Two-Step Estimation of the

Nonlinear Autoregressive Distributed Lag Model[*]

JIN SEO CHO

School of Humanities and Social Sciences, Beijing Institute of Technology, Haidan, Beijing 100081, China
School of Economics, Yonsei University, Seodaemun-gu, Seoul 03722, Korea
jinseocho@yonsei.ac.kr

MATTHEW GREENWOOD-NIMMO

Faculty of Business and Economics, University of Melbourne, Carlton, VIC 3053, Australia
Centre for Applied Macroeconomic Analysis, Australian National University, Canberra, ACT 2600, Australia
matthew.greenwood@unimelb.edu.au

YONGCHEOL SHIN

Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, U.K.
yongcheol.shin@york.ac.uk

September 18, 2020

Abstract

We consider estimation of and inference on the nonlinear autoregressive distributed lag (NARDL) model, which is a single-equation error correction model that allows for asymmetry with respect to positive and negative changes in the explanatory variable(s). We show that the NARDL model exhibits an asymptotic singularity issue that frustrates efforts to derive the asymptotic properties of the single-step OLS estimator. Consequently, we propose a two-step estimation framework, in which the parameters of the long-run relationship are estimated first using the fully-modified least squares estimator before the dynamic parameters are estimated by OLS in the next step. We show that our two-step estimators are consistent for the parameters of the NARDL model under stated regularity conditions and we derive their limit distributions. We develop Wald test statistics for the hypotheses of short-run and long-run parameter asymmetry. We demonstrate the utility of our framework with an application to postwar dividend-smoothing in the US.

Key Words: Nonlinear Autoregressive Distributed Lag (NARDL) Model; Fully-Modified Least Squares Estimator; Two-step Estimation; Wald Test Statistic; Dividend-Smoothing.

JEL Classifications: C22, G35.

1 Introduction

The Nonlinear Autoregressive Distributed Lag (NARDL) model of Shin, Yu, and Greenwood-Nimmo (2014, hereafter SYG) is an asymmetric generalization of the ARDL model of Pesaran and Shin (1998) and Pesaran, Shin, and Smith (2001). Specifically, the NARDL model is a single-equation error correction model that can accommodate asymmetry in the long-run equilibrium relationship and/or the short-run dynamic coefficients via the use of partial sum decompositions of the independent variable(s). Due to its simplicity and ease of interpretation, uptake of the NARDL model in applied research has been rapid, with applications in diverse fields including criminology (Box, Gratzer, and Lin, 2018), economic growth (Eberhardt and Presbitero, 2015), energy economics (Greenwood-Nimmo and Shin, 2013; Hammoudeh, Lahiani, Nguyen, and Sousa, 2015), exchange rates and trade (Verheyen, 2013; Brun-Aguerre, Fuertes, and Greenwood-Nimmo, 2017), financial economics (He and Zhou, 2018), health economics (Barati and Fariditavana, 2018), the economics of tourism (Süssmuth and Woitek, 2013) and political science (Ferris, Winer, and Olmstead, 2020), to list only a few.[1] However, despite its growing popularity, the theoretical foundations for estimation of and inference on the NARDL model have yet to be fully developed. It is this issue that we address.

SYG show that the parameters of the NARDL model can be estimated in a single step by ordinary least squares (OLS), as is the case in the linear ARDL model. However, the authors note that the positive and negative partial sums of the independent variables in the NARDL model are dominated by deterministic time trend terms that are asymptotically perfectly collinear. These collinear trend terms introduce an asymptotic singularity problem that represents a substantial barrier to the development of asymptotic theory for the single-step estimation framework, frustrating efforts to derive the limit distribution of the estimator. Consequently, SYG do not provide asymptotic theory but rather conduct Monte Carlo simulations to validate the properties of the single-step OLS estimator in finite samples.

In order to address this asymptotic singularity problem, we first consider the simple case of a bivariate model with a scalar dependent variable, y_t, and a scalar explanatory variable, x_t. In this case, the asymmetric long-run relationship is usually expressed among the level of the dependent variable and the positive and negative cumulative partial sum processes of the dependent variable, x_t^+ and x_t^-, respectively, the latter of which share asymptotically collinear time trends. Note, however, that the long-run relationship can be expressed equivalently by making use of a simple one-to-one transformation as a relationship between y_t, x_t and x_t^+. By reparameterizing the asymmetric long-run relationship to exclude one partial sum process, the asymptotic singularity issue in the long-run levels relationship is resolved. It is important to realize, however, that although this reparameterization is sufficient to resolve the singularity in the long-run relationship and will play an important role in our estimation strategy, it is insufficient to resolve the singularity problem associated with the single-step NARDL estimator. In fact, we show that it introduces a further asymptotic singularity problem, once again frustrating efforts to obtain the necessary limit theory (see Appendix A.2 for details).

In the bivariate case, our solution is to adopt a two-step estimation framework for the NARDL model. In the first step, the parameters of the transformed long-run relationship are estimated using any consistent estimator with a convergence

---

[1] At the time of writing, SYG has been cited more than 1,100 times according to Google Scholar.

rate faster than the square root of the sample size, $T^{1/2}$. We demonstrate that it is possible to consistently estimate the long-run parameters in the first step by OLS but that this approach suffers several drawbacks, most notably that the limit distribution of the OLS estimators is asymptotically non-normal and depends on nuisance parameters. Consequently, we advocate the use of the fully-modified OLS (FM-OLS) estimator of Phillips and Hansen (1990) in the first step, which we show to follow an asymptotic mixed normal distribution that facilitates standard inference on the long-run parameters. Furthermore, unlike OLS, FM-OLS is known to be robust to potential endogeneity among the regressors and to serial correlation in the error terms. Given the super-consistency of the long-run parameter estimator from the first step, the error correction term can be treated as known in the second step regression, where OLS provides a consistent and asymptotically normal estimator for the short-run dynamic parameters.

The two-step estimator described above will often be sufficient for practical use, as many applications of the NARDL framework focus on bivariate relationships. A good example is the analysis of Okun's Law presented by SYG. However, in general, a NARDL model may contain one or more explanatory variables, i.e., $x_t \in \mathbb{R}^k$ ($k \geq 1$). The two-step procedure is not appropriate for the estimation of NARDL models with multiple explanatory variables (i.e. $k > 1$) because, in this case, a further singular matrix problem arises at the limit when estimating the reparameterized long-run equation due to the collinearity of the trends of x_t and x_t^+. To resolve this issue, we propose to first detrend x_t^+ by OLS and then to use the implied residuals as an explanatory variable together with a time trend and x_t. This modified two-step procedure allows for estimation of the long-run relationship without any singularity problem, even for $k > 1$. The short-run parameters can then be estimated in a final step by OLS.

Because the NARDL model allows for asymmetry in both the long-run equilibrium relationship and the short-run adjustment parameters, testing restrictions on the long- and short-run parameters is an important aspect of inference on the NARDL model. We develop Wald tests for this purpose. In both the short- and long-run cases, we demonstrate that the null distribution of the Wald statistics weakly converges to a chi-squared distribution.

We conduct a suite of Monte Carlo simulations to investigate the properties of our estimators and test statistics. We find that the finite sample bias of the estimators of both the long- and short-run parameters is modest and diminishes rapidly as the sample size increases. Likewise, the mean squared error of the estimators quickly falls as the sample size grows. The Wald tests of both the short- and long-run parameters have high power and exhibit only mild size distortions in small samples that are rapidly corrected at larger sample sizes. Overall, our simulation results lend robust support to our theoretical findings.

We apply our technique to the analysis of postwar dividend-smoothing in the US. Following the seminal study on dividend policy by Lintner (1956), it is widely believed that firms gradually adjust their dividends in response to changes in earnings toward their long-run target payout ratio. Compelling evidence of this effect has been documented by Brav, Graham, Harvey, and Michaely (2005) on the basis of a survey of 384 financial executives, the results of which show that the link between dividends and earnings is relatively weak, with payout policy being subject to strategic considerations including signalling effects.

Our contribution is to test whether dividend policy may be asymmetric with respect to positive and negative changes in earnings. We fit a fourth order NARDL model to quarterly data on real dividends and real earnings for the S&P 500 over the period 1946Q1 to 2006Q4. Our model allows for asymmetry in the long-run equilibrium relationship and in the short-run dynamics. We consider both the single-step estimation procedure proposed by SYG and our newly-developed two-step procedure. We find that, in long-run equilibrium, executives pass earnings increases through to dividends slightly more strongly than earnings decreases, although neither the single-step nor the two-step estimation results provide any support for the existence of short-run dynamic asymmetry. The magnitude of the long-run asymmetry is relatively small but it is economically significant, which is consistent with existing evidence of asymmetric aggregate payout policy documented by Brav et al. (2005), among others.

Furthermore, our estimation results shed light on the performance of the single-step estimation procedure of SYG relative to our two-step framework. In practice, we find that both procedures yield qualitatively and quantitatively similar estimation and testing results. This indicates that they may be used interchangeably in practice. However, the two-step framework yields greater precision in the estimation of the long-run parameters, as it is not subject to the influence of nuisance parameters, unlike the single-step procedure. This may improve one's ability to detect long-run asymmetry, particularly in small samples. This represents an important practical benefit of our two-step estimation framework, particularly given that NARDL models are often used in macroeconomic applications, where a low sampling frequency and relatively short time period necessitate the use of small samples.

This paper proceeds in 7 sections. In Section 2, we introduce the NARDL model in its original form and demonstrate how the asymptotic singularity problem arises. In Section 3, we introduce our estimation framework and derive the asymptotic properties of the estimators. In Section 4, we develop Wald tests for the null hypotheses of short- and long-run symmetry against the alternative hypotheses of asymmetry. In Section 5, we scrutinize the finite sample properties of the estimators and test statistics using Monte Carlo simulations. Section 6 is devoted to our empirical application. We conclude in Section 7. Additional proofs are collected in an Appendix.

2 The NARDL Model in the Prior Literature

Consider the NARDL(p, q) process:

$$y_t = \sum_{j=1}^{p} \phi_{j*} y_{t-j} + \sum_{j=0}^{q} (\boldsymbol{\theta}_{j*}^{+'} \boldsymbol{x}_{t-j}^{+} + \boldsymbol{\theta}_{j*}^{-'} \boldsymbol{x}_{t-j}^{-}) + e_t, \tag{1}$$

where $\boldsymbol{x}_t \in \mathbb{R}^k$, $\boldsymbol{x}_t^{+} := \sum_{j=1}^{t} \Delta \boldsymbol{x}_j^{+}$, $\boldsymbol{x}_t^{-} := \sum_{j=1}^{t} \Delta \boldsymbol{x}_j^{-}$, $\Delta \boldsymbol{x}_t^{+} := \max[\boldsymbol{0}, \Delta \boldsymbol{x}_t]$, and $\Delta \boldsymbol{x}_t^{-} := \min[\boldsymbol{0}, \Delta \boldsymbol{x}_t]$, such that $\Delta \boldsymbol{x}_t$ is a stationary process. Note that SYG rewrite (1) in error-correction form as:

$$\Delta y_t = \rho_* y_{t-1} + \boldsymbol{\theta}_*^{+'} \boldsymbol{x}_{t-1}^{+} + \boldsymbol{\theta}_*^{-'} \boldsymbol{x}_{t-1}^{-} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^{+'} \Delta \boldsymbol{x}_{t-j}^{+} + \boldsymbol{\pi}_{j*}^{-'} \Delta \boldsymbol{x}_{t-j}^{-} \right) + e_t, \tag{2}$$

4

for some ρ_*, $\boldsymbol{\theta}_*^+$, $\boldsymbol{\theta}_*^{-1}$, γ_*, φ_{j*} ($j = 1, 2, \ldots, p-1$), $\boldsymbol{\pi}_{j*}^+$, and $\boldsymbol{\pi}_{j*}^-$ ($j = 0, 1, \ldots, q-1$), where $\{e_t, \mathcal{F}_t\}$ is a martingale difference sequence and $\mathcal{F}_t$ is the smallest σ-algebra driven by $\{y_{t-1}, \boldsymbol{x}_t^+, \boldsymbol{x}_t^-, y_{t-2}, \boldsymbol{x}_{t-1}^+, \boldsymbol{x}_{t-1}^-, \ldots\}$. If y_t is cointegrated with $(\boldsymbol{x}_t^{+\prime}, \boldsymbol{x}_t^{-\prime})'$, then we may rewrite (2) as:

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^{+\prime} \Delta \boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_{j*}^{-\prime} \Delta \boldsymbol{x}_{t-j}^- \right) + e_t, \tag{3}$$

where $u_{t-1} := y_{t-1} - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_{t-1}^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_{t-1}^-$ is the cointegrating error, $\boldsymbol{\beta}_*^+ := -(\boldsymbol{\theta}_*^+/\rho_*)$ and $\boldsymbol{\beta}_*^- := -(\boldsymbol{\theta}_*^-/\rho_*)$. Note that u_t is a stationary process that may be correlated with $\Delta \boldsymbol{x}_t$.

The NARDL process is able to capture a cointegrating relationship between a deterministic time trend process driven by a unit-root process and other unit-root processes, possibly associated with a time trend. Suppose that $\mathbb{E}[\Delta \boldsymbol{x}_t] \equiv \boldsymbol{0}$ and that $\boldsymbol{\mu}_*^+ := \mathbb{E}[\Delta \boldsymbol{x}_t^+]$ and $\boldsymbol{\mu}_*^- := \mathbb{E}[\Delta \boldsymbol{x}_t^-]$. It follows that $\boldsymbol{\mu}_*^+ + \boldsymbol{\mu}_*^- \equiv \boldsymbol{0}$ by construction. Therefore, if we further let $\boldsymbol{s}_t^+ := \Delta \boldsymbol{x}_t^+ - \boldsymbol{\mu}_*^+$ and $\boldsymbol{s}_t^- := \Delta \boldsymbol{x}_t^- - \boldsymbol{\mu}_*^-$, then:

$$\boldsymbol{x}_t^+ = \boldsymbol{\mu}_*^+ t + \sum_{j=1}^t \boldsymbol{s}_j^+ \quad \text{and} \quad \boldsymbol{x}_t^- = \boldsymbol{\mu}_*^- t + \sum_{j=1}^t \boldsymbol{s}_j^-. \tag{4}$$

It is clear from (4) that $\boldsymbol{x}_t^+$ and $\boldsymbol{x}_t^-$ are deterministic time-trend processes driven by unit-root processes. It follows that Δy_t is not necessarily distributed around zero even if $\boldsymbol{x}_t$ is a unit-root process without a deterministic trend. Note that $\rho_* := 1 - \sum_{j=1}^p \phi_{j*}$. From (1), we find that $\delta_* := \mathbb{E}[\Delta y_t] = -\frac{1}{\rho_*}[\sum_{j=0}^q (\boldsymbol{\theta}_{j*}^+)' \boldsymbol{\mu}_*^+ + \sum_{j=0}^q (\boldsymbol{\theta}_{j*}^-)' \boldsymbol{\mu}_*^-]$. Therefore, if we define $d_t := \Delta y_t - \delta_*$, then:

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

which shows that y_t is a deterministic time-trend process driven by a unit-root process, if $\delta_* \neq 0$. It is evident that the NARDL model captures a cointegrating relationship between a deterministic time-trend process driven by a unit-root process and a unit-root process without a deterministic time trend.

SYG propose to estimate the unknown parameters of (2) in a single step by OLS, and obtain the properties of the OLS estimator by simulation because it is not straightforward to derive the limit distributions of the single-step OLS estimator. To demonstrate this, we make the following assumptions:

Assumption 1.

(i) $\{(\Delta \boldsymbol{x}_t', u_t)'\}$ is a globally covariance stationary mixing process of $(k+1) \times 1$ vectors of ϕ of size $-r/(2(r-1))$ or α of size $-r/(r-2)$ and $r > 2$;

(ii) $\mathbb{E}[\Delta \boldsymbol{x}_t] = \boldsymbol{0}$, $\mathbb{E}[|\Delta \boldsymbol{x}_{ti}|^r] < \infty$ ($i = 1, 2, \ldots, k$), $\mathbb{E}[|u_t|^r] < \infty$, and $\mathbb{E}[|e_t|^2] < \infty$;

(iii) $\lim_{T \to \infty} var[T^{-1/2} \sum_{t=1}^T (\Delta \boldsymbol{x}_t', u_t)']$ exists and is positive definite; and

(iv) for some $(\rho_*, \boldsymbol{\theta}_*^{+\prime}, \boldsymbol{\theta}_*^{-\prime}, \gamma_*, \varphi_{1*}, \ldots, \varphi_{p-1*}, \boldsymbol{\pi}_{0*}^{+\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{-\prime})'$, Δy_t is generated by (2) such that $\{e_t, \mathcal{F}_t\}$ is a martingale difference sequence and $\mathcal{F}_t$ is the smallest σ-algebra driven by $\{y_{t-1}, \boldsymbol{x}_t^+, \boldsymbol{x}_t^-, y_{t-2}, \boldsymbol{x}_{t-1}^+,$

$$\boldsymbol{x}_{t-1}^-, \dots\}. \qquad \qquad \qquad \square$$

Furthermore, for notational simplicity, we define:

$$\boldsymbol{z}_t := [\; \boldsymbol{z}_{1t}' \; \vdots \; \boldsymbol{z}_{2t}' \;]' := [\; y_{t-1} \quad \boldsymbol{x}_{t-1}^{+\prime} \quad \boldsymbol{x}_{t-1}^{-\prime} \; \vdots \; 1 \quad \Delta \boldsymbol{y}_{t-1}' \quad \Delta \boldsymbol{x}_t^{+\prime} \quad \dots \quad \Delta \boldsymbol{x}_{t-q+1}^{+\prime} \quad \Delta \boldsymbol{x}_t^{-\prime} \quad \dots \quad \Delta \boldsymbol{x}_{t-q+1}^{-\prime} \;]',$$

where $\Delta \boldsymbol{y}_{t-1} := [\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p+1}]'$. Note that $\boldsymbol{z}_t$ is partitioned into nonstationary and stationary variables. Next, $\boldsymbol{z}_{2t}$ is further partitioned as:

$$\boldsymbol{z}_{2t} := [\; 1 \; \vdots \; \boldsymbol{w}_t' \;]' := [\; 1 \; \vdots \; \boldsymbol{w}_{1t}' \; \vdots \; \boldsymbol{w}_{2t}' \; \vdots \; \boldsymbol{w}_{3t}' \;]' := [\; 1 \; \vdots \; \Delta \boldsymbol{y}_{t-1}' \; \vdots \; \Delta \boldsymbol{x}_t^{+\prime} \quad \dots \quad \Delta \boldsymbol{x}_{t-q+1}^{+\prime} \; \vdots \; \Delta \boldsymbol{x}_t^{-\prime} \quad \dots \quad \Delta \boldsymbol{x}_{t-q+1}^{-\prime} \;]'.$$

In addition, we define:

$$\boldsymbol{\alpha}_* := [\; \boldsymbol{\alpha}_{1*}' \; \vdots \; \boldsymbol{\alpha}_{2*}' \;]' := [\; \rho_* \quad \boldsymbol{\theta}_*^{+\prime} \quad \boldsymbol{\theta}_*^{-\prime} \; \vdots \; \gamma_* \quad \boldsymbol{\varphi}_*' \quad \boldsymbol{\pi}_{0*}^{+\prime} \quad \dots \quad \boldsymbol{\pi}_{q-1*}^{+\prime} \quad \boldsymbol{\pi}_{0*}^{-\prime} \quad \dots \quad \boldsymbol{\pi}_{q-1*}^{-\prime} \;]',$$

where $\boldsymbol{\varphi}_* := [\varphi_{1*}, \varphi_{2*}, \dots, \varphi_{p-1*}]'$. With this notation in hand, the OLS estimator can be expressed as

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

Inference on the unknown parameters using $\widehat{\boldsymbol{\alpha}}_T$ is challenging, because $\sum_{t=1}^T \boldsymbol{z}_t \boldsymbol{z}_t'$ is asymptotically singular. This is shown in the following lemma:

Lemma 1. *Given Assumption 1:*

(i)
$$\frac{1}{T^3} \sum_{t=1}^T \boldsymbol{z}_{t1} \boldsymbol{z}_{t1}' \xrightarrow{\mathbb{P}} \mathbf{M}_{11} := \frac{1}{3} \begin{bmatrix} \delta_*^2 & \delta_* \boldsymbol{\mu}_*^{+\prime} & \delta_* \boldsymbol{\mu}_*^{-\prime} \\ \delta_* \boldsymbol{\mu}_*^+ & \boldsymbol{\mu}_*^+ \boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\mu}_*^+ \boldsymbol{\mu}_*^{-\prime} \\ \delta_* \boldsymbol{\mu}_*^- & \boldsymbol{\mu}_*^- \boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\mu}_*^- \boldsymbol{\mu}_*^{-\prime} \end{bmatrix};$$

(ii)
$$\frac{1}{T^2} \sum_{t=1}^T \boldsymbol{z}_{1t} \boldsymbol{z}_{2t}' \xrightarrow{\mathbb{P}} \mathbf{M}_{12} := \frac{1}{2} \begin{bmatrix} \delta_* & \delta_*^2 \boldsymbol{\iota}_{p-1}' & \delta_* \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^{+\prime} & \delta_* \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^{-\prime} \\ \boldsymbol{\mu}_*^+ & \delta_* \boldsymbol{\mu}_*^+ \boldsymbol{\iota}_{p-1}' & \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^+ \boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^+ \boldsymbol{\mu}_*^{-\prime} \\ \boldsymbol{\mu}_*^- & \delta_* \boldsymbol{\mu}_*^- \boldsymbol{\iota}_{p-1}' & \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^- \boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^- \boldsymbol{\mu}_*^{-\prime} \end{bmatrix}; \quad and$$

(iii)
$$\frac{1}{T} \sum_{t=1}^T \boldsymbol{z}_{2t} \boldsymbol{z}_{2t}' \xrightarrow{\mathbb{P}} \mathbf{M}_{22} := \begin{bmatrix} 1 & \vdots & \delta_* \boldsymbol{\iota}_{p-1}' & \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^{-\prime} \\ \cdots & & & & \\ \delta_* \boldsymbol{\iota}_{p-1} & \vdots & & & \\ \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_*^+ & \vdots & & \mathbb{E}[\boldsymbol{w}_t \boldsymbol{w}_t'] & \\ \boldsymbol{\iota}_q \otimes \boldsymbol{\mu}_*^- & \vdots & & & \end{bmatrix},$$

where $\boldsymbol{\iota}_a$ is an $a \times 1$ vector of ones. $\qquad \square$

6

Lemma 1 implies that, if we let $\mathbf{D}_T := \mathrm{diag}[T^{3/2}\mathbf{I}_{2+2k}, T^{1/2}\mathbf{I}_{p+2qk}]$, then:

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

which is singular. Due to this singularity, it is difficult to derive the limit distribution of $\widehat{\boldsymbol{\alpha}}_T$ directly. To do so would require one to derive the limit distribution of the determinant of $\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'$, which is analytically challenging. In practice, a higher-order approximation of $\left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'\right)^{-1}$ would be necessary to derive the limit distribution of the OLS estimator, and the limit distribution of the OLS estimator may depend on the estimated parameter values.

3 NARDL Estimation and Limit Distribution

In this section, we propose an analytically tractable two-step estimation procedure that draws on Engle and Granger (1987) and Phillips and Hansen (1990) and derive the relevant limit distributions. For clarity of exposition, we divide this section into two subsections, the first focusing on the estimation of the model with $k = 1$ and the second with $k > 1$. As detailed below, we provide a separate estimation framework for each case.

3.1 Estimation and Limit Distribution with $k = 1$

3.1.1 Estimation of the Long-Run Parameters

First Step Estimation by OLS: In keeping with the two-step estimation framework of Engle and Granger (1987), one may attempt to estimate the long-run parameters by OLS. Recall that the long-run relationship may be written as follows:

$$y_t = \alpha_* + \beta_*^+ x_t^+ + \beta_*^- x_t^- + u_t. \tag{6}$$

Now, define $\bar{\mathbf{D}}_T := \mathrm{diag}[T^{1/2}, T^{3/2}\mathbf{I}_2]$ and $\boldsymbol{v}_t := (1, x_t^+, x_t^-)'$ such that:

$$\bar{\mathbf{D}}_T^{-1}\left(\sum_{t=1}^{T} \boldsymbol{v}_t \boldsymbol{v}_t'\right)\bar{\mathbf{D}}_T^{-1} \xrightarrow{\mathbb{P}} \begin{bmatrix} 1 & \frac{1}{2}\mu_*^+ & \frac{1}{2}\mu_*^- \\ \frac{1}{2}\mu_*^+ & \frac{1}{3}\mu_*^+\mu_*^+ & \frac{1}{3}\mu_*^+\mu_*^- \\ \frac{1}{2}\mu_*^- & \frac{1}{3}\mu_*^-\mu_*^+ & \frac{1}{3}\mu_*^-\mu_*^- \end{bmatrix}.$$

By applying Lemma 1(*i*) and (*ii*), it is straightforward to show that this is a singular matrix, which frustrates analytical efforts to obtain the limit distribution of the OLS estimator. Note that this singularity arises from the collinearity of the trends in x_t^+ and x_t^-. We proceed by reparameterizing (6) in the following form, which facilitates estimation of the long-run parameters:

$$y_t = \alpha_* + \lambda_* x_t^+ + \eta_* x_t + u_t, \tag{7}$$

where $x_t \equiv x_t^+ + x_t^-$, $\lambda_* = \beta_*^+ - \beta_*^-$ and $\eta_* = \beta_*^-$. It follows that $\beta_*^+ = \lambda_* + \eta_*$ and $\beta_*^- = \eta_*$. It is possible to estimate $\boldsymbol{\varrho}_* := (\alpha_*, \lambda_*, \eta_*)'$ by OLS as follows:

$$\widehat{\boldsymbol{\varrho}}_T := (\widehat{\alpha}_T, \widehat{\lambda}_T, \widehat{\eta}_T)' := \arg\min_{\alpha,\,\lambda,\,\eta} \sum_{t=1}^{T} (y_t - \alpha - \lambda x_t^+ - \eta x_t)^2,$$

where we can recover $\widehat{\beta}_T^+ := \widehat{\lambda}_T + \widehat{\eta}_T$ and $\widehat{\beta}_T^- = \widehat{\eta}_T$. Notice that $(\widehat{\beta}_T^+, \widehat{\beta}_T^-)$ is identical to the OLS estimator obtained by regressing y_t on $(1, x_t^+, x_t^-)$. Now:

$$\widehat{\boldsymbol{\varrho}}_T = \boldsymbol{\varrho}_* + \left(\sum_{t=1}^{T} \boldsymbol{q}_t \boldsymbol{q}_t' \right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{q}_t u_t \right), \tag{8}$$

where $\boldsymbol{q}_t := (1, x_t^+, x_t)'$. For the analysis of the components in (8), we define:

$$\boldsymbol{\Sigma}_* := \begin{bmatrix} \sigma_*^{(1,1)} & \sigma_*^{(1,2)} \\ \sigma_*^{(2,1)} & \sigma_*^{(2,2)} \end{bmatrix} := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}[\boldsymbol{g}_t \boldsymbol{g}_s] \quad \text{and} \quad [\mathcal{B}_x(\cdot), \mathcal{B}_u(\cdot)]' := \boldsymbol{\Sigma}_*^{1/2} [\mathcal{W}_x(\cdot), \mathcal{W}_u(\cdot)]',$$

where $\boldsymbol{g}_t := [\Delta x_t, u_t]'$ and $[\mathcal{W}_x(\cdot), \mathcal{W}_u(\cdot)]'$ is a 2×1 vector of independent Wiener processes. If $\{u_t\}$ is serially uncorrelated and independent of $\{\Delta x_t\}$, then $\boldsymbol{\Sigma}_*$ and $[\mathcal{B}_x(\cdot), \mathcal{B}_u(\cdot)]'$ simplify to $\mathrm{diag}\,[\sigma_x^2, \sigma_u^2]$ and $[\sigma_x \mathcal{W}_x(\cdot), \sigma_u \mathcal{W}_u(\cdot)]'$, respectively, where $\sigma_x^2 := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}[\Delta x_t \Delta x_s]$ and $\sigma_u^2 := \mathbb{E}[u_t^2]$. The following lemma provides the limit behaviors of the components constituting the OLS estimator:

Lemma 2. *Given Assumption 1, if $k = 1$ and $\boldsymbol{\Sigma}_*$ is positive definite:*

(i)

$$\widehat{\mathbf{Q}}_T := \widetilde{\mathbf{D}}_T^{-1} \left(\sum_{t=1}^{T} \boldsymbol{q}_t \boldsymbol{q}_t' \right) \widetilde{\mathbf{D}}_T^{-1} \Rightarrow \mathcal{Q} := \begin{bmatrix} 1 & \frac{1}{2}\mu_*^+ & \int \mathcal{B}_x \\ \frac{1}{2}\mu_*^+ & \frac{1}{3}\mu_*^+ \mu_*^+ & \mu_*^+ \int r\mathcal{B}_x \\ \int \mathcal{B}_x & \int r\mathcal{B}_x \mu_*^+ & \int \mathcal{B}_x \mathcal{B}_x \end{bmatrix},$$

where $\widetilde{\mathbf{D}}_T := \mathrm{diag}[T^{1/2}, T^{3/2}, T]$; and

(ii) if $\upsilon_* := \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{t} \mathbb{E}[\Delta x_i u_t]$ is finite, then $\widehat{\mathbf{U}}_T := \widetilde{\mathbf{D}}_T^{-1} (\sum_{t=1}^{T} \boldsymbol{q}_t u_t) \Rightarrow \mathcal{U} := [\int d\mathcal{B}_u, \mu_*^+ \int r d\mathcal{B}_u, \int \mathcal{B}_x d\mathcal{B}_u + \upsilon_*]$ $\qquad \square$

Here, all integrals are computed with respect to $r \in [0, 1]$. For example, $\int r\mathcal{B}_x$ denotes $\int_0^1 r\mathcal{B}_x(r)dr$. Note that $\mathcal{Q}$ is nonsingular with probability 1, so the limit distribution of $\widehat{\boldsymbol{\varrho}}_T$ is obtained as a product of $\mathcal{Q}^{-1}$ and $\mathcal{U}$, as stated in the following corollary:

Corollary 1. *Given Assumption 1, if $k = 1$ and $\boldsymbol{\Sigma}_*$ is positive definite, $\widetilde{\mathbf{D}}_T(\widehat{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) \Rightarrow \mathcal{Q}^{-1}\mathcal{U}$.* $\qquad \square$

Corollary 1 has important implications, as summarized in the following remarks:

Remarks.

(a) By virtue of the reparameterization in (7), the collinearity between x_t^+ and x_t^- is removed. The fact that $\sum_{t=1}^{T} x_t = O_{\mathbb{P}}(T^{3/2})$ and $\sum_{t=1}^{T} x_t^+ = O_{\mathbb{P}}(T^2)$ leads to different convergence rates for $\widehat{\lambda}_T$ and $\widehat{\eta}_T$, viz., $\widehat{\lambda}_T - \lambda_* = O_{\mathbb{P}}(T^{-3/2})$ and $\widehat{\eta}_T - \eta_* = O_{\mathbb{P}}(T^{-1})$.

(b) Although x_t^+ is one of the regressors, the limit distribution of $\widehat{\varrho}_T$ is obtained without applying the functional central limit theorem (FCLT) to $\sum_{t=1}^{[T(\cdot)]}(x_t^+ - \mu_*^+)$. We apply the FCLT only to $\sum_{t=1}^{T(\cdot)} \boldsymbol{g}_t$ and obtain the limit distribution.

(c) Using the definition of $\widehat{\lambda}_T$, we have: $T\{(\widehat{\beta}_T^+ - \widehat{\beta}_T^-) - (\beta_*^+ - \beta_*^-)\} = O_{\mathbb{P}}(T^{-1/2})$. This implies that $T(\widehat{\beta}_T^+ - \beta_*^+) = T(\widehat{\beta}_T^- - \beta_*^-) + o_{\mathbb{P}}(1)$, such that the limit distributions of $T(\widehat{\beta}_T^+ - \beta_*^+)$ and $T(\widehat{\beta}_T^- - \beta_*^-)$ are equivalent. Because the convergence rate of the long-run parameter estimator is faster than $T^{1/2}$, $\widehat{\beta}_T^+$ and $\widehat{\beta}_T^-$ can be treated as known when estimating the short-run dynamic parameters in the second step. $\qquad\square$

The following theorem presents the limit distribution of the OLS estimator of the long-run parameters:

Theorem 1. *Given Assumption 1, if $k = 1$, $T[(\widehat{\beta}_T^+ - \beta_*^+), (\widehat{\beta}_T^- - \beta_*^-)]' \Rightarrow \iota_2 \otimes \mathbf{S}\mathcal{Q}^{-1}\mathcal{U}$, where $\mathbf{S} := [\mathbf{0}_{1\times 2}, 1]$.* $\qquad\square$

First Step Estimation by FM-OLS: Note that the limit distribution in Theorem 1 is non-normal and depends on the nuisance parameters, $\mathbf{\Sigma}_*$ and υ_*. Due to the presence of nuisance parameters, the limit distribution cannot be readily exploited for inference on the long-run parameters. Furthermore, except in the special case in which $\{u_t\}$ is independent of $\{\Delta x_t\}$ and/or serially uncorrelated, the OLS estimator of the long-run parameter exhibits an asymptotic bias determined by υ_*. The FM-OLS estimator developed by Phillips and Hansen (1990) overcomes these problems. It is free from asymptotic bias even in the presence of endogenous regressors and/or serial correlation and it follows an asymptotic mixed normal distribution. We therefore advocate the use of FM-OLS to estimate the long-run cointegrating parameters in the first step.

Suppose that $\mathbf{\Sigma}_*$ can be consistently estimated by a heteroskedasticity and autocorrelation consistent covariance matrix estimator. For example, following Newey and West (1987), we have:

$$\widetilde{\mathbf{\Sigma}}_T := \begin{bmatrix} \widetilde{\sigma}_T^{(1,1)} & \widetilde{\sigma}_T^{(1,2)} \\ \widetilde{\sigma}_T^{(2,1)} & \widetilde{\sigma}_T^{(2,2)} \end{bmatrix} := \frac{1}{T}\sum_{t=1}^{T} \widehat{\boldsymbol{g}}_t \widehat{\boldsymbol{g}}_t' + \frac{1}{T}\sum_{k=1}^{\ell} \omega_{\ell k} \sum_{t=k+1}^{T} \left\{ \widehat{\boldsymbol{g}}_{t-k}\widehat{\boldsymbol{g}}_t' + \widehat{\boldsymbol{g}}_t \widehat{\boldsymbol{g}}_{t-k}' \right\},$$

where $\widehat{\boldsymbol{g}}_t := [\Delta x_t, \widehat{u}_t]'$, $\omega_{\ell k} := 1 - k/(1+\ell)$, $\ell = O(T^{1/4})$ and $\widehat{u}_t := y_t - \widehat{\alpha}_T - \widehat{\beta}_T^+ x_t^+ - \widehat{\beta}_T^- x_t^-$. In addition, under mild regularity conditions, it is straightforward to show that the asymptotic bias, υ_*, in $\mathcal{U}$ can be consistently estimated by:

$$\widetilde{\mathbf{\Pi}}_T := \begin{bmatrix} \widetilde{\pi}_T^{(1,1)} & \widetilde{\pi}_T^{(1,2)} \\ \widetilde{\pi}_T^{(2,1)} & \widetilde{\pi}_T^{(2,2)} \end{bmatrix} := \frac{1}{T}\sum_{k=0}^{\ell}\sum_{t=k+1}^{T} \widehat{\boldsymbol{g}}_{t-k}\widehat{\boldsymbol{g}}_t'.$$

Now, define the following long-run parameter estimator:

$$\widetilde{\boldsymbol{\varrho}}_T := (\widetilde{\alpha}_T, \widetilde{\lambda}_T, \widetilde{\eta}_T)' := \left(\sum_{t=1}^{T} \boldsymbol{q}_t \boldsymbol{q}_t'\right)^{-1}\left(\sum_{t=1}^{T} \boldsymbol{q}_t \widetilde{y}_t - T\mathbf{S}'\widetilde{\upsilon}_T\right),$$

9

where $\widetilde{y}_t := y_t - \Delta x_t \left(\widetilde{\sigma}_T^{(1,1)}\right)^{-1} \widetilde{\sigma}_T^{(1,2)}$ and $\widetilde{v}_T := \widetilde{\pi}_T^{(1,2)} - \widetilde{\pi}_T^{(1,1)} \left(\widetilde{\sigma}_T^{(1,1)}\right)^{-1} \widetilde{\sigma}_T^{(1,2)}$. Finally, the FM-OLS estimators of the long-run parameters are obtained as $\widetilde{\beta}_T^+ := \widetilde{\lambda}_T + \widetilde{\eta}_T$ and $\widetilde{\beta}_T^- := \widetilde{\eta}_T$. Note that these estimators are designed to remove the asymptotic bias, as in Phillips and Hansen (1990). To derive the limit distribution of the FM-OLS estimator, we add the following regularity conditions:

Assumption 2.

(i) $\mathbf{\Sigma}_*$ is finite and positive definite and $\widetilde{\mathbf{\Sigma}}_T \xrightarrow{\mathbb{P}} \mathbf{\Sigma}_*$; and

(ii) $\mathbf{\Pi}_*$ is finite and $\widetilde{\mathbf{\Pi}}_T \xrightarrow{\mathbb{P}} \mathbf{\Pi}_*$, where:

$$\mathbf{\Pi}_* := \begin{bmatrix} \pi_*^{(1,1)} & \pi_*^{(1,2)} \\ \pi_*^{(2,1)} & \pi_*^{(2,2)} \end{bmatrix} := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{t} \mathbb{E}[\boldsymbol{g}_i \boldsymbol{g}_t']. \qquad \square$$

Note that the definition of $\pi_*^{(1,2)}$ is identical to υ_*. The following lemma provides the limit behavior of the constituent components of the FM-OLS estimator:

Lemma 3. *Given Assumptions 1 and 2, , if $k = 1$, $\widetilde{\mathbf{U}}_T := \widetilde{\mathbf{D}}_T^{-1}\{\sum_{t=1}^{T} \boldsymbol{q}_t(u_t - \Delta x_t(\widetilde{\sigma}_T^{(1,1)})^{-1}\widetilde{\sigma}_T^{(1,2)}) - T\mathbf{S}'\widetilde{v}_T\} \Rightarrow \widetilde{\mathcal{U}} := \tau_*[\int d\mathcal{W}_u, \mu_*^+ \int r d\mathcal{W}_u, \int \mathcal{B}_x d\mathcal{W}_u]'$, where $\tau_*^2 := plim_{T \to \infty}\widetilde{\tau}_T^2$ and $\widetilde{\tau}_T^2 := \widetilde{\sigma}_T^{(2,2)} - \widetilde{\sigma}_T^{(2,1)}(\widetilde{\sigma}_T^{(1,1)})^{-1}\widetilde{\sigma}_T^{(1,2)}$.* $\qquad \square$

Given Lemma 2(*i*), $\widehat{\mathbf{Q}}_T \Rightarrow \mathcal{Q}$, which is nonsingular with probability 1. The limit distribution of $\widetilde{\varrho}_T$ can therefore be obtained as the product of $\mathcal{Q}^{-1}$ and $\widetilde{\mathcal{U}}$, as stated in the following corollary:

Corollary 2. *Given Assumption 1, if $k = 1$, $\widetilde{\mathbf{D}}_T(\widetilde{\varrho}_T - \varrho_*) \Rightarrow \mathcal{Q}^{-1}\widetilde{\mathcal{U}}$.* $\qquad \square$

Corollary 2 has a number of important implications, as outlined in the following Remarks:

Remarks.

(a) The limit distribution of the FM-OLS estimator is mixed normal. Conditional on $\sigma\{\mathcal{B}_x(r), r \in (0, 1]\}$, the limit distribution of $\widetilde{\mathbf{D}}_T(\widetilde{\varrho}_T - \varrho_*)$ is $N(\mathbf{0}, \tau_*^2 \mathcal{Q}^{-1})$. Consequently, if a Wald test statistic is constructed using the FM-OLS estimator, its null limit distribution will be chi-squared.

(b) As in the case of the 2-step OLS estimator, we have: $T(\widetilde{\beta}_T^+ - \beta_*^+) = T(\widetilde{\beta}_T^- - \beta_*^-) + o_{\mathbb{P}}(1)$, such that the limit distribution of $\widetilde{\beta}_T^+$ is equivalent to that of $\widetilde{\beta}_T^-$. Furthermore, the limit distribution of $\widetilde{\beta}_T^-$ is given by that of $\widetilde{\eta}_T$.

(c) The convergence rates of $\widetilde{\beta}_T^+$ and $\widetilde{\beta}_T^-$ are both T. Because their convergence rates exceed $T^{1/2}$, we can estimate the short-run parameters in the second stage regression by replacing u_{t-1} with $\widetilde{u}_{t-1} := y_{t-1} - \widetilde{\alpha}_T - \widetilde{\beta}_T^+ x_{t-1}^+ - \widetilde{\beta}_T^- x_{t-1}^-$. $\qquad \square$

The following theorem formally presents the limit distribution of the FM-OLS estimator:

Theorem 2. *Given Assumptions 1 and 2, if $k = 1$, $T[(\widetilde{\beta}_T^+ - \beta_*^+), (\widetilde{\beta}_T^- - \beta_*^-)]' \Rightarrow \boldsymbol{\iota}_2 \otimes \mathbf{S}\mathcal{Q}^{-1}\widetilde{\mathcal{U}}$.* $\qquad \square$

3.1.2 Estimation of the Short-Run Parameters

As we have shown above, the long-run coefficients can be estimated by an estimator with a convergence rate faster than $T^{1/2}$, so we can treat them as known in order to estimate the short-run parameters. Let:

$$u_{t-1} := y_{t-1} - \beta_*^{+'} x_{t-1}^+ - \beta_*^{-'} x_{t-1}^-. \tag{9}$$

Assuming that β_*^+ and β_*^- are known, we can rewrite (2) as:

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^{+'} \Delta \boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_{j*}^{-'} \Delta \boldsymbol{x}_{t-j}^- \right) + e_t. \tag{10}$$

Note that all variables in (10) are stationary, so the unknown parameters can be estimated by OLS. If we define $\boldsymbol{\zeta}_* := (\rho_*, \boldsymbol{\beta}_{2*}')'$ and $\boldsymbol{h}_t := (u_{t-1}, \boldsymbol{z}_{2t}')'$, where $\boldsymbol{\beta}_{2*} := (\gamma_*, \varphi_{1*}, \dots, \varphi_{p-1*}, \boldsymbol{\pi}_{0*}^{+'}, \dots, \boldsymbol{\pi}_{q-1*}^{+'}, \boldsymbol{\pi}_{0*}^{-'}, \dots, \boldsymbol{\pi}_{q-1*}^{-'},)'$, then (10) can be rewritten as $\Delta y_t = \boldsymbol{\zeta}_*' \boldsymbol{h}_t + e_t$, and we can obtain the OLS estimator as follows:

$$\widehat{\boldsymbol{\zeta}}_T := \left(\sum_{t=1}^{T} \boldsymbol{h}_t \boldsymbol{h}_t' \right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{h}_t \Delta y_t \right) = \boldsymbol{\zeta}_* + \left(\sum_{t=1}^{T} \boldsymbol{h}_t \boldsymbol{h}_t' \right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{h}_t e_t \right). \tag{11}$$

The following lemma shows the limit behaviors of the constituent components of $\widehat{\boldsymbol{\zeta}}_T$:

Lemma 4. *Given Assumption 1:*

(i)
$$\widehat{\boldsymbol{\Gamma}}_T := \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{h}_t \boldsymbol{h}_t' \xrightarrow{\mathbb{P}} \boldsymbol{\Gamma}_* := \begin{bmatrix} \mathbb{E}[u_t^2] & \mathbb{E}[u_t \boldsymbol{z}_{2t}'] \\ \mathbb{E}[\boldsymbol{z}_{2t} u_t] & \mathbf{M}_{22} \end{bmatrix};$$

(ii) $T^{-1/2} \sum_{t=1}^{T} \boldsymbol{h}_t e_t \overset{A}{\sim} N[\mathbf{0}, \boldsymbol{\Omega}_*]$, *where* $\boldsymbol{\Omega}_* := \mathbb{E}[e_t^2 \boldsymbol{h}_t \boldsymbol{h}_t']$; *and*

(iii) *in the special case where* $\mathbb{E}[e_t^2 | \boldsymbol{h}_t] = \sigma_*^2$, $\boldsymbol{\Omega}_*$ *simplifies to* $\sigma_*^2 \boldsymbol{\Gamma}_*$. □

We omit the proof of Lemma 4, because it is straightforward. Using Lemma 4, it is possible to derive the limit distribution of $\widehat{\boldsymbol{\zeta}}_T$, which is provided in the following theorem:

Theorem 3. *Given Assumption 1, if* $\boldsymbol{\Gamma}_*$ *and* $\boldsymbol{\Omega}_*$ *are positive definite:*

(i) $\sqrt{T}(\widehat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}_*) \overset{A}{\sim} N(\mathbf{0}, \boldsymbol{\Gamma}_*^{-1} \boldsymbol{\Omega}_* \boldsymbol{\Gamma}_*^{-1})$; *and*

(ii) *if it further holds that* $\mathbb{E}[e_t^2 | \boldsymbol{h}_t] = \sigma_*^2$, *then* $\sqrt{T}(\widehat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}_*) \overset{A}{\sim} N(\mathbf{0}, \sigma_*^2 \boldsymbol{\Gamma}_*^{-1})$. □

Theorem 3 shows that, if there is any estimator converging to the cointegrating coefficient faster than $T^{1/2}$ such as the first-step OLS and FM-OLS estimators, then we can use the resulting parameter estimate as if it is known.

3.2 Estimation and Limit Distribution with $k > 1$

3.2.1 Estimation of the Long-Run Parameters

First-Step Transformed OLS: If more than a single explanatory variable enters the long-run relationship, then the two-step estimation procedure outlined above is not appropriate. This can be seen by letting $x_t \equiv x_t^+ + x_t^-$, $\lambda_* = \beta_*^+ - \beta_*^-$ and $\eta_* = \beta_*^-$ with $k > 1$, such that:

$$y_t = \alpha_* + \lambda_*' x_t^+ + \eta_*' x_t + u_t. \tag{12}$$

If we further let $q_t := (1, x_t^{+\prime}, x_t')'$ then, by extending Lemma 2, it follows that:

$$\widehat{\mathbf{Q}}_T := \widetilde{\mathbf{D}}_T^{-1} \left(\sum_{t=1}^{T} q_t q_t' \right) \widetilde{\mathbf{D}}_T^{-1} \Rightarrow \mathcal{Q} := \begin{bmatrix} 1 & \frac{1}{2}\mu_*^+ & \int \mathcal{B}_x' \\ \frac{1}{2}\mu_*^+ & \frac{1}{3}\mu_*^+ \mu_*^{+\prime} & \mu_*^+ \int r\mathcal{B}_x' \\ \int \mathcal{B}_x dr & \int r\mathcal{B}_x \mu_*^{+\prime} & \int \mathcal{B}_x \mathcal{B}_x' \end{bmatrix},$$

where $\widetilde{\mathbf{D}}_T := \mathrm{diag}[T^{1/2}, T^{3/2}\mathbf{I}_k, T\mathbf{I}_k]$, and $\mathcal{B}_x(\cdot)$ is a $k \times 1$ vector of Brownian motions, such that:

$$\left[\mathcal{B}_x(\cdot)', \mathcal{B}_u(\cdot) \right]' := \Sigma_*^{1/2} \left[\mathcal{W}_x(\cdot)', \mathcal{W}_u(\cdot)' \right]' \quad \text{and} \quad \Sigma_* := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}[g_t g_s],$$

with $g_t := [\Delta x_t, u_t]'$ and $[\mathcal{W}_x(\cdot)', \mathcal{W}_u(\cdot)]'$ being a $(k+1) \times 1$ vector of independent Wiener processes. Note that the blocks on the second-row of $\mathcal{Q}$ now form a sub-matrix with rank equal to unity. That is:

$$\left[\begin{array}{ccc} \frac{1}{2}\mu_*^+ & \frac{1}{3}\mu_*^+ \mu_*^{+\prime} & \mu_*^+ \int r\mathcal{B}_x' \end{array} \right] = \mu_*^+ \left[\begin{array}{ccc} \frac{1}{2} & \frac{1}{3}\mu_*^{+\prime} & \int r\mathcal{B}_x' \end{array} \right],$$

which implies that $\mathcal{Q}$ is a singular matrix with probability 1, so the two-step procedure that we develop above cannot be used to estimate the long-run parameters. To overcome this issue, we first note that, if $m_t := \sum_{j=1}^{t} s_j^+$ in (4), then m_t is a unit-root process, the increments of which have a population mean of zero. Therefore, if x_t^+ is regressed against t, μ_*^+ can be estimated by $\widehat{\mu}_T^+ := \left(\sum_{t=1}^{T} t^2 \right)^{-1} \sum_{t=1}^{T} t x_t^+ = \mu_*^+ + \left(\sum_{t=1}^{T} t^2 \right)^{-1} \sum_{t=1}^{T} t m_t$, where $\widehat{m}_t := x_t^+ - \widehat{\mu}_T^+ t$ is the regression residual. Consequently, $m_t = \widehat{m}_t + t d_T$, where $d_T := \left(\sum_{t=1}^{T} t^2 \right)^{-1} \sum_{t=1}^{T} t m_t$, so that:

$$x_t^+ = \widehat{m}_t + (\mu_*^+ + d_T)t. \tag{13}$$

Under the regularity conditions given above, $d_T = O_{\mathbb{P}}(T^{-1/2})$. Consequently, if we let $\delta_{*T} := \mu_*^+ + d_T$, then $\delta_{*T} = \mu_* + O_{\mathbb{P}}(T^{-1/2})$. Now, we rewrite (12) using (13) to obtain:

$$y_t = \alpha_* + \lambda_*'(\mu_*^+ + d_T)t + \lambda_*' \widehat{m}_t + \eta_*' x_t + u_t. \tag{14}$$

Therefore, if y_t is regressed against $\boldsymbol{r}_t := (1, t, \widehat{\boldsymbol{m}}_t', \boldsymbol{x}_t')'$, then $\boldsymbol{\lambda}_*$ and $\boldsymbol{\eta}_*$ can be consistently estimated. Let $\dddot{\boldsymbol{\varpi}}_T :=$ $(\ddot{\alpha}_T, \ddot{\xi}_{*T}, \ddot{\boldsymbol{\lambda}}_T', \ddot{\boldsymbol{\eta}}_T')'$ be the OLS estimate for $\boldsymbol{\varpi}_{*T} := (\alpha_*, \xi_{*T}, \boldsymbol{\lambda}_*', \boldsymbol{\eta}_*')'$, where $\xi_{*T} := \boldsymbol{\lambda}_*' \boldsymbol{\delta}_{*T}$. The long-run coefficients are obtained by $\ddot{\boldsymbol{\beta}}_T^+ := \ddot{\boldsymbol{\lambda}}_T + \ddot{\boldsymbol{\eta}}_T$ and $\ddot{\boldsymbol{\beta}}_T^- := \ddot{\boldsymbol{\eta}}_T$ for $\boldsymbol{\beta}_*^+$ and $\boldsymbol{\beta}_*^-$, respectively, similarly to $(\widehat{\boldsymbol{\beta}}_T^+, \widehat{\boldsymbol{\beta}}_T^-)$. If:

$$
\ddot{\mathbf{S}} := \begin{bmatrix} \mathbf{0}_{k \times 1} & \mathbf{0}_{k \times 1} & \mathbf{I}_k & \mathbf{I}_k \\ \mathbf{0}_{k \times 1} & \mathbf{0}_{k \times 1} & \mathbf{0}_{k \times k} & \mathbf{I}_k \end{bmatrix}, \quad \text{then} \quad \begin{bmatrix} \widehat{\boldsymbol{\beta}}_T^+ \\ \widehat{\boldsymbol{\beta}}_T^- \end{bmatrix} = \ddot{\mathbf{S}} \dddot{\boldsymbol{\varpi}}_T.
$$

We refer to this estimator as the *first-step transformed OLS (TOLS) estimator*. The intuition of the TOLS estimator is straightforward—because it is the collinear trend in $\boldsymbol{x}_t^+$ that results in the singularity of $\mathcal{Q}$, which, in turn, renders first-step estimation by OLS and FM-OLS inoperable, we de-trend $\boldsymbol{x}_t^+$ prior to estimation.

The limit distribution of the first-step TOLS estimator is obtained in a similar fashion to the limit distribution of the first-step OLS estimator. First, note that:

$$
\dddot{\boldsymbol{\varpi}}_T = \boldsymbol{\varpi}_{*T} + \left(\sum_{t=1}^T \boldsymbol{r}_t \boldsymbol{r}_t' \right)^{-1} \sum_{t=1}^T \boldsymbol{r}_t u_t.
$$

To characterize the limit behaviors of the components on the right side, we first define:

$$
\ddot{\boldsymbol{\Sigma}}_* := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}[\ddot{\boldsymbol{g}}_t \ddot{\boldsymbol{g}}_t'] \quad \text{and} \quad [\mathcal{B}_m(\cdot)', \mathcal{B}_x(\cdot)', \mathcal{B}_u(\cdot)]' := \ddot{\boldsymbol{\Sigma}}_*^{1/2} [\mathcal{W}_m(\cdot)', \mathcal{W}_x(\cdot)', \mathcal{W}_u(\cdot)]',
$$

where $\ddot{\boldsymbol{g}}_t := [\Delta \boldsymbol{m}_t', \Delta \boldsymbol{x}_t, u_t]'$ and $[\mathcal{W}_m(\cdot)', \mathcal{W}_x(\cdot)', \mathcal{W}_u(\cdot)]'$ is a $(2k+1) \times 1$ vector of independent Wiener processes.

Lemma 5. *Given Assumption 1, if $\ddot{\boldsymbol{\Sigma}}_*$ is positive definite:*

(i) $\ddot{\mathbf{R}}_T := \ddot{\mathbf{D}}_T^{-1} \left(\sum_{t=1}^T \boldsymbol{r}_t \boldsymbol{r}_t' \right) \ddot{\mathbf{D}}_T^{-1} \Rightarrow \mathcal{R}$, *where:*

$$
\mathcal{R} := \begin{bmatrix} 1 & \frac{1}{2} & \int (1 - \frac{3}{2}r) \mathcal{B}_m' & \int \mathcal{B}_x' \\ \frac{1}{2} & \frac{1}{3} & \mathbf{0}_{1 \times k} & \int r \mathcal{B}_x' \\ \int (1 - \frac{3}{2}r) \mathcal{B}_m & \mathbf{0}_{k \times 1} & \int \mathcal{B}_m \mathcal{B}_m' - 3 \int r \mathcal{B}_m \int r \mathcal{B}_m' & \int \mathcal{B}_m \mathcal{B}_x' - 3 \int r \mathcal{B}_m \int r \mathcal{B}_x' \\ \int \mathcal{B}_x & \int r \mathcal{B}_x & \int \mathcal{B}_x \mathcal{B}_m' - 3 \int r \mathcal{B}_x \int r \mathcal{B}_m' & \int \mathcal{B}_x \mathcal{B}_x' \end{bmatrix}
$$

and $\ddot{\mathbf{D}}_T := diag[T^{1/2}, T^{3/2}, T\mathbf{I}_{2k}]$; and

(ii) if $\boldsymbol{v}_{x*} := \lim_{T \to \infty} T^{-1} \sum_{t=1}^T \sum_{i=1}^t \mathbb{E}[\Delta \boldsymbol{x}_i u_t]$ and $\boldsymbol{v}_{m*} := \lim_{T \to \infty} T^{-1} \sum_{t=1}^T \sum_{i=1}^t \mathbb{E}[\Delta \boldsymbol{m}_i u_t]$ are finite, then
$$
\ddot{\mathbf{U}}_T := \ddot{\mathbf{D}}_T^{-1} \left(\sum_{t=1}^T \boldsymbol{r}_t u_t \right) \Rightarrow \ddot{\mathcal{U}} := [\int d\mathcal{B}_u, \int r d\mathcal{B}_u, \int \mathcal{B}_m' d\mathcal{B}_u - 3 \int r d\mathcal{B}_u \int r \mathcal{B}_m' + \boldsymbol{v}_{m*}', \int \mathcal{B}_x' d\mathcal{B}_u + \boldsymbol{v}_{x*}']'. \quad \square
$$

Note that $\mathcal{R}$ is no longer singular because $\sum_{t=1}^{[T(\cdot)]} \ddot{\boldsymbol{g}}_t$ obeys the FCLT using partially correlated increments. The following corollary shows the limit behavior of $\dddot{\boldsymbol{\varpi}}_T$:

Corollary 3. *Given Assumption 1, $\ddot{\mathbf{D}}_T(\dddot{\boldsymbol{\varpi}}_T - \boldsymbol{\varpi}_{*T}) \Rightarrow \mathcal{R}^{-1} \ddot{\mathcal{U}}$ and $T^{1/2}(\widehat{\xi}_T - \boldsymbol{\lambda}_*' \boldsymbol{\mu}_*^+) \Rightarrow 3 \boldsymbol{\lambda}_*' \int r \mathcal{B}_m$.* $\quad \square$

The first part of Corollary 3 is straightforward from the structure of the first-step TOLS estimator. For the second part, notice that $\widehat{\xi}_T$ is not of primary interest. Although the convergence rate of $(\widehat{\xi}_T - \xi_{*T})$ is $T^{3/2}$, as given in the first part, $\xi_{*T} := \boldsymbol{\lambda}_*'\boldsymbol{\delta}_{*T}$ converges to $\boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+$ at the rate $T^{1/2}$. This implies that $T^{1/2}(\widehat{\xi}_T - \boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+)$ is asymptotically bounded in probability. The following theorem provides the limit distribution of the long-run parameter estimator.

Theorem 4. *Given Assumption 1, if $\ddot{\boldsymbol{\Sigma}}_*$ is positive definite, then $T[(\ddot{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)', (\ddot{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)']' \Rightarrow \ddot{\mathbf{S}}\mathcal{R}^{-1}\ddot{\mathcal{U}}.$* ☐

First-Step FM-TOLS: Despite the simplicity of the TOLS estimator of the long-run parameters, the limit distribution in Theorem 4 does not provide a straightforward basis for inference because it is non-standard and it exhibits asymptotic bias driven by $\boldsymbol{v}_{*m}$ and $\boldsymbol{v}_{*x}$. We therefore provide an alternative estimator in the same spirit as the first-step FM-OLS estimator introduced in Section 3.1.1 for the case where $k = 1$. We begin by making the following assumption:

Assumption 3.

(i) $\ddot{\boldsymbol{\Sigma}}_*$ *is finite and positive definite, such that there is a consistent estimator for* $\ddot{\boldsymbol{\Sigma}}_*$:

$$
\bar{\boldsymbol{\Sigma}}_T := \left[\begin{array}{cc} \bar{\boldsymbol{\Sigma}}_T^{(1,1)} & \bar{\boldsymbol{\sigma}}_T^{(1,2)} \\ \bar{\boldsymbol{\sigma}}_T^{(2,1)} & \bar{\sigma}_T^{(2,2)} \end{array} \right] \xrightarrow{\mathbb{P}} \ddot{\boldsymbol{\Sigma}}_* := \left[\begin{array}{cc} \ddot{\boldsymbol{\Sigma}}_*^{(1,1)} & \ddot{\boldsymbol{\sigma}}_*^{(1,2)} \\ \ddot{\boldsymbol{\sigma}}_*^{(2,1)} & \sigma_*^{(2,2)} \end{array} \right]; \quad and
$$

(ii) *if we let* $\bar{\boldsymbol{\Pi}}_T := T^{-1}\sum_{k=0}^{\ell}\sum_{t=k+1}^{T}\bar{\boldsymbol{g}}_{t-k}\bar{\boldsymbol{g}}_t'$:

$$
\left[\begin{array}{cc} \bar{\boldsymbol{\Pi}}_T^{(1,1)} & \bar{\boldsymbol{\pi}}_T^{(1,2)} \\ \bar{\boldsymbol{\pi}}_T^{(2,1)} & \bar{\pi}_T^{(2,2)} \end{array} \right] := \bar{\boldsymbol{\Pi}}_T \xrightarrow{\mathbb{P}} \ddot{\boldsymbol{\Pi}}_* := \left[\begin{array}{cc} \ddot{\boldsymbol{\Pi}}_*^{(1,1)} & \ddot{\boldsymbol{\pi}}_*^{(1,2)} \\ \ddot{\boldsymbol{\pi}}_*^{(2,1)} & \pi_*^{(2,2)} \end{array} \right] := \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\mathbb{E}[\ddot{\boldsymbol{g}}_t\ddot{\boldsymbol{g}}_t'],
$$

which is finite, where $\bar{\boldsymbol{g}}_t := [\Delta\widehat{\boldsymbol{m}}_t', \Delta\boldsymbol{x}_t', \ddot{u}_t]'$ *and* $\ddot{u}_t := y_t - \ddot{\alpha}_T - \ddot{\boldsymbol{\beta}}_T^{+'}\boldsymbol{x}_t^+ - \ddot{\boldsymbol{\beta}}_T^{-'}\boldsymbol{x}_t^-.$ ☐

Note that Assumption 3 corresponds to Assumption 2 in the case where $k > 1$. The *fully-modified TOLS (FM-TOLS) estimator* is defined as follows:

$$
\bar{\boldsymbol{\varpi}}_T := (\bar{\alpha}_T, \bar{\xi}_T, \bar{\boldsymbol{\beta}}_T^{+'}, \bar{\boldsymbol{\beta}}_T^{-'})' := \left(\sum_{t=1}^{T}\boldsymbol{r}_t\boldsymbol{r}_t' \right)^{-1} \left(\sum_{t=1}^{T}\boldsymbol{r}_t\bar{y}_t - T\bar{\mathbf{S}}'\bar{\boldsymbol{v}}_T \right),
$$

where $\bar{y}_t := y_t - \boldsymbol{\ell}_t'(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)}$, $\boldsymbol{\ell}_t := (\Delta\widehat{\boldsymbol{m}}_t', \Delta\boldsymbol{x}_t')'$, $\bar{\boldsymbol{v}}_T := \bar{\pi}_T^{(1,2)} - \bar{\boldsymbol{\Pi}}_T^{(1,1)}(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)}$ and $\bar{\mathbf{S}} := [\mathbf{0}_{2k\times 2}, \mathbf{I}_{2k}]$. The limit distribution of the FM-TOLS estimator is obtained in a similar way to the limit distribution of the FM-OLS estimator. The following lemma corresponds to Lemma 3 in the case where $k > 1$:

Lemma 6. *Given Assumptions 1 and 3, $\bar{\mathbf{U}}_T := \ddot{\mathbf{D}}_T^{-1}\{\sum_{t=1}^{T}\boldsymbol{r}_t(u_t - \boldsymbol{\ell}_t'(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)}) - T\bar{\mathbf{S}}'\bar{\boldsymbol{v}}_T\} \Rightarrow \bar{\mathcal{U}} := \ddot{\tau}[\int d\mathcal{W}_u,$*
$\int r d\mathcal{W}_u, \int \mathcal{B}_m' d\mathcal{W}_u - 3\int r d\mathcal{W}_u \int r\mathcal{B}_m', \int \mathcal{B}_x' d\mathcal{W}_u]'$, where $\ddot{\tau}^2 := plim_{T\to\infty}\bar{\tau}_T^2$ and $\bar{\tau}_T^2 := \bar{\sigma}_T^{(2,2)} - \bar{\boldsymbol{\sigma}}_T^{(2,1)}(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)}.$ ☐

Given Lemmas 5 and 6, the limit distribution of the FM-TOLS estimator is obtained as the product of $\mathcal{R}^{-1}$ and $\bar{\mathcal{U}}$, which is formally stated in the following theorem:

Theorem 5. *Given Assumptions 1 and 3, $\ddot{\mathbf{D}}_T(\bar{\boldsymbol{\varpi}}_T - \bar{\boldsymbol{\varpi}}_{*T}) \Rightarrow \mathcal{R}^{-1}\bar{\mathcal{U}}$ and $T[(\bar{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)', (\bar{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)']' \Rightarrow \ddot{\mathbf{S}}\mathcal{R}^{-1}\bar{\mathcal{U}}$.* □

The limit distribution of the FM-TOLS estimator is mixed normal. Conditional on $\sigma\{(\mathcal{B}_m(r)', \mathcal{B}_x(r)')', r \in (0, 1]\}$, the limit distribution of $\bar{\mathbf{D}}_T(\bar{\boldsymbol{\varpi}}_T - \bar{\boldsymbol{\varpi}}_{*T})$ is $N(\mathbf{0}, \ddot{\tau}^2 \mathcal{R}^{-1})$.

3.2.2 Estimation of the Short-Run Parameters

As in Section 3.1.2, suppose that the cointegrating coefficient is known or can be estimated by an estimator with a convergence rate faster than $T^{1/2}$. That is:

$$u_{t-1} := y_{t-1} - \boldsymbol{\beta}_*^{+'}\boldsymbol{x}_{t-1}^+ - \boldsymbol{\beta}_*^{-'}\boldsymbol{x}_{t-1}^- = y_{t-1} - \boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+(t-1) - \boldsymbol{\lambda}_*'\boldsymbol{m}_{t-1} - \boldsymbol{\eta}_*'\boldsymbol{x}_{t-1}. \tag{15}$$

Assuming that $\boldsymbol{\lambda}_*$, $\boldsymbol{\mu}_*^+$, and $\boldsymbol{\eta}_*$ are known, we can rewrite (2) as (10), so that the unknown parameters can be estimated by OLS as in (11) and Theorem 3 applies. If $\boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+$, $\boldsymbol{\lambda}_*$ and $\boldsymbol{\eta}_*$ are estimated by the FM-OLS or FM-TOLS estimators, we can use the resulting parameter estimate as if it is known. Note that this result is parallel to that in Section 3.1.2.

4 Hypothesis Testing

The NARDL model differs from the linear ARDL model advanced by Pesaran and Shin (1998) and Pesaran et al. (2001) in its use of partial sum decompositions to accommodate asymmetries. Consequently, it is important to test whether any asymmetries in the short-run or the long-run are statistically significant. In this section, we develop a testing methodology based on Wald's (1943) testing principle. As before, we separate our discussion into two cases by letting $k = 1$ and $k > 1$.

4.1 Hypothesis Testing with $k = 1$

4.1.1 Testing for Symmetry of the Long-Run Parameters

Consider the following hypotheses for the long-run parameters: $H_0' : (\beta_*^+ - \beta_*^-) = r$ vs. $H_1' : (\beta_*^+ - \beta_*^-) \neq r$ for some $r \in \mathbb{R}$. By setting $r = 0$, we can test whether $\beta_*^+ = \beta_*^-$. In models with multiple independent variables, we can also test the partial equality of β_*^+ and β_*^- by selecting r appropriately.

Recall that $\lambda_* := \beta_*^+ - \beta_*^-$ in (7). Consequently, we can restate H_0' as follows: $H_0'' : \lambda_* = r$ vs. $H_1'' : \lambda_* \neq r$, from which the long-run symmetry restriction, $\beta_*^+ = \beta_*^-$, is equivalent to the restriction that $\lambda_* = 0$. It is straightforward to test this restriction if λ_* is estimated by FM-OLS, because the FM-OLS estimator is asymptotically mixed-normally distributed, so the Wald test statistic follows an asymptotic chi-squared distribution under the null. This is an important advantage of FM-OLS over OLS, which yields a non-standard limit distribution for the long-run parameter.

Corollary 2 provides the limit distribution of $\widetilde{\lambda}_T$. If we let $\mathbf{S}_\ell := [0, 1, 0]$, then $T^{3/2}(\widetilde{\lambda}_T - \lambda_*) = \mathbf{S}_\ell\widetilde{\mathbf{D}}_T(\widetilde{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) \Rightarrow \mathbf{S}_\ell\mathcal{Q}^{-1}\widetilde{\mathcal{U}}$, implying that $T^{3/2}(\widetilde{\lambda}_T - \lambda_*) \Rightarrow \mathbf{S}_\ell\mathcal{Q}^{-1}\widetilde{\mathcal{U}}$, so that $T^{3/2}(\widetilde{\lambda}_T - r) \Rightarrow \mathbf{S}_\ell\mathcal{Q}^{-1}\widetilde{\mathcal{U}}$ under H_0''. The Wald test

statistic is constructed in the usual manner:

$$\mathcal{W}_T^{(\ell)} := \frac{T^3(\widetilde{\lambda}_T - r)^2}{\left(\widetilde{\tau}_T^2 \mathbf{S}_\ell \widehat{\mathbf{Q}}_T^{-1} \mathbf{S}_\ell'\right)}.$$

Note that the Wald statistic above may be inappropriate to test other forms of hypothesis. For example, consider the following hypotheses: $H_0''' : \mathbf{R}\boldsymbol{\beta}_* = \mathbf{r}$ vs. $H_1''' : \mathbf{R}\boldsymbol{\beta}_* \neq \mathbf{r}$ for some $\mathbf{R} \in \mathbb{R}^{r \times 2}$ and $\mathbf{r} \in \mathbb{R}^r$ $(r \in \{1, 2\})$, where $\boldsymbol{\beta}_* := (\beta_*^+, \beta_*^-)'$. Define:

$$\widetilde{\mathbf{R}}_\ell := \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

such that these hypotheses can be rewritten as follows: $H_0''' : \widetilde{\mathbf{R}}\boldsymbol{\varrho}_* = \mathbf{r}$ vs. $H_1''' : \widetilde{\mathbf{R}}\boldsymbol{\varrho}_* \neq \mathbf{r}$, where $\widetilde{\mathbf{R}}\boldsymbol{\varrho}_* = \boldsymbol{\beta}_*$ and $\widetilde{\mathbf{R}} := \mathbf{R}\widetilde{\mathbf{R}}_\ell$. Define the following Wald test statistic: $\widetilde{\mathcal{W}}_T^{(\ell)} := (\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r})'(\widetilde{\tau}_T^2 \widetilde{\mathbf{R}}\mathbf{Q}_T^{-1}\widetilde{\mathbf{R}}')^{-1}(\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r})$, where $\mathbf{Q}_T := \sum_{t=1}^T \boldsymbol{q}_t \boldsymbol{q}_t'$. The following theorem describes the limit behavior of the Wald test statistics:

Theorem 6. *Given Assumptions 1 and 2, $\mathcal{W}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_1^2$ under H_0'' and $\widetilde{\mathcal{W}}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_2^2$ under H_0'''; and for any sequence, c_T and $\widetilde{c}_T$, such that $c_T = o(T^3)$ and $\widetilde{c}_T = o(T^2)$, $\mathbb{P}(\mathcal{W}_T^{(\ell)} > c_T) \to 1$ under H_1'' and $\mathbb{P}(\widetilde{\mathcal{W}}_T^{(\ell)} > \widetilde{c}_T) \to 1$ under H_1'''.* $\square$

The null limit distribution of $\mathcal{W}_T^{(\ell)}$ can also be generated by simulation, as in SYG.

4.1.2   Testing for Symmetry of the Short-Run Parameters

We begin by examining the test for additive symmetry of the short-run dynamic parameters. Consider the following null and alternative hypotheses: $H_0 : \mathbf{R}_s \boldsymbol{\zeta}_* = \mathbf{r}$ vs. $H_1 : \mathbf{R}_s \boldsymbol{\zeta}_* \neq \mathbf{r}$, where $\mathbf{R}_s \in \mathbb{R}^{r \times (1+p+2q)}$, and $\mathbf{r} \in \mathbb{R}^r$ $(r \in \mathbb{N})$ are selection matrices. If we define $\mathbf{R}_s := [\mathbf{0}_{1+p}', \boldsymbol{\iota}_q', -\boldsymbol{\iota}_q']$ and $\mathbf{r} = 0$, then we can test the null hypothesis of additive short-run symmetry against the alternative hypothesis of additive short-run asymmetry: $H_0 : \sum_{j=0}^{q-1} \pi_{j*}^+ = \sum_{j=0}^{q-1} \pi_{j*}^-$ vs. $H_1 : \sum_{j=0}^{q-1} \pi_{j*}^+ \neq \sum_{j=0}^{q-1} \pi_{j*}^-$.[2]

We construct a Wald test statistic as: $\mathcal{W}_T^{(s)} := T(\mathbf{R}_s \widehat{\boldsymbol{\zeta}}_T - \mathbf{r})'(\mathbf{R}_s \widehat{\boldsymbol{\Gamma}}_T^{-1} \widehat{\boldsymbol{\Omega}}_T \widehat{\boldsymbol{\Gamma}}_T^{-1} \mathbf{R}_s')^{-1} (\mathbf{R}_s \widehat{\boldsymbol{\zeta}}_T - \mathbf{r})$, where $\widehat{\boldsymbol{\Omega}}_T$ is a consistent estimator for $\boldsymbol{\Omega}_*$: $\widehat{\boldsymbol{\Omega}}_T := T^{-1} \sum_{t=1}^T \widehat{e}_t^2 \boldsymbol{h}_t \boldsymbol{h}_t'$. We let $\widehat{e}_t := \Delta y_t - \widehat{\boldsymbol{\zeta}}_T' \boldsymbol{h}_t$, where $\widehat{\boldsymbol{\zeta}}_T$ can be constructed from the first step regression using FM-OLS. Furthermore, if the condition in Lemma 4(*iii*) holds, then the Wald test statistic simplifies to $\mathcal{W}_T^{(s)} := T(\mathbf{R}_s \widehat{\boldsymbol{\zeta}}_T - \mathbf{r})'(\widehat{\sigma}_{e,T}^2 \mathbf{R}_s \widehat{\boldsymbol{\Gamma}}_T^{-1} \mathbf{R}_s')^{-1}(\mathbf{R}_s \widehat{\boldsymbol{\zeta}}_T - \mathbf{r})$, where $\widehat{\sigma}_{e,T}^2 := T^{-1} \sum_{t=1}^T \widehat{e}_t^2$. In Theorem 7, we establish that the null and alternative limit distributions of the Wald test statistic are standard.

Theorem 7. *Given Assumption 1, if $\boldsymbol{\Gamma}_*$ and $\boldsymbol{\Omega}_*$ are positive definite, then $\mathcal{W}_T^{(s)} \overset{A}{\sim} \mathcal{X}_r^2$ under H_0 and, for any sequence, c_T, such that $c_T = o(T)$, $\mathbb{P}(\mathcal{W}_T^{(s)} > c_T) \to 1$ under H_1.* $\square$

We omit the proof of Theorem 7 as it is straightforward.

---

[2] Studies in the NARDL literature test several different forms of short-run symmetry restrictions, including the additive form, as well as pairwise symmetry between $\boldsymbol{\pi}_{j*}^+$ and $\boldsymbol{\pi}_{j*}^-$ for $j = 0, \dots, q - 1$ (e.g. SYG) and impact symmetry between $\boldsymbol{\pi}_{0*}^+$ and $\boldsymbol{\pi}_{0*}^-$ (e.g. Greenwood-Nimmo and Shin, 2013). It is straightforward to test these alternative symmetry restrictions by specifying appropriate selection matrices, $\mathbf{R}_s$ and $\mathbf{r}$.

4.2 Hypotheses Testing with $k > 1$

4.2.1 Testing for Symmetry of the Long-Run Parameters

Suppose that $k > 1$ and consider hypothesis testing on the long-run parameters. We define $\boldsymbol{\beta}_* := (\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$ as a natural extension of the corresponding definition in Section 4.1.1. Consider the following hypotheses: $H_0^{(4)} : \mathbf{R}\boldsymbol{\beta}_* = \boldsymbol{r}$ vs. $H_1^{(4)} : \mathbf{R}\boldsymbol{\beta}_* \neq \boldsymbol{r}$ for some $\mathbf{R} \in \mathbb{R}^{r \times 2k}$ and $\boldsymbol{r} \in \mathbb{R}^r$ ($r \leq 2k$). Define $\ddot{\mathcal{W}}_T := T^2 (\mathbf{R}\bar{\boldsymbol{\beta}}_T - \boldsymbol{r})' \left\{ \bar{\tau}_T^2 \mathbf{R}\ddot{\mathbf{S}}\ddot{\mathbf{R}}_T^{-1}\ddot{\mathbf{S}}'\mathbf{R}' \right\}^{-1} (\mathbf{R}\bar{\boldsymbol{\beta}}_T - \boldsymbol{r})$, where $\bar{\boldsymbol{\beta}}_T := (\bar{\boldsymbol{\beta}}_T^{+\prime}, \bar{\boldsymbol{\beta}}_T^{-\prime})'$. The following theorem describes the limit behavior of the Wald test statistic.

Theorem 8. *Given Assumptions 1 and 3, $\ddot{\mathcal{W}}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_r^2$ under $H_0^{(4)}$. Further, for any sequence c_T such that $c_T = o(T^2)$, $\mathbb{P}(\ddot{\mathcal{W}}_T^{(\ell)} > c_T) \to 1$ under $H_1^{(4)}$.* $\qquad\square$

4.2.2 Testing for Symmetry of the Short-Run Parameters

To examine the test for additive symmetry of the short-run dynamic parameters, consider the following null and alternative hypotheses: $H_0 : \mathbf{R}_s\boldsymbol{\zeta}_* = \mathbf{r}$ vs. $H_1 : \mathbf{R}_s\boldsymbol{\zeta}_* \neq \mathbf{r}$, where $\mathbf{R}_s \in \mathbb{R}^{r \times (1+p+2qk)}$ and $\mathbf{r} \in \mathbb{R}^r$ are selection matrices, and $\boldsymbol{\zeta}_* := (\rho_*, \gamma_*, \varphi_{1*}, \ldots, \varphi_{p-1*}, \boldsymbol{\pi}_{0*}^{+\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{-\prime})$, which generalizes our prior definition of $\boldsymbol{\zeta}_*$ for $k = 1$ in Section 4.1.2. If we let $\mathbf{R}_s := \left[\mathbf{0}_{k \times (1+p)}, \boldsymbol{\iota}_q' \otimes \mathbf{I}_k, -\boldsymbol{\iota}_q \otimes \mathbf{I}_k \right]$ and $\mathbf{r} = \mathbf{0}$, then we can test the null hypothesis of additive short-run symmetry as for $k = 1$: $H_0 : \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}^+ = \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}^-$ vs. $H_1 : \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}^+ \neq \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}^-$. If the cointegration residuals are obtained as in Section 3.2.2, then we can employ the same Wald test statistic introduced in Section 4.1.2 in the case where $k > 1$, because the convergence rates of both the TOLS and FM-TOLS estimators are T.

5 Monte Carlo Simulations

In this section, we examine the estimation and inferential properties of the estimators and test statistics defined in Sections 3 and 4 by simulation. We treat the cases of $k = 1$ and $k > 1$ separately. First, we study the finite sample bias and mean squared error (MSE) of the parameters estimated by our two-step procedure, where the estimator for the long-run parameters is either OLS or FM-OLS (TOLS or FM-TOLS) and the estimator for the short-run parameters is OLS. We then examine the properties of the Wald test statistics.

5.1 Monte Carlo Simulations with $k = 1$

We generate simulated data using the following NARDL(1,0) data generating process (DGP):

$$\Delta y_t = \gamma_* + \rho_* u_{t-1} + \varphi_* \Delta y_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t, \tag{16}$$

where $u_{t-1} := y_{t-1} - \alpha_* - \beta_*^+ x_{t-1}^+ - \beta_*^- x_{t-1}^-$, $\Delta x_t := \kappa_* \Delta x_{t-1} + \sqrt{1 - \kappa_*^2} v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. We set $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 2, 1, 0, -2/3, \varphi_*, 1, 1/2, 1/2)$ and we allow the sample size, T, and the

parameter φ_* to vary. Note that Δx_t is generated by an AR(1) process with normally distributed disturbances and that u_t is both serially correlated and contemporaneously correlated with Δx_t.

5.1.1 Finite Sample Performance of the Two-step Estimators

Next, we specify the following long-run and short-run models:

$$y_t = \alpha + \lambda x_t^+ + \eta x_t + u_t \quad \text{and} \quad \Delta y_t = \gamma + \rho \widehat{u}_{t-1} + \varphi_1 \Delta y_{t-1} + \pi_0^+ \Delta x_t^+ + \pi_0^- \Delta x_t^- + e_t,$$

where $\widehat{u}_t := y_t - \widehat{\alpha}_T - \widehat{\lambda}_T x_t^+ - \widehat{\eta}_T x_t$. We estimate these models in two steps. In the first step, we estimate the parameters of the long-run relationship using either OLS or FM-OLS. In the second step, we estimate the short-run parameters by OLS. In each case, we evaluate the performance of the estimators by comparing their finite sample bias and MSE. We calculate the bias as follows:

$$\text{Bias}_T(\beta_*^+) := R^{-1} \sum_{j=1}^{R} (\widehat{\beta}_{T,j}^+ - \beta_*^+) \quad \text{and} \quad \text{Bias}_T(\varphi_*) := R^{-1} \sum_{j=1}^{R} (\widehat{\varphi}_{T,j} - \varphi_*),$$

where R is the number of replications used in the simulation experiment, $\widehat{\beta}_T^+$ is obtained in the first step by OLS or FM-OLS and $\widehat{\varphi}_T$ is obtained in the second-step by OLS. Likewise, we calculate the finite sample MSE of $\widehat{\beta}_T^+$ and $\widehat{\varphi}_T$ as:

$$\text{MSE}_T(\beta_*^+) := R^{-1} \sum_{j=1}^{R} (\widehat{\beta}_{T,j}^+ - \beta_*^+)^2 \quad \text{and} \quad \text{MSE}_T(\varphi_*) := R^{-1} \sum_{j=1}^{R} (\widehat{\varphi}_{T,j} - \varphi_*)^2.$$

The finite sample bias and MSE of the estimated parameters based on $R = 5,000$ replications of the simulation experiments are reported in Tables 1 and 2, respectively. To conserve space, we do not report the finite sample bias or MSE for the intercepts, α and γ, but these results are available from the authors on request.

— Insert Tables 1 and 2 Here —

First, consider the long-run parameter estimators obtained in the first step. The finite sample bias of the FM-OLS estimator is substantially smaller than that of the first step OLS estimator. Recall that FM-OLS yields normally distributed estimators for the long-run parameters, β_*^+ and β_*^-. Consequently, in most cases, we find that the finite sample bias of the FM-OLS estimator is close to zero, because $T(\widetilde{\beta}_T^+ - \beta_*^+)$ and $T(\widetilde{\beta}_T^- - \beta_*^-)$ are asymptotically mixed-normally distributed around zero. By contrast, the OLS estimator is not asymptotically distributed around zero and exhibits non-negligible bias. In addition, our simulation results indicate that the FM-OLS estimator is often more efficient than its OLS counterpart, resulting in a smaller MSE as the sample size increases. This tendency is particularly apparent for small and/or negative values of φ_*, although it is likely to be different for other nuisance parameters. Taken as a whole, these results strongly favor the use of FM-OLS in the first step.

Now, consider the short-run parameter estimators obtained by OLS in the second step. We find that the finite sample biases of the second step OLS estimators of the dynamic parameters become negligible as the sample size increases. This is true irrespective of whether we use OLS or FM-OLS in the first step, although the smallest biases are obtained in almost all cases when the first step estimator is FM-OLS. The MSEs of the second step OLS estimators are similar irrespective of the use of OLS or FM-OLS in the first step. Even for a small sample of just 50 observations, the bias is minor in all cases. This is an encouraging observation, because many existing applications of the NARDL model rely on small datasets, constrained by the low sampling frequency and limited history of many macroeconomic databases.

5.1.2 Finite Sample Performance of the Wald Statistics

Testing Restrictions on the Long-Run Parameters: Here, we confine our attention to the case where the FM-OLS estimator is used in the first step. We generate data using (16) and set $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 1, 1, 0, -2/3, \varphi_*, 1/3, 1/2, 1/2)$, as in Section 5.1.1. We test the following hypotheses: $H_0^{(\ell)} : \beta_*^+ - \beta_*^- = 0$ vs. $H_1^{(\ell)} : \beta_*^+ - \beta_*^- \neq 0$. The simulation results reported in Table 3 reveal some mis-sizing in small samples, particularly for negative values of φ_*. Nonetheless, as the sample size increases, the distribution of the Wald test statistic becomes increasingly well-approximated by the chi-squared distribution with one degree of freedom. However, in practical applications where the sample size is smaller than 500, the use of resampling techniques to obtain an empirical p-value may be advisable.

— Insert Table 3 Here —

To examine the power of the Wald test, we generate data from (16) with $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 1.01, 1, 0, -2/3, \varphi_*, 1/3, 1/2, 1/2)$ and we allow φ_* to vary over -0.50, -0.25, 0, 0.25 and 0.50, as before. The simulation results for $\mathcal{W}_T^{(\ell)}$ are reported in Table 4. We find that the Wald test statistic is consistent under the alternative hypothesis. Irrespective of the value of φ_*, the empirical rejection rates of the Wald test statistic converge to 100%. Furthermore, the power patterns of the Wald test statistic are largely insensitive to the value of φ_*.

— Insert Table 4 Here —

Testing Restrictions on the Short-Run Parameters: To examine the empirical level properties of the Wald test statistic, we generate data using (16), with $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 2, 1, 0, -2/3, \varphi_*, 1/2, 1/2, 1/2)$ and allow φ_* to vary over -0.50, -0.25, 0, 0.25 and 0.50, as in Section 5.1.1. We first estimate the long-run parameters by FM-OLS and compute $\widehat{u}_t$ before we estimate the short-run parameters by OLS. We then test the following hypotheses: $H_0^{(s)} : \pi_*^+ - \pi_*^- = 0$ vs. $H_1^{(s)} : \pi_*^+ - \pi_*^- \neq 0$, using $\mathcal{W}_T^{(s)}$ with the heteroskedasticity consistent covariance estimator $\widehat{\boldsymbol{\Omega}}_T$. The value of the Wald test statistic is then compared against the critical values of the chi-squared distribution with one degree of freedom at the 1%, 5% and 10% levels of significance.

The simulation results reported in Table 5 reveal that the finite sample distribution of the Wald test statistic is well-approximated by the chi-squared distribution. For each level of significance, the empirical level of the test statistic

is approximately correct, particularly once the number of observations reaches 500. Interestingly, the empirical levels display little sensitivity to the value of φ_*, even for moderate T.

— Insert Table 5 Here —

Next, we examine the empirical power properties of the Wald test statistic. For this exercise, we maintain the same hypotheses but update the parameters, setting $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 2, 1, 0, -2/3, \varphi_*, 1, 1/2, 1/2)$. As before, we allow φ_* to vary over -0.50, -0.25, 0, 0.25 and 0.50 and we compute $\mathcal{W}_T^{(s)}$ using the heteroskedasticity consistent covariance matrix estimator. The simulation results are reported in Table 6. Two points are noteworthy. First, the empirical power of the Wald test statistic increases with T, indicating that the test statistic is consistent. Second, the power of the Wald test statistic exhibits little sensitivity to the degree of autocorrelation, captured by the value of φ_*.

— Insert Table 6 Here —

5.2 Monte Carlo Simulations when $k = 2$

This section examines the finite sample performance of the Wald test statistics derived in Section 4.2 by letting $k = 2$. We first generate simulated data using the following NARDL(1,0) DGP:

$$\Delta y_t = \gamma_* + \rho_* u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t, \tag{17}$$

where $u_{t-1} := y_{t-1} - \alpha_* - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_{t-1}^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_{t-1}^-$, $\Delta \boldsymbol{x}_t := \kappa_* \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - \kappa_*^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t)' \sim \text{IID}N(\boldsymbol{0}_3, \mathbf{I}_3)$. We set $(\alpha_*, \gamma_*, \rho_*, \varphi_*, \kappa_*) = (0, 0, -1, \varphi_*, 0.5)$, $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})' = (-1, 0.5, 0.75, -1.5)'$, and $(\boldsymbol{\pi}_{0*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime})' = (0.5, -0.5, -1, 1)'$. As before, we allow T and φ_* to vary and examine the effect of differing degrees of serial correlation.

5.2.1 Finite Sample Performance of the Two-Step Transformed Estimators

We specify the long-run and short-run models as: $y_t = \alpha + \boldsymbol{\lambda}' \boldsymbol{x}_t^+ + \boldsymbol{\eta} \boldsymbol{x}_t + u_t$ and $\Delta y_t = \gamma + \rho \widehat{u}_{t-1} + \varphi_1 \Delta y_{t-1} + \boldsymbol{\pi}_0^+ \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_0^- \Delta \boldsymbol{x}_t^- + e_t$, where $\widehat{u}_t$ is the regression residual obtained from the first step estimation. In the first step, we estimate the long-run parameter by TOLS or FM-TOLS, and then subsequently estimate the short-run parameter by OLS. For $k = 2$, we cannot estimate the unknown parameters using the OLS or FM-OLS estimators in the first step. As in the case with $k = 1$, we evaluate the performance of the estimators by studying their respective finite sample bias and MSE, which are obtained using $R = 5,000$ replications. The results are reported in Tables 7 and 8.

— Insert Tables 7 and 8 Here —

We first examine the long-run parameter estimators obtained in the first step. As T increases, the finite sample bias of the FM-TOLS estimator becomes smaller than that of the TOLS estimator. Recall that FM-OLS yields normally distributed estimators for the long-run parameters, $\boldsymbol{\beta}_*^+$ and $\boldsymbol{\beta}_*^-$. Consequently, we find that the finite sample bias of the

20

FM-TOLS estimator approaches zero much faster than the TOLS estimator as T increases. In addition, the FM-TOLS estimator becomes more efficient than the TOLS estimator as T increases, resulting in a smaller MSE at larger sample sizes. These results favor the use of the FM-TOLS estimator in samples of small to moderate size.

We next consider the short-run dynamic parameter estimators obtained by OLS. We find that the finite sample biases of the OLS estimator become negligible as the sample size rises. This is true irrespective of the use of either TOLS or FM-TOLS in the preceding step. Likewise, the MSEs of the OLS estimators of the dynamic parameters are similar irrespective of whether we use TOLS or FM-TOLS in the first step, particularly as the sample size becomes larger.

5.2.2 Finite Sample Performance of the Wald Statistics

Testing Restrictions on the Long-Run Parameters: Here, we confine our attention to the case where the FM-TOLS estimator is used in the first step. We generate data using (17) and set $(\alpha_*, \gamma_*, \rho_*, \varphi_*, \kappa_*) = (0, 0, -1, \varphi_*, 0.5)$, $(\beta_*^{+\prime}, \beta_*^{-\prime})' = (-1, 0.5, 0.75, -1.5)'$, and $(\pi_{0*}^{+\prime}, \pi_{0*}^{-\prime})' = (0.5, -0.5, -1, 1)'$. We test the following hypotheses: $\dot{H}_0^{(\ell)} : \iota_2' \beta_*^+ = -0.50$ and $\iota_2' \beta_*^- = -0.75$ versus $\dot{H}_1^{(\ell)} : \iota_2' \beta_*^+ \neq -0.50$ or $\iota_2' \beta_*^- \neq -0.75$ using the Wald test by allowing φ_* to vary over $-0.3, -0.1, 0, 0.1$ and 0.3. The simulation results are reported in Table 9. As the sample size increases, we find that the distribution of the Wald test statistic becomes well approximated by the chi-squared distribution with two degrees of freedom. In the case of $\varphi_* = 0.3$, however, a larger sample size is required to achieved a satisfactory approximation using the chi-square distribution. In this setting, the use of resampling methods may be advisable.

— Insert Table 9 Here —

To examine the empirical power of the Wald test statistic, we generate the same data from (17) and allow φ_* to vary as before. We test the following hypotheses: $\dot{H}_0^{(\ell)} : \iota_2' \beta_*^+ = -0.40$ and $\iota_2' \beta_*^- = -0.65$ versus $\dot{H}_1^{(\ell)} : \iota_2' \beta_*^+ \neq -0.40$ or $\iota_2' \beta_*^- \neq -0.65$. The results reported in Table 10 reveal that the Wald test statistic is consistent under the alternative hypothesis. Irrespective of the value of φ_*, the empirical rejection rates of the Wald test statistic converge to 100%.

— Insert Table 10 Here —

Testing Restrictions on the Short-Run Parameters: To examine the empirical level properties of the Wald test statistic in this case, we generate data using (17), with $(\alpha_*, \gamma_*, \rho_*, \varphi_*, \kappa_*) = (0, 0, -1, \varphi_*, 0.5)$, $(\beta_*^{+\prime}, \beta_*^{-\prime})' = (-1, 0.5, 0.75, -1.5)'$, and $(\pi_{0*}^{+\prime}, \pi_{0*}^{-\prime})' = (0.5, 0.2, 0.5, 0.2)'$, while allowing φ_* to vary as before. For this exercise, we limit our attention to the case in which the long-run parameters are estimated using the FM-TOLS estimator and we compute $\widehat{u}_t$ prior to estimation of the short-run parameters by OLS. We then test the following hypotheses: $\dot{H}_0^{(s)} : \pi_{0*}^+ - \pi_{0*}^- = \mathbf{0}$ vs. $\dot{H}_1^{(s)} : \pi_{0*}^+ - \pi_{0*}^- \neq \mathbf{0}$, using the Wald test with the heteroskedasticity consistent covariance estimator $\widehat{\Omega}_T$. The resulting value of the Wald test statistic is then compared against the critical values of the chi-squared distribution with two degrees of freedom at the 1%, 5% and 10% levels of significance.

The simulation results reported in Table 11 reveal that the finite sample distribution of the Wald test statistic is well-approximated by the chi-squared distribution. For each level of significance, the empirical level of the test statistic

is approximately correct, particularly once the number of observations reaches 1,000. Interestingly, the empirical levels display little sensitivity to the value of φ_*, even for moderate T.

— Insert Table 11 Here —

Next, we examine the empirical power of the Wald test statistic. For this exercise, we work with the same DGP and test the following hypotheses: $\ddot{H}_0^{(s)} : \boldsymbol{\pi}_{0*}^+ - \boldsymbol{\pi}_{0*}^- = 0.3\boldsymbol{\iota}$ vs. $\ddot{H}_1^{(s)} : \boldsymbol{\pi}_{0*}^+ - \boldsymbol{\pi}_{0*}^- \neq 0.3\boldsymbol{\iota}$. We use the same Wald test statistic and report the simulation results in Table 12. The empirical power of the Wald test statistic increases with T to reach 100% and the Wald test statistic once again exhibits little sensitivity to the degree of autocorrelation measured by φ_*.

— Insert Table 12 Here —

6   Empirical Application: Post-war Dividend Smoothing in the US

To illustrate the use of our two-step estimation procedure, we analyze the relationship between real dividends and real earnings in the US. Among firms that pay dividends, the common practice is to adjust the dividend gradually in response to earnings news. The seminal study of dividend smoothing behavior was conducted by Lintner (1956), based on interviews with managers from twenty-eight companies. A key finding from these interviews is that managers are reluctant to announce dividend changes that they may subsequently be obliged to reverse. Consequently, Lintner contends that firms only adjust their dividends in response to non-transitory earnings changes, with the goal of achieving a desired long-run target payout ratio. A substantial body of empirical work supports this view (e.g. Fama and Babiak, 1968; Marsh and Merton, 1987; Garrett and Priestley, 2000; Andres, Betzer, Goergen, and Renneboog, 2009).

A more recent study by Brav et al. (2005) focusing on the factors that determine dividend and share repurchase decisions largely corroborates Lintner's findings. Specifically, Brav et al. find that 93.8% of managers agree that executives strive to avoid reducing dividends, while 89.6% agree that executives smooth the dividend stream. 77.9% agree that executives are reluctant to announce dividend changes that will subsequently be reversed, because 88.1% of managers perceive that there are negative consequences to cutting dividends. Indeed, such is the reluctance to cut dividends that Brav et al. (2005) find that managers would first consider liquidating assets, reducing the workforce or even deferring profitable investments. The importance that managers attach to dividends supports DeAngelo and DeAngelo's (2006) view that dividends matter to investors, contrary to the classic irrelevance theorem of Miller and Modigliani (1961).

To capture the gradualism with which firms approach their target dividend, Lintner (1956) proposes the following partial adjustment model: $\Delta D_t = a_* - \zeta_* \left(D_t^* - D_{t-1} \right) + \epsilon_t$, where D_t and D_t^* denote the current level and the target level of dividends at time t, respectively, and where $|\zeta_*|$ measures the speed with which the dividend is adjusted toward the target. As noted by Cho, Kim, and Shin (2015), it is widely believed that an equilibrium relation exists between the dividend target and current earnings, E_t. Writing this equilibrium relation as $D_t^* = \beta_* E_t$, where β_* captures the target

payout ratio, we rewrite Lintner's partial adjustment model in the following form:

$$\Delta D_t = a_* + \zeta_* D_{t-1} + \theta_* E_t + \epsilon_t, \tag{18}$$

where $\theta_* = -\zeta_* \beta_*$. As a linear partial adjustment process, (18) implies that the dividend is adjusted symmetrically with respect to both positive and negative earnings news. This is incompatible with the behavior documented in the surveys of Lintner (1956) and Brav et al. (2005). In particular, it is difficult to reconcile with the survey respondents' insistence that managers tend to smooth the dividend stream and avoid cutting dividends where possible. To allow for differential adjustment with respect to positive and negative earnings news, we first define the following partial sum decomposition of real earnings: $E_t = E_0 + E_t^+ + E_t^-$, where the initial value, E_0, can be set to zero without loss of generality, $E_t^+ = \sum_{j=1}^{t} \left(\Delta E_j \mathbb{1}_{\{\Delta E_j \geq 0\}} \right)$, $E_t^- = \sum_{j=1}^{t} \left(\Delta E_j \mathbb{1}_{\{\Delta E_j < 0\}} \right)$ and $\mathbb{1}_{\{\cdot\}}$ is a Heaviside function taking the value 1 if the condition in braces is satisfied and zero otherwise. Now, we propose the following asymmetric generalization of the equilibrium relation between the target dividend and real earnings: $D_t^* = \beta_*^+ E_t^+ + \beta_*^- E_t^-$, where β_*^+ and β_*^- capture the target payout ratios with respect to positive and negative earnings news, respectively. We now rewrite (18) as:

$$\Delta D_t = a_* + \zeta_* D_{t-1} + \theta_*^+ E_t^+ + \theta_*^- E_t^- + \epsilon_t, \tag{19}$$

where $\theta_*^+ = -\zeta_* \beta_*^+$ and $\theta_*^- = -\zeta_* \beta_*^-$. Unit root testing reveals that D_t and E_t are first difference stationary time series, implying that there is an asymmetric cointegrating relation between these variables provided that their linear combination is stationary.[3] To account for serial correlation in ϵ_t, (19) may be embedded within a NARDL(p,q) model as follows:

$$\Delta D_t = \alpha_* + \zeta_* (D_{t-1} - \beta_*^+ E_{t-1}^+ - \beta_*^- E_{t-1}^-) + \sum_{j=1}^{p-1} \lambda_{j*} \Delta D_{t-j} + \sum_{j=0}^{q-1} d_{j*}^+ \Delta E_{t-j}^+ + \sum_{j=0}^{q-1} d_{j*}^- \Delta E_{t-j}^- + \varepsilon_t, \tag{20}$$

where the use of a sufficiently rich lag structure will ensure that ε_t is serially uncorrelated. Noting that this a NARDL model with a single explanatory variable, it follows that (20) can be estimated either by the single-step procedure advanced by SYG or by the two-step procedure that we propose above, without the need for detrending prior to estimation. We will take the opportunity to compare both estimation procedures.

Using data from the *Irrational Exuberance* dataset maintained by Robert Shiller, we construct a quarterly dataset of real earnings and real dividends for the S&P 500 index over the period 1946Q1–2006Q4.[4] Our sample period starts after World War II because there is evidence of a substantial change in payout policy at approximately this time. For example, Chen, Da, and Priestley (2012) find that dividends adjust to earnings news four times slower in the post-war period (1946–2006 in their analysis) compared to a pre-war sample period (1871–1945). We choose to end our sample in 2006Q4, immediately prior to the period of extreme earnings volatility associated with the global financial crisis.

In Table 13, we report descriptive statistics for both the level and first difference of real earnings and real dividends.

---

[3]Unit root testing results are available from the authors on request.
[4]Shiller's dataset is available from `http://www.econ.yale.edu/~shiller/data/ie_data.xls`.

The descriptive statistics demonstrate that real earnings are considerably more volatile than real dividends, with greater tail mass. The standard deviation of real earnings is almost four times larger than that of real dividends. Furthermore, unlike the real dividends data, which is approximately symmetrically distributed with little excess kurtosis, real earnings display a notable right skew and notable excess kurtosis. Similar patterns are also evident in the first-differenced data, although neither series displays a notable skew after differencing. These observations are collectively consistent with the notion that executives smooth the time path of dividends relative to earnings news. This tendency can be easily discerned by eye in Figure 1, which presents time series plots of the level of real earnings and real dividends.

— Insert Table 13 and Figure 1 Here —

In light of the quarterly sampling frequency of our data, we estimate a NARDL(4,4) model using both the single-step estimation routine devised by SYG and the two-step procedure that we develop above for models where $k = 1$, using FM-OLS in the first step. In Table 14, we report the long-run parameter estimates obtained in each case. To facilitate comparisons between the two estimation strategies, we transform the estimated parameters to obtain estimated values of β^+ and β^-, the corresponding standard errors of which are computed via the delta method. The point estimates obtained from the two different estimation frameworks are remarkably similar in both cases, although the two-step estimation procedure yields more precise estimates, with standard errors approximately half as large as those obtained from the single-step procedure. We conjecture that the relative imprecision of the long-run parameters obtained from the single-step estimator may arise due to the way in which the long-run parameters are constructed as ratios. For example, the standard error of the long-run parameter estimator may be inflated if the numerator and demoninator share a negative covariance. In addition, the precision of the single-step estimates of the long-run parameters may deteriorate for values of the error correction coefficient close to zero. No such issue arises in the case of two-step estimation.

The difference in precision of the long-run parameter estimates has an important practical implication. Based on the results of the single-step procedure, we are unable to reject the null hypothesis of long-run symmetry: the Wald test of $H_0 : \beta^+ = \beta^-$ versus $H_1 : \beta^+ \neq \beta^-$ returns a p-value of 0.412. By contrast, the increased precision of the two-step estimation procedure allows us to reject the null hypothesis of long-run symmetry at the 5% level (the Wald test of $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$ returns a p-value of 0.0126). A comparison of the magnitude of the long-run parameters associated with positive and negative earnings reveals that dividends respond slightly more strongly to earnings increases than to earnings decreases in long-run equilibrium. This phenomenon offers a simple explanation for the growing gap between real earnings and real dividends in Figure 1 and is consistent with the evidence that executives are loathe to cut dividends for fear of sending adverse signals regarding corporate performance.

— Insert Table 14 Here —

Following SYG, support for the existence of an asymmetric cointegrating relationship between real dividends and real earnings can be obtained using either the ECM-based t_{BDM}-test of Banerjee, Dolado, and Mestre (1998) or F_{PSS}-test

proposed by Pesaran et al. (2001) in the case of single-step estimation. However, the lagged levels terms D_{t-1}, E_{t-1}^{+} and E_{t-1}^{-} are not included in the second-step of our two-step estimation framework, so only the t_{BDM}-test is applicable in this case. Based on the single-step estimation results, we obtain a t_{BDM}-test statistic of -2.935; in the two-step case, we obtain a value of -3.086. Both exceed the relevant 10% critical value of -2.91 tabulated by Pesaran et al. (2001), indicating a rejection of the null hypothesis of no asymmetric cointegration at the 10% level.[5]

Given the similarity of the point estimates of the long-run parameters obtained from the single-step and two-step estimation frameworks, we expect that the long-run disequilibrium errors obtained from each method should track one-another closely. Figure 2 reveals that this is the case, with both displaying almost identical dynamics.

— Insert Figure 2 Here —

The similarity of the long-run disequilibrium errors, in turn, suggests that the speed of error correction implied by each model should also be very similar. Table 15 reveals this to be the case. The single-step parameter estimates imply that disequilibrium errors are corrected at a rate of 3.1% per quarter, while the corresponding value based on the two-step approach is 3.2%. Likewise, given the similarities documented to this point, we expect the dynamic parameter estimates to be very similar across both estimation methods. In practice, the degree of similarity revealed by Table 15 is striking.

— Insert Table 15 Here —

Neither the single-step nor the two-step estimation results provide support for the hypothesis of short-run asymmetry at any horizon. For example, the Wald test of the null hypothesis of impact symmetry, $H_0 : \delta_{0*}^{+} = \delta_{0*}^{-}$, versus the two-sided alternative $H_1 : \delta_{0*}^{+} \neq \delta_{0*}^{-}$ returns a p-value of 0.127 in the single-step case and 0.145 in the two-step case. Likewise, the null hypothesis of additive short-run symmetry, $H_0 : \sum_{j=0}^{q-1} \delta_{j*}^{+} = \sum_{j=0}^{q-1} \delta_{j*}^{-}$, is not rejected against the alternative, $H_1 : \sum_{j=0}^{q-1} \delta_{j*}^{+} \neq \sum_{j=0}^{q-1} \delta_{j*}^{-}$, in both cases, with p-values of 0.251 (single-step) and 0.236 (two-step).

Overall, our empirical results suggest that executives pass earnings increases through to dividends slightly more strongly than earnings decreases in long-run equilibrium. The magnitude of this asymmetry is relatively small but nonetheless it is economically significant and it is consistent with existing evidence of asymmetric aggregate payout policy (e.g. Brav et al., 2005). Both the single-step and two-step estimation procedures yield qualitatively and quantitatively similar results, indicating that both procedures may be used in practice, particularly in large samples, where their asymptotic equivalence should become apparent. However, when working with small samples, the two-step approach may yield greater precision in the estimation of the long-run parameters and this may improve one's ability to detect long-run asymmetry.

---

[5]This critical value is obtained from Table CII(iii) in Pesaran et al. (2001). Following the conservative rule-of-thumb advocated by SYG, we select the critical value for a model with a single explanatory variable (i.e. we count the number of explanatory variables prior to their decomposition into positive and negative cumulative partial sums).

7 Concluding Remarks

In this paper, we revisit the NARDL model developed by SYG. In the existing literature, it is typically estimated in a single step by OLS. Support for the efficacy of the single-step OLS estimator based on Monte Carlo simulations has been provided by SYG. However, efforts to develop asymptotic theory for the single-step estimator have been impeded by the presence of an asymptotic singularity problem caused by the presence of asymptotically perfectly collinear time trends in the positive and negative cumulative partial sum processes that are used to introduce asymmetry in the NARDL model.

We develop a two-step estimation procedure that makes use of a one-to-one transformation of the asymmetric long-run relationship in the NARDL model to overcome this asymptotic singularity issue. In models with a single explanatory variable, the first step involves estimating the parameters of the transformed asymmetric long-run relationship using any consistent estimator with a convergence rate faster than the square root of the sample size, $T^{1/2}$. In practice, we advocate the use of OLS or the FM-OLS estimator of Phillips and Hansen (1990) in the first step, the latter of which accounts for serial correlation and potential endogeneity of the explanatory variables, while facilitating standard inference by virtue of its asymptotic mixed normality. In models with multiple explanatory variables, it is necessary to transform the partial sum processes of the explanatory variables to remove their trends prior to estimating the parameters of the long-run relationship. In the second step, the dynamic coefficients can be estimated consistently by OLS treating the error correction term obtained from the first step as given, in light of the super-consistency of the first step estimator. Unlike the single-step estimation procedure, our two-step procedure is analytically tractable. Consequently, we are able to derive the asymptotic properties of the estimators and to characterize their limit distributions. We also develop Wald tests that can be used to evaluate restrictions on the short- and long-run parameters. In both cases, we demonstrate that the null distribution of the Wald statistic weakly converges to a chi-squared distribution. A suite of Monte Carlo simulations indicate that our asymptotic results offer satisfactory approximations even in finite samples.

We illustrate our methodology with an application to dividend-smoothing in the postwar period in the US. Our results are consistent with a large body of research that finds that managers smooth the time path of dividends relative to earnings. We document evidence of asymmetry in long-run equilibrium, where we find that managers allow real dividends to respond slightly more strongly to positive earnings news than to negative earnings news. By contrast, we find no evidence of asymmetry in the short-run dynamic parameters.

References

ANDRES, C., A. BETZER, M. GOERGEN, AND L. RENNEBOOG (2009): "Dividend Policy of German Firms: A Panel Data Analysis of Partial Adjustment Models," *Journal of Empirical Finance*, 16, 175–187.

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.

BARATI, M. AND H. FARIDITAVANA (2018): "Asymmetric Effect of Income on the US Healthcare Expenditure: Evidence from the Nonlinear Autoregressive Distributed Lag (ARDL) Approach," *Empirical Economics*, in press.

BOX, M., K. GRATZER, AND X. LIN (2018): "The Asymmetric Effect of Bankruptcy Fraud in Sweden: A Long-Term Perspective," *Journal of Quantitative Criminology*, in press.

BRAV, A., J. GRAHAM, C. HARVEY, AND R. MICHAELY (2005): "Payout Policy in the 21st Century," *Journal of Financial Economics*, 77, 483–527.

BRUN-AGUERRE, R. X., A.-M. FUERTES, AND M. J. GREENWOOD-NIMMO (2017): "Heads I Win; Tails You Lose: Asymmetry in Exchange Rate Pass-Through into Import Prices," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180, 587–612.

CHEN, L., Z. DA, AND R. PRIESTLEY (2012): "Dividend Smoothing and Predictability," *Management Science*, 58, 1834–1853.

CHO, J. S., T.-H. KIM, AND Y. SHIN (2015): "Quantile Cointegration in the Autoregressive Distributed Lag Modeling Framework," *Journal of Econometrics*, 188, 281–300.

DEANGELO, H. AND L. DEANGELO (2006): "The Irrelevance of the MM Dividend Irrelevance Theorem," *Journal of Financial Economics*, 79, 293–315.

EBERHARDT, M. AND A. F. PRESBITERO (2015): "Public Debt and Growth: Heterogeneity and Non-Linearity," *Journal of International Economics*, 97, 45–58.

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

FAMA, E. AND H. BABIAK (1968): "Dividend Policy: An Empirical Analysis," *Journal of the American Statistical Association*, 63, 1132–1161.

FERRIS, J. S., S. L. WINER, AND D. E. OLMSTEAD (2020): "A Dynamic Model of Political Party Equilibrium: The Evolution of ENP in Canada, 1870 – 2015," Working Paper 8387, CESifo, Munich.

GARRETT, I. AND R. PRIESTLEY (2000): "Dividend Behavior and Dividend Signaling," *Journal of Financial and Quantitative Analysis*, 35, 173–189.

GREENWOOD-NIMMO, M. J. AND Y. SHIN (2013): "Taxation and the Asymmetric Adjustment of Selected Retail Energy Prices in the UK," *Economics Letters*, 121, 411–416.

HAMMOUDEH, S., A. LAHIANI, D. K. NGUYEN, AND R. M. SOUSA (2015): "An Empirical Analysis of Energy Cost Pass-Through to CO2 Emission Prices," *Energy Economics*, 49, 149–156.

HE, Z. AND F. ZHOU (2018): "Time-Varying and Asymmetric Eeffects of the Oil-Specific Demand Shock on Investor Sentiment," *PLOS ONE*, 31.

LINTNER, J. (1956): "Distribution of Incomes of Corporations among Dividends, Retained Earnings, and Taxes," *American Economic Review*, 46, 97–113.

MARSH, T. AND R. MERTON (1987): "Dividend Behaviour for the Aggregate Stock Market," *Journal of Business*, 60, 1–40.

MILLER, M. AND F. MODIGLIANI (1961): "Dividend Policy, Growth, and the Valuation of Shares," *Journal of Business*, 34, 411–433.

NEWEY, W. K. AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

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. AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variable Regression with I(1) Processes," *Review of Economic Studies*, 57, 99–125.

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.

SÜSSMUTH, B. AND U. WOITEK (2013): "Estimating Dynamic Asymmetries in Demand at the Munich Oktoberfest," *Tourism Economics*, 19, 653–674.

VERHEYEN, F. (2013): "Exchange Rate Nonlinearities in EMU Exports to the US," *Economic Modelling*, 32, 66–76.

WALD, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

A Appendix

A.1 Proofs

Proof of Lemma 1. (*i*) By (4) and (5), we obtain the following:

- $T^{-3}\sum_{t=1}^{T}y_{t-1}^{2} = \frac{1}{3}\delta_{*}^{2} + o_{\mathbb{P}}(1)$;
- $T^{-3}\sum_{t=1}^{T}y_{t-1}\boldsymbol{x}_{t}^{+\prime} = \frac{1}{3}\delta_{*}\boldsymbol{\mu}_{*}^{+\prime} + o_{\mathbb{P}}(1)$;
- $T^{-3}\sum_{t=1}^{T}y_{t-1}\boldsymbol{x}_{t}^{-\prime} = \frac{1}{3}\delta_{*}\boldsymbol{\mu}_{*}^{-\prime} + o_{\mathbb{P}}(1)$;

- $T^{-3}\sum_{t=1}^{T}\boldsymbol{x}_{t}^{+}\boldsymbol{x}_{t}^{+\prime} = \frac{1}{3}\boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{+\prime} + o_{\mathbb{P}}(1)$;
- $T^{-3}\sum_{t=1}^{T}\boldsymbol{x}_{t}^{+}\boldsymbol{x}_{t}^{-\prime} = \frac{1}{3}\boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{-\prime} + o_{\mathbb{P}}(1)$; and
- $T^{-3}\sum_{t=1}^{T}\boldsymbol{x}_{t}^{-}\boldsymbol{x}_{t}^{-\prime} = \frac{1}{3}\boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{-\prime} + o_{\mathbb{P}}(1)$.

These limits imply that $T^{-3}\sum_{t=1}^{T}\boldsymbol{z}_{1t}\boldsymbol{z}_{1t}' = \mathbf{M}_{11} + o_{\mathbb{P}}(1)$.

(*ii*) By (4) and (5), we note that:

- $T^{-2}\sum_{t=1}^{T}y_{t-1} = \frac{1}{2}\delta_{*} + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}y_{t-1}\boldsymbol{w}_{1t}' = T^{-2}\sum_{t=1}^{T}[\delta_{*}^{2}t, \delta_{*}^{2}t, \ldots, \delta_{*}^{2}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\delta_{*}^{2}\boldsymbol{\iota}_{p-1}' + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}y_{t-1}\boldsymbol{w}_{2t}' = T^{-2}\sum_{t=1}^{T}[\delta_{*}\boldsymbol{\mu}_{*}^{+\prime}t, \delta_{*}\boldsymbol{\mu}_{*}^{+\prime}t, \ldots, \delta_{*}\boldsymbol{\mu}_{*}^{+\prime}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\delta_{*}\boldsymbol{\iota}_{q}' \otimes \boldsymbol{\mu}_{*}^{+\prime} + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}y_{t-1}\boldsymbol{w}_{3t}' = T^{-2}\sum_{t=1}^{T}[\delta_{*}\boldsymbol{\mu}_{*}^{-\prime}t, \delta_{*}\boldsymbol{\mu}_{*}^{-\prime}t, \ldots, \delta_{*}\boldsymbol{\mu}_{*}^{-\prime}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\delta_{*}\boldsymbol{\iota}_{q}' \otimes \boldsymbol{\mu}_{*}^{-\prime} + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}^{+} = \frac{1}{2}\boldsymbol{\mu}_{*}^{+} + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}^{+}\boldsymbol{w}_{1t}' = T^{-2}\sum_{t=1}^{T}[\delta_{*}\boldsymbol{\mu}_{*}^{+}t, \delta_{*}\boldsymbol{\mu}_{*}^{+}t, \ldots, \delta_{*}\boldsymbol{\mu}_{*}^{+}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\delta_{*}\boldsymbol{\mu}_{*}^{+}\boldsymbol{\iota}_{p-1}' + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}^{+}\boldsymbol{w}_{2t}' = T^{-2}\sum_{t=1}^{T}[\boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{+\prime}t, \boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{+\prime}t, \ldots, \boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{+\prime}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\boldsymbol{\iota}_{q}' \otimes \boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{+\prime} + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}^{+}\boldsymbol{w}_{3t}' = T^{-2}\sum_{t=1}^{T}[\boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{-\prime}t, \boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{-\prime}t, \ldots, \boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{-\prime}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\boldsymbol{\iota}_{q}' \otimes \boldsymbol{\mu}_{*}^{+}\boldsymbol{\mu}_{*}^{-\prime} + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}^{-} = -\frac{1}{2}\boldsymbol{\mu}_{*}^{+} + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}^{-}\boldsymbol{w}_{1t}' = T^{-2}\sum_{t=1}^{T}[\delta_{*}\boldsymbol{\mu}_{*}^{-}t, \delta_{*}\boldsymbol{\mu}_{*}^{-}t, \ldots, \delta_{*}\boldsymbol{\mu}_{*}^{-}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\delta_{*}\boldsymbol{\mu}_{*}^{-}\boldsymbol{\iota}_{p-1}' + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}^{-}\boldsymbol{w}_{2t}' = T^{-2}\sum_{t=1}^{T}[\boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{+\prime}t, \boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{+\prime}t, \ldots, \boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{+\prime}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\boldsymbol{\iota}_{q}' \otimes \boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{+\prime} + o_{\mathbb{P}}(1)$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}^{-}\boldsymbol{w}_{3t}' = T^{-2}\sum_{t=1}^{T}[\boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{-\prime}t, \boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{-\prime}t, \ldots, \boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{-\prime}t] + o_{\mathbb{P}}(1) = \frac{1}{2}\boldsymbol{\iota}_{q}' \otimes \boldsymbol{\mu}_{*}^{-}\boldsymbol{\mu}_{*}^{-\prime} + o_{\mathbb{P}}(1)$.

These limit results imply that $T^{-1}\sum_{t=1}^{T}\boldsymbol{z}_{1t}\boldsymbol{z}_{2t}' = \mathbf{M}_{12} + o_{\mathbb{P}}(1)$.

(*iii*) We note that:

- $T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{1t}' = \mathbb{E}[\Delta\boldsymbol{y}_{t-1}]' + o_{\mathbb{P}}(1) = \delta_{*}\boldsymbol{\iota}_{p-1}' + o_{\mathbb{P}}$;
- $T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{2t}' = [\mathbb{E}[\Delta\boldsymbol{x}_{t}^{+\prime}], \mathbb{E}[\Delta\boldsymbol{x}_{t-1}^{+\prime}], \ldots, \mathbb{E}[\Delta\boldsymbol{x}_{t-q+1}^{+\prime}]] + o_{\mathbb{P}}(1) = [\boldsymbol{\mu}_{*}^{+\prime}, \boldsymbol{\mu}_{*}^{+\prime}, \ldots, \boldsymbol{\mu}_{*}^{+\prime}] + o_{\mathbb{P}} = \boldsymbol{\iota}_{q}' \otimes \boldsymbol{\mu}_{*}^{+}$;
- $T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{3t}' = [\mathbb{E}[\Delta\boldsymbol{x}_{t}^{-\prime}], \mathbb{E}[\Delta\boldsymbol{x}_{t-1}^{-\prime}], \ldots, \mathbb{E}[\Delta\boldsymbol{x}_{t-q+1}^{-\prime}]] + o_{\mathbb{P}}(1) = [\boldsymbol{\mu}_{*}^{-\prime}, \boldsymbol{\mu}_{*}^{-\prime}, \ldots, \boldsymbol{\mu}_{*}^{-\prime}] + o_{\mathbb{P}} = \boldsymbol{\iota}_{q}' \otimes \boldsymbol{\mu}_{*}^{-}$; and
- $T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{t}\boldsymbol{w}_{t}' = \mathbb{E}[\boldsymbol{w}_{t}\boldsymbol{w}_{t}'] + o_{\mathbb{P}}(1)$.

These limits imply that $T^{-1}\sum_{t=1}^{T}\boldsymbol{z}_{2t}\boldsymbol{z}_{2t}' = \mathbf{M}_{22} + o_{\mathbb{P}}(1)$, as desired. ∎

Proof of Lemma 2.

(*i*) We note that:

- $T^{-2}\sum_{t=1}^{T}x_{t}^{+} = T^{-1}\sum_{t=1}^{T}\mu_{*}^{+}(t/T) + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} \frac{1}{2}\mu_{*}^{+}$;
- $T^{-3/2}\sum_{t=1}^{T}x_{t} = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta x_{i}) \Rightarrow \int\mathcal{B}_{x}$ using that $T^{-1/2}\sum_{i=1}^{[T(\cdot)]}\Delta x_{i} \Rightarrow \int_{0}^{(\cdot)}d\mathcal{B}_{x}$;
- $T^{-3}\sum_{t=1}^{T}x_{t}^{+}x_{t}^{+} = T^{-1}\sum_{t=1}^{T}\mu_{*}^{+}\mu_{*}^{+}(t/T)^{2} + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} \frac{1}{3}\mu_{*}^{+}\mu_{*}^{+}$;

29

- $T^{-5/2} \sum_{t=1}^{T} x_t^+ x_t = T^{-1} \sum_{t=1}^{T} \mu_*^+(t/T)(T^{-1/2} \sum_{i=1}^{t} \Delta x_t) + o_{\mathbb{P}}(1) \Rightarrow \mu_*^+ \int r\mathcal{B}_x$; and
- $T^{-2} \sum_{t=1}^{T} x_t x_t = T^{-1} \sum_{t=1}^{T} (T^{-1/2} \sum_{i=1}^{t} \Delta x_t)(T^{-1/2} \sum_{i=1}^{t} \Delta x_t) \Rightarrow \int \mathcal{B}_x^2$.

Therefore, $\widehat{\mathbf{Q}}_T \Rightarrow \mathcal{Q}$, as desired.

(*ii*) We note that:

- $T^{-1/2} \sum_{t=1}^{T} u_t \Rightarrow \int d\mathcal{B}_u$ using that $T^{-1/2} \sum_{t=1}^{[T(\cdot)]} u_t \Rightarrow \int_0^{(\cdot)} d\mathcal{B}_u$;
- $T^{-3/2} \sum_{t=1}^{T} x_t^+ u_t = T^{-1/2} \sum_{t=1}^{T} \mu_*^+(t/T) u_t + o_{\mathbb{P}}(1) \Rightarrow \mu_*^+ \int r d\mathcal{B}_u$; and
- $T^{-1} \sum_{t=1}^{T} x_t u_t = T^{-1/2} \sum_{t=1}^{T} (T^{-1/2} \sum_{i=1}^{t} \Delta x_i) u_t \Rightarrow \int \mathcal{B}_x d\mathcal{B}_u + \upsilon_*$ using the fact that $\upsilon_* := \lim_{T \to \infty}$ $T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{t} \mathbb{E}[\Delta x_i u_t]$ is finite.

Therefore, $\widehat{\mathbf{U}}_T \Rightarrow \mathcal{U}$. ∎

Proof of Corollary 1.

Given (8), the desired result follows from Lemma 4. ∎

Proof of Theorem 1.

We note that $T\{(\widehat{\beta}_T^+ - \widehat{\beta}_T^-) - (\beta_*^+ - \beta_*^-)\} = O_{\mathbb{P}}(T^{-1/2})$ by the definition of $\widehat{\lambda}_T$. Therefore, the weak limit of $T(\widehat{\beta}_T^+ - \beta_*^+)$ is equivalent to that of $T(\widehat{\beta}_T^- - \beta_*^-)$. Furthermore, Corollary 1 implies that $T(\widehat{\beta}_T^- - \beta_*^-) \Rightarrow \mathbf{S}\mathcal{Q}^{-1}\mathcal{U}$, leading to the desired result. ∎

Proof of Lemma 3.

Given Assumption 2, we note that $\widetilde{\upsilon}_T \xrightarrow{\mathbb{P}} \upsilon_*$ and $(\widetilde{\sigma}_T^{(1,1)})^{-1}\widetilde{\sigma}_T^{(1,2)} \xrightarrow{\mathbb{P}} \boldsymbol{\nu}_* := (\sigma_*^{(1,1)})^{-1}\sigma_*^{(1,2)}$. Therefore, if we let $\dot{u}_t := u_t - \Delta x_t \boldsymbol{\nu}_*$, $\widetilde{\mathbf{U}}_T = \widetilde{\mathbf{D}}_T^{-1} \sum_{t=1}^{T} \{\boldsymbol{q}_t \dot{u}_t - \mathbf{S}'\upsilon_*\} + o_{\mathbb{P}}(1)$, then $\widetilde{\mathbf{U}}_T \Rightarrow [\int d\mathcal{B}_{\dot{u}}, \mu_*^+ \int r d\mathcal{B}_{\dot{u}}, \int \mathcal{B}_x d\mathcal{B}_{\dot{u}}]'$, where $\mathcal{B}_{\dot{u}}(\cdot) := \tau_* \mathcal{W}_u(\cdot)$. Therefore, $\widetilde{\mathbf{U}}_T \Rightarrow \widetilde{\mathcal{U}}$. ∎

Proof of Corollary 2.

Given that $\widetilde{\mathbf{D}}_T(\widetilde{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) = [\widetilde{\mathbf{D}}_T^{-1}(\sum_{t=1}^{T} \boldsymbol{q}_t \boldsymbol{q}_t')\widetilde{\mathbf{D}}_T^{-1}]^{-1}\widetilde{\mathbf{U}}_T$, the desired result follows from Lemmas 2(*i*) and 3. ∎

Proof of Theorem 2.

Given that $(\widetilde{\beta}_T^+ - \beta_*^+) - (\widetilde{\beta}_T^- - \beta_*^-) = \widetilde{\lambda}_T - \lambda_* = O_{\mathbb{P}}(T^{-3/2})$ and $(\widetilde{\beta}_T^- - \beta_*^-) = O_{\mathbb{P}}(T^{-1})$, it follows that $(\widetilde{\beta}_T^+ - \beta_*^+) = O_{\mathbb{P}}(T^{-1})$, implying that the weak limit of $T(\widetilde{\beta}_T^+ - \beta_*^+)$ is equivalent to that of $T(\widetilde{\beta}_T^- - \beta_*^-)$. Furthermore, Corollary 1 implies that $T(\widetilde{\eta}_T^- - \eta_*^-) = T(\widetilde{\beta}_T^- - \beta_*^-) \Rightarrow \mathbf{S}\mathcal{Q}^{-1}\widetilde{\mathcal{U}}$, leading to the desired result. ∎

Proof of Lemma 4.

This result is easily obtained using the ergodic theorem and the multivariate central limit theorem. ∎

Proof of Theorem 3.

(*i*) Given (11), we can combine Lemmas 4 (*i* and *ii*) to obtain the desired result.

(*ii*) If it further holds that $\mathbb{E}[e_t^2|\boldsymbol{h}_t] = \sigma_*^2$, then Lemma 4(*iii*) implies that $\boldsymbol{\Omega}_* = \sigma_*^2 \boldsymbol{\Gamma}_*$. Therefore, Theorem 3(*i*) now implies that $\sqrt{T}(\widehat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}_*) \overset{A}{\sim} N(\mathbf{0}, \sigma_*^2 \boldsymbol{\Gamma}_*^{-1})$. ∎

30

Proof of Lemma 5.

(*i*) We note that:

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

- $T^{-3/2}\sum_{t=1}^{T}\widehat{m}_t = T^{-3/2}\sum_{t=1}^{T} m_t - (T^{-2}\sum_{t=1}^{T}t)(T^{-3}\sum_{t=1}^{T}t^2)^{-1}(T^{-5/2}\sum_{t=1}^{T}tm_t) = T^{-1}\sum_{t=1}^{T}T^{-1/2}$
  $(\sum_{i=1}^{t}\Delta m_i) - (T^{-2}\sum_{t=1}^{T}t)(T^{-3}\sum_{t=1}^{T}t^2)^{-1}(T^{-1}\sum_{t=1}^{T}(t/T)T^{-1/2}\sum_{i=1}^{t}\Delta m_i) \Rightarrow \int \mathcal{B}_m - \frac{3}{2}\int r\mathcal{B}_m$ using
  the fact that $T^{-1/2}\sum_{i=1}^{[T(\cdot)]}\Delta m_i \Rightarrow \int_0^{(\cdot)}d\mathcal{B}_m$;

- $T^{-3/2}\sum_{t=1}^{T} x_t = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta x_i) \Rightarrow \int \mathcal{B}_x$ using the fact that $T^{-1/2}\sum_{i=1}^{[T(\cdot)]}\Delta x_i \Rightarrow \int_0^{(\cdot)}d\mathcal{B}_x$;

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

- $T^{-5/2}\sum_{t=1}^{T}t\widehat{m}_t = T^{-5/2}\sum_{t=1}^{T}tm_t - T^{-5/2}(\sum_{t=1}^{T}t^2)(\sum_{t=1}^{T}t^2)^{-1}(\sum_{t=1}^{T}tm_t) = \mathbf{0}$;

- $T^{-5/2}\sum_{t=1}^{T}tx_t = T^{-1}\sum_{t=1}^{T}(t/T)(T^{-1/2}\sum_{i=1}^{t}\Delta x_i) \Rightarrow \int r\mathcal{B}_x$;

- $T^{-2}\sum_{t=1}^{T}\widehat{m}_t\widehat{m}_t' = T^{-2}\sum_{t=1}^{T}m_tm_t' - T^{-2}\sum_{t=1}^{T}tm_t(\sum_{t=1}^{T}t^2)^{-1}\sum_{t=1}^{T}tm_t' = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta m_i)$
  $(T^{-1/2}\sum_{i=1}^{t}\Delta m_i)' - T^{-1}\sum_{t=1}^{T}((t/T)T^{-1/2}\sum_{i=1}^{t}\Delta m_i)(T^{-3}\sum_{t=1}^{T}t^2)^{-1}T^{-1}\sum_{t=1}^{T}((t/T)T^{-1/2}\sum_{i=1}^{t}\Delta m_i)'$
  $\Rightarrow \int \mathcal{B}_m\mathcal{B}_m' - 3\int r\mathcal{B}_m\int r\mathcal{B}_m'$;

- $T^{-2}\sum_{t=1}^{T}x_t\widehat{m}_t' = T^{-2}\sum_{t=1}^{T}x_tm_t' - T^{-2}\sum_{i=1}^{t}tx_t(\sum_{i=1}^{t}t^2)^{-1}\sum_{t=1}^{T}tm_t' = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta x_i)$
  $(T^{-1/2}\sum_{i=1}^{t}\Delta m_i) - T^{-1}\sum_{t=1}^{T}((t/T)T^{-1/2}\sum_{i=1}^{t}\Delta x_i)(T^{-3}\sum_{t=1}^{T}t^2)^{-1}((t/T)T^{-1/2}\sum_{i=1}^{t}\Delta m_i') \Rightarrow \int \mathcal{B}_x$
  $\mathcal{B}_m' - 3\int r\mathcal{B}_x\int r\mathcal{B}_m'$;

- $T^{-2}\sum_{t=1}^{T}x_tx_t' = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta x_t)(T^{-1/2}\sum_{i=1}^{t}\Delta x_t) \Rightarrow \int \mathcal{B}_x\mathcal{B}_x'$.

Therefore, $\ddot{\mathbf{R}}_T \Rightarrow \mathcal{R}$, as desired.

(*ii*) We note that:

- $T^{-1/2}\sum_{t=1}^{T}u_t \Rightarrow \int d\mathcal{B}_u$ using the fact that $T^{-1/2}\sum_{t=1}^{[T(\cdot)]}u_t \Rightarrow \int_0^{(\cdot)}d\mathcal{B}_u$;

- $T^{-3/2}\sum_{t=1}^{T}tu_t = T^{-1/2}\sum_{t=1}^{T}(t/T)u_t + o_{\mathbb{P}}(1) \Rightarrow \int r d\mathcal{B}_u$;

- $T^{-1}\sum_{t=1}^{T}\widehat{m}_tu_t = T^{-1}\sum_{t=1}^{T}u_t(m_t - t(\sum_{t=1}^{T}t^2)^{-1}\sum_{t=1}^{T}tm_t) = T^{-1}\sum_{t=1}^{T}u_tm_t - T^{-3/2}\sum_{t=1}^{T}u_tt(T^{-3}$
  $\sum_{t=1}^{T}t^2)^{-1}T^{-5/2}\sum_{t=1}^{T}tm_t \Rightarrow \int \mathcal{B}_md\mathcal{B}_u + v_{m*} - 3\int r d\mathcal{B}_u\int r\mathcal{B}_m$ using the fact that $v_{m*} := \lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}$
  $\sum_{i=1}^{t}\mathbb{E}[\Delta m_iu_t]$ is finite.

- $T^{-1}\sum_{t=1}^{T}x_tu_t = T^{-1/2}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta x_i)u_t \Rightarrow \int \mathcal{B}_xd\mathcal{B}_u + v_{x*}$ using the fact that $v_{x*} := \lim_{T\to\infty}$
  $T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t}\mathbb{E}[\Delta x_iu_t]$ is finite.

Therefore, $\ddot{\mathbf{U}}_T \Rightarrow \ddot{\mathcal{U}}$. ∎

Proof of Corollary 3.

As the first claim follows in straightforward fashion from Lemma 5, we instead focus on the proof of the second claim. Note that $T^{3/2}(\widehat{\xi}_T - \xi_{*T}) = O_{\mathbb{P}}(1)$ and $\xi_{*T} = \lambda_*'\mu_*^+ + \lambda_*'\sum tm_t(\sum t^2)^{-1}$, so that $T^{3/2}(\widehat{\xi}_T - \lambda_*'\mu_*^+ - \lambda_*'\sum tm_t(\sum t^2)^{-1}) = O_{\mathbb{P}}(1)$. Here, we note that $T^{1/2}\sum tm_t(\sum t^2)^{-1} \Rightarrow \frac{1}{3}\int r\mathcal{B}_m$. Therefore, it now follows that $T^{1/2}(\widehat{\xi}_T - \lambda_*'\mu_*^+) = T^{1/2}\lambda_*'\sum tm_t(\sum t^2)^{-1}) + O_{\mathbb{P}}(T^{-1}) \Rightarrow 3\lambda_*'\int r\mathcal{B}_m$. ∎

Proof of Theorem 4.

This result is easily obtained from Corollary 3. ∎

Proof of Lemma 6.

Given Assumption 3, note that $\bar{\boldsymbol{v}}_T \xrightarrow{\mathbb{P}} \ddot{\boldsymbol{v}}_* := [\boldsymbol{v}'_{m*}, \boldsymbol{v}'_{x*}]'$ and $(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)} \xrightarrow{\mathbb{P}} \ddot{\boldsymbol{\nu}}_* := (\boldsymbol{\Sigma}_*^{(1,1)})^{-1}\boldsymbol{\sigma}_*^{(1,2)}$. Therefore, if we let $\mathring{u}_t := u_t - \boldsymbol{\ell}'_t\ddot{\boldsymbol{\nu}}_*$, $\bar{\mathbf{U}}_T = \ddot{\mathbf{D}}_T^{-1}\sum_{t=1}^T \{\boldsymbol{r}_t\mathring{u}_t - \bar{\mathbf{S}}'\ddot{\boldsymbol{v}}_*\} + o_{\mathbb{P}}(1)$, then $\bar{\mathbf{U}}_T \Rightarrow [\int d\mathcal{B}_{\mathring{u}}, \int r d\mathcal{B}_{\mathring{u}}, \int \boldsymbol{\mathcal{B}}'_m d\mathcal{B}_{\mathring{u}} - 3\int r d\mathcal{B}_{\mathring{u}}\int r\boldsymbol{\mathcal{B}}'_m, \int \boldsymbol{\mathcal{B}}'_x d\mathcal{B}_{\mathring{u}}]'$, where $\mathcal{B}_{\mathring{u}}(\cdot) := \ddot{\tau}\mathcal{W}_u(\cdot)$. Therefore, $\bar{\mathbf{U}}_T \Rightarrow \bar{\boldsymbol{\mathcal{U}}}$. ∎

Proof of Theorem 5.

Given that $\ddot{\mathbf{D}}_T(\bar{\boldsymbol{\varpi}}_T - \bar{\boldsymbol{\varpi}}_{*T}) = (\ddot{\mathbf{D}}_T^{-1}(\sum_{t=1}^T \boldsymbol{r}_t\boldsymbol{r}'_t)\ddot{\mathbf{D}}_T^{-1})^{-1}\bar{\mathbf{U}}_T$, Lemma 6 now proves the first claim. Second, note that $T[(\bar{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)', (\bar{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)']' = \ddot{\mathbf{S}}\ddot{\mathbf{D}}_T(\bar{\boldsymbol{\varpi}}_T - \bar{\boldsymbol{\varpi}}_{*T}) \Rightarrow \ddot{\mathbf{S}}\boldsymbol{\mathcal{R}}^{-1}\bar{\boldsymbol{\mathcal{U}}}$, as desired. ∎

Proof of Theorem 6.

Corollary 2 implies that $T^{3/2}(\widetilde{\boldsymbol{\lambda}}_T - r) \Rightarrow \mathbf{S}\boldsymbol{\mathcal{Q}}^{-1}\widetilde{\boldsymbol{\mathcal{U}}}$ under H_0'', while Lemma 2(i) implies that $\widehat{\mathbf{Q}}_T := \widetilde{\mathbf{D}}_T^{-1}(\sum_{t=1}^T \boldsymbol{q}_t\boldsymbol{q}'_t)\widetilde{\mathbf{D}}_T^{-1} \Rightarrow \boldsymbol{\mathcal{Q}}$. Furthermore, Assumption 2(i) implies that $\widetilde{\tau}_T^2 = \tau_*^2 + o_{\mathbb{P}}(1)$. Given the mixed normal distribution of the FM-OLS estimator for the long-run parameter in Corollary 2, it follows that $\mathcal{W}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_1^2$ under $\mathcal{H}_0''$.

In addition, we note that $\widetilde{\mathcal{W}}_T^{(\ell)} = (\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r})'\widetilde{\mathbf{D}}_T(\widetilde{\tau}_T^2\widetilde{\mathbf{R}}\widehat{\mathbf{Q}}_T^{-1}\widetilde{\mathbf{R}}')^{-1}\widetilde{\mathbf{D}}_T(\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r})$. Furthermore, Theorem 2 implies that $\widetilde{\mathbf{D}}_T(\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r}) \overset{A}{\sim} N(\mathbf{0}, \tau_*^2\widetilde{\mathbf{R}}\boldsymbol{\mathcal{Q}}^{-1}\widetilde{\mathbf{R}}')$ conditional on $\sigma\{\mathcal{B}_x(r) : r \in (0,1]\}$ under H_0'''. Given the condition that $\widehat{\mathbf{Q}}_T \Rightarrow \boldsymbol{\mathcal{Q}}$ and $\widetilde{\tau}_T^2 \xrightarrow{\mathbb{P}} \tau_*^2$, it now follows that $\widetilde{\mathcal{W}}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_2^2$ under H_0'''.

Given that $(\widetilde{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda}_*) = O_{\mathbb{P}}(T^{-3/2})$, $\mathcal{W}_T^{(\ell)} = O_{\mathbb{P}}(T^3)$ under H_1''. Therefore, for any $c_T = o(T^3)$, $\mathbb{P}(\mathcal{W}_T^{(\ell)} > c_T) \to 1$. Furthermore, $(\widetilde{\beta}_T - \beta_*) = O_{\mathbb{P}}(T^{-1})$, implying that $\widetilde{\mathcal{W}}_T^{(\ell)} = O_{\mathbb{P}}(T^2)$ under H_1'''. Therefore, for any $\widetilde{c}_T = o(T^2)$, $\mathbb{P}(\widetilde{\mathcal{W}}_T^{(\ell)} > \widetilde{c}_T) \to 1$. This completes the proof. ∎

Proof of Theorem 7.

Due to its similarity to the standard case, we omit the proof. ∎

Proof of Theorem 8.

Due to its similarity to the standard case, we omit the proof. ∎

A.2 A Further Singularity Problem under Single-Step Estimation

It is important to realize that the re-parameterization of the long-run relationship that we propose to resolve the singularity issue under 2-step estimation in (7) is insufficient to resolve the singularity issue involved in single-step NARDL estimation. In fact, efforts to estimate the short-run and the long-run parameters in a single step by combining (9) with (3) will encounter a further singularity problem, even if $k = 1$. Using the definitions of $\lambda_* := \beta_*^+ - \beta_*^-$ and $\eta_* := \beta_*^-$, it follows that $u_{t-1} = y_{t-1} - \lambda_* x_{t-1}^+ - \beta_* x_{t-1}$, such that:

$$\Delta y_t = \rho_* y_{t-1} + (\theta_*^+ - \theta_*^-)x_{t-1}^+ + \theta_*^- x_{t-1} + \gamma_* + \sum_{j=1}^{p-1}\varphi_{j*}\Delta y_{t-j} + \sum_{j=0}^{q-1}\left(\pi_{j*}^+\Delta x_{t-j}^+ + \pi_{j*}^-\Delta x_{t-j}^-\right) + e_t, \quad (21)$$

where $\beta_*^+ := -\theta_*^+/\rho_*$ and $\beta_*^- := -\theta_*^-/\rho_*$. Let:

$$\boldsymbol{\xi}_* := \begin{bmatrix} \boldsymbol{\xi}_{1*}' & \vdots & \boldsymbol{\xi}_{2*}' \end{bmatrix}' := \begin{bmatrix} \rho_* & \theta_* & \theta_*^- & \vdots & \boldsymbol{\alpha}_{2*}' \end{bmatrix}' \quad \text{and} \quad \boldsymbol{p}_t := \begin{bmatrix} \boldsymbol{p}_{1t}' & \vdots & \boldsymbol{p}_{2t}' \end{bmatrix}' := \begin{bmatrix} y_{t-1} & x_{t-1}^+ & x_{t-1} & \vdots & \boldsymbol{z}_{2t}' \end{bmatrix}'.$$

Note that $\boldsymbol{\xi}_{2*}$ and $\boldsymbol{p}_{2t}$ are identical to $\boldsymbol{\alpha}_{2*}$ and $\boldsymbol{z}_{2t}$, respectively, where $\theta_* := \theta_*^+ - \theta_*^-$. If we attempt to estimate the vector of unknown parameters, $\boldsymbol{\xi}_*$, in (21) by OLS, we obtain:

$$\widehat{\boldsymbol{\xi}}_T := \left(\sum_{t=1}^T \boldsymbol{p}_t \boldsymbol{p}_t' \right)^{-1} \left(\sum_{t=1}^T \boldsymbol{p}_t \Delta y_t \right).$$

We demonstrate that the inverse matrix in $\widehat{\boldsymbol{\xi}}_T$ is asymptotically singular in the following lemma:

Lemma 7. *Given Assumption 1:*

(i) $\ddot{\mathbf{D}}_{1,T}^{-1} \left(\sum_{t=1}^T \boldsymbol{p}_{1t} \boldsymbol{p}_{1t}' \right) \ddot{\mathbf{D}}_{1,T}^{-1} \Rightarrow \boldsymbol{\mathcal{P}}_{11}$, *where* $\ddot{\mathbf{D}}_{1,T} := diag[T^{3/2}\mathbf{I}_2, T]$ *and:*

$$\boldsymbol{\mathcal{P}}_{11} := \begin{bmatrix} \frac{1}{3}\delta_*^2 & \frac{1}{3}\delta_*\mu_*^+ & \delta_* \int r\mathcal{B}_x \\ \frac{1}{3}\delta_*\mu_*^+ & \frac{1}{3}\mu_*^+\mu_*^+ & \mu_*^+ \int r\mathcal{B}_x \\ \delta_* \int r\mathcal{B}_x & \int r\mathcal{B}_x\mu_*^+ & \int \mathcal{B}_x^2 \end{bmatrix};$$

(ii) $\ddot{\mathbf{D}}_{1,T}^{-1} \left(\sum_{t=1}^T \boldsymbol{p}_{1t} \boldsymbol{p}_{2t}' \right) \ddot{\mathbf{D}}_{2,T}^{-1} \Rightarrow \boldsymbol{\mathcal{P}}_{12}$, *where* $\ddot{\mathbf{D}}_{2,T} := diag[T^{1/2}\mathbf{I}_{1+p+2q}]$ *and:*

$$\boldsymbol{\mathcal{P}}_{12} := \begin{bmatrix} \frac{1}{2}\delta_* & \frac{1}{2}\delta_*^2\boldsymbol{\iota}_{p-1}' & \frac{1}{2}\delta_*\boldsymbol{\iota}_q' \otimes \mu_*^+ & \frac{1}{2}\delta_*\boldsymbol{\iota}_q' \otimes \mu_*^- \\ \frac{1}{2}\mu_*^+ & \frac{1}{2}\delta_*\mu_*^+\boldsymbol{\iota}_{p-1}' & \frac{1}{2}\boldsymbol{\iota}_q' \otimes \mu_*^+\mu_*^+ & \frac{1}{2}\boldsymbol{\iota}_q' \otimes \mu_*^+\mu_*^- \\ \int \mathcal{B}_x & \delta_* \int \mathcal{B}_x\boldsymbol{\iota}_{p-1}' & \boldsymbol{\iota}_q' \otimes \int \mathcal{B}_x\mu_*^+ & \boldsymbol{\iota}_q' \otimes \int \mathcal{B}_x\mu_*^- \end{bmatrix}; \quad \text{and}$$

(iii) $\ddot{\mathbf{D}}_{2,T}^{-1} \left(\sum_{t=1}^T \boldsymbol{p}_{2t} \boldsymbol{p}_{2t}' \right) \ddot{\mathbf{D}}_{2,T}^{-1} \overset{\mathbb{P}}{\to} \mathbf{P}_{22} := \mathbf{M}_{22}.$ $\qquad\qquad\square$

We omit the proof of Lemma 7, as it can be easily derived from the proof of Lemma 1. Let $\ddot{\mathbf{D}}_T := diag[T^{3/2}\mathbf{I}_2, T, T^{1/2}\mathbf{I}_{1+p+2q}]$, then:

$$\ddot{\mathbf{D}}_T^{-1} \left(\sum_{t=1}^T \boldsymbol{p}_t \boldsymbol{p}_t' \right) \ddot{\mathbf{D}}_T^{-1} \Rightarrow \boldsymbol{\mathcal{P}} := \begin{bmatrix} \boldsymbol{\mathcal{P}}_{11} & \boldsymbol{\mathcal{P}}_{12} \\ \boldsymbol{\mathcal{P}}_{21} & \mathbf{P}_{22} \end{bmatrix},$$

where $\boldsymbol{\mathcal{P}}_{21} := \boldsymbol{\mathcal{P}}_{12}'$. Note that $\boldsymbol{\mathcal{P}}$ is singular, so it is difficult to obtain the limit distribution of $\widehat{\boldsymbol{\xi}}_T$ using the one-step approach even after applying the re-parameterization of the long-run levels relationship in (7).

	Sample Size	50		100		150		200		250	
	First Step	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS
φ_*	Second Step	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
-0.50	β_*^+	-0.263	-0.130	-0.140	-0.038	-0.095	-0.017	-0.072	-0.010	-0.058	-0.006
	β_*^-	-0.269	-0.038	-0.140	-0.009	-0.095	-0.004	-0.071	-0.002	-0.058	-0.001
	ρ_*	-0.101	-0.083	-0.036	-0.029	-0.022	-0.017	-0.015	-0.012	-0.012	-0.009
	φ_*	0.112	0.074	0.055	0.028	0.036	0.016	0.028	0.012	0.023	0.009
	π_*^+	-0.062	-0.022	-0.026	0.004	-0.018	0.007	-0.012	0.005	-0.010	0.006
	π_*^-	-0.107	-0.038	-0.044	-0.008	-0.032	-0.008	-0.024	-0.006	-0.020	-0.005
-0.25	β_*^+	-0.185	-0.073	-0.098	-0.020	-0.066	-0.007	-0.050	-0.002	-0.040	-0.001
	β_*^-	-0.192	0.004	-0.099	0.002	-0.066	0.003	-0.050	0.004	-0.041	0.003
	ρ_*	-0.084	-0.070	-0.030	-0.026	-0.018	-0.017	-0.012	-0.012	-0.011	-0.010
	φ_*	0.088	0.044	0.048	0.016	0.033	0.008	0.025	0.006	0.020	0.004
	π_*^+	-0.042	-0.002	-0.024	0.002	-0.017	0.003	-0.010	0.005	-0.009	0.005
	π_*^-	-0.069	-0.011	-0.032	-0.005	-0.026	-0.007	-0.019	-0.004	-0.015	-0.003
0.00	β_*^+	-0.112	-0.033	-0.057	-0.010	-0.037	-0.001	-0.027	0.000	-0.022	0.002
	β_*^-	-0.117	0.023	-0.057	0.004	-0.037	0.005	-0.027	0.005	-0.022	0.004
	ρ_*	-0.081	-0.069	-0.035	-0.034	-0.024	-0.023	-0.017	-0.016	-0.013	-0.012
	φ_*	0.051	0.016	0.028	0.007	0.019	0.003	0.015	0.001	0.011	0.000
	π_*^+	-0.035	-0.010	-0.018	-0.006	-0.009	0.001	-0.008	0.000	-0.006	0.002
	π_*^-	-0.047	-0.007	-0.026	-0.011	-0.017	-0.007	-0.015	-0.007	-0.009	-0.003
0.25	β_*^+	-0.024	0.000	-0.009	-0.006	-0.006	-0.003	-0.004	-0.001	-0.004	-0.001
	β_*^-	-0.027	0.024	-0.010	-0.003	-0.007	-0.001	-0.004	0.000	-0.004	-0.001
	ρ_*	-0.072	-0.068	-0.035	-0.036	-0.022	-0.023	-0.017	-0.018	-0.014	-0.014
	φ_*	0.015	0.001	0.006	0.002	0.005	0.001	0.003	0.000	0.003	0.000
	π_*^+	-0.001	-0.004	-0.004	-0.011	0.001	-0.004	-0.003	-0.007	-0.001	-0.005
	π_*^-	-0.024	-0.022	-0.007	-0.015	-0.007	-0.012	-0.004	-0.008	-0.005	-0.008
0.50	β_*^+	0.065	0.015	0.034	-0.023	0.024	-0.016	0.018	-0.013	0.014	-0.011
	β_*^-	0.062	-0.004	0.034	-0.035	0.024	-0.021	0.017	-0.016	0.014	-0.012
	ρ_*	-0.046	-0.046	-0.022	-0.019	-0.015	-0.013	-0.011	-0.010	-0.008	-0.007
	φ_*	-0.016	-0.008	-0.009	0.002	-0.005	0.003	-0.004	0.002	-0.003	0.002
	π_*^+	0.023	-0.017	0.011	-0.021	0.010	-0.012	0.006	-0.012	0.005	-0.009
	π_*^-	0.026	-0.019	0.013	-0.021	0.007	-0.017	0.005	-0.012	0.004	-0.010

Table 1: FINITE SAMPLE BIAS. This table reports the finite sample bias associated with our two-step estimation procedure, both in the case where OLS is used in the first step and in the case where FM-OLS is used in the first step. In all cases, OLS is used in the second step. The data is generated as follows: $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + \Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, where $u_t := y_t - 2x_t^+ - x_t^-$, $\Delta x_t = 0.5\Delta x_{t-1} + \sqrt{1 - 0.5^2}v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	Sample Size	50		100		150		200		250	
	First Step	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS
	Second Step	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
-0.50	β_*^+	0.104	0.057	0.029	0.009	0.013	0.003	0.007	0.001	0.005	0.001
	β_*^-	0.120	0.129	0.029	0.010	0.014	0.003	0.007	0.001	0.005	0.001
	ρ_*	0.026	0.024	0.006	0.006	0.003	0.003	0.002	0.002	0.002	0.002
	φ_*	0.020	0.013	0.006	0.004	0.003	0.002	0.002	0.001	0.002	0.001
	π_*^+	0.153	0.134	0.062	0.053	0.037	0.032	0.025	0.022	0.019	0.017
	π_*^-	0.180	0.152	0.062	0.053	0.038	0.032	0.026	0.022	0.020	0.017
-0.25	β_*^+	0.060	0.034	0.016	0.006	0.007	0.002	0.004	0.001	0.003	0.001
	β_*^-	0.068	0.096	0.016	0.007	0.007	0.002	0.004	0.001	0.003	0.001
	ρ_*	0.023	0.022	0.007	0.007	0.004	0.004	0.003	0.003	0.002	0.002
	φ_*	0.016	0.011	0.007	0.004	0.004	0.003	0.003	0.002	0.002	0.002
	π_*^+	0.136	0.125	0.055	0.051	0.034	0.031	0.023	0.021	0.018	0.017
	π_*^-	0.140	0.130	0.056	0.050	0.033	0.030	0.023	0.021	0.018	0.017
0.00	β_*^+	0.032	0.023	0.007	0.004	0.003	0.001	0.002	0.001	0.001	0.000
	β_*^-	0.036	0.045	0.007	0.005	0.003	0.001	0.002	0.001	0.001	0.000
	ρ_*	0.022	0.021	0.008	0.008	0.005	0.005	0.003	0.003	0.003	0.003
	φ_*	0.011	0.010	0.005	0.004	0.003	0.003	0.002	0.002	0.002	0.002
	π_*^+	0.126	0.121	0.050	0.048	0.031	0.030	0.022	0.022	0.018	0.017
	π_*^-	0.128	0.126	0.050	0.048	0.030	0.029	0.023	0.022	0.018	0.017
0.25	β_*^+	0.015	0.018	0.002	0.003	0.001	0.001	0.001	0.001	0.000	0.000
	β_*^-	0.015	0.034	0.003	0.004	0.001	0.001	0.001	0.001	0.000	0.000
	ρ_*	0.019	0.019	0.007	0.007	0.004	0.004	0.003	0.003	0.002	0.002
	φ_*	0.007	0.009	0.004	0.004	0.002	0.002	0.002	0.002	0.001	0.001
	π_*^+	0.116	0.117	0.047	0.046	0.030	0.029	0.021	0.021	0.017	0.017
	π_*^-	0.112	0.114	0.047	0.047	0.030	0.030	0.022	0.022	0.017	0.017
0.50	β_*^+	0.022	0.020	0.004	0.004	0.002	0.002	0.001	0.001	0.001	0.001
	β_*^-	0.022	0.033	0.004	0.007	0.002	0.002	0.001	0.001	0.001	0.001
	ρ_*	0.011	0.011	0.005	0.004	0.003	0.003	0.002	0.002	0.002	0.002
	φ_*	0.007	0.007	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001
	π_*^+	0.116	0.115	0.045	0.046	0.030	0.030	0.021	0.021	0.017	0.017
	π_*^-	0.113	0.111	0.046	0.047	0.029	0.029	0.021	0.021	0.017	0.017

Table 2: FINITE SAMPLE MEAN SQUARED ERROR (MSE) OF THE ESTIMATORS. This table reports the finite sample MSE associated with our two-step estimation procedure, both in the case where OLS is used in the first step and in the case where FM-OLS is used in the first step. In all cases, OLS is used in the second step. The data is generated as follows: $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + \Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, where $u_t := y_t - 2x_t^+ - x_t^-$, $\Delta x_t = 0.5\Delta x_{t-1} + \sqrt{1-0.5^2}v_t$, and $(e_t, v_t)' \sim \mathrm{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	250	500	750	1,000
-0.50	1%	12.40	5.06	2.38	1.82	1.44
	5%	23.34	12.88	7.96	7.70	5.90
	10%	31.50	21.04	14.08	13.22	11.46
-0.25	1%	8.74	3.80	2.54	1.74	1.18
	5%	19.06	11.10	8.44	6.62	5.48
	10%	27.22	18.36	14.96	11.68	10.48
0.00	1%	4.96	2.92	1.84	1.72	1.42
	5%	13.86	9.76	7.20	6.60	6.12
	10%	21.40	16.28	13.02	12.20	11.34
0.25	1%	3.32	1.62	1.34	1.22	1.06
	5%	10.38	6.24	5.56	5.66	5.60
	10%	17.28	11.42	10.96	11.22	10.70
0.50	1%	1.70	0.86	0.78	0.66	0.72
	5%	5.82	4.06	4.60	4.30	4.22
	10%	10.74	8.20	9.88	9.08	9.12

Table 3: EMPIRICAL LEVELS THE WALD TEST FOR LONG-RUN SYMMETRY. This table reports the empirical level of the Wald test statistic for the long-run parameter, where FM-OLS is used in the first step. The data is generated as follows: $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + (1/3)\Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, where $u_t := y_t - x_t^+ - x_t^-$, $\Delta x_t = 0.5\Delta x_{t-1} + \sqrt{1 - 0.5^2}v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. $H_0^{(\ell)} : \beta_*^+ - \beta_*^- = 0$ vs. $H_1^{(\ell)} : \beta_*^+ - \beta_*^- \neq 0$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	250	500	750	1,000
-0.50	1%	9.80	20.76	83.66	97.34	99.76
	5%	20.00	35.78	89.36	98.54	99.96
	10%	27.66	44.70	91.98	98.92	99.96
-0.25	1%	7.90	26.04	88.16	98.44	99.82
	5%	17.98	41.74	92.88	99.32	99.92
	10%	25.08	51.30	94.58	99.60	99.94
0.00	1%	5.74	32.24	91.22	99.20	99.86
	5%	14.72	49.46	95.12	99.68	99.96
	10%	22.40	58.54	96.66	99.80	99.98
0.25	1%	4.4	34.46	92.36	99.38	99.96
	5%	12.08	52.82	96.04	99.70	100.0
	10%	19.2	61.96	97.22	99.82	100.0
0.50	1%	2.92	25.40	90.96	99.16	99.98
	5%	9.44	47.06	95.40	99.70	100.0
	10%	16.08	58.68	96.98	99.82	100.0

Table 4: EMPIRICAL POWER THE WALD TEST FOR LONG-RUN SYMMETRY (IN PERCENT). This table shows the empirical power of the Wald test statistic for the long-run parameter, where FM-OLS is used in the first step. The data is generated as follows: $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + (1/3)\Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, where $u_t := y_t - 1.01x_t^+ - x_t^-$, $\Delta x_t = 0.5\Delta x_{t-1} + \sqrt{1 - 0.5^2}v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. $H_0^{(\ell)} : \beta_*^+ - \beta_*^- = 0$ vs. $H_1^{(\ell)} : \beta_*^+ - \beta_*^- \neq 0$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	250	500	750	1,000
-0.50	1%	2.44	1.60	0.98	1.18	1.12
	5%	8.06	6.06	5.42	5.46	6.02
	10%	13.82	11.00	10.90	10.36	10.94
-0.25	1%	2.38	1.74	1.46	1.30	1.14
	5%	7.38	6.44	6.02	5.36	5.30
	10%	12.90	11.38	11.28	10.54	10.48
0.00	1%	2.12	1.18	1.22	1.26	0.98
	5%	7.30	5.86	5.76	6.00	5.20
	10%	13.40	11.26	10.94	11.16	10.22
0.25	1%	2.28	1.42	1.36	0.96	0.82
	5%	7.32	6.14	5.84	5.12	4.68
	10%	13.42	11.40	10.96	9.76	9.50
0.50	1%	2.02	1.80	0.98	1.10	1.22
	5%	6.64	6.44	5.44	5.52	5.54
	10%	11.84	11.54	10.62	10.60	10.74

Table 5: EMPIRICAL LEVELS THE WALD TEST FOR SHORT-RUN SYMMETRY (IN PERCENT). This table reports the empirical levels of testing the short-run parameters, where FM-OLS is used in the first step and OLS is used in the second step. The data is generated as follows: $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + (1/2)\Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, where $u_t := y_t - 2x_t^+ - x_t^-$, $\Delta x_t = 0.5\Delta x_{t-1} + \sqrt{1-0.5^2}v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. $H_0^{(s)} : \pi_*^+ - \pi_*^- = 0$ vs. $H_1^{(s)} : H_0 : \pi_*^+ - \pi_*^- \neq 0$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	250	500	750	1,000
-0.50	1%	17.86	44.00	78.48	93.50	98.40
	5%	35.38	66.66	91.82	98.44	99.76
	10%	45.56	76.70	95.76	99.34	99.92
-0.25	1%	17.58	44.84	79.40	93.70	98.32
	5%	34.96	66.66	91.90	98.26	99.72
	10%	46.04	76.64	95.96	99.14	99.86
0.00	1%	17.26	43.16	78.56	93.28	98.60
	5%	35.66	66.68	92.38	98.14	99.72
	10%	46.62	76.14	96.06	99.18	99.90
0.25	1%	17.90	43.02	78.76	93.34	98.72
	5%	35.02	66.14	92.12	98.32	99.68
	10%	45.80	76.24	95.48	99.34	99.98
0.50	1%	17.50	42.82	77.82	92.94	98.54
	5%	34.20	65.78	91.32	98.28	99.78
	10%	44.90	76.06	94.90	99.26	99.92

Table 6: EMPIRICAL POWER THE WALD TEST FOR SHORT-RUN SYMMETRY (IN PERCENT). This table reports the empirical rejection rates of testing the short-run parameters, where FM-OLS is used in the first step and OLS is used in the second step. The data is generated as follows: $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + \Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, where $u_t := y_t - 2x_t^+ - x_t^-$, $\Delta x_t = 0.5\Delta x_{t-1} + \sqrt{1-0.5^2}v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. $H_0^{(s)} : \pi_*^+ - \pi_*^- = 0$ vs. $H_1^{(s)} : H_0 : \pi_*^+ - \pi_*^- \neq 0$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	Sample Size	50		100		200		300		400		500	
	First Step	TOLS	FM-TOLS	TOLS	FM-TOLS	TOLS	FM-TOLS	TOLS	FM-TOLS	TOLS	FM-TOLS	TOLS	FM-TOLS
	Second Step	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
-0.30	β_{1*}^+	0.416	0.723	0.208	0.246	0.102	0.061	0.067	0.035	0.051	0.019	0.040	0.011
	β_{2*}^+	-0.334	-0.737	-0.168	-0.242	-0.082	-0.064	-0.057	-0.039	-0.039	-0.024	-0.033	-0.013
	β_{1*}^-	-0.431	-0.816	-0.212	-0.289	-0.107	-0.071	-0.069	-0.043	-0.053	-0.023	-0.042	-0.014
	β_{2*}^-	0.568	1.065	0.289	0.365	0.144	0.095	0.097	0.057	0.071	0.032	0.057	0.018
	ρ_*	-0.190	0.024	-0.076	-0.031	-0.032	-0.032	-0.021	-0.020	-0.015	-0.016	-0.011	-0.011
	φ_*	0.137	0.036	0.067	0.043	0.033	0.023	0.022	0.016	0.016	0.010	0.013	0.007
	π_{1*}^+	0.091	0.226	0.023	0.083	0.004	0.015	0.003	0.012	0.004	0.007	0.000	0.003
	π_{2*}^+	-0.120	-0.312	-0.057	-0.099	-0.030	-0.024	-0.021	-0.013	-0.008	-0.008	-0.012	-0.004
	π_{1*}^-	-0.047	-0.222	0.002	-0.076	0.003	-0.014	0.008	-0.008	0.003	-0.002	0.005	-0.001
	π_{2*}^-	-0.011	0.219	-0.027	0.066	-0.022	0.017	-0.017	0.012	-0.017	0.006	-0.014	0.004
-0.10	β_{1*}^+	0.390	0.694	0.190	0.259	0.090	0.075	0.059	0.040	0.044	0.021	0.034	0.015
	β_{2*}^+	-0.318	-0.683	-0.153	-0.265	-0.074	-0.078	-0.048	-0.043	-0.037	-0.024	-0.029	-0.016
	β_{1*}^-	-0.414	-0.779	-0.198	-0.293	-0.098	-0.086	-0.062	-0.047	-0.047	-0.025	-0.037	-0.018
	β_{2*}^-	0.522	0.999	0.252	0.384	0.121	0.111	0.080	0.062	0.059	0.034	0.047	0.022
	ρ_*	-0.158	0.019	-0.063	-0.016	-0.026	-0.026	-0.016	-0.017	-0.012	-0.013	-0.009	-0.011
	φ_*	0.124	0.050	0.068	0.042	0.034	0.024	0.024	0.015	0.017	0.010	0.013	0.007
	π_{1*}^+	0.073	0.227	0.017	0.079	0.003	0.024	0.001	0.013	0.003	0.008	-0.001	0.006
	π_{2*}^+	-0.109	-0.280	-0.048	-0.111	-0.021	-0.031	-0.010	-0.012	-0.010	-0.009	-0.008	-0.003
	π_{1*}^-	-0.030	-0.222	0.005	-0.071	0.004	-0.022	0.010	-0.009	0.004	-0.008	0.007	-0.005
	π_{2*}^-	-0.017	0.233	-0.036	0.088	-0.032	0.033	-0.026	0.016	-0.018	0.011	-0.015	0.008
0.00	β_{1*}^+	0.379	0.698	0.173	0.264	0.083	0.078	0.053	0.042	0.040	0.026	0.031	0.016
	β_{2*}^+	-0.317	-0.686	-0.146	-0.267	-0.071	-0.081	-0.047	-0.045	-0.034	-0.026	-0.028	-0.017
	β_{1*}^-	-0.409	-0.811	-0.186	-0.310	-0.092	-0.090	-0.058	-0.049	-0.044	-0.030	-0.034	-0.019
	β_{2*}^-	0.501	1.014	0.232	0.391	0.111	0.116	0.075	0.065	0.055	0.038	0.044	0.025
	ρ_*	-0.141	0.033	-0.057	-0.007	-0.025	-0.021	-0.014	-0.015	-0.011	-0.013	-0.008	-0.010
	φ_*	0.112	0.044	0.061	0.040	0.032	0.023	0.021	0.015	0.017	0.010	0.013	0.007
	π_{1*}^+	0.084	0.237	0.015	0.089	0.003	0.027	0.002	0.016	0.001	0.011	-0.002	0.006
	π_{2*}^+	-0.106	-0.287	-0.043	-0.118	-0.023	-0.029	-0.014	-0.014	-0.009	-0.012	-0.007	-0.007
	π_{1*}^-	-0.041	-0.239	0.011	-0.086	0.008	-0.025	0.010	-0.013	0.007	-0.009	0.007	-0.008
	π_{2*}^-	-0.027	0.226	-0.045	0.094	-0.033	0.037	-0.023	0.019	-0.019	0.016	-0.015	0.009
0.10	β_{1*}^+	0.345	0.729	0.169	0.272	0.077	0.083	0.049	0.048	0.036	0.030	0.029	0.019
	β_{2*}^+	-0.306	-0.706	-0.148	-0.291	-0.067	-0.085	-0.042	-0.049	-0.032	-0.030	-0.026	-0.020
	β_{1*}^-	-0.389	-0.829	-0.182	-0.311	-0.086	-0.095	-0.053	-0.054	-0.040	-0.034	-0.033	-0.022
	β_{2*}^-	0.476	1.059	0.223	0.411	0.101	0.122	0.066	0.071	0.049	0.043	0.039	0.029
	ρ_*	-0.125	0.051	-0.050	0.005	-0.020	-0.018	-0.014	-0.012	-0.010	-0.011	-0.007	-0.009
	φ_*	0.101	0.035	0.055	0.038	0.029	0.022	0.021	0.013	0.015	0.010	0.012	0.007
	π_{1*}^+	0.056	0.258	0.018	0.092	0.004	0.036	0.001	0.019	-0.001	0.014	-0.001	0.008
	π_{2*}^+	-0.105	-0.294	-0.042	-0.129	-0.019	-0.035	-0.008	-0.020	-0.009	-0.014	-0.007	-0.007
	π_{1*}^-	-0.022	-0.239	0.013	-0.081	0.010	-0.029	0.012	-0.014	0.009	-0.015	0.006	-0.008
	π_{2*}^-	-0.027	0.265	-0.041	0.100	-0.033	0.042	-0.022	0.026	-0.019	0.018	-0.014	0.011
0.30	β_{1*}^+	0.306	0.774	0.140	0.303	0.063	0.103	0.040	0.054	0.029	0.034	0.023	0.025
	β_{2*}^+	-0.308	-0.770	-0.128	-0.306	-0.062	-0.101	-0.040	-0.052	-0.028	-0.035	-0.023	-0.023
	β_{1*}^-	-0.369	-0.886	-0.160	-0.348	-0.071	-0.117	-0.047	-0.061	-0.034	-0.039	-0.027	-0.027
	β_{2*}^-	0.438	1.137	0.181	0.437	0.082	0.148	0.052	0.077	0.038	0.052	0.030	0.036
	ρ_*	-0.093	0.086	-0.039	0.022	-0.017	-0.007	-0.010	-0.009	-0.008	-0.008	-0.006	-0.006
	φ_*	0.069	0.012	0.041	0.028	0.022	0.019	0.015	0.012	0.011	0.008	0.009	0.006
	π_{1*}^+	0.044	0.269	0.013	0.114	0.001	0.046	-0.001	0.024	0.000	0.018	-0.002	0.012
	π_{2*}^+	-0.105	-0.323	-0.039	-0.133	-0.019	-0.040	-0.011	-0.023	-0.006	-0.013	-0.006	-0.011
	π_{1*}^-	-0.013	-0.268	0.018	-0.104	0.016	-0.045	0.014	-0.023	0.010	-0.016	0.009	-0.013
	π_{2*}^-	-0.032	0.304	-0.042	0.122	-0.034	0.050	-0.022	0.027	-0.021	0.021	-0.016	0.019

Table 7: FINITE SAMPLE BIAS. This table reports the finite sample bias associated with our two-step estimation procedure, both in the case where the TOLS is used in the first step and in the case where the FM-TOLS is used in the first step. In all cases, OLS is used in the second step. The data is generated as follows: $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5 \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \mathrm{IID} N(\boldsymbol{0}_3, \mathbf{I}_3)$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	Sample Size	50		100		200		300		400		500	
	First Step	TOLS	FM-TOLS	TOLS	FM-TOLS	TOLS	FM-TOLS	TOLS	FM-TOLS	TOLS	FM-TOLS	TOLS	FM-TOLS
	Second Step	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
-0.30	β_{1*}^+	0.461	1.373	0.113	0.183	0.027	0.017	0.012	0.007	0.007	0.002	0.004	0.001
	β_{2*}^+	0.406	1.443	0.096	0.188	0.024	0.019	0.011	0.007	0.006	0.003	0.004	0.001
	β_{1*}^-	0.472	1.573	0.111	0.221	0.028	0.020	0.012	0.008	0.007	0.003	0.004	0.001
	β_{2*}^-	0.543	2.052	0.139	0.279	0.035	0.025	0.016	0.009	0.008	0.003	0.006	0.001
	ρ_*	0.059	0.060	0.012	0.012	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001
	φ_*	0.027	0.020	0.007	0.006	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.000
	π_{1*}^+	0.423	0.586	0.117	0.132	0.040	0.031	0.021	0.018	0.015	0.011	0.011	0.009
	π_{2*}^+	0.412	0.673	0.115	0.142	0.039	0.031	0.023	0.017	0.015	0.011	0.011	0.009
	π_{1*}^-	0.408	0.595	0.117	0.132	0.039	0.032	0.022	0.018	0.015	0.012	0.011	0.008
	π_{2*}^-	0.345	0.505	0.112	0.124	0.040	0.032	0.023	0.018	0.016	0.012	0.012	0.009
-0.10	β_{1*}^+	0.445	1.257	0.098	0.192	0.023	0.019	0.010	0.006	0.005	0.002	0.003	0.001
	β_{2*}^+	0.378	1.234	0.087	0.192	0.020	0.019	0.008	0.007	0.005	0.002	0.003	0.001
	β_{1*}^-	0.438	1.438	0.098	0.224	0.024	0.022	0.010	0.008	0.006	0.003	0.003	0.001
	β_{2*}^-	0.509	1.812	0.114	0.289	0.027	0.028	0.011	0.010	0.006	0.003	0.004	0.002
	ρ_*	0.043	0.041	0.010	0.010	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001
	φ_*	0.024	0.018	0.008	0.006	0.003	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	π_{1*}^+	0.382	0.549	0.102	0.130	0.036	0.031	0.020	0.017	0.014	0.011	0.010	0.009
	π_{2*}^+	0.378	0.569	0.111	0.133	0.035	0.032	0.020	0.017	0.014	0.011	0.010	0.009
	π_{1*}^-	0.355	0.565	0.102	0.126	0.034	0.031	0.020	0.017	0.013	0.011	0.010	0.009
	π_{2*}^-	0.341	0.507	0.105	0.126	0.037	0.032	0.021	0.018	0.014	0.012	0.011	0.009
0.00	β_{1*}^+	0.429	1.254	0.088	0.190	0.020	0.020	0.008	0.006	0.005	0.003	0.003	0.001
	β_{2*}^+	0.371	1.350	0.077	0.196	0.019	0.020	0.008	0.007	0.004	0.003	0.003	0.001
	β_{1*}^-	0.440	1.516	0.093	0.235	0.021	0.024	0.009	0.008	0.005	0.003	0.003	0.001
	β_{2*}^-	0.476	1.933	0.099	0.305	0.023	0.030	0.010	0.010	0.006	0.004	0.004	0.002
	ρ_*	0.037	0.042	0.009	0.009	0.003	0.003	0.001	0.002	0.001	0.001	0.001	0.001
	φ_*	0.021	0.018	0.007	0.006	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	π_{1*}^+	0.365	0.549	0.099	0.127	0.034	0.030	0.019	0.017	0.013	0.012	0.010	0.009
	π_{2*}^+	0.352	0.609	0.096	0.132	0.033	0.031	0.019	0.017	0.013	0.012	0.009	0.009
	π_{1*}^-	0.343	0.559	0.102	0.124	0.033	0.030	0.020	0.017	0.013	0.012	0.010	0.009
	π_{2*}^-	0.332	0.505	0.098	0.122	0.035	0.031	0.020	0.017	0.014	0.012	0.010	0.009
0.10	β_{1*}^+	0.397	1.286	0.084	0.209	0.017	0.021	0.007	0.007	0.004	0.003	0.003	0.001
	β_{2*}^+	0.372	1.294	0.078	0.219	0.017	0.022	0.007	0.007	0.004	0.003	0.002	0.001
	β_{1*}^-	0.416	1.551	0.087	0.238	0.019	0.026	0.008	0.008	0.004	0.003	0.003	0.002
	β_{2*}^-	0.461	2.088	0.097	0.323	0.020	0.032	0.009	0.011	0.005	0.004	0.003	0.002
	ρ_*	0.030	0.041	0.007	0.009	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	φ_*	0.018	0.016	0.006	0.006	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	π_{1*}^+	0.340	0.539	0.095	0.131	0.032	0.031	0.018	0.017	0.013	0.012	0.010	0.009
	π_{2*}^+	0.336	0.567	0.091	0.143	0.032	0.031	0.018	0.017	0.013	0.012	0.010	0.009
	π_{1*}^-	0.317	0.553	0.091	0.126	0.032	0.032	0.020	0.018	0.013	0.011	0.010	0.009
	π_{2*}^-	0.318	0.532	0.095	0.126	0.032	0.032	0.020	0.018	0.014	0.012	0.011	0.009
0.30	β_{1*}^+	0.360	1.553	0.071	0.225	0.014	0.027	0.006	0.008	0.003	0.003	0.002	0.002
	β_{2*}^+	0.391	1.511	0.066	0.232	0.014	0.026	0.006	0.008	0.003	0.003	0.002	0.002
	β_{1*}^-	0.413	1.866	0.074	0.276	0.015	0.032	0.006	0.009	0.003	0.004	0.002	0.002
	β_{2*}^-	0.436	2.386	0.076	0.357	0.016	0.043	0.006	0.012	0.003	0.005	0.002	0.002
	ρ_*	0.021	0.041	0.006	0.009	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	φ_*	0.012	0.015	0.005	0.005	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	π_{1*}^+	0.305	0.567	0.084	0.134	0.029	0.034	0.018	0.017	0.012	0.012	0.009	0.009
	π_{2*}^+	0.320	0.611	0.083	0.140	0.029	0.034	0.017	0.018	0.012	0.011	0.009	0.009
	π_{1*}^-	0.313	0.603	0.081	0.131	0.029	0.034	0.017	0.017	0.012	0.012	0.009	0.009
	π_{2*}^-	0.282	0.573	0.089	0.130	0.031	0.035	0.019	0.018	0.013	0.012	0.010	0.009

Table 8: FINITE SAMPLE MEAN SQUARED ERROR (MSE) OF THE ESTIMATORS. This table reports the finite sample MSE associated with our two-step estimation procedure, both in the case where the TOLS is used in the first step and in the case where the FM-TOLS is used in the first step. In all cases, OLS is used in the second step. The data is generated as follows: $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5 \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \text{IID} N(\boldsymbol{0}_3, \mathbf{I}_3)$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	250	500	1,000	3,000	5,000
-0.30	1%	45.20	17.06	7.46	2.62	1.10	1.32
	5%	58.92	30.68	17.62	9.50	5.64	5.86
	10%	65.82	40.08	25.76	16.40	11.72	10.62
-0.10	1%	42.96	18.12	8.12	3.32	1.48	1.24
	5%	57.82	31.10	19.04	10.76	6.02	5.48
	10%	65.34	39.20	26.66	17.48	11.64	10.34
0.00	1%	42.50	16.78	8.44	4.30	1.42	1.28
	5%	56.38	29.90	19.20	11.62	5.80	5.60
	10%	64.06	38.70	28.20	18.74	11.36	10.76
0.10	1%	41.24	16.18	8.56	4.02	1.68	1.18
	5%	56.14	30.00	19.30	11.32	6.68	5.56
	10%	63.30	38.78	27.12	19.24	12.50	11.26
0.30	1%	40.10	15.72	6.94	3.92	2.16	1.84
	5%	54.52	28.24	17.16	12.52	8.00	7.34
	10%	63.44	36.76	24.94	19.10	14.24	12.88

Table 9: EMPIRICAL LEVELS THE WALD TEST FOR LONG-RUN SYMMETRY. This table reports the empirical level of the Wald test statistic for the long-run parameter, where the FM-TOLS estimator is used in the first step. The data is generated as follows: $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5\Delta \boldsymbol{x}_{t-1} + \sqrt{1-0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \text{IID}N(\boldsymbol{0}_3, \mathbf{I}_3)$. $\dot{H}_0^{(\ell)} : \boldsymbol{\iota}_2' \boldsymbol{\beta}_*^+ = -0.5$ and $\boldsymbol{\iota}_2' \boldsymbol{\beta}_*^- = -0.75$ vs. $\dot{H}_1^{(\ell)} : \boldsymbol{\iota}_2' \boldsymbol{\beta}_*^+ \neq -0.5$ or $\boldsymbol{\iota}_2' \boldsymbol{\beta}_*^- \neq -0.75$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	200	300	400	500
-0.30	1%	47.46	70.16	90.44	98.10	99.74
	5%	62.40	80.76	94.26	99.00	99.88
	10%	70.16	85.42	96.18	99.32	99.92
-0.10	1%	48.98	71.36	91.30	98.24	99.70
	5%	63.08	81.32	95.26	99.10	99.88
	10%	70.12	86.22	96.76	99.36	99.94
0.00	1%	47.32	70.54	90.56	98.20	99.78
	5%	61.72	81.70	94.02	99.12	99.94
	10%	69.78	86.56	95.44	99.46	99.94
0.10	1%	46.36	69.20	90.22	98.06	99.86
	5%	60.26	79.82	94.00	98.98	99.96
	10%	67.16	84.56	95.86	99.36	99.98
0.30	1%	43.10	63.52	88.70	97.58	99.44
	5%	56.98	75.22	93.74	98.88	99.78
	10%	65.50	80.66	95.68	99.18	99.88

Table 10: EMPIRICAL POWER THE WALD TEST FOR LONG-RUN SYMMETRY (IN PERCENT). This table shows the empirical power of the Wald test statistic for the long-run parameter, where the FM-TOLS estimator is used in the first step. The data is generated as follows: $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5\Delta \boldsymbol{x}_{t-1} + \sqrt{1-0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \text{IID}N(\boldsymbol{0}_3, \mathbf{I}_3)$. $\ddot{H}_0^{(\ell)} : \boldsymbol{\iota}_2' \boldsymbol{\beta}_*^+ = -0.4$ and $\boldsymbol{\iota}_2' \boldsymbol{\beta}_*^- = -0.65$ vs. $\ddot{H}_1^{(\ell)} : \boldsymbol{\iota}_2' \boldsymbol{\beta}_*^+ \neq -0.4$ or $\boldsymbol{\iota}_2' \boldsymbol{\beta}_*^- \neq -0.65$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	200	400	600	800	1,000
-0.30	1%	10.62	4.16	1.58	1.66	1.28	1.56
	5%	23.24	11.90	7.28	6.78	5.88	5.74
	10%	31.96	18.88	13.34	12.26	11.34	10.80
-0.10	1%	10.58	3.72	2.02	1.34	1.20	1.22
	5%	22.50	11.12	7.56	6.36	6.06	5.84
	10%	32.18	18.66	13.38	12.28	11.38	10.94
0.00	1%	10.90	3.92	2.00	1.54	1.48	1.28
	5%	23.34	11.80	7.90	6.36	6.64	5.50
	10%	31.46	19.30	13.96	11.36	12.46	10.50
0.10	1%	10.80	3.98	2.00	1.36	1.44	1.26
	5%	22.60	11.96	7.36	6.54	6.46	5.40
	10%	30.78	18.68	12.74	12.38	12.16	10.82
0.30	1%	11.04	4.14	1.84	1.72	1.38	1.10
	5%	22.68	11.56	7.60	7.20	6.74	5.14
	10%	32.22	18.98	13.06	12.56	12.34	10.62

Table 11: EMPIRICAL LEVELS THE WALD TEST FOR SHORT-RUN SYMMETRY (IN PERCENT). This table reports the empirical levels of testing the short-run parameters, where the FM-TOLS estimator is used in the first step and OLS is used in the second step. The data is generated as follows: $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5\Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \text{IID} N(\boldsymbol{0}_3, \mathbf{I}_3)$. $\dot{H}_0^{(s)} : \boldsymbol{\pi}_{0*}^+ - \boldsymbol{\pi}_{0*}^- = \boldsymbol{0}_2$ vs. $\dot{H}_1^{(s)} : \boldsymbol{\pi}_{0*}^+ - \boldsymbol{\pi}_{0*}^- \neq \boldsymbol{0}_2$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	200	400	600	800	1,000
-0.30	1%	20.96	21.94	39.98	60.80	77.26	86.48
	5%	36.20	39.78	62.62	80.76	91.16	95.32
	10%	45.90	51.10	73.52	87.68	94.76	97.76
-0.10	1%	21.56	22.94	41.36	61.54	76.68	86.96
	5%	37.70	41.56	63.86	80.14	91.18	95.40
	10%	46.98	52.92	74.76	87.70	95.16	97.72
0.00	1%	21.04	23.38	42.32	61.96	76.86	86.66
	5%	36.70	42.28	64.08	81.12	90.50	95.58
	10%	46.52	53.74	74.38	88.60	94.46	97.34
0.10	1%	20.04	24.42	40.92	61.14	76.30	86.52
	5%	35.20	42.46	63.34	80.60	90.32	95.54
	10%	45.34	53.40	73.86	87.96	94.84	97.78
0.30	1%	20.48	23.32	42.08	62.02	76.58	86.82
	5%	36.56	42.84	63.60	80.88	90.20	95.28
	10%	46.76	53.20	73.78	87.98	94.48	97.42

Table 12: EMPIRICAL POWER THE WALD TEST FOR SHORT-RUN SYMMETRY (IN PERCENT). This table reports the empirical rejection rates of testing the short-run parameters, where the FM-TOLS estimator is used in the first step and OLS is used in the second step. The data is generated as follows: $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5\Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \text{IID} N(\boldsymbol{0}_3, \mathbf{I}_3)$. $\ddot{H}_0^{(s)} : \boldsymbol{\pi}_{0*}^+ - \boldsymbol{\pi}_{0*}^- = 0.3\boldsymbol{\iota}_2$ vs. $\ddot{H}_1^{(s)} : \boldsymbol{\pi}_{0*}^+ - \boldsymbol{\pi}_{0*}^- \neq 0.3\boldsymbol{\iota}_2$. The simulation results are obtained using $R = 5,000$ replications.

	Real Earnings		Real Dividends	
	Level	Difference	Level	Difference
Mean	41.059	0.374	19.289	0.091
Median	38.462	0.375	18.988	0.064
Maximum	103.475	14.383	31.585	1.257
Minimum	11.177	-13.117	8.422	-1.289
Standard Deviation	15.951	2.458	4.296	0.352
Skewness	1.353	-0.435	-0.239	0.319
Excess Kurtosis	2.275	9.909	0.326	2.223

Table 13: COMMON SAMPLE DESCRIPTIVE STATISTICS. Descriptive statistics are computed over 243 quarters from 1946Q2–2006Q4. Both real earnings and real dividends are measured in US Dollars at January 2000 prices. We convert from the original monthly sampling frequency used by Shiller to quarterly frequency by taking end-of-period values.

	One-step NARDL		Two step FM-OLS	
	Estimate	S.E.	Estimate	S.E.
Intercept	–	–	17.774	1.558
β^+	0.168	0.075	0.170	0.038
β^-	0.143	0.099	0.135	0.049

Table 14: LONG-RUN PARAMETER ESTIMATES. This table reports the long-run parameter estimates obtained from the single-step estimation procedure of SYG as well as our two-step estimation procedure, where FM-OLS is used in the first step and OLS is used in the second step. The long-run parameters are obtained from the single-step estimation results as $\hat{\beta}^+ = -\hat{\theta}^+/\hat{\rho}$ and $\hat{\beta}^- = -\hat{\theta}^-/\hat{\rho}$ and the corresponding analytical standard errors are computed via the Delta method. Note that the intercept of the cointegrating equation is not identified in the single-step estimation procedure. The long-run parameters are obtained from first stage FM-OLS estimation results as $\hat{\beta}^+ = \hat{\lambda} + \hat{\eta}$ and $\hat{\beta}^- = \hat{\eta}$. The standard error of $\hat{\beta}^-$ is obtained directly, while the analytical standard error of $\hat{\beta}^+$ is computed via the Delta method.

	One-step NARDL		Two-step FM-OLS	
	Estimate	S.E.	Estimate	S.E.
Intercept	0.406	0.146	0.008	0.030
D_{t-1}	-0.031	0.010	–	–
E_{t-1}^{+}	0.005	0.003	–	–
E_{t-1}^{-}	0.004	0.003	–	–
ECM_{t-1}	–	–	-0.032	0.010
ΔD_{t-1}	0.245	0.066	0.245	0.066
ΔD_{t-2}	0.161	0.067	0.159	0.067
ΔD_{t-3}	0.139	0.066	0.136	0.065
ΔE_{t}^{+}	0.052	0.017	0.049	0.015
ΔE_{t-1}^{+}	-0.012	0.018	-0.014	0.017
ΔE_{t-2}^{+}	0.004	0.018	0.003	0.018
ΔE_{t-3}^{+}	0.006	0.017	0.004	0.017
ΔE_{t}^{-}	0.007	0.021	0.010	0.020
ΔE_{t-1}^{-}	0.005	0.026	0.006	0.026
ΔE_{t-2}^{-}	0.014	0.026	0.014	0.026
ΔE_{t-3}^{-}	-0.024	0.021	-0.021	0.020
Adjusted R^2	0.285		0.291	
$\mathcal{X}^2_{\text{S.Corr.}}$	0.092		0.103	
$\mathcal{X}^2_{\text{Hetero.}}$	0.064		0.094	

Table 15: DYNAMIC PARAMETER ESTIMATES. This table reports parameter estimates for the NARDL(4,4) model in error correction form, estimated in a single-step following SYG and using our two-step procedure, where FM-OLS is used in the first step and OLS is used in the second step. $\chi^2_{\text{S.Corr.}}$ denotes the Breusch–Godfrey Lagrange multiplier test for serial correlation up to order four. $\chi^2_{\text{Hetero.}}$ denotes the Breusch–Pagan–Godfrey Lagrange multiplier test for residual heteroskedasticity. The values reported for these two tests are asymptotic p-values.
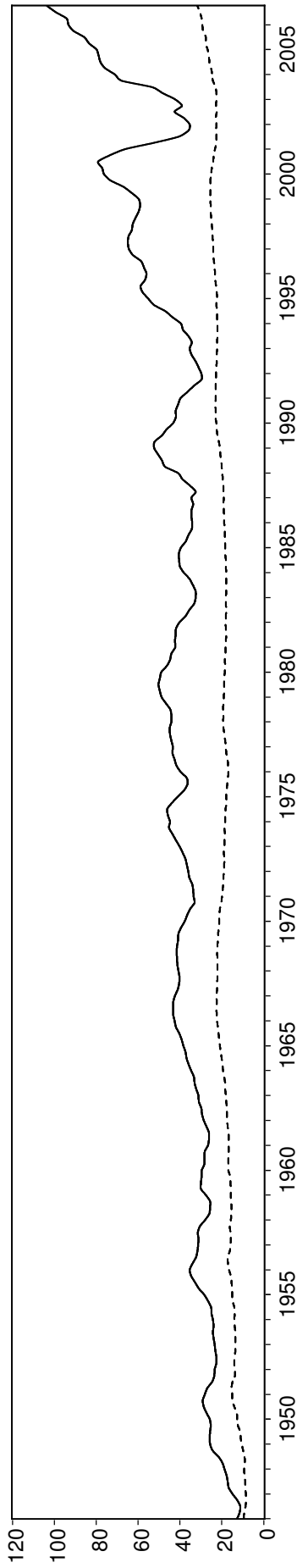
Figure 1: REAL EARNINGS VS. REAL DIVIDENDS. The solid line represents real earnings and the dashed line real dividends. Both series are measured in US Dollars at January 2000 prices. We convert from the original monthly sampling frequency to quarterly frequency by taking the end-of-period value.
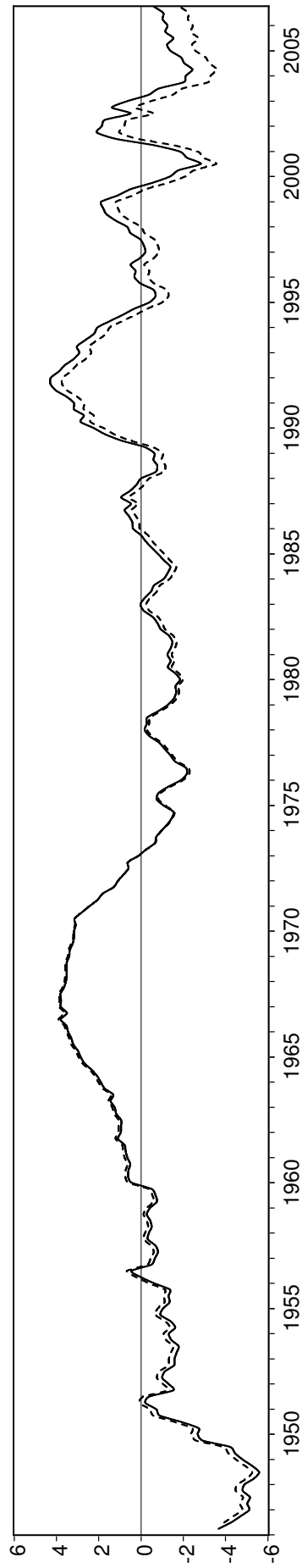

Figure 2: LONG-RUN DISEQUILIBRIUM ERROR. The long-run disequilibrium error obtained from the single-step estimation procedure is shown as a solid line. The values are obtained as $\xi_t = D_t - \beta^+ E_t^+ - \beta^- E_t^-$. In the figure, we de-mean ξ_t before we plot it. The long-run disequilibrium error obtained from the two-step estimation procedure using FM-OLS in the first step and OLS in the second step is shown as a dashed line. The values in this case are obtained simply as the residual from the first-stage FM-OLS regression.