as follows:

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

Remarks. **(a)** If the long-run parameter estimator $\widehat{\boldsymbol{v}}_T$ and the primitive estimator $\dot{\boldsymbol{\tau}}_T$ are consistent for their target parameters, $\boldsymbol{v}_*$ and $\boldsymbol{\tau}_*$, respectively, then

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

because $\boldsymbol{\eta}_* = \mathbf{0}$. In addition, the final limit is identical to $[\rho_*, \boldsymbol{\theta}_*', \xi_*, \alpha_*, \boldsymbol{\varphi}_*', \boldsymbol{\pi}_*']'$, by the definitions of $\boldsymbol{\theta}_*, \xi_*, \alpha_*,$ and $\boldsymbol{\tau}_{2*}$. From this, the OLS estimator is consistent for its target parameter $\boldsymbol{\alpha}_*$.

(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: $\widehat{\boldsymbol{\mu}}_T$, $\widetilde{\boldsymbol{v}}_T$, and $\dot{\boldsymbol{\tau}}_T$; 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 $\widehat{\boldsymbol{\theta}}_T = -\dot{\rho}_T\widehat{\boldsymbol{\beta}}_T + \dot{\boldsymbol{\eta}}_T$ from (10), and the convergence rates of $\dot{\rho}_T$ and $(\widehat{\boldsymbol{\beta}}_T, \dot{\boldsymbol{\eta}}_T)$ are $T^{1/2}$ and T, respectively, by Lemmas 3 and 4. From this, the convergence rate of the long-run OLS estimator $\widehat{\boldsymbol{\theta}}_T$ is determined as $T^{1/2}$, which is slower than $T^{3/2}$ given by $\mathbf{D}$. Similar arguments apply to the other elements of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$, making its convergence rate slower than $\mathbf{D}$. $\qquad \square$

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

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

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 $\boldsymbol{\mathscr{B}}(\cdot)$:

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

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

The following lemma shows that neither $\dot{\boldsymbol{\tau}}_T$ nor $\widetilde{\boldsymbol{v}}_T$ suffers from an asymptotically singular matrix problem. As such, the long- and short-run parameter estimators, $\widehat{\boldsymbol{v}}_T$ 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) if $\dot{\mathbf{D}}_1 := diag[T\mathbf{I}_{2k}, T^{3/2}, T^{1/2}]$, $T^{-1/2} \left(\sum_{t=1}^T \widehat{u}_{t-1}\dot{\boldsymbol{r}}_{t-1} \right) \dot{\mathbf{D}}_1^{-1} \Rightarrow \dot{\mathbf{M}}_{1u} := \mathbf{0}_{(2k+2)\times 1}$;

(ii) $\dot{\mathbf{D}}_1^{-1} \left(\sum_{t=1}^T \dot{\boldsymbol{r}}_{t-1}\dot{\boldsymbol{r}}_{t-1}' \right) \dot{\mathbf{D}}_1^{-1} \Rightarrow \dot{\boldsymbol{\mathscr{M}}}_{11}$, where

$$\dot{\boldsymbol{\mathscr{M}}}_{11} := \begin{bmatrix} \int \bar{\boldsymbol{\mathscr{B}}}_m \bar{\boldsymbol{\mathscr{B}}}_m' & \mathbf{0}_{k\times 1} & \int \bar{\boldsymbol{\mathscr{B}}}_m \\ \mathbf{0}_{1\times k} & \frac{1}{3} & \frac{1}{2} \\ \int \bar{\boldsymbol{\mathscr{B}}}_m' & \frac{1}{2} & 1 \end{bmatrix}$$

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

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

(iv) $\dot{\mathbf{D}}_2^{-1} \left(\sum_{t=1}^T \boldsymbol{z}_{2t}\dot{\boldsymbol{r}}_{t-1}' \right) \dot{\mathbf{D}}_1^{-1} \Rightarrow \dot{\boldsymbol{\mathscr{M}}}_{21}$, where

$$\dot{\boldsymbol{\mathscr{M}}}_{21} := \begin{bmatrix} \delta_*\boldsymbol{\iota}_{p-1} \int \bar{\boldsymbol{\mathscr{B}}}_m' & \frac{1}{2}\delta_*\boldsymbol{\iota}_{p-1} & \delta_*\boldsymbol{\iota}_{p-1} \\ \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* \int \bar{\boldsymbol{\mathscr{B}}}_m' & \frac{1}{2}\boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* & \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* \end{bmatrix};$$

(v) $\dot{\mathbf{D}}_2^{-1} \left(\sum_{t=1}^T \boldsymbol{z}_{2t}\boldsymbol{z}_{2t}' \right) \dot{\mathbf{D}}_2^{-1} \overset{\mathbb{P}}{\to} \dot{\mathbf{M}}_{22} := \mathbf{M}_{22}$;

(vi) if $\dot{\mathbf{D}} := diag[T^{1/2}, \dot{\mathbf{D}}_1, \dot{\mathbf{D}}_2]$, $\dot{\mathbf{D}}^{-1} \left(\sum_{t=1}^T \dot{\boldsymbol{z}}_t\dot{\boldsymbol{z}}_t' \right) \dot{\mathbf{D}}^{-1} \Rightarrow \dot{\boldsymbol{\mathscr{M}}}$, where

$$\dot{\boldsymbol{\mathscr{M}}} := \begin{bmatrix} \sigma_u^2 & \dot{\mathbf{M}}_{u1} & \dot{\mathbf{M}}_{u2} \\ \dot{\mathbf{M}}_{1u} & \dot{\boldsymbol{\mathscr{M}}}_{11} & \dot{\boldsymbol{\mathscr{M}}}_{12} \\ \dot{\mathbf{M}}_{2u} & \dot{\boldsymbol{\mathscr{M}}}_{21} & \dot{\mathbf{M}}_{22} \end{bmatrix},$$

$\dot{\boldsymbol{\mathscr{M}}}_{12} := \dot{\boldsymbol{\mathscr{M}}}_{21}'$, $\dot{\mathbf{M}}_{u1} := \dot{\mathbf{M}}_{1u}'$, $\dot{\mathbf{M}}_{u2} := \dot{\mathbf{M}}_{2u}'$, and $\sigma_u^2 := \mathbb{E}[u_t^2]$; and

(vii) $\dot{\mathbf{D}}_1^{-1} \left(\sum_{t=1}^T \dot{\boldsymbol{r}}_{t-1}\dot{\boldsymbol{r}}_{t-1}' \right) \dot{\mathbf{D}}_1^{-1} \Rightarrow \dot{\boldsymbol{\mathscr{M}}}_{11}$. □

Both weak limits $\dot{\boldsymbol{\mathscr{M}}}$ and $\dot{\boldsymbol{\mathscr{M}}}_{11}$ given in Lemmas 2 (*vi* and *vii*) are nonsingular almost surely, mean-

ing that the weak limits of $(\sum_{t=1}^{T} \dot{z}_t \dot{z}_t')^{-1}$ and $(\sum_{t=1}^{T} \dot{r}_{t-1} \dot{r}_{t-1}')^{-1}$ are obtained as $\dot{\mathcal{M}}^{-1}$ and $\dot{\mathcal{M}}_{11}^{-1}$, respectively, where the inverse matrices are those in the short- and long-run parameter estimators, $\dot{\boldsymbol{\tau}}_T :=$ $(\sum_{t=1}^{T} \dot{z}_t \dot{z}_t')^{-1} \sum_{t=1}^{T} \dot{z}_t \Delta y_t$ and $\widetilde{\boldsymbol{v}}_T := (\sum_{t=1}^{T} \dot{r}_{t-1} \dot{r}_{t-1}')^{-1} \sum_{t=1}^{T} \dot{r}_{t-1} y_{t-1}$. From this, neither $\dot{\boldsymbol{\tau}}_T$ nor $\widetilde{\boldsymbol{v}}_T$ 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 $(\widehat{\boldsymbol{v}}_T, \dot{\boldsymbol{\tau}}_T)$ for their target parameters $(\boldsymbol{v}_*, \boldsymbol{\tau}_*)$, 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 $\widehat{\boldsymbol{v}}_T$ and show its consistency. Second, we conduct the same procedure for the primitive estimator $\dot{\boldsymbol{\tau}}_T$. 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 estimator $\widehat{\boldsymbol{v}}_T$. For this, note that $\widehat{\boldsymbol{v}}_T = \mathbf{P}_T \widetilde{\boldsymbol{v}}_T$, from Proposition 2. The weak limit of the long-run parameter estimator $\widehat{\boldsymbol{v}}_T$ 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 $\widehat{\boldsymbol{\mu}}_T$ for $\boldsymbol{\mu}_*$. Next, we obtain the limit distribution of the primitive estimator $\widetilde{\boldsymbol{v}}_T$ by noting that $\mathbf{m}_t = \widehat{\mathbf{m}}_t + (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)t$. If we rewrite (8) as

$$y_t = \boldsymbol{\beta}_*' \widehat{\mathbf{m}}_t + \vartheta_{T*} t + \nu_* + u_t = \dot{\boldsymbol{r}}_t' \bar{\boldsymbol{v}}_{T*} + u_t, \quad \text{where} \quad \bar{\boldsymbol{v}}_{T*} := [\boldsymbol{\beta}_*', \vartheta_{T*}, \nu_*]' \quad \text{with} \quad \vartheta_{T*} := \boldsymbol{\beta}_*'(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) + \vartheta_*, \tag{12}$$

it follows that

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

Using this arrangement, we provide the weak limit of the primitive estimator $\widetilde{\boldsymbol{v}}_T$ 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) $\dot{\mathbf{D}}_1(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) \Rightarrow [\dot{\mathscr{L}}_1', \dot{\mathscr{L}}_2, \dot{\mathscr{L}}_3]' := \dot{\boldsymbol{\mathscr{L}}} := \dot{\boldsymbol{\mathscr{M}}}_{11}^{-1} \dot{\boldsymbol{\mathscr{S}}}$, where

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

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

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

Lemma 3 (ii) also provides the weak limit of the long-run parameter estimator $\widehat{\boldsymbol{v}}_T$.

Remarks. (a) In the online supplement, we note that $\dot{\mathbf{D}}_1(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) = (\dot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \dot{\boldsymbol{r}}_{t-1} \dot{\boldsymbol{r}}_{t-1}' \dot{\mathbf{D}}_1^{-1})^{-1}$ $\dot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \dot{\boldsymbol{r}}_{t-1} u_{t-1}$ and prove Lemma 3 (i). We focus on deriving the weak limit of $\dot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \dot{\boldsymbol{r}}_{t-1} u_{t-1}$, because Lemma 2 (vii) already provides the weak limit of $(\dot{\mathbf{D}}_1^{-1} \sum_{t=1}^T \dot{\boldsymbol{r}}_{t-1} \dot{\boldsymbol{r}}_{t-1}' \dot{\mathbf{D}}_1^{-1})^{-1}$. We next note that $\widehat{\boldsymbol{v}}_T - \boldsymbol{v}_* = (\mathbf{P}_T - \mathbf{P})(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) + \mathbf{P}(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) + (\mathbf{P}_T - \mathbf{P})(\bar{\boldsymbol{v}}_{T*} - \bar{\boldsymbol{v}}_*) +$ $\mathbf{P}(\bar{\boldsymbol{v}}_{T*} - \bar{\boldsymbol{v}}_*) + (\mathbf{P}_T - \mathbf{P})\bar{\boldsymbol{v}}_*$, and prove Lemma 3 (ii) by deriving the weak limit of each component on the right side.

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

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

Second Step We derive the weak limit of the primitive estimator $\dot{\boldsymbol{\tau}}_T$. We first rewrite (9) as $\Delta y_t = \boldsymbol{\tau}_{T*}' \dot{\boldsymbol{z}}_t + e_t$, where

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

This equality is established by noting that $u_t = \widehat{u}_t + (\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)'\widehat{\mathbf{m}}_t + (\widetilde{\vartheta}_T - (\boldsymbol{\beta}_*'\widehat{\boldsymbol{\mu}}_T + \zeta_*))t + (\widetilde{\nu}_T - \nu_*)$, which follows from (12) by noting that $\widehat{u}_t := y_t - \widetilde{\boldsymbol{r}}_t'\widehat{\boldsymbol{v}}_T = y_t - \dot{\boldsymbol{r}}_t'\widetilde{\boldsymbol{v}}_T$. Using this representation for Δy_t, we decompose the primitive estimator $\dot{\boldsymbol{\tau}}_T$ into $\boldsymbol{\tau}_{T*}$ and a remainder:

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

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

Lemma 4. *Given Assumption 1,*

(i) $\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) \Rightarrow [\dot{\mathscr{D}}_1, \dot{\boldsymbol{\mathscr{D}}}_2', \dot{\mathscr{D}}_3, \dot{\mathscr{D}}_4, \dot{\boldsymbol{\mathscr{D}}}_5']' := \dot{\boldsymbol{\mathscr{D}}} := \dot{\boldsymbol{\mathscr{M}}}^{-1}\dot{\boldsymbol{\mathscr{J}}}$, where

$$\dot{\boldsymbol{\mathscr{J}}} := \left[\begin{array}{c|c|cc} \dot{\mathscr{J}}_1 & \dot{\boldsymbol{\mathscr{J}}}_2' & \dot{\mathscr{J}}_3 \quad \dot{\mathscr{J}}_4 & \dot{\boldsymbol{\mathscr{J}}}_5' \end{array} \right]' := \left[\begin{array}{c|c|cc} \int d\mathscr{B}_{ue} & \int \bar{\boldsymbol{\mathscr{B}}}_m' d\mathscr{B}_e & \int r d\mathscr{B}_e \quad \int d\mathscr{B}_e & \int d\boldsymbol{\mathscr{B}}_{ze}' \end{array} \right]',$$

such that $\dot{\boldsymbol{\mathscr{D}}}_2$ and $\dot{\boldsymbol{\mathscr{J}}}_2 \in \mathbb{R}^{2k}$, and $\dot{\boldsymbol{\mathscr{D}}}_5$ and $\dot{\boldsymbol{\mathscr{J}}}_5 \in \mathbb{R}^{p+2kq-1}$; and

(ii) $\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) \Rightarrow \dot{\boldsymbol{\mathscr{D}}} + \rho_*[0, \dot{\boldsymbol{\mathscr{L}}}_1', \dot{\mathscr{L}}_2, \dot{\mathscr{L}}_3, \mathbf{0}']'$. $\qquad\square$

Remarks. **(a)** From the definition of $\dot{\boldsymbol{\mathscr{D}}}$, the primitive estimator $\dot{\boldsymbol{\tau}}_T$ asymptotically follows a mixed normal distribution.

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

(c) Unlike the weak limit of the long-run parameter estimator $\widehat{\upsilon}_T$ given in Lemma 3 (*ii*), the weak limit of the primitive estimator $\dot{\boldsymbol{\tau}}_T$ cannot be written as a linear combination of other weak limits. This confirms that the primitive estimator $\dot{\boldsymbol{\tau}}_T$ 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 $\rho_* = 0$. $\qquad\square$

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

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

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

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

Here, the weak limits of $\mathbf{T}_T$ and $\dot{\boldsymbol{\tau}}_T$ are already provided in Lemmas 3 and 4, respectively. Using these weak limits, we can 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

as

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

where $\boldsymbol{c}_* := [1, -\boldsymbol{\beta}'_*, -\zeta_*, -\nu_*, \mathbf{0}']'$ and $\boldsymbol{d}_T := [0, \mathbf{0}'_{2k \times 1}, 0, \{(\dot{\gamma}_T - \gamma_*) - \rho_*(\widetilde{\nu}_T - \nu_*)\}, (\dot{\boldsymbol{\tau}}_{2T} - \boldsymbol{\tau}_{2*})']'$. This representation implies that the weak limit of the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ is determined by the weak limits of $\dot{\rho}_T$ and $\boldsymbol{d}_T$. As another example, 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 by (16). That is, $\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_*) = o_{\mathbb{P}}(1)$ or $\sqrt{T}(\widehat{\xi}_T - \xi_*) = o_{\mathbb{P}}(1)$. 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 show 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 value conditions. For notational simplicity, we let $\dot{\boldsymbol{\mathscr{D}}}_2 \equiv [\dot{\boldsymbol{\mathscr{D}}}_2^{+'}, \dot{\boldsymbol{\mathscr{D}}}_2^{-'}]'$, such that $\dot{\boldsymbol{\mathscr{D}}}_2^+$ and $\dot{\boldsymbol{\mathscr{D}}}_2^- \in \mathbb{R}^k$. Recall that $\dot{\boldsymbol{\mathscr{D}}}_2$ is the weak limit of $\dot{\boldsymbol{\eta}}_T - (\boldsymbol{\eta}_* + \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*))$.

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 $\sqrt{T}(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) \Rightarrow \boldsymbol{c}_* \dot{\mathscr{D}}_1 + [0, \mathbf{0}'_{2k \times 1}, 0, \dot{\mathscr{D}}_4, \dot{\boldsymbol{\mathscr{D}}}'_5]'$;*

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

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

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

Remarks. **(a)** 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. Thus, the 2SNARDL estimator must be asymptotically more efficient than the OLS estimator. The long-run parameter estimator has a convergence rate of T for 2SNARDL.

(b) 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 $\widehat{\boldsymbol{\theta}}_T$:

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

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

17

(c) By Theorem 1 (*i*), the weak limit of $(\widehat{\rho}_T, \widehat{\boldsymbol{\theta}}_T', \widehat{\xi}_T)'$ is $(1, \boldsymbol{\beta}_*', -\zeta_*)'\dot{\mathscr{D}}_1$, 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 $\dot{\mathscr{D}}_1$ is not associated with an asymptotically singular matrix problem, the form of the transformation means that the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ suffers from the singularity problem.

(d) 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 $\dot{\mathscr{D}}_1$, $\dot{\mathscr{D}}_4$, and $\dot{\boldsymbol{\mathscr{D}}}_5$. This implies the following. First, the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ follows a mixed normal distribution, by Lemma 4. 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*). Second, both $\dot{\mathscr{D}}_1$ and $\dot{\boldsymbol{\mathscr{D}}}_5$ are the weak limits of the last two OLS estimators obtained by regressing Δy_t against $(1, u_{t-1}, \boldsymbol{z}_{2t}')'$ or $(1, \widehat{u}_{t-1}, \boldsymbol{z}_{2t}')'$. 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_*.

(e) If the zero coefficient conditions in Theorem 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). For instance, if $\boldsymbol{\beta}_*^+ = \boldsymbol{0}$, (17) implies that $T(\widehat{\boldsymbol{\theta}}_T^+ - \boldsymbol{\theta}_*^+) = T\{(\dot{\boldsymbol{\eta}}_T^+ - \boldsymbol{\eta}_*^+) - \rho_*(\widetilde{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)\} \Rightarrow \dot{\boldsymbol{\mathscr{D}}}_2^+$ by Lemma 4 (*i*). In parallel, if $\boldsymbol{\beta}_*^- = \boldsymbol{0}$, then $T(\widehat{\boldsymbol{\theta}}_T^- - \boldsymbol{\theta}_*^-) = T\{(\dot{\boldsymbol{\eta}}_T^- - \boldsymbol{\eta}_*^-) - \rho_*(\widetilde{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)\} \Rightarrow \dot{\boldsymbol{\mathscr{D}}}_2^-$; and if $\zeta_* = 0$, then $T(\widehat{\xi}_T - \xi_*) = -\boldsymbol{\mu}_*'T\{(\dot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)\} + O_\mathbb{P}(T^{-1/2}) \Rightarrow -\boldsymbol{\mu}_*'\dot{\boldsymbol{\mathscr{D}}}_2$, meaning that the weak limits of $\widehat{\boldsymbol{\theta}}_T$ and $\widehat{\xi}_T$ are determined by $\dot{\boldsymbol{\mathscr{D}}}_2$. Given that $\dot{\boldsymbol{\mathscr{D}}}_2$ follows a mixed normal distribution, if the standard t-test applies to $\widehat{\boldsymbol{\theta}}_T$ or $\widehat{\xi}_T$, its null weak limit follows a mixed normal distribution.

(f) There is a caveat to Theorems 1 (*ii*, *iii*, and *iv*). If $\boldsymbol{\beta}_* = \boldsymbol{0}$ and $\zeta_* = 0$, then $\rho_* = 0$, by the remark below Assumption 1, which contradicts the assumption that $\rho_* < 0$. Theorems 1 (*ii*, *iii*, and *iv*) assume an environment in which none of $\boldsymbol{\beta}_*$ and ζ_* are zero.

(g) As a remark relevant to **(f)**, 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 $\widehat{\boldsymbol{\alpha}}_T$ under the conditions in Theorem 1 are different from that under $\boldsymbol{\beta}_* = \boldsymbol{0}$ and $\zeta_* = 0$. $\qquad\qquad\square$

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.

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:

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

$\dot{\sigma}_{e,T}^2 := T^{-1} \sum_{t=1}^T (\Delta y_t - \dot{\boldsymbol{z}}_t' \dot{\boldsymbol{\tau}}_T)^2$, and $\dot{\mathbf{R}}_1 := [\mathbf{0}_{k \times 1}, \mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k \times (1+p+2kq)}]$. Note that $\dot{\mathbf{R}}_1 \widehat{\boldsymbol{\alpha}}_T = (\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-)$, so that $\sqrt{T}(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) \Rightarrow (\boldsymbol{\beta}_*^- - \boldsymbol{\beta}_*^+)\dot{\mathscr{D}}_1$. Thus, the weak limit in Theorem 1 (*i*) 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

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

under $\mathscr{H}_0'$, where we let $\boldsymbol{\eta}_{T*} := \boldsymbol{\eta}_* + \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)$ 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 $\widehat{\boldsymbol{\beta}}_T = \widetilde{\boldsymbol{\beta}}_T$. Therefore, the null limit distribution of the Wald test $W_T^{(1)}$ is determined not by Theorem 1 (*i*), but by Lemma 4. All these imply $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = O_{\mathbb{P}}(T^{-1})$. However, its alternative behavior is different. From $\dot{\mathbf{R}}_1 \widehat{\boldsymbol{\alpha}}_T = (\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-)$,

its divergence rate is determined by Theorem 1 (*i*), meaning that $\dot{\mathbf{R}}_1\widehat{\boldsymbol{\alpha}}_T = O_{\mathbb{P}}(T^{-1/2})$. Therefore, 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 $\widetilde{\boldsymbol{v}}_T$. This is mainly because the primitive estimator $\widetilde{\boldsymbol{v}}_T$ 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 $\widetilde{\boldsymbol{\beta}}_T$. If we let $\ddot{\mathbf{R}}_1 := [\mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}_{k\times 2}]$, $\ddot{\mathbf{R}}_1\bar{\boldsymbol{v}}_{T*} = \boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-$, which we can estimate by $\widetilde{\boldsymbol{\beta}}_T^+ - \widetilde{\boldsymbol{\beta}}_T^- = \ddot{\mathbf{R}}_1\widetilde{\boldsymbol{v}}_T$. Using this feature, we define the following supplementary Wald test:

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

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

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

and $\dot{\mathbf{R}}_2 := [\mathbf{0}_{kq\times(2+p+2k)}, \mathbf{I}_{kq}, -\mathbf{I}_{kq}]$.

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

define the following Wald test:

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

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

Testing $\mathscr{H}_0'''$ versus $\mathscr{H}_1'''$ Finally, we apply the Wald test principle to test $\mathscr{H}_0'''$. For this, we apply the Wald test principle to the OLS estimator $\widehat{\boldsymbol{\alpha}}_T$ and define the following test:

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

and $\dot{\mathbf{R}}_3 := \dot{\mathbf{R}}_3 := \text{diag}[\dot{\mathbf{R}}_1, \dot{\mathbf{R}}_2]$.

We also define another supplementary Wald test. 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 $\ddot{\boldsymbol{\delta}}_T := (\widetilde{\boldsymbol{\upsilon}}_T', \dot{\boldsymbol{\tau}}_T')'$, $\bar{\mathbf{D}} := \text{diag}[\dot{\mathbf{D}}_1, \dot{\mathbf{D}}]$, $\ddot{\mathbf{R}}_3 := \text{diag}[\ddot{\mathbf{R}}_1, \ddot{\mathbf{R}}_2]$, 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, $\widetilde{\boldsymbol{\upsilon}}_T$ and $\dot{\boldsymbol{\tau}}_T$, and by noting that $\ddot{\mathbf{R}}_3 \bar{\mathbf{D}} \ddot{\boldsymbol{\delta}}_T = (T(\widetilde{\boldsymbol{\beta}}_T^+ - \widetilde{\boldsymbol{\beta}}_T^-)', \sqrt{T}(\dot{\boldsymbol{\pi}}_T^+ - \dot{\boldsymbol{\pi}}_T^-)')'$. 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.

We now provide the limit behaviors of the Wald tests under the null and alternative using the weak limits of the primitive estimators, $\widetilde{\boldsymbol{\upsilon}}_T$ and $\dot{\boldsymbol{\tau}}_T$. 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 $\mathbb{H}_{11}''' : \boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$ and $\mathbb{H}_{12}''' : \boldsymbol{\pi}_*^+ \neq \boldsymbol{\pi}_*^-$. We create this partition because the power behavior of the Wald test $\mathscr{W}_T^{(3)}$ depends on the sub-hypotheses. We also let $\sigma_e^2 := \mathbb{E}[e_t^2]$. We report 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) $W_T^{(1)} \Rightarrow \dot{\boldsymbol{\mathscr{D}}}'\mathbf{R}_1'(\sigma_e^2 \mathbf{R}_1 \dot{\boldsymbol{\mathscr{M}}}^{-1}\mathbf{R}_1')^{-1}\mathbf{R}_1 \dot{\boldsymbol{\mathscr{D}}}$ *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) $W_T^{(2)} \Rightarrow \dot{\boldsymbol{\mathscr{D}}}'\dot{\mathbf{R}}_2'(\sigma_e^2 \dot{\mathbf{R}}_2 \dot{\boldsymbol{\mathscr{M}}}^{-1}\dot{\mathbf{R}}_2')^{-1}\dot{\mathbf{R}}_2 \dot{\boldsymbol{\mathscr{D}}}$ *under* $\mathscr{H}_0''$;

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

(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) $\mathscr{W}_T^{(1)} \Rightarrow \dot{\boldsymbol{\mathscr{L}}}'\ddot{\mathbf{R}}_1'(\sigma_u^2 \ddot{\mathbf{R}}_1 \dot{\boldsymbol{\mathscr{M}}}_{11}^{-1}\ddot{\mathbf{R}}_1')^{-1}\ddot{\mathbf{R}}_1 \dot{\boldsymbol{\mathscr{L}}}$ *under* $\mathbb{H}_0'$;

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

    (c) $\mathscr{W}_T^{(3)} \Rightarrow \dot{\boldsymbol{\mathscr{L}}}'\ddot{\mathbf{R}}_1'(\sigma_u^2 \ddot{\mathbf{R}}_1 \dot{\boldsymbol{\mathscr{M}}}_{11}^{-1}\ddot{\mathbf{R}}_1')^{-1}\ddot{\mathbf{R}}_1 \dot{\boldsymbol{\mathscr{L}}} + \dot{\boldsymbol{\mathscr{D}}}'\ddot{\mathbf{R}}_2'(\sigma_e^2 \ddot{\mathbf{R}}_2 \dot{\boldsymbol{\mathscr{M}}}^{-1}\ddot{\mathbf{R}}_2')^{-1}\ddot{\mathbf{R}}_2 \dot{\boldsymbol{\mathscr{D}}}$ *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* $\mathbb{H}_{11}''' \bigcap \mathbb{H}_{12}'''$. $\qquad\square$

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 $(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*})$. For example, it follows from (18) that $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-) - (\dot{\rho}_T - \rho_*)(\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-) + (\dot{\boldsymbol{\eta}}_T^+ - \boldsymbol{\eta}_{T*}^+) - (\dot{\boldsymbol{\eta}}_T^- - \boldsymbol{\eta}_{T*}^-) + o_{\mathbb{P}}(T^{-1})$. Therefore, if we impose $\mathscr{H}_0'$, $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\dot{\boldsymbol{\eta}}_T^+ - \boldsymbol{\eta}_{T*}^+) - (\dot{\boldsymbol{\eta}}_T^- - \boldsymbol{\eta}_{T*}^-) + o_{\mathbb{P}}(T^{-1})$. In other words, $\dot{\mathbf{R}}_1 \widehat{\boldsymbol{\alpha}}_T = \mathbf{R}_1(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) + o_{\mathbb{P}}(T^{-1})$. This fact implies that the null weak limit of the Wald test $W_T^{(1)}$ is determined by that of $(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*})$. 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 $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-) - (\dot{\rho}_T - \rho_*)(\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-) + o_{\mathbb{P}}(T^{-1/2})$ and $(\dot{\rho}_T - \rho_*) = O_{\mathbb{P}}(T^{-1/2})$. 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.

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

(**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 $\dot{\mathscr{S}}'\dot{\mathscr{M}}_{11}^{-1}\ddot{\mathbf{R}}_1'(\sigma_u^2\ddot{\mathbf{R}}_1\dot{\mathscr{M}}_{11}^{-1}\ddot{\mathbf{R}}_1')^{-1}\ddot{\mathbf{R}}_1\dot{\mathscr{M}}_{11}^{-1}\dot{\mathscr{S}}$ using the definition of $\dot{\mathscr{L}}$, meaning that the Wald test $\mathscr{W}_T^{(1)}$ does not follow a mixed chi-squared distribution, because the weak limit $\dot{\mathscr{S}}$ does not follow a mixed normal distribution. The same result holds for the Wald test $\mathscr{W}_T^{(3)}$ from $\mathscr{W}_T^{(3)} = \mathscr{W}_T^{(1)} + \mathscr{W}_T^{(2)}$. On the other hand, the null weak limit of the Wald test $\mathscr{W}_T^{(2)}$ is $\dot{\mathscr{J}}'\dot{\mathscr{M}}^{-1}\ddot{\mathbf{R}}_1'(\sigma_e^2\ddot{\mathbf{R}}_1\dot{\mathscr{M}}^{-1}\ddot{\mathbf{R}}_1')^{-1}\ddot{\mathbf{R}}_1\dot{\mathscr{M}}^{-1}\dot{\mathscr{J}}$, meaning that its null weak limit follows a mixed chi-squared distribution, because the weak limit $\dot{\mathscr{J}}$ follows a mixed normal distribution. Therefore, tabulated asymptotic critical values are available only 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: $\beta_*^+ \neq \beta_*^-$, the divergence speed is T^2. In contrast, if only the short-run parameters are asymmetric: $\pi_*^+ \neq \pi_*^-$, its divergence speed is T. If both the long- and short-run parameters are asymmetric, the divergence speed is determined by T^2. Theorem 2 (iv.c) summarizes these results. □

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 data-generating process (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 $\Delta x_t = v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. 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 $\bar{\upsilon}_{T*}$ and τ_{T*} separately by specifying the following models:

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

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

where we set $\widehat{m}_t^+ := x_t^+ - t\widehat{\mu}_T^+$, $\widehat{m}_t^- := x_t^- - t\widehat{\mu}_T^-$, and $\widehat{u}_t := y_t - \widetilde{r}_t'\widehat{v}_T$, with $\widehat{\mu}_T^+ := (\sum_{t=1}^{T-1} t^2)^{-1} \sum_{t=1}^{T-1} t x_t^+$,

$\widehat{\mu}_T^- := (\sum_{t=1}^{T-1} t^2)^{-1} \sum_{t=1}^{T-1} t x_t^-$, $\widehat{\boldsymbol{v}}_T := (\sum_{t=1}^{T} \widetilde{\boldsymbol{r}}_{t-1} \widetilde{\boldsymbol{r}}_{t-1}')^{-1} \sum_{t=1}^{T} \widetilde{\boldsymbol{r}}_{t-1} y_{t-1}$, and $\widetilde{\boldsymbol{r}}_{t-1} := [x_{t-1}^+, x_{t-1}^-, (t-1), 1]'$. 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 applies even when the null weak limit of the Wald test follows a mixed chi-squared distribution.

24

We conduct our simulation following the above residual bootstrap procedure. Assuming that $B = 500$, we independently iterate the above experiment 5,000 times for $T = 100, 200, 300, 400,$ and 500. The simulation results are reported in Tables 1 and 2. We also evaluate the Wald tests by mixed chi-squared distributions and report the empirical rejection rates. The simulation results are summarized as follows:

(a) Table 1 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 $\dot{\pi}_T^+$ and $\dot{\pi}_T^-$ 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 2 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 1. 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 $\Delta x_t = v_t$, we set $\Delta x_t = \kappa_* \Delta x_{t-1} + v_t$ and apply the same residual bootstrap method. Then, assuming $\kappa_* = 1/2$, we conduct simulations by setting $B = 500$ and independently iterating the experiment 5,000 times for $T = 100, 200, 300, 400,$ and 500. The simulation results are presented in Tables 3 and 4 in the same format as in Tables 1 and 2, respectively. The simulation results are summarized as follows:

(a) Table 3 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 1.

(b) Table 4 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. $\square$

7 Empirical Application

This section examines the empirical data provided by Romer and Romer (2010) to measure the fiscal policy impact on the U.S. GDP. 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. 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 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 in (21) by OLS, 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. Although the NARDL model assumptions do not perfectly align with the characteristics of the exogenous

---

[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}.

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.1 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 5. The columns marked "Exo" give the parameter estimates obtained by OLS for the NARDL and ARDL models. In other words, 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 5 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 $\mathscr{W}_T^{(1)}$ and $W_T^{(1)}$ provide p-values that make it difficult to reject the symmetry hypothesis.

(d) Short-run symmetry is confirmed by the Wald tests $\mathscr{W}_T^{(2)}$ and $W_T^{(2)}$, because they do not reject the symmetry hypothesis. Moreover, neither the long-run nor the short-run symmetry hypothesis is challenged by the Wald test $\mathscr{W}_T^{(3)}$ or $W_T^{(3)}$. As such, we conclude that the ARDL model is appropriate for studying the relationship between the log real GDP and the log of an exogenous tax shock.

(e) We present the estimation results for the log of an endogenous tax shock, calculated in the same way as for the log of an exogenous tax shock. Columns labeled "Endo" give the estimation and inference results obtained using the log of the endogenous tax shock data. Similarly, the columns labeled "Sum" show the estimation and inference results obtained using the logs of both exogenous and endogenous tax shocks. The estimated signs of τ_{t-1}^{+}, τ_{t-1}^{-}, and τ_{t-1} are inconsistent with the parameter values posited by standard economic theory. When using the logs of both exogenous and endogenous tax shocks jointly, there is little evidence of cointegration. These estimation results indicate that only the log of an exogenous tax shock can be used to estimate the coefficients having signs consistent with standard economic theory. □

Next, we estimate the NARDL and ARDL models using the 2SNARDL method proposed by Cho, Greenwood-Nimmo, and Shin (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 6, which is structured similarly to Table 5. This separate investigation is conducted to validate the inference results presented in Table 5. We summarize the results in Table 6.

(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. Given that 2SNARDL is super-consistent, these estimates are likely to be more precise than those estimated by OLS.

(c) Our findings for the endogenous and aggregate log tax shocks align with the results presented in Table 5. 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 nonsignificant. We believe these inconsistent results are due to the correlation between an endogenous shock and a structural error. □

The results in Tables 5 and 6 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 5 and 6, we estimate the models using OLS and 2SNARDL. The estimation and inference results are presented in Tables 7 and 8, respectively. We summarize the results as follows:

(a) For the exogenous tax change, the coefficient of y_{t-1} in Table 7 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 8 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 8 than to those in Table 7, because 2SNARDL is more efficient than OLS. 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 7 rejects the hypothesis of no cointegration. Furthermore, none of the coefficients of u_{t-1} in Table 8 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 allow us to derive the weak limit of the OLS.

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.

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 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.

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.

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.

CHESNES, M. (2016): "Asymmetric Pass-Through in U.S. Gasoline Prices," *Energy Journal*, 37, 153–180.

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.

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

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

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.

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

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

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

DE JONG, R. M. AND J. DAVIDSON (2000): "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes," *Econometric Theory*, 16, 621–642.

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

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

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

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.

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

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.

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.

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.

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

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

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.

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

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.

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.

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

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

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%	10.70	10.78	10.80	9.58	10.54	38.16	40.26	40.48	40.24	41.02
	5%	5.52	5.48	5.38	4.68	5.12	30.38	31.30	32.50	31.52	32.46
	1%	1.28	1.00	0.94	1.12	0.98	17.46	18.38	19.48	18.34	19.24
$\mathscr{W}_T^{(2)}$	10%	9.46	10.26	9.94	9.46	10.26	11.12	11.24	10.58	9.96	10.52
	5%	4.74	5.10	5.00	4.90	4.94	5.80	5.68	5.38	5.20	5.36
	1%	1.00	1.02	1.06	0.98	1.04	1.44	1.28	1.26	0.94	1.02
$\mathscr{W}_T^{(3)}$	10%	10.26	11.02	10.54	9.84	10.80	33.96	35.14	36.18	35.62	36.64
	5%	5.40	5.62	4.98	4.62	5.44	26.32	27.08	27.90	26.74	27.94
	1%	1.24	1.04	1.00	1.00	1.04	14.28	14.92	15.84	14.28	15.70
$W_T^{(1)}$	10%	10.94	9.90	10.18	9.74	10.44	14.98	12.14	11.48	10.68	11.36
	5%	5.18	4.84	4.98	4.86	4.96	8.70	6.24	5.92	5.46	5.68
	1%	0.88	0.82	1.14	0.84	0.86	2.18	1.42	1.44	1.16	1.04
$W_T^{(2)}$	10%	10.74	9.16	9.58	9.22	10.02	12.68	10.04	10.10	9.60	10.62
	5%	5.06	4.62	5.02	4.58	5.14	6.86	5.20	5.28	4.96	5.28
	1%	1.06	0.96	1.20	1.08	1.14	1.80	1.24	1.30	1.12	1.22
$W_T^{(3)}$	10%	10.78	8.92	9.80	9.42	10.06	14.92	11.46	11.04	10.52	10.90
	5%	4.98	4.46	4.82	4.92	4.94	8.36	5.82	5.64	5.48	5.44
	1%	0.98	0.90	1.16	0.90	1.20	2.42	1.52	1.42	1.24	1.46

Table 1: 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 5,000, and the bootstrap iteration is 500. DGP: $\Delta y_t = \rho_* u_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t$, $u_t = y_t - \beta_*^+ x_t^+ - \beta_*^- x_t^-$, $\Delta x_t = v_t$, 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%	49.52	88.04	97.98	99.88	99.96	70.78	96.20	99.48	99.96	100.0
	5%	37.34	81.12	96.62	99.60	99.92	65.14	94.50	99.30	99.96	99.98
	1%	15.54	61.22	90.30	98.22	99.64	52.50	90.24	98.64	99.88	99.98
$\mathscr{W}_T^{(2)}$	10%	17.62	25.96	33.14	40.40	48.24	20.18	27.28	34.70	41.24	48.94
	5%	10.58	17.32	21.98	28.90	36.24	12.70	18.60	23.42	30.02	37.34
	1%	2.96	6.00	8.44	11.88	16.30	4.60	6.98	9.44	12.96	17.54
$\mathscr{W}_T^{(3)}$	10%	47.82	86.70	97.84	99.84	99.96	67.46	94.86	99.22	99.96	99.98
	5%	35.50	80.26	96.38	99.54	99.92	60.74	92.78	99.00	99.96	99.98
	1%	15.22	59.92	89.70	97.98	99.62	48.80	87.72	97.88	99.86	99.96
$W_T^{(1)}$	10%	44.88	86.62	98.28	99.62	99.98	53.84	89.18	98.54	99.68	99.98
	5%	33.16	79.72	96.54	99.42	99.94	43.76	83.78	97.32	99.56	99.98
	1%	14.12	59.88	90.36	98.16	99.76	25.30	68.94	92.64	98.56	99.80
$W_T^{(2)}$	10%	17.06	24.66	33.86	40.02	48.20	19.80	26.00	34.62	41.44	48.92
	5%	9.66	15.64	22.94	28.38	35.82	11.72	17.24	24.26	29.34	37.16
	1%	2.90	4.74	8.90	11.78	16.50	4.18	6.04	10.06	13.12	17.64
$W_T^{(3)}$	10%	39.80	82.48	97.08	99.42	99.98	48.72	85.30	97.70	99.50	99.98
	5%	28.60	74.36	94.76	99.02	99.94	38.42	78.44	95.74	99.08	99.94
	1%	11.76	53.86	86.66	97.42	99.52	20.60	61.92	89.80	97.94	99.66

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 5,000, and the bootstrap iteration is 500. DGP: $\Delta y_t = \rho_* u_{t-1} + \pi^+_* \Delta x^+_t + \pi^-_* \Delta x^-_t + e_t$, $u_t = y_t - \beta^+_* x^+_t - \beta^-_* x^-_t$, $\Delta x_t = v_t$, 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%	10.72	10.98	10.44	10.24	9.58	50.64	53.24	52.44	52.16	51.76
	5%	5.72	6.04	5.20	5.16	4.80	42.80	45.64	44.82	45.54	44.54
	1%	1.00	1.14	0.90	0.86	1.16	29.24	32.36	32.20	32.72	31.04
$\mathscr{W}_T^{(2)}$	10%	10.18	10.30	10.04	9.90	10.38	12.70	11.16	10.86	10.62	10.90
	5%	4.84	5.36	4.88	4.76	5.12	6.78	6.22	5.42	5.26	5.48
	1%	0.84	1.08	0.92	0.84	1.04	1.70	1.44	1.12	0.98	1.10
$\mathscr{W}_T^{(3)}$	10%	10.26	11.12	9.98	10.12	9.20	47.14	49.32	48.42	48.74	47.70
	5%	5.40	5.78	5.28	5.28	4.62	38.56	41.22	40.00	41.20	39.10
	1%	0.94	1.00	0.86	0.80	1.26	25.36	28.38	27.32	27.82	26.70
$W_T^{(1)}$	10%	9.76	9.70	9.48	9.62	9.78	14.04	11.92	10.68	10.82	10.46
	5%	4.82	5.02	4.78	4.60	5.14	8.08	6.66	5.46	5.50	5.90
	1%	1.08	0.98	0.74	1.08	0.98	2.36	1.54	1.20	1.32	1.20
$W_T^{(2)}$	10%	10.14	10.04	9.96	10.26	9.80	12.80	11.56	10.76	10.82	10.12
	5%	5.36	5.08	4.88	5.12	5.14	7.14	6.04	5.78	5.54	5.34
	1%	1.20	1.00	0.92	0.88	0.86	1.96	1.30	1.18	1.16	1.12
$W_T^{(3)}$	10%	10.48	10.14	9.92	9.64	9.96	15.02	12.58	11.18	10.72	11.12
	5%	5.10	4.72	4.90	5.04	4.82	9.14	6.48	6.00	5.98	5.58
	1%	0.90	0.96	0.84	0.92	0.84	2.44	1.46	1.34	1.24	1.00

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 5,000, and the bootstrap iteration is 500. DGP: $\Delta y_t = \rho_* u_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t$, $u_t = y_t - \beta_*^+ x_t^+ - \beta_*^- x_t^-$, $\Delta x_t = \kappa_* \Delta x_{t-1} + v_t$, 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%	70.42	98.08	99.96	100.0	100.0	86.72	99.54	100.0	100.0	100.0
	5%	58.52	96.24	99.94	99.96	100.0	83.04	99.36	100.0	100.0	100.0
	1%	31.88	88.28	99.36	99.92	99.98	74.30	98.56	99.98	100.0	100.0
$\mathscr{W}_T^{(2)}$	10%	19.34	31.56	42.16	51.48	60.40	23.44	34.30	43.90	53.00	62.24
	5%	11.94	21.60	30.82	38.88	48.20	15.34	24.10	33.22	41.02	49.90
	1%	4.06	7.56	12.20	18.60	25.84	6.20	9.86	14.74	21.34	28.06
$\mathscr{W}_T^{(3)}$	10%	68.38	97.80	99.96	99.98	100.0	83.84	99.40	100.0	100.0	100.0
	5%	56.20	95.96	99.94	99.96	100.0	79.80	99.00	100.0	100.0	100.0
	1%	30.42	87.80	99.34	99.90	99.98	70.72	97.96	99.98	99.96	100.0
$W_T^{(1)}$	10%	68.52	97.96	99.92	100.0	100.0	76.54	98.62	99.94	100.0	100.0
	5%	56.32	95.98	99.80	100.0	100.0	68.12	97.34	99.90	100.0	100.0
	1%	32.50	89.18	99.38	100.0	99.98	48.98	92.80	99.62	100.0	100.0
$W_T^{(2)}$	10%	19.96	31.52	41.64	51.20	60.28	23.56	33.88	43.52	52.62	61.46
	5%	12.58	21.60	30.30	39.14	48.26	15.88	24.22	32.42	41.20	49.70
	1%	3.98	8.30	13.18	18.26	25.12	6.28	10.00	15.90	21.02	27.66
$W_T^{(3)}$	10%	61.92	96.50	99.86	100.0	100.0	70.90	97.36	99.90	100.0	100.0
	5%	49.76	94.32	99.74	100.0	100.0	62.10	95.48	99.82	100.0	100.0
	1%	27.10	85.10	99.14	100.0	99.96	42.90	89.64	99.44	100.0	99.98

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 5,000, and the bootstrap iteration is 500. DGP: $\Delta y_t = \rho_* u_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t$, $u_t = y_t - \beta_*^+ x_t^+ - \beta_*^- x_t^-$, $\Delta x_t = \kappa_* \Delta x_{t-1} + v_t$, 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 5: 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 100,000 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 6: 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}	$-0.0609^{\dagger}$	-0.0633	-0.0460	y_{t-1}	$-0.0552^{\dagger}$	-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	$-3.3087^{\dagger}$	-2.5198	-2.5917	t-test	$-3.3867^{\dagger}$	-2.6602	-2.4960
F-test	3.9276	3.5933	3.4873	F-test	5.7417	5.2684	4.8210
$\mathscr{W}_T^{(1)}$					52.7574	2.1358	48.9740
					(0.1294)	(0.7290)	(0.2581)
$\mathscr{W}_T^{(2)}$					0.0722	0.8121	0.4782
					(0.8015)	(0.4021)	(0.4981)
$\mathscr{W}_T^{(3)}$					52.8296	2.9480	49.4522
					(0.1340)	(0.7595)	(0.2610)
$W_T^{(1)}$					0.5384	0.3344	0.8601
					(0.5697)	(0.6463)	(0.5035)
$W_T^{(2)}$					0.4079	0.9926	1.5118
					(0.5278)	(0.3177)	(0.2279)
$W_T^{(3)}$					1.0869	1.2200	2.1652
					(0.6618)	(0.6150)	(0.4681)

Table 7: 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 100,000 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 8: 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.
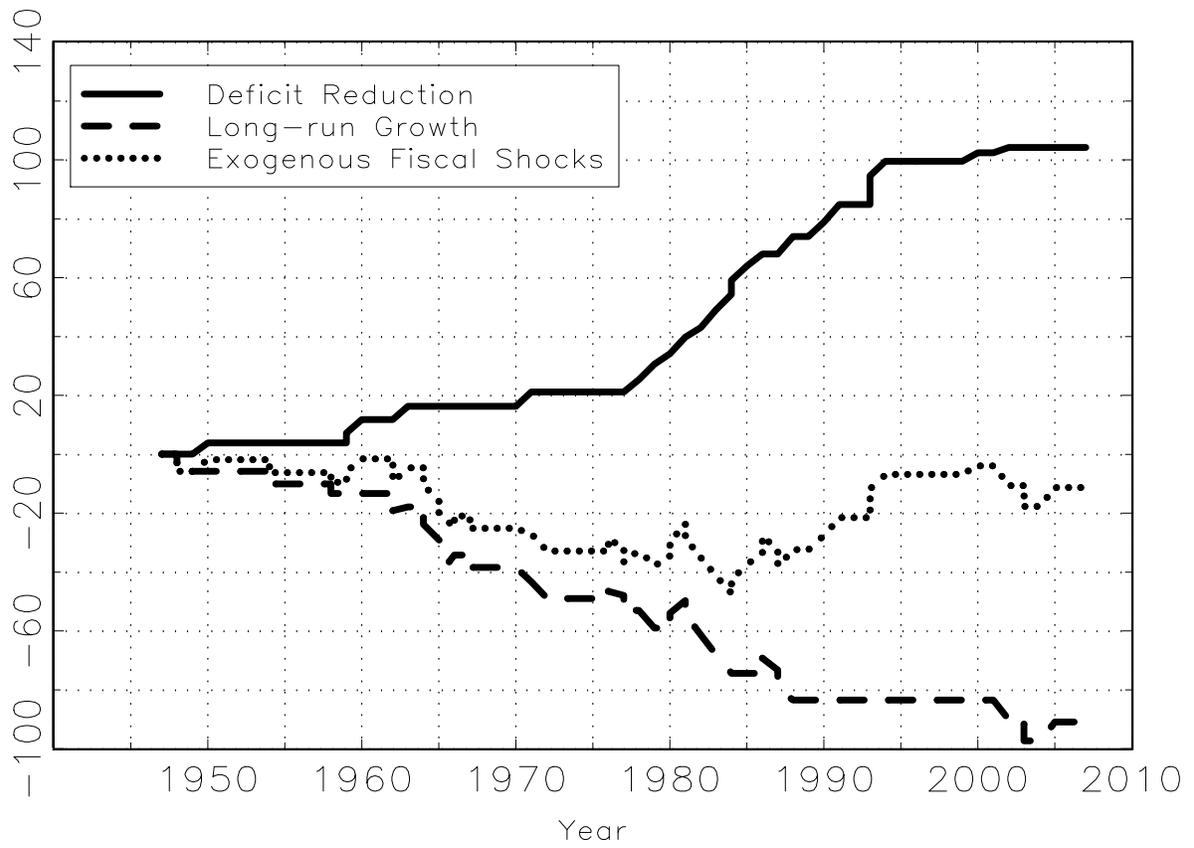
Figure 1: PARTIAL SUM PROCESS FORMED BY EXOGENOUS FISCAL SHOCKS. The solid, dashed, and dotted lines represent deficit reduction, long-run growth, and exogenous fiscal shocks, respectively.

<div align="center">

Online Supplement for

'Estimation and Inference of Nonlinear Autoregressive Distributed Lag Models

by Ordinary Least Squares'[1]

by

Jin Seo Cho

Yonsei University

</div>

This Online Supplement is an Appendix that provides proofs of all the results in the paper, including the lemmas.

A Supplements

A.1 Supplementary Lemmas

Before proving the main claims in the text, we first provide some preliminary lemmas to prove the main claims efficiently.

The following two lemmas provide alternative representations of the OLS estimators, which we define under different environments. We first suppose that $y_t \in \mathbb{R}$ is a dependent variable and $(\boldsymbol{x}_t', \boldsymbol{z}_t')' \in \mathbb{R}^{(s+k)}$ is an explanatory variable, and the OLS estimator is obtained by regressing y_t against $(\boldsymbol{x}_t', \boldsymbol{z}_t')'$. Given this, we provide alternative forms of the OLS estimator in the following lemmas:

Lemma A.1. *Suppose that* $\{(y_t, \boldsymbol{x}_t', \boldsymbol{z}_t')' \in \mathbb{R}^{1+s+k} : t = 1, 2, \ldots, T\}$. *If the OLS estimators are obtained as follows: for* $j = 1, 2, \ldots, s$,

$$(\widehat{\boldsymbol{\beta}}_T, \widehat{\boldsymbol{\gamma}}_T) := \arg\min_{\boldsymbol{\beta}, \boldsymbol{\gamma}} \sum_{t=1}^{T} \left(y_t - \boldsymbol{x}_t \boldsymbol{\beta} - \boldsymbol{z}_t' \boldsymbol{\gamma}\right)^2, \quad \widehat{\boldsymbol{\phi}}_{jT} := \arg\min_{\boldsymbol{\phi}_j} \sum_{t=1}^{T} \left(x_{jt} - \boldsymbol{z}_t' \boldsymbol{\phi}_j\right)^2, \quad and$$

$$(\widehat{\boldsymbol{\xi}}_T, \widehat{\boldsymbol{\delta}}_T) := \arg\min_{\boldsymbol{\xi}, \boldsymbol{\delta}} \sum_{t=1}^{T} \left(y_t - \widehat{\boldsymbol{v}}_t' \boldsymbol{\xi} - \boldsymbol{z}_t' \boldsymbol{\delta}\right)^2,$$

where for each t, $\widehat{\boldsymbol{v}}_t := \boldsymbol{x}_t - \widehat{\boldsymbol{\phi}}_T' \boldsymbol{z}_t$ *and* $\widehat{\boldsymbol{\phi}}_T := (\widehat{\boldsymbol{\phi}}_{1T}, \ldots, \widehat{\boldsymbol{\phi}}_{sT})$, *then* $\widehat{\boldsymbol{\beta}}_T = \widehat{\boldsymbol{\xi}}_T$ *and* $\widehat{\boldsymbol{\gamma}}_T = \widehat{\boldsymbol{\delta}}_T - \widehat{\boldsymbol{\phi}}_T \widehat{\boldsymbol{\xi}}_T.$ □

Therefore, we can obtain the OLS estimator $(\widehat{\boldsymbol{\beta}}_T, \widehat{\boldsymbol{\gamma}}_T)$ by combining the two OLS estimators $\widehat{\boldsymbol{\phi}}_T$ and $(\widehat{\boldsymbol{\xi}}_T, \widehat{\boldsymbol{\delta}}_T)$ obtained from the first- and second-step OLS estimators, respectively.

The following lemma considers a different environment. We let the OLS estimator be obtained by regressing y_t against $(\boldsymbol{x}_t, \boldsymbol{z}_t', \boldsymbol{w}_t')' \in \mathbb{R}^{s+k+d}$ and provide an alternative form of the OLS estimator:

Lemma A.2. *Suppose that* $\{(y_t, \boldsymbol{x}_t', \boldsymbol{z}_t', \boldsymbol{w}_t')' \in \mathbb{R}^{1+s+k+d} : t = 1, 2, \ldots, T\}$. *If the OLS estimators are obtained as follows: for* $j = 1, 2, \ldots, s$,

$$(\widehat{\boldsymbol{\beta}}_T, \widehat{\boldsymbol{\gamma}}_T, \widehat{\boldsymbol{\alpha}}_T) := \arg\min_{\beta, \gamma, \alpha} \sum_{t=1}^{T} \left(y_t - \boldsymbol{x}_t'\boldsymbol{\beta} - \boldsymbol{z}_t'\boldsymbol{\gamma} - \boldsymbol{w}_t'\boldsymbol{\alpha}\right)^2, \quad \widehat{\boldsymbol{\phi}}_{jT} := \arg\min_{\phi} \sum_{t=1}^{T} \left(x_{jt} - \boldsymbol{z}_t'\boldsymbol{\phi}_j\right)^2, \quad and$$

$$(\widehat{\boldsymbol{\xi}}_T, \widehat{\boldsymbol{\delta}}_T, \widehat{\boldsymbol{\theta}}_T) := \arg\min_{\xi, \delta, \theta} \sum_{t=1}^{T} \left(y_t - \widehat{\boldsymbol{v}}_t'\boldsymbol{\xi} - \boldsymbol{z}_t'\boldsymbol{\delta} - \boldsymbol{w}_t'\boldsymbol{\theta}\right)^2,$$

where for each t, $\widehat{\boldsymbol{v}}_t := \boldsymbol{x}_t - \widehat{\boldsymbol{\phi}}_T'\boldsymbol{z}_t$ *and* $\widehat{\boldsymbol{\phi}}_T := (\widehat{\boldsymbol{\phi}}_{1T}, \ldots, \widehat{\boldsymbol{\phi}}_{sT})$, *then* $\widehat{\boldsymbol{\beta}}_T = \widehat{\boldsymbol{\xi}}_T$, $\widehat{\boldsymbol{\gamma}}_T = \widehat{\boldsymbol{\delta}}_T - \widehat{\boldsymbol{\phi}}_T\widehat{\boldsymbol{\xi}}_T$, *and* $\widehat{\boldsymbol{\alpha}}_T = \widehat{\boldsymbol{\theta}}_T$. □

Note that $\boldsymbol{w}_t$ is added as an additional regressor to the regressors given in Lemma A.1 and that the nuisance parameter estimator $\widehat{\boldsymbol{\alpha}}_T$ is the same as the nuisance parameter estimator $\widehat{\boldsymbol{\theta}}_T$ obtained in the second step.

We now prove Lemmas A.1 and A.2. For notational simplicity, we let

$$\mathbf{Y} := [y_1, y_2, \ldots, y_T]', \quad \mathbf{X} := [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T]', \quad \mathbf{Z} := [\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_T]', \quad \widehat{\mathbf{V}} := [\widehat{v}_1, \widehat{v}_2, \ldots, \widehat{v}_T]',$$

and $\mathbf{W} := [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_T]'$.

Proof of Lemma A.1. From the definition of $(\widehat{\boldsymbol{\beta}}_T, \widehat{\boldsymbol{\gamma}}_T)$, we first note that

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}}_T \\ \widehat{\boldsymbol{\gamma}}_T \end{bmatrix} = \begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X'Y} \\ \mathbf{Z'Y} \end{bmatrix} = \begin{bmatrix} (\mathbf{X'QX})^{-1}\mathbf{X'QY} \\ (\mathbf{Z'Z})^{-1}\mathbf{Z'}[\mathbf{I} - \mathbf{X}(\mathbf{X'QX})^{-1}\mathbf{X'Q}]\mathbf{Y} \end{bmatrix}, \quad \text{(A.1)}$$

where $\mathbf{Q} := \mathbf{I} - \mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'}$. Next, we note that

$$\widehat{\mathbf{V}} = \mathbf{X} - \mathbf{Z}\widehat{\boldsymbol{\phi}}_T = \mathbf{X} - \mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'X} = \mathbf{QX}. \quad \text{(A.2)}$$

Therefore, $\mathbf{Z'}\widehat{\mathbf{V}} = \mathbf{0}$ by noting that $\mathbf{Z'Q} = \mathbf{0}$. Third, we note that

$$\begin{bmatrix} \widehat{\boldsymbol{\xi}}_T \\ \widehat{\boldsymbol{\delta}}_T \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{V}}'\widehat{\mathbf{V}} & \widehat{\mathbf{V}}'\mathbf{Z} \\ \mathbf{Z'}\widehat{\mathbf{V}} & \mathbf{Z'Z} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\mathbf{V}}'\mathbf{Y} \\ \mathbf{Z'Y} \end{bmatrix} = \begin{bmatrix} (\widehat{\mathbf{V}}'\widehat{\mathbf{V}})^{-1}\widehat{\mathbf{V}}'\mathbf{Y} \\ (\mathbf{Z'Z})^{-1}\mathbf{Z'Y} \end{bmatrix}$$

2

using the fact that $\mathbf{Z}'\widehat{\mathbf{V}} = \mathbf{0}$. Therefore,

$$\widehat{\boldsymbol{\xi}}_T = (\widehat{\mathbf{V}}'\widehat{\mathbf{V}})^{-1}\widehat{\mathbf{V}}'\mathbf{Y} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{Y} \tag{A.3}$$

using (A.2). This shows that $\widehat{\boldsymbol{\beta}}_T = \widehat{\boldsymbol{\xi}}_T$. Finally, we note that

$$\widehat{\boldsymbol{\delta}}_T - \widehat{\boldsymbol{\phi}}_T\widehat{\boldsymbol{\xi}}_T = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} - (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}(\widehat{\mathbf{V}}'\widehat{\mathbf{V}})^{-1}\widehat{\mathbf{V}}'\mathbf{Y} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}]\mathbf{Y} = \widehat{\boldsymbol{\gamma}}_T,$$

which follows from (A.1), where the second equality follows from (A.3). Thus, $\widehat{\boldsymbol{\delta}}_T - \widehat{\boldsymbol{\phi}}_T\widehat{\boldsymbol{\xi}}_T = \widehat{\boldsymbol{\gamma}}_T$. This completes the proof. ∎

Proof of Lemma A.2. To prove the claim, we represent the OLS estimators in different forms. If we let

$$(\widehat{\boldsymbol{\beta}}_T(\boldsymbol{\alpha}), \widehat{\boldsymbol{\gamma}}_T(\boldsymbol{\alpha})) := \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{t=1}^{T} \left(y_t - \boldsymbol{x}_t'\boldsymbol{\beta} - \boldsymbol{z}_t'\boldsymbol{\gamma} - \boldsymbol{w}_t'\boldsymbol{\alpha}\right)^2,$$

then

$$\widehat{\boldsymbol{\alpha}}_T = \arg\min_{\boldsymbol{\alpha}} \sum_{t=1}^{T} \left(y_t - \boldsymbol{x}_t\widehat{\boldsymbol{\beta}}_T(\boldsymbol{\alpha}) - \boldsymbol{z}_t'\widehat{\boldsymbol{\gamma}}_T(\boldsymbol{\alpha}) - \boldsymbol{w}_t'\boldsymbol{\alpha}\right)^2,$$

and $(\widehat{\boldsymbol{\beta}}_T(\widehat{\boldsymbol{\alpha}}_T), \widehat{\boldsymbol{\gamma}}_T(\widehat{\boldsymbol{\alpha}}_T)) = (\widehat{\boldsymbol{\beta}}_T, \widehat{\boldsymbol{\gamma}}_T)$. Likewise, if we let

$$(\widehat{\boldsymbol{\xi}}_T(\boldsymbol{\theta}), \widehat{\boldsymbol{\delta}}_T(\boldsymbol{\theta})) := \arg\min_{\boldsymbol{\xi},\boldsymbol{\delta}} \sum_{t=1}^{T} \left(y_t - \widehat{\boldsymbol{v}}_t'\boldsymbol{\xi} - \boldsymbol{z}_t'\boldsymbol{\delta} - \boldsymbol{w}_t'\boldsymbol{\theta}\right)^2,$$

then

$$\widehat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left(y_t - \widehat{\boldsymbol{v}}_t'\widehat{\boldsymbol{\xi}}_T(\boldsymbol{\theta}) - \boldsymbol{z}_t'\widehat{\boldsymbol{\delta}}_T(\boldsymbol{\theta}) - \boldsymbol{w}_t'\boldsymbol{\theta}\right)^2,$$

and $(\widehat{\boldsymbol{\xi}}_T(\widehat{\boldsymbol{\theta}}_T), \widehat{\boldsymbol{\delta}}_T(\widehat{\boldsymbol{\theta}}_T)) = (\widehat{\boldsymbol{\xi}}_T, \widehat{\boldsymbol{\delta}}_T)$.

Here, for each $\boldsymbol{\alpha}$, if we let $y_t(\boldsymbol{\alpha}) := y_t - \boldsymbol{w}_t'\boldsymbol{\alpha}$,

$$(\widehat{\boldsymbol{\beta}}_T(\cdot), \widehat{\boldsymbol{\gamma}}_T(\cdot)) := \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{t=1}^{T} \left(y_t(\cdot) - \boldsymbol{x}_t'\boldsymbol{\beta} - \boldsymbol{z}_t'\boldsymbol{\gamma}\right)^2, \quad \text{and}$$

$$(\widehat{\boldsymbol{\xi}}_T(\cdot), \widehat{\boldsymbol{\delta}}_T(\cdot)) := \arg\min_{\boldsymbol{\xi},\boldsymbol{\delta}} \sum_{t=1}^{T} \left(y_t(\cdot) - \widehat{\boldsymbol{v}}_t'\boldsymbol{\xi} - \boldsymbol{z}_t'\boldsymbol{\delta}\right)^2,$$

so that Lemma A.1 implies that $\widehat{\boldsymbol{\beta}}_T(\cdot) = \widehat{\boldsymbol{\xi}}_T(\cdot)$ and $\widehat{\boldsymbol{\gamma}}_T(\cdot) = \widehat{\boldsymbol{\delta}}_T(\cdot) - \widehat{\boldsymbol{\phi}}_T \widehat{\boldsymbol{\xi}}_T(\cdot)$. Therefore,

$$
\sum_{t=1}^{T} \left(y_t(\cdot) - \widehat{\boldsymbol{v}}_t' \widehat{\boldsymbol{\xi}}_T(\cdot) - \boldsymbol{z}_t' \widehat{\boldsymbol{\delta}}_T(\cdot) \right)^2 = \sum_{t=1}^{T} \left(y_t(\cdot) - (\boldsymbol{x}_t - \widehat{\boldsymbol{\phi}}_T' \boldsymbol{z}_t)' \widehat{\boldsymbol{\xi}}_T(\cdot) - \boldsymbol{z}_t' \widehat{\boldsymbol{\delta}}_T(\cdot) \right)^2
$$

$$
= \sum_{t=1}^{T} \left(y_t(\cdot) - \boldsymbol{x}_t' \widehat{\boldsymbol{\xi}}_T(\cdot) - \boldsymbol{z}_t' (\widehat{\boldsymbol{\delta}}_T(\cdot) - \widehat{\boldsymbol{\phi}}_T \widehat{\boldsymbol{\xi}}_T(\cdot)) \right)^2 = \sum_{t=1}^{T} \left(y_t(\cdot) - \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}_T(\cdot) - \boldsymbol{z}_t' \widehat{\boldsymbol{\gamma}}_T(\cdot) \right)^2,
$$

implying that

$$
\arg\min_{\boldsymbol{\alpha}} \sum_{t=1}^{T} \left(y_t(\boldsymbol{\alpha}) - \widehat{\boldsymbol{v}}_t' \widehat{\boldsymbol{\xi}}_T(\boldsymbol{\alpha}) - \boldsymbol{z}_t' \widehat{\boldsymbol{\delta}}_T(\boldsymbol{\alpha}) \right)^2 = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left(y_t(\boldsymbol{\theta}) - \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}_T(\boldsymbol{\theta}) - \boldsymbol{z}_t' \widehat{\boldsymbol{\gamma}}_T(\boldsymbol{\theta}) \right)^2,
$$

viz., $\widehat{\boldsymbol{\alpha}}_T = \widehat{\boldsymbol{\theta}}_T$. Thus, it follows that $\widehat{\boldsymbol{\beta}}_T(\widehat{\boldsymbol{\alpha}}_T) = \widehat{\boldsymbol{\xi}}_T(\widehat{\boldsymbol{\theta}}_T)$ and $\widehat{\boldsymbol{\gamma}}_T(\widehat{\boldsymbol{\alpha}}_T) = \widehat{\boldsymbol{\delta}}_T(\widehat{\boldsymbol{\theta}}_T) - \widehat{\boldsymbol{\phi}}_T \widehat{\boldsymbol{\xi}}_T(\widehat{\boldsymbol{\theta}}_T)$. That is, $\widehat{\boldsymbol{\beta}}_T = \widehat{\boldsymbol{\xi}}_T$ and $\widehat{\boldsymbol{\gamma}}_T = \widehat{\boldsymbol{\delta}}_T - \widehat{\boldsymbol{\phi}}_T \widehat{\boldsymbol{\xi}}_T$. This completes the proof. ∎

The following lemma shows that $\widehat{\boldsymbol{v}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$ defined in Section 3 suffer from an asymptotically singular matrix problem.

Lemma A.3. *Given Assumption 1,*

(i) $T^{-1} \sum_{t=1}^{T} \widehat{u}_{t-1}^2 \xrightarrow{\mathbb{P}} \sigma_u^2 := \mathbb{E}[u_t^2];$

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

(iii) $\widetilde{\mathbf{D}}_1^{-1} \left(\sum_{t=1}^{T} \widetilde{\boldsymbol{r}}_{t-1} \widetilde{\boldsymbol{r}}_{t-1}' \right) \widetilde{\mathbf{D}}_1^{-1} \Rightarrow \widetilde{\mathscr{M}}_{11}$, *where*

$$
\widetilde{\mathscr{M}}_{11} := \begin{bmatrix} \frac{1}{3}\boldsymbol{\mu}_* \boldsymbol{\mu}_*' & \frac{1}{3}\boldsymbol{\mu}_* & \frac{1}{2}\boldsymbol{\mu}_* \\ \frac{1}{3}\boldsymbol{\mu}_*' & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2}\boldsymbol{\mu}_*' & \frac{1}{2} & 1 \end{bmatrix};
$$

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

(v) $\widetilde{\mathbf{D}}_2^{-1} \left(\sum_{t=1}^{T} \boldsymbol{z}_{2t} \widetilde{\boldsymbol{r}}_{t-1}' \right) \widetilde{\mathbf{D}}_1^{-1} \Rightarrow \widetilde{\mathscr{M}}_{21}$, *where*

$$
\widetilde{\mathscr{M}}_{21} := \begin{bmatrix} \frac{1}{2}\delta_* \boldsymbol{\iota}_{p-1} \boldsymbol{\mu}_*' & \frac{1}{2}\delta_* \boldsymbol{\iota}_{p-1} & \delta_* \boldsymbol{\iota}_{p-1} \\ \frac{1}{2}\boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* \boldsymbol{\mu}_*' & \frac{1}{2}\boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* & \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_* \end{bmatrix};
$$

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

4

(vii) if we let $\widetilde{\mathbf{D}} := diag[T^{1/2}, \widetilde{\mathbf{D}}_1, \widetilde{\mathbf{D}}_2]$, $\widetilde{\mathbf{D}}^{-1}\left(\sum_{t=1}^{T} \widetilde{\boldsymbol{z}}_t \widetilde{\boldsymbol{z}}_t'\right)\widetilde{\mathbf{D}}^{-1} \Rightarrow \widetilde{\mathscr{M}}$, *where*

$$\widetilde{\mathscr{M}} := \begin{bmatrix} \sigma_u^2 & \widetilde{\mathbf{M}}_{u1} & \widetilde{\mathbf{M}}_{u2} \\ \widetilde{\mathbf{M}}_{1u} & \widetilde{\mathscr{M}}_{11} & \widetilde{\mathscr{M}}_{12} \\ \widetilde{\mathbf{M}}_{2u} & \widetilde{\mathscr{M}}_{21} & \widetilde{\mathbf{M}}_{22} \end{bmatrix},$$

which is singular, $\widetilde{\mathscr{M}}_{12} := \widetilde{\mathscr{M}}_{21}'$, $\widetilde{\mathbf{M}}_{u1} := \widetilde{\mathbf{M}}_{1u}'$, *and* $\widetilde{\mathbf{M}}_{u2} := \widetilde{\mathbf{M}}_{2u}'$; *and*

(viii) $\widetilde{\mathbf{D}}_1^{-1}\left(\sum_{t=1}^{T} \widetilde{\boldsymbol{r}}_{t-1}\widetilde{\boldsymbol{r}}_{t-1}'\right)\widetilde{\mathbf{D}}_1^{-1} \xrightarrow{\mathbb{P}} \widetilde{\mathscr{M}}_{11}$, *which is singular.* $\qquad\square$

Both $\widehat{\boldsymbol{v}}_T$ and $\widetilde{\boldsymbol{\omega}}_T$ suffer from the asymptotically singular matrix problem by Lemma A.3 (*vii* and *viii*). Specifically, every column from the second to $(1 + k)$-th columns of $\widetilde{\mathscr{M}}$ is proportional to the $(2 + 2k)$-th column. Likewise, every column from the first to k-th columns in $\widetilde{\mathscr{M}}_{11}$ is proportional to the $(1 + 2k)$-th column of $\widetilde{\mathscr{M}}_{11}$.

We now prove Lemma A.3.

Proof of Lemma A.3. (*i*) This follows from Lemma B.6 (*iv*) given in Section A.2.

(*ii*) This follows from Lemmas B.3 (*v*), B.4 (*v*), and B.5 (*iv*) given in Section A.2.

(*iii*) This follows from Lemmas B.3 (*i, ii*), B.4 (*i, ii*), and B.5 (*i*) given in Section A.2.

(*iv*) This follows from Lemma B.2 (*vii*) given in Section A.2.

(*v*) This follows from Lemma B.2 (*ii, iii,* and *iv*) given in Section A.2.

(*vi*) This follows from Lemma B.2 (*i*) given in Section A.2.

(*vii*) The given weak convergence follows from Lemma A.3 (*i, ii, iii, iv, v,* and *vi*) given in Section A.2, and the singularity follows from the structure of $\widetilde{\mathscr{M}}$.

(*viii*) This follows from Lemma A.3 (*iii*) given in Section A.2. $\qquad\blacksquare$

A.2 Preliminary Lemmas

We next provide preliminary lemmas to prove the main claims efficiently.

Lemma B.1. *Given Assumption 1,* $\mathbf{B}_T(\cdot) := T^{-1/2} \sum_{t=1}^{\lfloor (\cdot)T \rfloor} \boldsymbol{w}_t \Rightarrow \mathscr{B}(\cdot)$. $\qquad\square$

Lemma B.2. *Given Assumption 1,*

(*i*) $T^{-1} \sum_{t=1}^{T} \boldsymbol{z}_{2t}\boldsymbol{z}_{2t}' \xrightarrow{\mathbb{P}} \mathbb{E}[\boldsymbol{z}_{2t}\boldsymbol{z}_{2t}']$;

(*ii*) $T^{-1} \sum_{t=1}^{T} \boldsymbol{z}_{2t} \xrightarrow{\mathbb{P}} [\delta_* \boldsymbol{\iota}_{p-1}', \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*']'$;

(*iii*) $T^{-2} \sum_{t=1}^{T} (t-1)\boldsymbol{z}_{2t} \xrightarrow{\mathbb{P}} [\frac{1}{2}\delta_* \boldsymbol{\iota}_{p-1}', \frac{1}{2}\boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*']'$;

(*iv*) $T^{-2} \sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{t-1}\boldsymbol{z}_{2t}' \xrightarrow{\mathbb{P}} [\frac{1}{2}\delta_* \boldsymbol{\mu}_* \boldsymbol{\iota}_{p-1}', \frac{1}{2}\boldsymbol{\mu}_* \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*']'$;

5

(v) $T^{-3/2} \sum_{t=1}^{T} \widehat{\mathbf{m}}_{t-1} \mathbf{z}_{2t}' \Rightarrow [\delta_* \int \bar{\bar{\boldsymbol{\mathscr{B}}}}_m \boldsymbol{\iota}_{p-1}', \int \bar{\bar{\boldsymbol{\mathscr{B}}}}_m \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*']'$;

(vi) $T^{-2} \sum_{t=1}^{T} y_{t-1} \mathbf{z}_{2t}' \xrightarrow{\mathbb{P}} [\frac{1}{2}\delta_*^2 \boldsymbol{\iota}_{p-1}', \frac{1}{2}\delta_* \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*']'$;

(vii) $T^{-1} \sum_{t=1}^{T} \widehat{u}_{t-1} \mathbf{z}_{2t} \xrightarrow{\mathbb{P}} \mathbb{E}[u_{t-1}\mathbf{z}_{2t}]$. □

Lemma B.3. *Given Assumption 1,*

(i) $T^{-2} \sum_{t=1}^{T} t \to \frac{1}{2}$;

(ii) $T^{-2} \sum_{t=1}^{T} \ddot{\boldsymbol{x}}_t \xrightarrow{\mathbb{P}} \frac{1}{2}\boldsymbol{\mu}_*$;

(iii) $T^{-3/2} \sum_{t=1}^{T} \widehat{\mathbf{m}}_t \Rightarrow \int \bar{\boldsymbol{\mathscr{B}}}_m$;

(iv) $T^{-2} \sum_{t=1}^{T} y_t \xrightarrow{\mathbb{P}} \frac{1}{2}\delta_*$;

(v) $\sum_{t=1}^{T} \widehat{u}_{t-1} \equiv 0$. □

Lemma B.4. *Given Assumption 1,*

(i) $T^{-3} \sum_{t=1}^{T} t^2 \to \frac{1}{3}$;

(ii) $T^{-3} \sum_{t=1}^{T} t\ddot{\boldsymbol{x}}_t \xrightarrow{\mathbb{P}} \frac{1}{3}\boldsymbol{\mu}_*$;

(iii) $\sum_{t=1}^{T} (t-1)\widehat{\mathbf{m}}_{t-1} \equiv \mathbf{0}$;

(iv) $T^{-3} \sum_{t=1}^{T} ty_t \xrightarrow{\mathbb{P}} \frac{1}{3}\delta_*$;

(v) $\sum_{t=1}^{T} (t-1)\widehat{u}_{t-1} \equiv 0$. □

Lemma B.5. *Given Assumption 1,*

(i) $T^{-3} \sum_{t=1}^{T} \ddot{\boldsymbol{x}}_t \ddot{\boldsymbol{x}}_t' \xrightarrow{\mathbb{P}} \frac{1}{3}\boldsymbol{\mu}_*\boldsymbol{\mu}_*'$;

(ii) $T^{-2} \sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{t-1} \widehat{\mathbf{m}}_{t-1}' \Rightarrow \int \boldsymbol{\mathscr{B}}_m \bar{\boldsymbol{\mathscr{B}}}_m'$;

(iii) $T^{-3} \sum_{t=1}^{T} \ddot{\boldsymbol{x}}_t y_t \xrightarrow{\mathbb{P}} \frac{1}{3}\delta_* \boldsymbol{\mu}_*$;

(iv) $\sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{t-1} \widehat{u}_{t-1} \equiv \mathbf{0}$. □

Remark. $\int \boldsymbol{\mathscr{B}}_m \bar{\bar{\boldsymbol{\mathscr{B}}}}_m' = \int \bar{\boldsymbol{\mathscr{B}}}_m \bar{\boldsymbol{\mathscr{B}}}_m' = \int \boldsymbol{\mathscr{B}}_m \boldsymbol{\mathscr{B}}_m' - 3 \int r\boldsymbol{\mathscr{B}}_m \int r\boldsymbol{\mathscr{B}}_m'$ from the definition of $\bar{\boldsymbol{\mathscr{B}}}_m(s) = \boldsymbol{\mathscr{B}}_m(s) - 3s \int r\boldsymbol{\mathscr{B}}_m$. □

Lemma B.6. *Given Assumption 1,*

(i) $T^{-2} \sum_{t=1}^{T} \widehat{\mathbf{m}}_{t-1} \widehat{\mathbf{m}}_{t-1}' \Rightarrow \int \boldsymbol{\mathscr{B}}_m \bar{\boldsymbol{\mathscr{B}}}_m'$;

(ii) $\sum_{t=1}^{T} \widehat{\mathbf{m}}_{t-1} \widehat{u}_{t-1} \equiv \mathbf{0}$;

(iii) $T^{-3} \sum_{t=1}^{T} y_t^2 \xrightarrow{\mathbb{P}} \frac{1}{3}\delta_*^2$;

(iv) $\widehat{\sigma}_{u,T}^2 := T^{-1} \sum_{t=1}^{T} \widehat{u}_t^2 \xrightarrow{\mathbb{P}} \sigma_u^2 := \mathbb{E}[u_t^2]$. □

Lemma B.7. *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) $\sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \Rightarrow 3 \int r\boldsymbol{\mathscr{B}}_m$;

(ii) *for every* t, $\widetilde{\boldsymbol{r}}_t' \widehat{\boldsymbol{v}}_T = \dot{\boldsymbol{r}}_t' \widetilde{\boldsymbol{v}}_T$;

(iii) $T^{-1} \sum_{t=1}^{T} \mathbf{m}_{t-1} u_t \Rightarrow \int \boldsymbol{\mathscr{B}}_m d\mathscr{B}_u + \boldsymbol{\varrho}_{m*}$;

(iv) $T^{-1} \sum_{t=1}^{T} \mathbf{m}_{t-1} e_t \Rightarrow \int \boldsymbol{\mathscr{B}}_m d\mathscr{B}_e$;

(v) $\dot{\sigma}_{e,T}^2 := T^{-1} \sum_{t=1}^{T} (\Delta y_t - \dot{\boldsymbol{z}}_t' \dot{\boldsymbol{\tau}}_T)^2 \xrightarrow{\mathbb{P}} \sigma_e^2 := \mathbb{E}[e_t^2]$. □

We now prove the preliminary Lemmas B.1 to B.7.

Proof of Lemma B.1. This trivially follows theorem 7.30 of White (2001). ∎

Proof of Lemma B.2. (*i*) It holds by the ergodic theorem.

(*ii*) It holds by the ergodic theorem and the fact that $\mathbb{E}[\boldsymbol{z}_{2t}] = [\delta_* \boldsymbol{\iota}'_{p-1}, \boldsymbol{\iota}'_q \otimes \boldsymbol{\mu}'_*]'$ because $\delta_* \boldsymbol{\iota}_{p-1} = \mathbb{E}[\Delta \boldsymbol{y}_{t-}], \boldsymbol{\iota}'_q \otimes \boldsymbol{\mu}_*^{+'} = \mathbb{E}[(\Delta \boldsymbol{x}_t^{+'}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{+'})]$, and $\boldsymbol{\iota}'_q \otimes \boldsymbol{\mu}_*^{-'} = \mathbb{E}[(\Delta \boldsymbol{x}_t^{-'}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{-'})]$.

(*iii*) We note that $T^{-2} \sum_{t=1}^T (t-1) \boldsymbol{z}_{2t} = T^{-1} \sum_{t=1}^T ((t-1)/T) \boldsymbol{z}_{2t}$, and we let $\boldsymbol{q}_t := ((t-1)/T) \boldsymbol{z}_{2t}$ for notational simplicity, which is a heterogeneous process. Therefore, $T^{-1} \sum_{t=1}^T (\boldsymbol{q}_t - \mathbb{E}[\boldsymbol{q}_t]) \xrightarrow{\mathbb{P}} \mathbf{0}$. In addition, $T^{-1} \sum_{t=1}^T \mathbb{E}[\boldsymbol{q}_t] = T^{-2} \sum_{t=1}^T (t-1) \mathbb{E}[\boldsymbol{z}_{2t}] \to \frac{1}{2} \mathbb{E}[\boldsymbol{z}_{2t}]$, implying that $T^{-1} \sum_{t=1}^T \boldsymbol{q}_t \xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[\boldsymbol{z}_{2t}]$ by White (2001, theorem 3.47), given the DGP condition in Assumption 1. We further note that $\mathbb{E}[\boldsymbol{z}_{2t}] = [\delta_* \boldsymbol{\iota}'_{p-1}, \boldsymbol{\iota}'_q \otimes \boldsymbol{\mu}'_*]'$, leading to the desired result.

(*iv*) We first note that $\ddot{\boldsymbol{x}}_{t-1} = \boldsymbol{\mu}_*(t-1) + \mathbf{m}_{t-1}$. Therefore, $\sum_{t=1}^T \ddot{\boldsymbol{x}}_{t-1} \boldsymbol{z}'_{2t} = \sum_{t=1}^T \boldsymbol{\mu}_*(t-1) \boldsymbol{z}'_{2t} + \sum_{t=1}^T \mathbf{m}_{t-1} \boldsymbol{z}'_{2t}$. The proof of Lemma B.2 (*iii*) already shows that $T^{-2} \sum_{t=1}^T (t-1) \boldsymbol{z}_{2t} \xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[\boldsymbol{z}_{2t}]$. Furthermore, $\sum_{t=1}^T \mathbf{m}_{t-1} \boldsymbol{z}'_{2t} = O_{\mathbb{P}}(T^{3/2})$ as shown in the proof of Lemma B.2 (*v*). Therefore, it follows that $T^{-2} \sum_{t=1}^T \ddot{\boldsymbol{x}}_{t-1} \boldsymbol{z}'_{2t} \xrightarrow{\mathbb{P}} \frac{1}{2} \boldsymbol{\mu}_* \mathbb{E}[\boldsymbol{z}_{2t}]$, as desired.

(*v*) Note that $\widehat{\mathbf{m}}_{t-1} = \mathbf{m}_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1)$. Therefore, $\sum_{t=1}^T \widehat{\mathbf{m}}_{t-1} \boldsymbol{z}'_{2t} = \sum_{t=1}^T \mathbf{m}_{t-1} \boldsymbol{z}'_{2t} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \sum_{t=1}^T (t-1) \boldsymbol{z}'_{2t}$. We here note that $\sum_{t=1}^T \mathbf{m}_{t-1} \boldsymbol{z}'_{2t} = \sum_{t=1}^T \mathbf{m}_{t-1} \mathbb{E}[\boldsymbol{z}'_{2t}] + \sum_{t=1}^T \mathbf{m}_{t-1} (\boldsymbol{z}'_{2t} - \mathbb{E}[\boldsymbol{z}'_{2t}])$. The proof of Lemma A.3 (*iv*) implies that $T^{-3/2} \sum_{t=1}^T \mathbf{m}_{t-1} \Rightarrow \int \boldsymbol{\mathcal{B}}_m$, and $\sum_{t=1}^T \mathbf{m}_{t-1} (\boldsymbol{z}'_{2t} - \mathbb{E}[\boldsymbol{z}'_{2t}]) = o_{\mathbb{P}}(T^{3/2})$ by noting that $\mathbf{m}_{t-1} = O_{\mathbb{P}}(T^{1/2})$ and $\sum_{t=1}^T (\boldsymbol{z}'_{2t} - \mathbb{E}[\boldsymbol{z}'_{2t}]) = O_{\mathbb{P}}(T^{1/2})$. Therefore, $T^{-3/2} \sum_{t=1}^T \mathbf{m}_{t-1} \boldsymbol{z}'_{2t} \Rightarrow \int \boldsymbol{\mathcal{B}}_m \mathbb{E}[\boldsymbol{z}'_{2t}]$. Next, $(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \sum_{t=1}^T (t-1) \boldsymbol{z}'_{2t} = \sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) T^{-2} \sum_{t=1}^T (t-1) \boldsymbol{z}'_{2t}$, and Lemma B.7 (*i*) implies that $\sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \Rightarrow 3 \int r \boldsymbol{\mathcal{B}}_m$. In addition to this, Lemma B.2 (*iii*) shows that $T^{-2} \sum_{t=1}^T (t-1) \boldsymbol{z}'_{2t} \xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[\boldsymbol{z}'_{2t}]$. Therefore, $T^{-3/2}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \sum_{t=1}^T (t-1) \boldsymbol{z}'_{2t} \Rightarrow \frac{3}{2} \int r \boldsymbol{\mathcal{B}}_m$. Hence, if we combine all these, it follows that $\sum_{t=1}^T \widehat{\mathbf{m}}_{t-1} \boldsymbol{z}'_{2t} \Rightarrow (\int \boldsymbol{\mathcal{B}}_m - \frac{3}{2} \int r \boldsymbol{\mathcal{B}}_m) \mathbb{E}[\boldsymbol{z}'_{2t}] = \int \bar{\boldsymbol{\mathcal{B}}}_m \mathbb{E}[\boldsymbol{z}'_{2t}]$.

(*vi*) Note that $y_{t-1} = \delta_*(t-1) + \sum_{j=1}^{t-1} d_j$. Therefore, $\sum_{t=1}^T y_{t-1} \boldsymbol{z}'_{2t} = \sum_{t=1}^T \delta_*(t-1) \boldsymbol{z}'_{2t} + \sum_{t=1}^T (\sum_{j=1}^{t-1} d_j) \boldsymbol{z}'_{2t}$. The proof of Lemma B.2 (*iii*) already showed that $T^{-2} \sum_{t=1}^T (t-1) \boldsymbol{z}_{2t} \xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[\boldsymbol{z}_{2t}]$. Furthermore, we note that $\sum_{t=1}^T (\sum_{j=1}^{t-1} d_j) \boldsymbol{z}'_{2t} = O_{\mathbb{P}}(T^{3/2})$. Therefore, $T^{-2} \sum_{t=1}^T y_{t-1} \boldsymbol{z}'_{2t} \xrightarrow{\mathbb{P}} \frac{1}{2} \delta_* \mathbb{E}[\boldsymbol{z}_{2t}]$, as desired.

(*vii*) In the proof of Lemma B.6 (*iv*), we show that $\widehat{u}_t = u_t + O_{\mathbb{P}}(T^{-1/2})$. Therefore, it follows that $T^{-1} \sum_{t=1}^T \widehat{u}_{t-1} \boldsymbol{z}_{2t} = T^{-1} \sum_{t=1}^T u_{t-1} \boldsymbol{z}_{2t} + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \mathbb{E}[u_{t-1} \boldsymbol{z}_{2t}]$ by the ergodic theorem. This completes the proof. ∎

Proof of Lemma B.3. (*i*) $\sum_{t=1}^T t = T(T+1)/2$, leading to the desired result.

(*ii*) Note that $\ddot{\boldsymbol{x}}_t = \boldsymbol{\mu}_* t + \mathbf{m}_t$, so that $\sum_{t=1}^T \ddot{\boldsymbol{x}}_t = \boldsymbol{\mu}_* \sum_{t=1}^T t + \sum_{t=1}^T \mathbf{m}_t$. Here, $T^{-2} \sum_{t=1}^T t \to \frac{1}{2}$

by Lemma B.3 (i), and Lemma B.1 implies that $T^{-3/2}\sum_{t=1}^T \mathbf{m}_t = T^{-1}\sum_{t=1}^T \frac{1}{\sqrt{T}}\mathbf{m}_t = \int \mathbf{B}_m(r)dr \Rightarrow$ $\int \boldsymbol{\mathcal{B}}_m$. Therefore, $T^{-3/2}\sum_{t=1}^T \ddot{\boldsymbol{x}}_t = \boldsymbol{\mu}_* T^{-3/2}\sum_{t=1}^T t + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} \frac{1}{2}\boldsymbol{\mu}_*$.

(iii) Note that $\widehat{\mathbf{m}}_t = \mathbf{m}_t + (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)t$. Thus, $T^{-3/2}\sum_{t=1}^T \widehat{\mathbf{m}}_t = T^{-3/2}\sum_{t=1}^T \mathbf{m}_t + \sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)T^{-2}\sum_{t=1}^T t$. Lemma B.7 (i) implies that $\sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \Rightarrow 3\int r\boldsymbol{\mathcal{B}}_m$. Lemma B.3 (i) implies that $T^{-2}\sum_{t=1}^T t \to \frac{1}{2}$. In addition to these, $T^{-3/2}\sum_{t=1}^T \mathbf{m}_t = T^{-1}\sum_{t=1}^T \frac{1}{\sqrt{T}}\mathbf{m}_t = \int \mathbf{B}_m(r)dr \Rightarrow \int \boldsymbol{\mathcal{B}}_m$ by Lemma B.1. Thus, $T^{-3/2}\sum_{t=1}^T \widehat{\mathbf{m}}_t \Rightarrow \int \boldsymbol{\mathcal{B}}_m(r) - 3\int rdr\int s\boldsymbol{\mathcal{B}}_m(s)ds = \int \bar{\boldsymbol{\mathcal{B}}}_m$ by the definition of $\bar{\boldsymbol{\mathcal{B}}}_m(\cdot)$, viz., $\bar{\boldsymbol{\mathcal{B}}}_m(\cdot) = \boldsymbol{\mathcal{B}}_m(\cdot) - 3(\cdot)\int s\boldsymbol{\mathcal{B}}_m(s)ds$.

(iv) From (5), $T^{-2}\sum_{t=1}^T y_t = \delta_* T^{-2}\sum_{t=1}^T t + T^{-2}\sum_{t=1}^T \sum_{j=1}^t d_j = \frac{1}{2}\delta_* + o_{\mathbb{P}}(1)$.

(v) We note that $\widehat{u}_{t-1} := y_{t-1} - \widetilde{\boldsymbol{r}}_{t-1}'\widehat{\boldsymbol{v}}_T$, so that $\sum_{t=1}^T \widehat{u}_{t-1}\widetilde{\boldsymbol{r}}_{t-1} \equiv \mathbf{0}$. We here note that $\widetilde{\boldsymbol{r}}_{t-1} := [\ddot{\boldsymbol{x}}_{t-1}', (t-1), 1]'$. This completes the proof. ∎

Proof of Lemma B.4. (i) $\sum_{t=1}^T t^2 = T(T+1)(2T+1)/6$, leading to the desired result.

(ii) $T^{-3}\sum_{t=1}^T t\ddot{\boldsymbol{x}}_t = \boldsymbol{\mu}_* T^{-3}\sum_{t=1}^T t^2 + T^{-3}\sum_{t=1}^T t\mathbf{m}_t$. Here, it follows that $T^{-3}\sum_{t=1}^T t^2 \to \frac{1}{3}$ and $T^{-5/2}\sum_{t=1}^T t\mathbf{m}_t = T^{-1}\sum_{t=1}^T (\frac{t}{T})\frac{1}{\sqrt{T}}\mathbf{m}_t = \int r\mathbf{B}_{mT}(r)dr \Rightarrow \int r\boldsymbol{\mathcal{B}}_m$. Therefore, $T^{-3}\sum_{t=1}^T t\ddot{\boldsymbol{x}}_t = \frac{1}{3}\boldsymbol{\mu}_* + o_{\mathbb{P}}(1)$.

(iii) Note that $\widehat{\mathbf{m}}_t = \ddot{\boldsymbol{x}}_t - t\widehat{\boldsymbol{\mu}}_T$ and $\widehat{\boldsymbol{\mu}}_T = (\sum_{t=1}^{T-1} t^2)^{-1}\sum_{t=1}^{T-1} t\ddot{\boldsymbol{x}}_t$. Therefore, $\sum_{t=1}^T (t-1)\widehat{\mathbf{m}}_{t-1} \equiv \mathbf{0}$.

(iv) From (5), $T^{-3}\sum_{t=1}^T ty_t = \delta_* T^{-3}\sum_{t=1}^T t^2 + T^{-3}\sum_{t=1}^T t\sum_{j=1}^t d_j = \frac{1}{3}\delta_* + o_{\mathbb{P}}(1)$.

(v) The proof of Lemma B.3 (v) already shows the given claim. ∎

Proof of Lemma B.5. (i) Using the fact that $\ddot{\boldsymbol{x}}_t = \boldsymbol{\mu}_* t + \mathbf{m}_t$, $\sum_{t=1}^T \ddot{\boldsymbol{x}}_t\ddot{\boldsymbol{x}}_t' = \sum_{t=1}^T (\boldsymbol{\mu}_* t + \mathbf{m}_t)(\boldsymbol{\mu}_* t + \mathbf{m}_t)' = \sum_{t=1}^T \boldsymbol{\mu}_*\boldsymbol{\mu}_*' t^2 + \sum_{t=1}^T \boldsymbol{\mu}_* \mathbf{m}_t' t + \sum_{t=1}^T \mathbf{m}_t\boldsymbol{\mu}_*' t + \sum_{t=1}^T \mathbf{m}_t\mathbf{m}_t'$. Here, $T^{-3}\sum_{t=1}^T \boldsymbol{\mu}_*\boldsymbol{\mu}_*' t^2 \to \frac{1}{3}\boldsymbol{\mu}_*\boldsymbol{\mu}_*'$ by Lemma B.4 (i), and $T^{-5/2}\sum_{t=1}^T \boldsymbol{\mu}_* \mathbf{m}_t' t \Rightarrow \boldsymbol{\mu}_*\int r\boldsymbol{\mathcal{B}}_m$ as shown in the proof of Lemma B.4 (ii). Furthermore, $T^{-2}\sum_{t=1}^T \mathbf{m}_t\mathbf{m}_t' = T^{-1}\sum_{t=1}^T \frac{1}{\sqrt{T}}\mathbf{m}_t\frac{1}{\sqrt{T}}\mathbf{m}_t' = \int \mathbf{B}_{mT}(r)\mathbf{B}_{mT}(r)'dr \Rightarrow \int \boldsymbol{\mathcal{B}}_n\boldsymbol{\mathcal{B}}_m'$. Thus, $T^{-3}\sum_{t=1}^T \ddot{\boldsymbol{x}}_t\ddot{\boldsymbol{x}}_t' \overset{\mathbb{P}}{\to} \frac{1}{3}\boldsymbol{\mu}_*\boldsymbol{\mu}_*'$.

(ii) We first note that $\sum_{t=1}^T \ddot{\boldsymbol{x}}_{t-1}\widehat{\mathbf{m}}_{t-1}' = \sum_{t=1}^T (\boldsymbol{\mu}_*(t-1) + \mathbf{m}_{t-1})\widehat{\mathbf{m}}_{t-1}' = \boldsymbol{\mu}_*\sum_{t=1}^T (t-1)\widehat{\mathbf{m}}_{t-1}' + \sum_{t=1}^T \mathbf{m}_{t-1}\widehat{\mathbf{m}}_{t-1}' = \sum_{t=1}^T \mathbf{m}_{t-1}\widehat{\mathbf{m}}_{t-1}'$ by Lemma B.4 (iii). We further note that $\sum_{t=1}^T \mathbf{m}_{t-1}\widehat{\mathbf{m}}_{t-1}' = \sum_{t=1}^T \mathbf{m}_{t-1}(\mathbf{m}_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1))'$ using the fact that $\widehat{\mathbf{m}}_t = \mathbf{m}_t - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)t$. Here, we note that $T^{-2}\sum_{t=1}^T \mathbf{m}_t\mathbf{m}_t' \Rightarrow \int \boldsymbol{\mathcal{B}}_m\boldsymbol{\mathcal{B}}_m'$ as shown in the proof of Lemma B.5 (i), and $T^{-2}\sum_{t=1}^T t\mathbf{m}_t(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)' = T^{-5/2}\sum_{t=1}^T t\mathbf{m}_t\sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)' \Rightarrow 3\int r\boldsymbol{\mathcal{B}}_m\int r\boldsymbol{\mathcal{B}}_m$ by Lemma B.7 (ii) and the fact that $T^{-5/2}\sum_{t=1}^T t\mathbf{m}_t = T^{-1}\sum_{t=1}^T \frac{t}{T}\frac{1}{\sqrt{T}}\mathbf{m}_t = \int r\mathbf{B}_m(r)dr \Rightarrow \int r\boldsymbol{\mathcal{B}}_m$ as shown in the proof of Lemma B.4 (ii). Therefore, $T^{-2}\sum_{t=1}^T \ddot{\boldsymbol{x}}_{t-1}\widehat{\mathbf{m}}_{t-1}' \Rightarrow \int \boldsymbol{\mathcal{B}}_m(\boldsymbol{\mathcal{B}}_m - 3r\int s\boldsymbol{\mathcal{B}}_m)' = \int \boldsymbol{\mathcal{B}}_m\bar{\boldsymbol{\mathcal{B}}}_m'$. We further note that $\int \boldsymbol{\mathcal{B}}_m\bar{\boldsymbol{\mathcal{B}}}_m' = \int \bar{\boldsymbol{\mathcal{B}}}_m\bar{\boldsymbol{\mathcal{B}}}_m'$ by the definition of $\boldsymbol{\mathcal{B}}_m\bar{\boldsymbol{\mathcal{B}}}_m$ and the fact that $\int r^2 = \frac{1}{3}$.

(iii) Using the fact that $\ddot{\boldsymbol{x}}_t = \boldsymbol{\mu}_* t + \mathbf{m}_t$ and $y_t = \delta_* t + \sum_{j=1}^t d_j$, $\sum_{t=1}^T \ddot{\boldsymbol{x}}_t y_t = \boldsymbol{\mu}_*\delta_*\sum_{t=1}^T t^2 + \delta_*\sum_{t=1}^T t\mathbf{m}_t + \boldsymbol{\mu}_*\sum_{t=1}^T t\sum_{j=1}^t d_j + \sum_{t=1}^T \mathbf{m}_t\sum_{j=1}^t d_j$. Here, $T^{-3}\sum_{t=1}^T t^2 \to \frac{1}{3}$ by Lemma B.4 (i),

and $T^{-5/2} \sum_{t=1}^{T} t\mathbf{m}_t \Rightarrow \int r\boldsymbol{\mathcal{B}}_m$ as shown in the proof of Lemma B.4 (*ii*). Furthermore, we note that $\sum_{t=1}^{T} t \sum_{j=1}^{t} d_j = O_{\mathbb{P}}(T^{5/2})$ and $\sum_{t=1}^{T} \mathbf{m}_t \sum_{j=1}^{t} d_j = O_{\mathbb{P}}(T^2)$. Therefore, $T^{-3} \sum_{t=1}^{T} \ddot{\boldsymbol{x}}_t y_t \xrightarrow{\mathbb{P}} \frac{1}{3} \delta_* \boldsymbol{\mu}_*$.

(*iv*) The proof of Lemma B.3 (*v*) already shows the given claim. ∎

Proof of Lemma B.6. (*i*) We first note that $\sum_{t=1}^{T} \widehat{\mathbf{m}}_{t-1} \widehat{\mathbf{m}}'_{t-1} = \sum_{t=1}^{T} (\mathbf{m}_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1)) \widehat{\mathbf{m}}'_{t-1} = \sum_{t=1}^{T} \mathbf{m}_{t-1} \widehat{\mathbf{m}}'_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \sum_{t=1}^{T} (t-1) \widehat{\mathbf{m}}'_{t-1} = \sum_{t=1}^{T} \mathbf{m}_{t-1} \widehat{\mathbf{m}}'_{t-1}$ by Lemma B.4 (*iii*). We next note that $\sum_{t=1}^{T} \mathbf{m}_{t-1} \widehat{\mathbf{m}}'_{t-1} = \sum_{t=1}^{T} \mathbf{m}_{t-1} (\mathbf{m}_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1))'$ using the fact that $\widehat{\mathbf{m}}_t = \mathbf{m}_t - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)t$. Here, $T^{-2} \sum_{t=1}^{T} \mathbf{m}_t \mathbf{m}'_t \Rightarrow \int \boldsymbol{\mathcal{B}}_m \boldsymbol{\mathcal{B}}'_m$ as shown in the proof of Lemma B.5 (*i*), and $T^{-2} \sum_{t=1}^{T} t\mathbf{m}_t (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)' \Rightarrow 3 \int r\boldsymbol{\mathcal{B}}_m \int r\boldsymbol{\mathcal{B}}_m$ as shown in the proof of Lemma B.5 (*ii*). Therefore, $T^{-2} \sum_{t=1}^{T} \widehat{\mathbf{m}}_{t-1} \widehat{\mathbf{m}}'_{t-1} \Rightarrow \int \boldsymbol{\mathcal{B}}_m (\boldsymbol{\mathcal{B}}_m - 3r \int s\boldsymbol{\mathcal{B}}_m)' = \int \boldsymbol{\mathcal{B}}_m \bar{\boldsymbol{\mathcal{B}}}'_m$.

(*ii*) As Lemma B.7 (*iii*) shows, for each t, $\widetilde{\boldsymbol{r}}'_t \widehat{\boldsymbol{v}}_T = \dot{\boldsymbol{r}}'_t \widetilde{\boldsymbol{v}}_T$, so that $\widehat{u}_t := y_t - \widetilde{\boldsymbol{r}}'_t \widehat{\boldsymbol{v}}_T = y_t - \dot{\boldsymbol{r}}'_t \widetilde{\boldsymbol{v}}_T$. We further note that $\widetilde{\boldsymbol{v}}_T := (\sum_{t=1}^{T} \dot{\boldsymbol{r}}_{t-1} \dot{\boldsymbol{r}}'_{t-1})^{-1} \sum_{t=1}^{T} \dot{\boldsymbol{r}}_{t-1} y_{t-1}$, so that $\sum_{t=1}^{T} \dot{\boldsymbol{r}}_{t-1} \widehat{u}_{t-1} = \mathbf{0}$. We now note that $\dot{\boldsymbol{r}}_{t-1} := [\widehat{\mathbf{m}}'_{t-1}, (t-1), 1]'$, leading to that $\sum_{t=1}^{T} \widehat{\mathbf{m}}_{t-1} \widehat{u}_{t-1} = \mathbf{0}$.

(*iii*) From (5), $y_t = \delta_* t + \sum_{j=1}^{t} d_j$. We here note that $\sum_{t=1}^{T} y_t^2 = \sum_{t=1}^{T} (\delta_* t + \sum_{j=1}^{t} d_j)^2 = \sum_{t=1}^{T} \delta_*^2 t^2 + 2 \sum_{t=1}^{T} \delta_* t \sum_{j=1}^{t} d_j + \sum_{t=1}^{T} (\sum_{j=1}^{t} d_j)^2$. Furthermore, $T^{-3} \sum_{t=1}^{T} t^2 \to \frac{1}{3}$ by Lemma B.4 (*i*), $\sum_{t=1}^{T} \delta_* t \sum_{j=1}^{t} d_j = O_{\mathbb{P}}(T^{5/2})$, and $\sum_{t=1}^{T} (\sum_{j=1}^{t} d_j)^2 = O_{\mathbb{P}}(T^2)$, implying that $T^{-3} \sum_{t=1}^{T} y_t^2 \xrightarrow{\mathbb{P}} \frac{1}{3} \delta_*^2$.

(*iv*) Note that $\widehat{u}_t := y_t - \widetilde{\boldsymbol{r}}'_t \widehat{\boldsymbol{v}}_T$, and $\widetilde{\boldsymbol{r}}'_t \widehat{\boldsymbol{v}}_T = \ddot{\boldsymbol{x}}'_t \widehat{\boldsymbol{\beta}}_T + t\widehat{\zeta}_T + \widehat{\nu}_T$ with $\ddot{\boldsymbol{x}}_t = \widehat{\mathbf{m}}_t + \widehat{\boldsymbol{\mu}}_T t$. Furthermore, $y_t = \boldsymbol{\beta}'_*(\widehat{\mathbf{m}}_t + \widehat{\boldsymbol{\mu}}_T t) + \zeta_* t + \nu_* + u_t$ using (6). Hence, $\widehat{u}_t = u_t - (\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)' \widehat{\mathbf{m}}_t - (\widehat{\vartheta}_T - \vartheta_{T*})t - (\widehat{\nu}_T - \nu_*)$. We now note that Lemma 3, $\widehat{\mathbf{m}}_t = O_{\mathbb{P}}(T^{1/2})$, and $t = O(T)$ imply that $\widehat{u}_t = u_t + O_{\mathbb{P}}(T^{-1/2})$. Therefore, $T^{-1} \sum_{t=1}^{T} \widehat{u}_t^2 = T^{-1} \sum_{t=1}^{T} u_t^2 + o_{\mathbb{P}}(1)$, and $T^{-1} \sum_{t=1}^{T} u_t^2 \xrightarrow{\mathbb{P}} \mathbb{E}[u_t^2]$ by the ergodic theorem, implying that $T^{-1} \sum_{t=1}^{T} \widehat{u}_t^2 \xrightarrow{\mathbb{P}} \sigma_u^2 := \mathbb{E}[u_t^2]$. ∎

Proof of Lemma B.7. (*i*) Note that $\widehat{\boldsymbol{\mu}}_T = (\sum_{t=1}^{T-1} t^2)^{-1} \sum_{t=1}^{T-1} t\ddot{\boldsymbol{x}}_t$ and $\ddot{\boldsymbol{x}}_t = \boldsymbol{\mu}_* t + \mathbf{m}_t$. Therefore, $\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_* = (\sum_{t=1}^{T-1} t^2)^{-1} \sum_{t=1}^{T-1} t\mathbf{m}_t$, so that $\sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) = (T^{-3} \sum_{t=1}^{T-1} t^2)^{-1} T^{-5/2} \sum_{t=1}^{T-1} t\mathbf{m}_t$. Lemma B.4 (*i*) implies that $T^{-3} \sum_{t=1}^{T} t^2 \to \frac{1}{3}$. Furthermore, $T^{-5/2} \sum_{t=1}^{T-1} t\mathbf{m}_t = T^{-1} \sum_{t=1}^{T-1} \frac{t}{T} \frac{1}{\sqrt{T}} \mathbf{m}_t = \int r\mathbf{B}_{mT}(r)dr \Rightarrow \int r\boldsymbol{\mathcal{B}}_m$. Hence, $\sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \Rightarrow 3 \int r\boldsymbol{\mathcal{B}}_m$.

(*ii*) Note that $\widetilde{\boldsymbol{r}}'_t \widehat{\boldsymbol{v}}_T = \ddot{\boldsymbol{x}}'_t \widehat{\boldsymbol{\beta}}_T + t\widehat{\zeta}_T + \widehat{\nu}_T$ with $\ddot{\boldsymbol{x}}_t = \widehat{\mathbf{m}}_t + \widehat{\boldsymbol{\mu}}_T t$. Therefore, $\widetilde{\boldsymbol{r}}'_t \widehat{\boldsymbol{v}}_T = \widehat{\mathbf{m}}'_t \widehat{\boldsymbol{\beta}}_T + t(\widehat{\boldsymbol{\mu}}'_T \widehat{\boldsymbol{\beta}}_T + \widehat{\zeta}_T) + \widehat{\nu}_T = \widehat{\mathbf{m}}'_t \widehat{\boldsymbol{\beta}}_T + t\widetilde{\vartheta}_T + \widetilde{\nu}_T = \dot{\boldsymbol{r}}'_t \widetilde{\boldsymbol{v}}_T$, where the second last equality holds by Proposition 2, and the last equality follows from the definition of $\dot{\boldsymbol{r}}_t$.

(*iii*) Note that $T^{-1} \sum_{t=1}^{T} \mathbf{m}_{t-1} u_t = T^{-1} \sum_{t=1}^{T} T^{-1/2} \mathbf{m}_{t-1} \sqrt{T}(B_{uT}(t/T) - B_{uT}((t-1)/T)) = \int \mathbf{B}_{mT}(r)dB_{uT}(r)$. We here note that $\int \mathbf{B}_{mT}(r)dB_{uT}(r) \Rightarrow \int \boldsymbol{\mathcal{B}}_m d\mathcal{B}_u + \boldsymbol{\varrho}_{m*}$ by applying theorem 4 of de Jong and Davidson (2000).

(*iv*) We first note that $T^{-1} \sum_{t=1}^{T} \mathbf{m}_{t-1} e_t = T^{-1} \sum_{t=1}^{T} T^{-1/2} \mathbf{m}_{t-1} \sqrt{T}(B_{eT}(t/T) - B_{eT}((t-1)/T)) = \int \mathbf{B}_{mT}(r)dB_{eT}(r)$. Note that $\int \mathbf{B}_{mT}(r)dB_{eT}(r) \Rightarrow \int \boldsymbol{\mathcal{B}}_m d\mathcal{B}_e$ by applying theorem 4 of de Jong and

Davidson (2000) and noting that $\mathbb{E}[s_\tau e_t] = 0$ for each $\tau < t$.

(*v*) Note that if we let $\dot{e}_t := \Delta y_t - \dot{z}_t \dot{\tau}_T$, it follows that $\dot{e}_t = -\dot{z}'_t(\dot{\tau}_T - \tau_{T*}) + e_t$ from the fact that $\Delta y_t = \dot{z}'_t \tau_{T*} + e_t$, implying that

$$\sum_{t=1}^{T} \dot{e}_t^2 = (\dot{\tau}_T - \tau_{T*})'\dot{\mathbf{D}}\left(\dot{\mathbf{D}}^{-1}\sum_{t=1}^{T}\dot{z}_t\dot{z}'_t\dot{\mathbf{D}}^{-1}\right)\dot{\mathbf{D}}(\dot{\tau}_T - \tau_{T*}) - 2\left(\sum_{t=1}^{T}e_t\dot{z}_t\right)\dot{\mathbf{D}}^{-1}\dot{\mathbf{D}}(\dot{\tau}_T - \tau_{T*}) + \sum_{t=1}^{T}e_t^2.$$

We examine the asymptotic behavior of each element on the right side. First, Lemmas 3 (*vi*) and 4 (*i*) imply that $\dot{\mathbf{D}}(\dot{\tau}_T - \tau_{T*}) = O_{\mathbb{P}}(1)$ and $\dot{\mathbf{D}}^{-1}\sum_{t=1}^{T}\dot{z}_t\dot{z}'_t\dot{\mathbf{D}}^{-1} = O_{\mathbb{P}}(1)$. Second, from the definitions of $\dot{z}_t$ and $\dot{\mathbf{D}}$, $\sum_{t=1}^{T}e_t\dot{z}_t\dot{\mathbf{D}}^{-1} = [T^{-1/2}\sum_{t=1}^{T}e_t\widehat{u}_{t-1}, T^{-1}\sum_{t=1}^{T}e_t\widehat{\mathbf{m}}'_{t-1}, T^{-3/2}\sum_{t=1}^{T}e_t(t-1), T^{-1/2}\sum_{t=1}^{T}e_t, T^{-1/2}\sum_{t=1}^{T}e_t z'_{2t}]'$. We verify that each element on the right side is $O_{\mathbb{P}}(1)$. We first note that $T^{-3/2}\sum_{t=1}^{T}e_t(t-1) = O_{\mathbb{P}}(1)$, $T^{-1/2}\sum_{t=1}^{T}e_t = O_{\mathbb{P}}(1)$, and $T^{-1/2}\sum_{t=1}^{T}e_t z_{2t} = O_{\mathbb{P}}(1)$ by the martingale difference CLT based upon the fact that $\{e_t, \mathscr{F}_t\}$ is an MDA. In addition, $T^{-1}\sum_{t=1}^{T}e_t\widehat{\mathbf{m}}_{t-1} = T^{-1}\sum_{t=1}^{T}e_t\mathbf{m}_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)T^{-1}\sum_{t=1}^{T}e_t(t-1)$ by noting that $\widehat{\mathbf{m}}_{t-1} = \mathbf{m}_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1)$ as given in the proof of Lemma B.6(*i*). Here, Lemmas B.7 (*i* and *iv*) imply that $(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) = O_{\mathbb{P}}(T^{-1/2})$ and $T^{-1}\sum_{t=1}^{T}e_t\mathbf{m}_{t-1} = O_{\mathbb{P}}(1)$, respectively, so that $T^{-1}\sum_{t=1}^{T}e_t\widehat{\mathbf{m}}_{t-1} = O_{\mathbb{P}}(1)$. Finally, $T^{-1/2}\sum_{t=1}^{T}e_t\widehat{u}_{t-1} = T^{-1/2}\sum_{t=1}^{T}e_t u_{t-1} + o_{\mathbb{P}}(1)$ using the fact that $\widehat{u}_{t-1} = u_{t-1} + O_{\mathbb{P}}(T^{-1/2})$ as given in the proof of Lemma B.6 (*iv*). All these facts imply that $\sum_{t=1}^{T}e_t\dot{z}_t\dot{\mathbf{D}}^{-1} = O_{\mathbb{P}}(1)$. By these two facts, it follows that $\sum_{t=1}^{T}\dot{e}_t^2 = \sum_{t=1}^{T}e_t^2 + O_{\mathbb{P}}(1)$, implying that $\dot{\sigma}_{e,T}^2 := T^{-1}\sum_{t=1}^{T}\dot{e}_t^2 = T^{-1}\sum_{t=1}^{T}e_t^2 + O_{\mathbb{P}}(T^{-1})$. The desired result follows from the ergodic theorem, and this completes the proof. ∎

A.3   Proofs

Proof of Lemma 1. (*i*) This follows from Lemmas B.3 (*i*, *ii*, *iv*), B.4 (*i*, *ii*, *iv*), B.5 (*i*, *iii*), B.6 (*iii*), and the remarks below Lemmas B.3, B.4, and B.5.

(*ii*) This follows from Lemma B.2 (*i*).

(*iii*) This follows from Lemmas B.2 (*ii*, *iii*, *iv*, *vi*) and the remark below Lemma B.2.

(*iv*) This follows from Lemmas 1 (*i*, *ii*, *iii*) and the structure of $\mathbf{M}$. ∎

Proof of Lemma 2. (*i*) This follows from Lemmas B.3 (*v*), B.4 (*v*), and B.6 (*ii*).

(*ii*) This follows from Lemmas B.3 (*i*, *ii*, *iii*), B.4 (*i*, *ii*, *iii*), and B.6 (*i*).

(*iii*) This follows from Lemma B.2 (*vii*).

(*iv*) This follows from Lemma B.2 (*ii*, *iii*, *iv*, and *vi*).

(*v*) This follows from Lemma B.2 (*i*).

(*vi*) This follows from Lemma A.3 (*i*) and Lemmas 2 (*i*, *ii*, *iii*, *iv*, *v*).

(*vii*) This follows from Lemma 2 (*ii*). ∎

Proof of Lemma 3. (*i*) We note that $\dot{\mathbf{D}}_1(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) = (\dot{\mathbf{D}}_1^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{r}}_{t-1}\dot{\boldsymbol{r}}_{t-1}'\dot{\mathbf{D}}_1^{-1})^{-1}\dot{\mathbf{D}}_1^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{r}}_{t-1}u_{t-1}$ from (13), and Lemma 2 (*vii*) implies that $\dot{\mathbf{D}}_1^{-1}(\sum_{t=1}^{T}\dot{\boldsymbol{r}}_{t-1}\dot{\boldsymbol{r}}_{t-1}')\dot{\mathbf{D}}_1^{-1} \Rightarrow \dot{\mathscr{M}}_{11}$. We therefore focus on the limit distribution of $\dot{\mathbf{D}}_1^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{r}}_{t-1}u_{t-1}$. Note that

$$\dot{\mathbf{D}}_1^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{r}}_{t-1}u_{t-1} = \left[T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{m}}_{t-1}'u_{t-1}, T^{-3/2}\sum_{t=1}^{T}(t-1)u_{t-1}, T^{-1/2}\sum_{t=1}^{T}u_{t-1}\right]'.$$

We now examine the asymptotic behavior of each element on the right side. First, we note that $\widehat{\mathbf{m}}_{t-1} = \mathbf{m}_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1)$. Therefore, $T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{m}}_{t-1}u_{t-1} = T^{-1}\sum_{t=1}^{T}\mathbf{m}_{t-2}u_{t-1} + T^{-1}\sum_{t=1}^{T}\boldsymbol{s}_{t-1}u_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)T^{-1}\sum_{t=1}^{T}(t-1)u_{t-1}$. We here note that $T^{-1}\sum_{t=1}^{T}\mathbf{m}_{t-2}u_{t-1} \Rightarrow \int \boldsymbol{\mathscr{B}}_m d\mathscr{B}_u + + \boldsymbol{\varrho}_{m*}$ by Lemma B.7 (*iii*), and $\sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \Rightarrow 3\int r\boldsymbol{\mathscr{B}}_m$ by Lemma B.7 (*i*). In addition to this, $T^{-3/2}\sum_{t=1}^{T}(t-1)u_{t-1} = T^{-1}\sum_{t=1}^{T}\frac{(t-1)}{T}\sqrt{T}(B_{uT}((t-1)/T) - B_{uT}((t-2)/T)) = \int r dB_{uT}(r) \Rightarrow \int r d\mathscr{B}_u$. Hence, it follows that $T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{m}}_{t-1}u_{t-1} \Rightarrow \dot{\mathcal{S}}_1 := \int \boldsymbol{\mathscr{B}}_m d\mathscr{B}_u + \boldsymbol{\varrho}_{m*} - 3\int r\boldsymbol{\mathscr{B}}_m \int r d\mathscr{B}_u$. Next, it is already showed that $T^{-3/2}\sum_{t=1}^{T}(t-1)u_{t-1} \Rightarrow \dot{\mathcal{S}}_2 := \int r d\mathscr{B}_u$. Third, note that $T^{-1/2}\sum_{t=1}^{T}u_{t-1} = T^{-1}\sum_{t=1}^{T}\sqrt{T}(B_{uT}((t-1)/T) - B_{uT}((t-2)/T)) = \int r dB_{uT}(r) \Rightarrow \dot{\mathcal{S}}_3 := \int d\mathscr{B}_u$. We now combine the first to third facts to obtain that $\dot{\mathbf{D}}_1^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{r}}_{t-1}u_{t-1} \Rightarrow \dot{\mathcal{S}}$, leading to that $\dot{\mathbf{D}}_1(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) \Rightarrow \dot{\mathscr{L}} := \dot{\mathscr{M}}_{11}^{-1}\dot{\mathcal{S}}$, as desired.

(*ii*) We first note that $\widehat{\boldsymbol{v}}_T - \boldsymbol{v}_* = \mathbf{P}_T\widetilde{\boldsymbol{v}}_T - \mathbf{P}\bar{\boldsymbol{v}}_*$. Therefore, $\widehat{\boldsymbol{v}}_T - \boldsymbol{v}_* = (\mathbf{P}_T - \mathbf{P})(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) + \mathbf{P}(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) + (\mathbf{P}_T - \mathbf{P})(\bar{\boldsymbol{v}}_{T*} - \bar{\boldsymbol{v}}_*) + \mathbf{P}(\bar{\boldsymbol{v}}_{T*} - \bar{\boldsymbol{v}}_*) + (\mathbf{P}_T - \mathbf{P})\bar{\boldsymbol{v}}_*$. From the definition of $\mathbf{P}_T$ and Lemma 3 (*i*), we note that $(\mathbf{P}_T - \mathbf{P}) = O_{\mathbb{P}}(T^{-1/2})$, $(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) = O_{\mathbb{P}}(\dot{\mathbf{D}}^{-1})$, and $(\mathbf{P}_T - \mathbf{P})(\bar{\boldsymbol{v}}_{T*} - \bar{\boldsymbol{v}}_*) = \mathbf{0}$. Furthermore, $\mathbf{P}(\bar{\boldsymbol{v}}_{T*} - \bar{\boldsymbol{v}}_*) = [\mathbf{0}', \mathbf{0}', (\vartheta_{T*} - \vartheta_*), 0]'$ such that $(\vartheta_{T*} - \vartheta_*) = \boldsymbol{\beta}_*'(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)$, and $(\mathbf{P}_T - \mathbf{P})\bar{\boldsymbol{v}}_* = [\mathbf{0}, \mathbf{0}, -\boldsymbol{\beta}_*'(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*), 0]'$, so that $\mathbf{P}(\bar{\boldsymbol{v}}_{T*} - \bar{\boldsymbol{v}}_*) + (\mathbf{P}_T - \mathbf{P})\bar{\boldsymbol{v}}_* = \mathbf{0}$. Hence, it now follows that $\widehat{\boldsymbol{v}}_T - \boldsymbol{v}_* = (\mathbf{P}_T - \mathbf{P})(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) + \mathbf{P}(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) = \mathbf{P}(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) + O_{\mathbb{P}}(T^{-3/2})$, so that

$$\dot{\mathbf{D}}_\dagger(\widehat{\boldsymbol{v}}_T - \boldsymbol{v}_*) = [\; T(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)' \;\vdots\; -\boldsymbol{\mu}_*'T(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*) \;\vdots\; \sqrt{T}(\widetilde{\nu}_T - \nu_*) \;]' + o_{\mathbb{P}}(1) \Rightarrow [\; \dot{\mathscr{L}}' \;\vdots\; -\boldsymbol{\mu}_*'\dot{\mathscr{L}} \;\vdots\; \mathcal{L}_3 \;]'$$

by Lemma 3 (*i*). ∎

Proof of Lemma 4. (*i*) We first note that $\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) = (\dot{\mathbf{D}}^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{z}}_t\dot{\boldsymbol{z}}_t'\dot{\mathbf{D}}^{-1})^{-1}\dot{\mathbf{D}}^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{z}}_t e_t$ from (14), and Lemma 2 (*vi*) implies that $\dot{\mathbf{D}}^{-1}(\sum_{t=1}^{T}\dot{\boldsymbol{z}}_t\dot{\boldsymbol{z}}_t')\dot{\mathbf{D}}^{-1} \Rightarrow \dot{\mathscr{M}}$. We therefore focus on the limit distribution of $\dot{\mathbf{D}}^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{z}}_t e_t$. We note that $\dot{\mathbf{D}}^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{z}}_t e_t = \dot{\mathbf{D}}^{-1}\sum_{t=1}^{T}[\widehat{u}_{t-1}e_r, \widehat{\mathbf{m}}_{t-1}'e_t, (t-1)e_t, e_t, \boldsymbol{z}_{2t}'e_t]'$.

We now investigate the asymptotic behavior of each element on the right side. First, we already showed in the proof of Lemma B.6 (*iv*) that $\widehat{u}_t = u_t + O_{\mathbb{P}}(T^{-1/2})$. Therefore, $T^{-1/2}\sum_{t=1}^{T}\widehat{u}_{t-1}e_t = T^{-1/2}\sum_{t=1}^{T}u_{t-1}e_t + o_{\mathbb{P}}(1)$. In addition, $T^{-1/2}\sum_{t=1}^{T}u_{t-1}e_t = T^{-1}\sum_{t=1}^{T}\sqrt{T}(B_{ueT}(t/T) - B_{ueT}((t-$

11

$1)/T)) = \int r dB_{ueT}(r) \Rightarrow \dot{\mathscr{J}}_1 := \int d\mathscr{B}_{ue}$. Second, we note that $\widehat{\mathbf{m}}_{t-1} = \mathbf{m}_{t-1} - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)(t-1)$. Therefore, $T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{m}}_{t-1}e_t = T^{-1}\sum_{t=1}^{T}\mathbf{m}_{t-1}e_t - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)T^{-1}\sum_{t=1}^{T}(t-1)e_t$. We here note that $T^{-1}\sum_{t=1}^{T}\mathbf{m}_{t-1}e_t \Rightarrow \int \mathscr{B}_m d\mathscr{B}_e$ by Lemma B.7 (*iv*) and $\sqrt{T}(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \Rightarrow 3\int r\mathscr{B}_m$ by Lemma B.7 (*i*). In addition to this, $T^{-3/2}\sum_{t=1}^{T}(t-1)e_t = T^{-1}\sum_{t=1}^{T}\frac{(t-1)}{T}\sqrt{T}(B_{eT}((t-1)/T) - B_{eT}((t-2)/T)) = \int r dB_{eT}(r) \Rightarrow \int r d\mathscr{B}_e$. Hence, it follows that $T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{m}}_{t-1}e_t \Rightarrow \dot{\mathscr{J}}_2 := \int \mathscr{B}_m d\mathscr{B}_e - 3\int r\mathscr{B}_m\int r d\mathscr{B}_u = \int \bar{\mathscr{B}}_m d\mathscr{B}_e$. Third, it is already showed that $T^{-3/2}\sum_{t=1}^{T}(t-1)e_{t-1} \Rightarrow \dot{\mathscr{J}}_3 := \int r d\mathscr{B}_e$. Fourth, we note that $T^{-1/2}\sum_{t=1}^{T}e_t = T^{-1}\sum_{t=1}^{T}\sqrt{T}(B_{eT}(t/T) - B_{eT}((t-1)/T)) = \int dB_{eT}(r) \Rightarrow \dot{\mathscr{J}}_4 := \int d\mathscr{B}_e$. Fifth, note that $T^{-1/2}\sum_{t=1}^{T}\mathbf{z}_{2t}e_t = T^{-1}\sum_{t=1}^{T}\sqrt{T}(\mathbf{B}_{zeT}(t/T) - \mathbf{B}_{zeT}((t-1)/T)) = \int d\mathbf{B}_{zeT}(r) \Rightarrow \dot{\mathscr{J}}_5 := \int d\mathscr{B}_{ze}$. We next combine the first to fifth facts to obtain that $\dot{\mathbf{D}}^{-1}\sum_{t=1}^{T}\dot{\mathbf{z}}_t e_t \Rightarrow \dot{\mathscr{J}}$, leading to that $\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) \Rightarrow \dot{\mathscr{M}}^{-1}\dot{\mathscr{J}}$, as desired.

(*ii*) From the definitions of $\boldsymbol{\tau}_{T*} := [\rho_*, \boldsymbol{\tau}'_{1T}, \boldsymbol{\tau}'_{2*}]'$ and $\boldsymbol{\tau}_{1T} := [(\boldsymbol{\eta}_* + \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*))', \varsigma_* + \boldsymbol{\eta}'_*(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) + \rho_*(\widetilde{\vartheta}_T - \vartheta_{T*}), \gamma_* + \rho_*(\widetilde{\nu}_T - \nu_*)]'$, we obtain that $\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) = \dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) + [0, -\rho_*T(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)', -\rho_*T^{3/2}(\widetilde{\vartheta}_T - \vartheta_{T*}), -\rho_*T^{1/2}(\widetilde{\nu}_T - \nu_*), \mathbf{0}']' \Rightarrow \dot{\mathscr{D}}$ by Lemma 4 (*i*) and noting that $\boldsymbol{\eta}_* = \mathbf{0}$. In addition, Lemma 3 implies that $T(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*) \Rightarrow \dot{\mathscr{L}}'_1$, and $T^{1/2}(\widetilde{\nu}_T - \nu_*) \Rightarrow \dot{\mathscr{L}}_3$. Furthermore, the proof of Lemma 3 (*i*) shows that $T^{3/2}(\widetilde{\vartheta}_T - \vartheta_{T*}) \Rightarrow \dot{\mathscr{L}}_2$. Therefore, $\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) \Rightarrow \dot{\mathscr{D}} + \rho_*[0, \dot{\mathscr{L}}'_1, \dot{\mathscr{L}}_2, \dot{\mathscr{L}}_3, \mathbf{0}']'$. This completes the proof. ∎

Proof of Theorem 1. (*i*) We note that $(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) = (\mathbf{T}_T - \mathbf{T})\boldsymbol{\tau}_* + \mathbf{T}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) + (\mathbf{T}_T - \mathbf{T})(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) = (\mathbf{T}_T - \mathbf{T})\boldsymbol{\tau}_* + \mathbf{T}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) + o_{\mathbb{P}}(\mathbf{T}_T - \mathbf{T})$ using (10) and (11). We further note that

$$(\mathbf{T}_T - \mathbf{T})\boldsymbol{\tau}_* + \mathbf{T}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*)$$

$$= \begin{bmatrix} (\dot{\rho}_T - \rho_*) \\ -\boldsymbol{\beta}_*(\dot{\rho}_T - \rho_*) + (\dot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*) \\ (\dot{\varsigma}_T - \varsigma_*) - \boldsymbol{\mu}'_*(\dot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \zeta_*(\dot{\rho}_T - \rho_*) - \rho_*(\widehat{\zeta}_T - \zeta_*) - \boldsymbol{\eta}'_*(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*) \\ (\dot{\gamma}_T - \gamma_*) - \rho_*(\widehat{\nu}_T - \nu_*) - \nu_*(\dot{\rho}_T - \rho_*) \\ (\dot{\boldsymbol{\tau}}_{2T} - \boldsymbol{\tau}_{*2}) \end{bmatrix}$$

and the fact that $\widehat{\boldsymbol{\beta}}_T = \widetilde{\boldsymbol{\beta}}_T$, $\widehat{\zeta}_T = -\widehat{\boldsymbol{\mu}}'_T\widetilde{\boldsymbol{\beta}}_T + \widetilde{\vartheta}_T$ and $\zeta_* = -\boldsymbol{\mu}'_*\boldsymbol{\beta}_* + \vartheta_*$, so that $(\widehat{\zeta}_T - \zeta_*) = (\widetilde{\vartheta}_T - \vartheta_{T*}) - \boldsymbol{\mu}'_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*) - (\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)'(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)$, where $(\widetilde{\vartheta}_T - \vartheta_{T*}) = O_{\mathbb{P}}(T^{-3/2})$ and $(\widehat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_*)'(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*) = $

$O_{\mathbb{P}}(T^{-3/2})$ by Lemmas 4 (*i*), B.7 (*i*), and 3 (*ii*). Therefore, it now follows that

$$(\mathbf{T}_T - \mathbf{T})\boldsymbol{\tau}_* + \mathbf{T}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*)$$

$$= (\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) - (\dot{\rho}_T - \rho_*) \begin{bmatrix} 0 \\ \boldsymbol{\beta}_* \\ \zeta_* \\ \nu_* \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{0} \\ -\boldsymbol{\mu}'_*\{(\dot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)\} + O_{\mathbb{P}}(T^{-3/2}) \\ 0 \\ \mathbf{0} \end{bmatrix}. \quad \text{(A.4)}$$

We here use the fact that $\boldsymbol{\eta}_* = \mathbf{0}$. We further note that $\sqrt{T}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) \Rightarrow [\dot{\mathscr{D}}_1, \mathbf{0}'_{2k\times 1}, 0, \dot{\mathscr{D}}_4, \dot{\boldsymbol{\mathscr{D}}}_5]'$ by Lemma 4 (*i*) and $-\sqrt{T}(\dot{\rho}_T - \rho_*)[0, \boldsymbol{\beta}'_*, \zeta_*, \nu_*, \mathbf{0}'] \Rightarrow -\dot{\mathscr{D}}_1[0, \boldsymbol{\beta}'_*, \zeta_*, \nu_*, \mathbf{0}']$. In addition, we note that $-\boldsymbol{\mu}'_*\sqrt{T}\{(\dot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)\} + O_{\mathbb{P}}(T^{-1}) = o_{\mathbb{P}}(1)$ because $\sqrt{T}\{(\dot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)\} = O_{\mathbb{P}}(T^{-1/2})$ by Lemma 4 (*i*). Therefore, it follows that

$$\sqrt{T}(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) \Rightarrow [\dot{\mathscr{D}}_1, -\boldsymbol{\beta}'_*\dot{\mathscr{D}}_1, -\zeta_*\dot{\mathscr{D}}_1, \dot{\mathscr{D}}_4 - \nu_*\dot{\mathscr{D}}_1, \dot{\boldsymbol{\mathscr{D}}}'_5]'. \quad \text{(A.5)}$$

We finally note that the derived weak limit is identical to $\boldsymbol{c}_*\dot{\mathscr{D}}_1 + [0, \mathbf{0}', \mathbf{0}', 0, \dot{\mathscr{D}}_4, \dot{\boldsymbol{\mathscr{D}}}'_5]'$.

(*ii*, *iii*, and *iv*) For an efficient proof, we prove (*ii*, *iii*, and *iv*) together by supposing that $\boldsymbol{\beta}_* = \mathbf{0}$ and $\zeta_* = 0$, although it is not permitted by $\rho_* < 0$. If so, it algebraically follows from (A.4) that

$$(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) = (\mathbf{T}_T - \mathbf{T})\boldsymbol{\tau}_* + \mathbf{T}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_*) + o_{\mathbb{P}}(\mathbf{T}_T - \mathbf{T})$$

$$= (\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) + \begin{bmatrix} 0 \\ \mathbf{0} \\ -\boldsymbol{\mu}'_*\{(\dot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)\} + O_{\mathbb{P}}(T^{-3/2}) \\ -\nu_*(\dot{\rho}_T - \rho_*) \\ \mathbf{0} \end{bmatrix} + o_{\mathbb{P}}(\mathbf{T}_T - \mathbf{T}).$$

Here, $\ddot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) \Rightarrow [\dot{\mathscr{D}}_1, \dot{\boldsymbol{\mathscr{D}}}'_2, 0, \dot{\mathscr{D}}_4, \dot{\boldsymbol{\mathscr{D}}}_5]'$ and $T\{(\dot{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_*) - \rho_*(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*)\} \Rightarrow \dot{\boldsymbol{\mathscr{D}}}_2$ by Lemma 4 (*i*), so that

$$\ddot{\mathbf{D}}(\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_*) \Rightarrow [\dot{\mathscr{D}}_1, \dot{\boldsymbol{\mathscr{D}}}'_2, -\boldsymbol{\mu}'_*\dot{\boldsymbol{\mathscr{D}}}_2, \dot{\mathscr{D}}_4 - \nu_*\dot{\mathscr{D}}_1, \dot{\boldsymbol{\mathscr{D}}}'_5]', \quad \text{(A.6)}$$

where $\ddot{\mathbf{D}} := \text{diag}[\sqrt{T}, T\mathbf{I}_{2k+1}, \sqrt{T}, \dot{\mathbf{D}}_2]$. From the weak limits in (A.5) and (A.6), (*ii*, *iii*, and *iv*) follow by applying the weak limits corresponding to the conditions in (*ii*, *iii*, and *iv*). For example, if the conditions in (*ii*) are imposed, the weak limit of $\widehat{\boldsymbol{\theta}}_T^+$ is determined by (A.6) as $\dot{\boldsymbol{\mathscr{D}}}_2^+$, but the others are determined by

(A.5), so that the weak limit in (*ii*) is delivered. As another example, if the conditions in (*iv*) are imposed, the weak limit of $\widehat{\xi}_T$ is determined by (A.6) as $\boldsymbol{\mu}'_*\dot{\mathscr{D}}_2$, but the others are determined by (A.5), so that the weak limit in (*iv*) is delivered. In this manner, the weak limits in (*ii*, *iii*, and *iv*) are obtained. ∎

Proof of Theorem 2. (*i*) We show the null limit distribution of each test using Lemma 4.

(*i.a*) Note that $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-) - (\dot{\rho}_T - \rho_*)(\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-) + (\dot{\boldsymbol{\eta}}_T^+ - \boldsymbol{\eta}_{T*}^+) - (\dot{\boldsymbol{\eta}}_T^- - \boldsymbol{\eta}_{T*}^-) + o_{\mathbb{P}}(T^{-1})$, so that it follows that $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\dot{\boldsymbol{\eta}}_T^+ - \boldsymbol{\eta}_{T*}^+) - (\dot{\boldsymbol{\eta}}_T^- - \boldsymbol{\eta}_{T*}^-) + o_{\mathbb{P}}(T^{-1})$ under $\mathscr{H}'_0$. Therefore, $\dot{\mathbf{R}}_1\widehat{\boldsymbol{\alpha}}_T = \mathbf{R}_1(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) + o_{\mathbb{P}}(T^{-1})$, and it follows from Lemmas 4 (*i*) and B.7 (*v*) that

$$W_T^{(1)} = (\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*})'\dot{\mathbf{D}}\mathbf{R}_1' \left(\widehat{\sigma}_{e,T}^2 \mathbf{R}_1 \left(\dot{\mathbf{D}}^{-1}\sum_{t=1}^T \dot{z}_t\dot{z}_t'\dot{\mathbf{D}}^{-1} \right)^{-1} \mathbf{R}_1' \right)^{-1} \mathbf{R}_1\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) + o_{\mathbb{P}}(1)$$

$$\Rightarrow \dot{\mathscr{D}}'\mathbf{R}_1' \left(\sigma_e^2\mathbf{R}_1\dot{\mathscr{M}}^{-1}\mathbf{R}_1' \right)^{-1} \mathbf{R}_1\dot{\mathscr{D}}.$$

(*i.b*) We note that $(\widehat{\boldsymbol{\pi}}_T^+, \widehat{\boldsymbol{\pi}}_T^-) = (\dot{\boldsymbol{\pi}}_T^+, \dot{\boldsymbol{\pi}}_T^-)$. Therefore, it follows from Lemmas 4 (*i*) and B.7 (*v*) that

$$W_T^{(2)} = (\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*})'\dot{\mathbf{D}}\dot{\mathbf{R}}_2' \left(\widehat{\sigma}_{e,T}^2 \dot{\mathbf{R}}_2 \left(\dot{\mathbf{D}}^{-1}\sum_{t=1}^T \dot{z}_t\dot{z}_t'\dot{\mathbf{D}}^{-1} \right)^{-1} \dot{\mathbf{R}}_2' \right)^{-1} \dot{\mathbf{R}}_2\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) + o_{\mathbb{P}}(1)$$

$$\Rightarrow \dot{\mathscr{D}}'\dot{\mathbf{R}}_2' \left(\sigma_e^2\dot{\mathbf{R}}_2\dot{\mathscr{M}}^{-1}\dot{\mathbf{R}}_2' \right)^{-1} \dot{\mathbf{R}}_2\dot{\mathscr{D}}.$$

(*i.c*) From the notice given in the proofs of (*i.a* and *i.b*), it follows from Lemmas 4 (*i*) and B.7 (*v*) that

$$W_T^{(3)} = (\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*})'\dot{\mathbf{D}}\mathbf{R}_3' \left(\widehat{\sigma}_{e,T}^2 \mathbf{R}_3 \left(\dot{\mathbf{D}}^{-1}\sum_{t=1}^T \dot{z}_t\dot{z}_t'\dot{\mathbf{D}}^{-1} \right)^{-1} \mathbf{R}_3' \right)^{-1} \mathbf{R}_3\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) + o_{\mathbb{P}}(1)$$

$$\Rightarrow \dot{\mathscr{D}}'\mathbf{R}_3' \left(\sigma_e^2\mathbf{R}_3\dot{\mathscr{M}}^{-1}\mathbf{R}_3' \right)^{-1} \mathbf{R}_3\dot{\mathscr{D}}.$$

(*ii*) We show the power behavior of each test using Lemma 4.

(*ii.a*) We show that $W_T^{(1)} = O_{\mathbb{P}}(T)$. Note that $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = (\boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-) - (\dot{\rho}_T - \rho_*)(\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-) + o_{\mathbb{P}}(T^{-1/2})$ under $\mathscr{H}'_1$ and $(\dot{\rho}_T - \rho_*) = O_{\mathbb{P}}(T^{-1/2})$, so that $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) - (\boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-) = O_{\mathbb{P}}(T^{-1/2})$. Therefore, $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = O_{\mathbb{P}}(T^{1/2})$, and this implies that $W_T^{(1)} = O_{\mathbb{P}}(T)$.

(*ii.b*) We show that $W_T^{(2)} = O_{\mathbb{P}}(T)$. Note that Lemma 4 (*i*) implies that $\dot{\boldsymbol{\pi}}_T^+ - \dot{\boldsymbol{\pi}}_T^- = \boldsymbol{\pi}_*^+ - \boldsymbol{\pi}_*^- + O_{\mathbb{P}}(T^{-1/2})$. Therefore, $\dot{\boldsymbol{\pi}}_T^+ - \dot{\boldsymbol{\pi}}_T^- = O_{\mathbb{P}}(T^{1/2})$, implying that $W_T^2 = O_{\mathbb{P}}(T)$.

(*ii.c*) From the proofs of (*ii.a*) and (*ii.b*), $(\widehat{\boldsymbol{\theta}}_T^+ - \widehat{\boldsymbol{\theta}}_T^-) = O_{\mathbb{P}}(T^{1/2})$ and $(\dot{\boldsymbol{\pi}}_T^+ - \dot{\boldsymbol{\pi}}_T^-) = O_{\mathbb{P}}(T^{1/2})$. Therefore, it trivially follows that $W_T^{(3)} = O_{\mathbb{P}}(T)$.

(*iii*) We show the null limit distribution of each test using Lemmas 3 and 4.

(*iii.a*) First, Lemma B.6 (*iv*) implies that $\widehat{\sigma}_{u,T}^2 \xrightarrow{\mathbb{P}} \sigma_u^2$, and $\dot{\mathbf{D}}_1^{-1}\sum_{t=1}^T \dot{\boldsymbol{r}}_{t-1}\dot{\boldsymbol{r}}_{t-1}'\dot{\mathbf{D}}_1^{-1} \Rightarrow \dot{\mathcal{M}}_{11}$ from Lemma 2 (*vii*). Therefore, $\ddot{\mathbf{W}}_T^{(1)} \Rightarrow \sigma_u^2\ddot{\mathbf{R}}_1\dot{\mathcal{M}}_{11}^{-1}\ddot{\mathbf{R}}_1'$. Second, $\mathbb{H}_0'$ implies that $\ddot{\mathbf{R}}_1\dot{\mathbf{D}}_1\widetilde{\boldsymbol{v}}_T = T(\widetilde{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+) - T(\widetilde{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-) \Rightarrow (\dot{\mathcal{L}}_1^+ - \dot{\mathcal{L}}_1^-)$. If we combine these two facts, it follows that $\mathscr{W}_T^{(1)} \Rightarrow (\dot{\mathcal{L}}_1^+ - \dot{\mathcal{L}}_1^-)'(\sigma_u^2\ddot{\mathbf{R}}_1\dot{\mathcal{M}}_{11}^{-1}\ddot{\mathbf{R}}_1')^{-1}(\dot{\mathcal{L}}_1^+ - \dot{\mathcal{L}}_1^-)$ that is identical to the given weak limit by the definitions of $\ddot{\mathbf{R}}_1$ and $\dot{\mathcal{L}}$.

(*iii.b*) First, Lemma B.7 (*v*) implies that $\widehat{\sigma}_{e,T}^2 \xrightarrow{\mathbb{P}} \sigma_e^2$, and $\dot{\mathbf{D}}^{-1}\sum_{t=1}^T \dot{\boldsymbol{z}}_t\dot{\boldsymbol{z}}_t'\dot{\mathbf{D}}^{-1} \Rightarrow \dot{\mathcal{M}}$ from Lemma 2 (*vi*). Therefore, $\ddot{\mathbf{W}}_T^{(2)} \Rightarrow \sigma_e^2\ddot{\mathbf{R}}_2\dot{\mathcal{M}}^{-1}\ddot{\mathbf{R}}_2'$. Second, $\ddot{\mathbf{R}}_2\dot{\mathbf{D}}\dot{\boldsymbol{\tau}}_T = \sqrt{T}(\dot{\boldsymbol{\pi}}_T^+ - \dot{\boldsymbol{\pi}}_T^-) \Rightarrow \dot{\mathcal{D}}_5^+ - \dot{\mathcal{D}}_5^- = \ddot{\mathbf{R}}_2\dot{\mathcal{D}}$. If we combine these two facts, it follows that $\mathscr{W}_T^{(2)} \Rightarrow \dot{\mathcal{D}}'\ddot{\mathbf{R}}_2'(\sigma_2^2\ddot{\mathbf{R}}_2\dot{\mathcal{M}}^{-1}\ddot{\mathbf{R}}_2')^{-1}\ddot{\mathbf{R}}_2\dot{\mathcal{D}}$.

(*iii.c*) We note that from the definition of $\mathscr{W}_T^{(3)}$, it follows that $\mathscr{W}_T^{(3)} = \mathscr{W}_T^{(1)} + \mathscr{W}_T^{(2)}$, and it follows from (*i.a* and *i.b*) that $\mathscr{W}_T^{(3)} \Rightarrow \dot{\mathcal{L}}'\ddot{\mathbf{R}}_1'(\sigma_u^2\ddot{\mathbf{R}}_1\dot{\mathcal{M}}_{11}^{-1}\ddot{\mathbf{R}}_1')^{-1}\ddot{\mathbf{R}}_1\dot{\mathcal{L}} + \dot{\mathcal{D}}'\ddot{\mathbf{R}}_2'(\sigma_2^2\ddot{\mathbf{R}}_2\dot{\mathcal{M}}^{-1}\ddot{\mathbf{R}}_2')^{-1}\ddot{\mathbf{R}}_2\dot{\mathcal{D}}$ under $\mathbb{H}_0'''$.

(*iv*) We next show the power behavior of each test.

(*iv.a*) To show the claim, we show that $\mathscr{W}_T^{(1)} = O_{\mathbb{P}}(T^2)$ under $\mathbb{H}_1'$. Given that $\ddot{\mathbf{W}}_T^{(1)} \Rightarrow \sigma_u^2\ddot{\mathbf{R}}_1\dot{\mathcal{M}}_{11}^{-1}\ddot{\mathbf{R}}_1'$, we focus on the limit behavior of $\ddot{\mathbf{R}}_1\dot{\mathbf{D}}_1\widetilde{\boldsymbol{v}}_T$. Note that $\ddot{\mathbf{R}}_1\dot{\mathbf{D}}_1(\widetilde{\boldsymbol{v}}_T - \bar{\boldsymbol{v}}_{T*}) \Rightarrow (\dot{\mathcal{L}}_1^+ - \dot{\mathcal{L}}_1^-)$ by Lemma 3 (*i*) and $\ddot{\mathbf{R}}_1\dot{\mathbf{D}}_1\bar{\boldsymbol{v}}_{T*} = T\boldsymbol{\beta}_* = O(T)$. Therefore, $\ddot{\mathbf{R}}_1\dot{\mathbf{D}}_1\widetilde{\boldsymbol{v}}_T = O_{\mathbb{P}}(T)$, implying that $\mathscr{W}_T^{(1)} = O_{\mathbb{P}}(T^2)$, leading to the desired result.

(*iv.b*) We show that $\mathscr{W}_T^{(2)} = O_{\mathbb{P}}(T)$ under $\mathscr{H}_1''$. Given that $\ddot{\mathbf{W}}_T^{(2)} \Rightarrow \sigma_e^2\ddot{\mathbf{R}}_2\dot{\mathcal{M}}^{-1}\ddot{\mathbf{R}}_2'$, we focus on the limit behavior of $\ddot{\mathbf{R}}_2\dot{\mathbf{D}}\dot{\boldsymbol{\tau}}_T$. We note that $\ddot{\mathbf{R}}_2\dot{\mathbf{D}}(\dot{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_{T*}) = \sqrt{T}[(\dot{\boldsymbol{\pi}}_T^+ - \dot{\boldsymbol{\pi}}_T^-) - (\boldsymbol{\pi}_*^+ - \boldsymbol{\pi}_*^-)] \Rightarrow \ddot{\mathbf{R}}_2\dot{\mathcal{D}}$ by Lemma 4 (*i*) and $\ddot{\mathbf{R}}_2\dot{\mathbf{D}}\boldsymbol{\tau}_{T*} = \sqrt{T}(\boldsymbol{\pi}_*^+ - \boldsymbol{\pi}_*^-) = O(T^{1/2})$. Therefore, $\ddot{\mathbf{R}}_2\dot{\mathbf{D}}\dot{\boldsymbol{\tau}}_T = O_{\mathbb{P}}(T^{1/2})$, implying that $\mathscr{W}_T^{(2)} = O_{\mathbb{P}}(T)$.

(*iv.c*) We show that $\mathscr{W}_T^{(3)} = O_{\mathbb{P}}(T^2)$ under $\mathbb{H}_1'''$. Note that $\mathscr{W}_T^{(3)} = \mathscr{W}_T^{(1)} + \mathscr{W}_T^{(2)}$, and $\mathscr{W}_T^{(1)}$ and $\mathscr{W}_T^{(2)}$ are $O_{\mathbb{P}}(T^2)$ and $O_{\mathbb{P}}(T)$ by (*ii.a*) and (*ii.b*), respectively. Therefore, $\mathscr{W}_T^{(3)} = O_{\mathbb{P}}(T^2)$ under $\mathbb{H}_{11}'''$ and $\mathscr{W}_T^{(3)} = O_{\mathbb{P}}(T)$ under $\mathbb{H}_{01}''' \bigcap \mathbb{H}_{12}'''$. This completes the proof. ∎

A.4 Additional Empirical Supplements

In this section, we provide additional empirical supplements.

Two tables are provided. First, Table A.1 provides the descriptive statistics of the variables examined in Sections 7.2.1 and 7.2.2. The sample period is from 1947Q1 to 2007:Q4.

Second, Table A.2 provides the testing results using Phillips and Perron's (1988) unit-root test applied to the partial sum processes for Tables 5 and 7. As we apply the unit-root testing by including both constant and trend or including only constant, two testing results are provided for each variable. Except r_t, the test results show that nonstationary data analysis has to be conducted for the other variables.

	Δy_t	Exo.			Endo.			Sum.		
		$\Delta\tau_{1t}$	$\Delta\tau_{2t}$	$\Delta\tau_t$	$\Delta\tau_{1t}$	$\Delta\tau_{2t}$	$\Delta\tau_t$	$\Delta\tau_{1t}$	$\Delta\tau_{2t}$	$\Delta\tau_t$
Mean	0.8256	0.4240	-0.4704	-0.0464	0.3773	-0.1841	0.1932	0.7564	-0.6228	0.1336
Median	0.7876	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maximum	4.0198	6.4312	0.0000	6.4312	6.7617	0.0000	6.7617	6.7617	0.0000	6.7617
Minimum	-2.7525	0.0000	-7.0965	-7.0965	0.0000	-7.1122	-7.1122	0.0000	-7.1122	-7.1122
Std. Dev.	0.9780	1.3755	1.5073	2.1364	1.3317	1.0117	1.7136	1.7775	1.7316	2.6653
Skewness	-0.0501	3.0654	-3.0598	-0.2762	3.4094	-5.4276	0.4394	2.0432	-2.5281	-0.1450
Kurtosis	4.3614	10.8514	10.8759	6.3787	13.2115	31.1514	10.4245	5.4941	7.7131	4.1776
Obs.	243	243	243	243	243	243	243	243	243	243

	Δy_t	Exo. ratio			Endo. ratio			Sum. ratio		
		Δr_{1t}	Δr_{2t}	Δr_t	Δr_{1t}	Δr_{2t}	Δr_t	Δr_{1t}	Δr_{2t}	Δr_t
Mean	0.8256	0.0246	-0.0517	-0.0271	0.0471	-0.0260	0.0211	0.0679	-0.0738	-0.0059
Median	0.7876	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maximum	4.0198	0.6977	0.0000	0.6977	1.9542	0.0000	1.9542	1.9542	0.0000	1.9542
Minimum	-2.7525	0.0000	-1.8706	-1.8706	0.0000	-2.8214	-2.8214	0.0000	-2.8214	-2.8214
Std. Dev.	0.9780	0.0963	0.2171	0.2428	0.2245	0.2028	0.3066	0.2351	0.2876	0.3847
Skewness	-0.0501	4.6231	-5.3926	-3.7013	6.0062	-11.5395	-1.0526	5.1164	-5.9045	-1.3783
Kurtosis	4.3614	25.4621	35.2975	25.0905	41.7414	152.4675	44.1557	33.0743	45.6366	21.6860
Obs.	243	243	243	243	243	243	243	243	243	243

Table A.1: DESCRIPTIVE STATISTICS. This table shows the descriptive statistics used in Sections 7.2.1 and 7.2.2.

PP test	y_t	Exo.			Endo.			Sum.		
		τ_{1t}	τ_{2t}	τ_t	τ_{1t}	τ_{2t}	τ_t	τ_{1t}	τ_{2t}	τ_t
PP test w/o trend	-1.4767	0.7551	-0.7188	-1.6243	-2.4701	-0.3499	-2.2743	-1.3064	-0.4173	-2.0303
p-value	0.5439	0.9931	0.8387	0.4686	0.1241	0.9140	0.1812	0.6270	0.9028	0.2738
PP test w/ trend	-2.3488	-1.8278	-1.3653	-1.3483	-0.7094	-2.2995	-0.8156	-0.9285	-2.0326	-2.7414
p-value	0.4056	0.6883	0.8686	0.8732	0.9707	0.4322	0.9618	0.9500	0.5801	0.2210

PP test	y_t	Exo. ratio			Endo. ratio			Sum. ratio		
		r_{1t}	r_{2t}	r_t	r_{1t}	r_{2t}	r_t	r_{1t}	r_{2t}	r_t
PP test w/o trend	-1.4767	0.7288	-0.7859	-1.7004	-2.5409	-0.9379	-3.0710	-1.7735	-1.8303	-2.9448
p-value	0.5439	0.9926	0.8210	0.4299	0.1071	0.7749	0.0301	0.3932	0.3652	0.0418
PP test w/ trend	-2.3488	-1.8466	-2.3558	-2.2056	-1.5667	-2.0336	-2.6499	-1.3375	-2.3690	-3.4114
p-value	0.4056	0.6790	0.4019	0.4840	0.8033	0.5796	0.2587	0.8761	0.3949	0.0521

Table A.2: PHILLIPS AND PERRON'S (1988) UNIT-ROOT TESTS APPLIED TO THE QUARTERLY DATA IN ROMER AND ROMER (2010). Two Phillips and Perron's test statistics are computed using the data in Table A.1 by including the time trend and/or the constant. The lag lengths are selected by BIC.