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

November 28, 2019

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 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 second step. We show that our two-step estimators are consistent for the parameters of the NARDL model and we derive their limit distributions. We also 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 U.S.

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.

*We are grateful to In Choi, Tae-Hwan Kim and Michael Thornton for constructive comments. Cho is grateful for financial support from the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2019S1A5A2A01035568). Greenwood-Nimmo acknowledges the hospitality of the Universities of York and Yonsei during extended visits in 2018 and 2019. The usual disclaimer applies.

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) and the economics of tourism (Süssmuth and Woitek, 2013), 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 variable 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 represent 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 overcome this asymptotic singularity problem, we begin by reparameterizing the asymmetric long-run relationship embedded in the NARDL model. In a bivariate model with a scalar dependent variable, y_t, and a scalar explanatory variable, x_t, the asymmetric long-run relationship is usually expressed among the level of the dependent variable, y_t, 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 making use of a simple

[1] At the time of writing, SYG has been cited more than 700 times according to Google Scholar.

one-to-one transformation as a relationship between y_t, x_t and x_t^+. By excluding one partial sum process in this way, 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, it is insufficient to resolve the singularity problem from 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). Our solution, therefore, 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 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 find follows 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.

Next, we develop Wald tests that can be used to evaluate restrictions on the short- and long-run parameters. 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, once again facilitating standard inference.

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. We show that the Wald tests of both the short- and long-run parameters have high power and generally exhibit only mild size distortions in small samples that are rapidly corrected at larger sample sizes. Overall, therefore, 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 to this literature 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 advanced by SYG and our newly-developed two-step procedure. Our estimation results suggest 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 this 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, we conjecture that, 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. 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 two-step 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:

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

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

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

such that $\Delta \boldsymbol{x}_t$ is a stationary process. Note that (1) can be re-written in error-correction form as:

$$\Delta y_t = \rho_* y_{t-1} + \boldsymbol{\theta}_*^{+\prime} \boldsymbol{x}_{t-1}^{+} + \boldsymbol{\theta}_*^{-\prime} \boldsymbol{x}_{t-1}^{-} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^{+\prime} \Delta \boldsymbol{x}_{t-j}^{+} + \boldsymbol{\pi}_{j*}^{-\prime} \Delta \boldsymbol{x}_{t-j}^{-} \right) + e_t, \tag{2}$$

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 re-write (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 unit-root process and other unit-root processes, possibly associated with a time trend. Suppose that $\mathbb{E}[\Delta \boldsymbol{x}_t] \equiv \mathbf{0}$ and that:

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

It follows that $\boldsymbol{\mu}_*^{+} + \boldsymbol{\mu}_*^{-} \equiv \mathbf{0}$ 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:

$$\delta_* := \mathbb{E}[\Delta y_t] = -\frac{1}{\rho_*}\left[\sum_{j=0}^{q}(\boldsymbol{\theta}_{j*}^+)'\boldsymbol{\mu}_*^+ + \sum_{j=0}^{q}(\boldsymbol{\theta}_{j*}^-)'\boldsymbol{\mu}_*^-\right].$$

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 $\delta_* \neq 0$.

Provided that $\mathbb{E}[\Delta \boldsymbol{x}_t] = \mathbf{0}$, then 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. Meanwhile, if $\mathbb{E}[\Delta \boldsymbol{x}_t] \neq \mathbf{0}$, then $\boldsymbol{x}_t$ is a deterministic time-trend process driven by a unit-root process. In this case, the model captures a cointegrating relationship between deterministic time-trend processes driven by unit-root processes.

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) $\mathbb{E}[\Delta \boldsymbol{x}_t] = \mathbf{0}$, $\mathbb{E}[|\Delta \boldsymbol{x}_{ti}|^r] < \infty$ $(i = 1, 2, \ldots, k)$, $\mathbb{E}[|u_t|^r] < \infty$, *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* $\{y_{t-1}, \boldsymbol{x}_t^+, \boldsymbol{x}_t^-, y_{t-2}, \boldsymbol{x}_{t-1}^+, \boldsymbol{x}_{t-1}^-, \ldots\}$. □

Furthermore, for notational simplicity, we define the following:

$$
\begin{aligned}
\boldsymbol{z}_t &:= \left[\begin{array}{c:c} \boldsymbol{z}_{1t}' & \boldsymbol{z}_{2t}' \end{array}\right]' \\
&:= \left[\begin{array}{ccc:cccccccc} y_{t-1} & \boldsymbol{x}_{t-1}^{+\prime} & \boldsymbol{x}_{t-1}^{-\prime} & 1 & \Delta \boldsymbol{y}_{t-1}' & \Delta \boldsymbol{x}_t^{+\prime} & \ldots & \Delta \boldsymbol{x}_{t-q+1}^{+\prime} & \Delta \boldsymbol{x}_t^{-\prime} & \ldots & \Delta \boldsymbol{x}_{t-q+1}^{-\prime} \end{array}\right]',
\end{aligned}
$$

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

$$
\begin{aligned}
\boldsymbol{z}_{2t} := \left[\begin{array}{c:c} 1 & \boldsymbol{w}_t' \end{array}\right]' &:= \left[\begin{array}{c:c:c:c} 1 & \boldsymbol{w}_{1t}' & \boldsymbol{w}_{2t}' & \boldsymbol{w}_{3t}' \end{array}\right]' \\
&:= \left[\begin{array}{c:c:ccc:ccc} 1 & \Delta \boldsymbol{y}_{t-1}' & \Delta \boldsymbol{x}_t^{+\prime} & \ldots & \Delta \boldsymbol{x}_{t-q+1}^{+\prime} & \Delta \boldsymbol{x}_t^{-\prime} & \ldots & \Delta \boldsymbol{x}_{t-q+1}^{-\prime} \end{array}\right]'.
\end{aligned}
$$

In addition, we define:

$$
\boldsymbol{\alpha}_* := \left[\begin{array}{c:c} \boldsymbol{\alpha}_{1*}' & \boldsymbol{\alpha}_{2*}' \end{array}\right]' := \left[\begin{array}{ccc:cccccccc} \rho_* & \boldsymbol{\theta}_*^{+\prime} & \boldsymbol{\theta}_*^{-\prime} & \gamma_* & \boldsymbol{\varphi}_*' & \boldsymbol{\pi}_{0*}^{+\prime} & \ldots & \boldsymbol{\pi}_{q-1*}^{+\prime} & \boldsymbol{\pi}_{0*}^{-\prime} & \ldots & \boldsymbol{\pi}_{q-1*}^{-\prime} \end{array}\right]',
$$

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

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

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)

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

(iii)

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{2t}\boldsymbol{z}_{2t}' \xrightarrow{\mathbb{P}} \mathbf{M}_{22} := \left[\begin{array}{c:ccc} 1 & \delta_*\boldsymbol{\iota}_{p-1}' & \boldsymbol{\iota}_q'\otimes\boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\iota}_q'\otimes\boldsymbol{\mu}_*^{-\prime} \\ \hdashline \delta_*\boldsymbol{\iota}_{p-1} & & & \\ \boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_*^{+} & & \mathbb{E}[\boldsymbol{w}_t\boldsymbol{w}_t'] & \\ \boldsymbol{\iota}_q\otimes\boldsymbol{\mu}_*^{-} & & & \end{array}\right],$$

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

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:

$$\mathbf{D}_T^{-1}\left(\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t'\right)\mathbf{D}_T^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_* := \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}, \tag{6}$$

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 $\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t'$, which is analytically challenging. In practice, a higher-order approximation of $\left(\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t'\right)^{-1}$ would be necessary to derive the limit distribution of the OLS estimator.

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 short-run parameters and the second on the estimation of the long-run parameters.

3.1 Estimation of the Short-Run Parameters

Suppose that the cointegrating coefficient is known or can be estimated by an estimator with a convergence rate faster than $T^{1/2}$. Specifically, let:

$$u_{t-1} := y_{t-1} - \boldsymbol{\beta}_*^{+\prime}\boldsymbol{x}_{t-1}^{+} - \boldsymbol{\beta}_*^{-\prime}\boldsymbol{x}_{t-1}^{-}, \tag{7}$$

where $\boldsymbol{\beta}_*^+ := -\boldsymbol{\theta}_*^+/\rho_*$ and $\boldsymbol{\beta}_*^- := -\boldsymbol{\theta}_*^-/\rho_*$. Assuming that $\boldsymbol{\beta}_*^+$ and $\boldsymbol{\beta}_*^-$ are known, we can re-write (2) as:

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^{+\prime} \Delta \boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_{j*}^{-\prime} \Delta \boldsymbol{x}_{t-j}^- \right) + e_t. \tag{8}$$

Note that all variables in (8) are stationary, so the unknown parameters can be estimated by OLS. If we define $\boldsymbol{\zeta}_* := (\rho_*, \boldsymbol{\beta}_{2*}')'$ and $\boldsymbol{h}_t := (u_t, \boldsymbol{z}_{2t}')'$, where $\boldsymbol{\beta}_{2*} := (\gamma_*, \varphi_{1*}, \ldots, \varphi_{p-1*}, \boldsymbol{\pi}_{0*}^{+\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{-\prime},)'$, then (8) can be re-written as:

$$\Delta y_t = \boldsymbol{\zeta}_*' \boldsymbol{h}_t + e_t,$$

and we can obtain the OLS estimator as follows:

$$\widehat{\boldsymbol{\zeta}}_T := \left(\sum_{t=1}^T \boldsymbol{h}_t \boldsymbol{h}_t' \right)^{-1} \left(\sum_{t=1}^T \boldsymbol{h}_t \Delta y_t \right) = \boldsymbol{\zeta}_* + \left(\sum_{t=1}^T \boldsymbol{h}_t \boldsymbol{h}_t' \right)^{-1} \left(\sum_{t=1}^T \boldsymbol{h}_t e_t \right). \tag{9}$$

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

Lemma 2. *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 addition, if $\mathbb{E}[e_t^2 | \boldsymbol{h}_t] = \sigma_*^2$, *then* $\boldsymbol{\Omega}_* = \sigma_*^2 \boldsymbol{\Gamma}_*$. □

We omit the proof of Lemma 2, because it is straightforward. Using Lemma 2, it is possible to derive the limit distribution of $\widehat{\boldsymbol{\zeta}}$, which is provided in the following theorem:

Theorem 1. *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 1 shows that, if there is any estimator converging to the cointegrating coefficient faster than $T^{1/2}$, then we can use the resulting parameter estimate as if it is known.

3.2 Estimation of the Long-Run Parameters

3.2.1 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

$$y_t = \alpha_* + \boldsymbol{\beta}_*^{+\prime}\boldsymbol{x}_t^+ + \boldsymbol{\beta}_*^{-\prime}\boldsymbol{x}_t^- + u_t. \tag{10}$$

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

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

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. We proceed by re-parameterizing (10) in the following form, which facilitates straightforward estimation of the long-run parameters:

$$y_t = \alpha_* + \boldsymbol{\lambda}_*'\boldsymbol{x}_t^+ + \boldsymbol{\eta}_*'\boldsymbol{x}_t + u_t, \tag{11}$$

where $\boldsymbol{x}_t \equiv \boldsymbol{x}_0 + \boldsymbol{x}_t^+ + \boldsymbol{x}_t^-$, $\boldsymbol{\lambda}_* = \boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-$ and $\boldsymbol{\eta}_* = \boldsymbol{\beta}_*^-$. It follows that $\boldsymbol{\beta}_*^+ = \boldsymbol{\lambda}_* + \boldsymbol{\eta}_*$ and $\boldsymbol{\beta}_*^- = \boldsymbol{\eta}_*$. It is possible to estimate $\boldsymbol{\varrho}_* := \left(\alpha_*, \boldsymbol{\lambda}_*', \boldsymbol{\eta}_*'\right)'$ by OLS as follows:

$$\widehat{\boldsymbol{\varrho}}_T := (\widehat{\alpha}_T', \widehat{\boldsymbol{\lambda}}_T', \widehat{\boldsymbol{\eta}}_T')' := \mathop{\arg\min}_{\alpha,\, \boldsymbol{\lambda},\, \boldsymbol{\eta}} \sum_{t=1}^{T}(y_t - \alpha - \boldsymbol{\lambda}'\boldsymbol{x}_t^+ - \boldsymbol{\eta}'\boldsymbol{x}_t)^2$$

where we can recover:

$$\widehat{\boldsymbol{\beta}}_T^+ := \widehat{\boldsymbol{\lambda}}_T + \widehat{\boldsymbol{\eta}}_T \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_T^- = \widehat{\boldsymbol{\eta}}_T.$$

Notice that $\left(\widehat{\boldsymbol{\beta}}_T^{+\prime}, \widehat{\boldsymbol{\beta}}_T^{-\prime}\right)$ is identical to the OLS estimator obtained by regressing y_t on $\left(1, \boldsymbol{x}_t^+, \boldsymbol{x}_t^-\right)$. Now, we have:

$$\widehat{\boldsymbol{\varrho}}_T = \boldsymbol{\varrho}_* + \left(\sum_{t=1}^{T}\boldsymbol{q}_t\boldsymbol{q}_t'\right)^{-1}\left(\sum_{t=1}^{T}\boldsymbol{q}_t u_t\right), \tag{12}$$

where $\boldsymbol{q}_t := \left(1, \boldsymbol{x}_t^{+\prime}, \boldsymbol{x}_t\right)'$. For the analysis of the components in (12), we define:

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

where $[\boldsymbol{\mathcal{W}}_x(\cdot)', \mathcal{W}_u]'$ is a $(k+1)\times 1$ vector of independent Wiener processes. If $\{u_t\}$ is serially uncorrelated and independent of $\{\Delta \boldsymbol{x}_t\}$, $\boldsymbol{\Sigma}_*$ simplifies to:

$$\boldsymbol{\Sigma}_* \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \mathbf{0} \\ \mathbf{0}' & \sigma_u^2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \boldsymbol{\mathcal{B}}_x(\cdot) \\ \mathcal{B}_u(\cdot) \end{bmatrix} := \begin{bmatrix} \boldsymbol{\Sigma}_{xx}^{1/2} \boldsymbol{\mathcal{W}}_x(\cdot) \\ \sigma_u \mathcal{W}_u(\cdot) \end{bmatrix},$$

where $\boldsymbol{\Sigma}_{xx} := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}[\Delta \boldsymbol{x}_t \Delta \boldsymbol{x}_s']$ and $\sigma_u^2 := \mathbb{E}[u_t^2]$. The following lemma provides the limit behaviors of the components constituting the OLS estimator:

Lemma 3. *Given Assumption 1:*

(i)

$$\widehat{\mathbf{Q}}_T := \widetilde{\mathbf{D}}_T^{-1} \left(\sum_{t=1}^{T} \boldsymbol{q}_t \boldsymbol{q}_t' \right) \widetilde{\mathbf{D}}_T^{-1} \Rightarrow \boldsymbol{\mathcal{Q}} := \begin{bmatrix} 1 & \frac{1}{2}\boldsymbol{\mu}_*^{+\prime} & \int_0^1 \boldsymbol{\mathcal{B}}_x(r)' dr \\ \frac{1}{2}\boldsymbol{\mu}_*^{+} & \frac{1}{3}\boldsymbol{\mu}_*^{+}\boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\mu}_*^{+} \int_0^1 r \boldsymbol{\mathcal{B}}_x(r)' dr \\ \int_0^1 \boldsymbol{\mathcal{B}}_x(r) dr & \int_0^1 r \boldsymbol{\mathcal{B}}_x(r) dr \boldsymbol{\mu}_*^{+\prime} & \int_0^1 \boldsymbol{\mathcal{B}}_x(r) \boldsymbol{\mathcal{B}}_x(r)' dr \end{bmatrix},$$

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

(ii) if $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{t} \mathbb{E}[\Delta \boldsymbol{x}_i u_t]$ *is finite, then:*

$$\widehat{\mathbf{U}}_T := \widetilde{\mathbf{D}}_T^{-1} \left(\sum_{t=1}^{T} \boldsymbol{q}_t u_t \right) \Rightarrow \boldsymbol{\mathcal{U}} := \begin{bmatrix} \int_0^1 d\mathcal{B}_u(r) \\ \boldsymbol{\mu}_*^{+} \int_0^1 r d\mathcal{B}_u(r) \\ \int_0^1 \boldsymbol{\mathcal{B}}_x(r) d\mathcal{B}_u(r) + \boldsymbol{\Lambda}_* \end{bmatrix},$$

where $\boldsymbol{\Lambda}_* := \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{t} \mathbb{E}[\Delta \boldsymbol{x}_i u_t]$*; and*

(iii) in the special case where $\{u_t\}$ *is serially uncorrelated and independent of* $\{\Delta \boldsymbol{x}_t\}$*, then* $\boldsymbol{\mathcal{U}}$ *simplifies*

to

$$\boldsymbol{\mathcal{U}}_0 := \sigma_u \begin{bmatrix} \int_0^1 d\mathcal{W}_u(r) \\ \boldsymbol{\mu}_*^+ \int_0^1 r d\mathcal{W}_u(r) \\ \boldsymbol{\Sigma}_{xx}^{1/2} \int_0^1 \boldsymbol{\mathcal{W}}_x(r) d\mathcal{W}_u(r) \end{bmatrix}. \qquad \square$$

Note that $\boldsymbol{\mathcal{Q}}$ is nonsingular with probability 1, so the limit distribution of $\widehat{\boldsymbol{\varrho}}_T$ is obtained as a product of $\boldsymbol{\mathcal{Q}}^{-1}$ and $\boldsymbol{\mathcal{U}}$, as stated in the following corollary:

Corollary 1. *Given Assumption 1,* $\widetilde{\mathbf{D}}_T(\widehat{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) \Rightarrow \boldsymbol{\mathcal{Q}}^{-1}\boldsymbol{\mathcal{U}}$. $\square$

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

Remarks.

(a) The convergence rates of $\widehat{\boldsymbol{\lambda}}_T$ and $\widehat{\boldsymbol{\eta}}_T$ are different; that is, $\widehat{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda}_* = O_{\mathbb{P}}(T^{-3/2})$ and $\widehat{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_* = O_{\mathbb{P}}(T^{-1})$.

(b) Using the definition of $\widehat{\boldsymbol{\lambda}}_T$, we have: $T\{(\widehat{\boldsymbol{\beta}}_T^+ - \widehat{\boldsymbol{\beta}}_T^-) - (\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-)\} = O_{\mathbb{P}}(T^{-1/2})$. This implies that:

$$T(\widehat{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+) = T(\widehat{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-) + o_{\mathbb{P}}(1) \tag{13}$$

such that the limit distributions of $T(\widehat{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)$ and $T(\widehat{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)$ are equivalent. Because the convergence rate of the long-run parameter estimator is faster than $T^{1/2}$, $\widehat{\boldsymbol{\beta}}_T^+$ and $\widehat{\boldsymbol{\beta}}_T^-$ can be treated as known when estimating the short-run dynamic parameters in the second step. $\square$

In the following theorem, we formally state the limit distributions of the OLS estimators of the long-run parameters:

Theorem 2. *Given Assumption 1,* $T[(\widehat{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)', (\widehat{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)']' \Rightarrow \boldsymbol{\iota}_2 \otimes \mathbf{S}\boldsymbol{\mathcal{Q}}^{-1}\boldsymbol{\mathcal{U}}$, *where* $\mathbf{S} := [\mathbf{0}_{k\times(1+k)}, \mathbf{I}_k]$. $\square$

3.2.2 First Step Estimation by FM-OLS

Note that the limit distribution in Theorem 2 is non-normal and depends on the nuisance parameters, $\boldsymbol{\Sigma}_*$ and $\boldsymbol{\Lambda}_*$. 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 where the $\{u_t\}$ is independent of $\{\Delta \boldsymbol{x}_t\}$ and/or serially uncorrelated, the OLS estimator of the long-run parameter exhibits an asymptotic bias determined by $\boldsymbol{\Lambda}_*$. The FM-OLS estimator developed by Phillips and Hansen (1990) overcomes these problems and it is free from asymptotic bias even in the presence of endogenous regressors and/or serial correlation. 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.

First, suppose that $\boldsymbol{\Sigma}_*$ can be consistently estimated by an autocorrelation consistent covariance matrix estimator. For example, the heteroskedasticity and autocorrelation consistent covariance matrix estimator of Newey and West (1987) can be applied as follows:

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

where $\omega_{\ell k} := 1 - k/(1+\ell)$, $\ell = O(T^{1/4})$ and $\widehat{u}_t := y_t - \widehat{\alpha}_T - \widehat{\boldsymbol{\beta}}_T^{+\prime} \boldsymbol{x}_t^+ - \widehat{\boldsymbol{\beta}}_T^{-\prime} \boldsymbol{x}_t^-$.

In addition, under mild regularity conditions, it is straightforward to show that the asymptotic bias, $\boldsymbol{\Lambda}_*$ in $\boldsymbol{\mathcal{U}}$ can be consistently estimated by the following estimator:

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

Now, define the following long-run parameter estimator:

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

where:

$$
\widetilde{y}_t := y_t - \Delta \boldsymbol{x}_t' \left(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)}\right)^{-1} \widetilde{\boldsymbol{\Sigma}}_T^{(1,2)} \quad \text{and} \quad \widetilde{\boldsymbol{\Lambda}}_T := \widetilde{\boldsymbol{\Pi}}_T^{(1,2)} - \widetilde{\boldsymbol{\Pi}}_T^{(1,1)} \left(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)}\right)^{-1} \widetilde{\boldsymbol{\Sigma}}_T^{(1,2)}.
$$

Finally, the FM-OLS estimators of the long-run parameters are obtained as follows:

$$\widetilde{\boldsymbol{\beta}}_T^+ := \widetilde{\boldsymbol{\lambda}}_T + \widetilde{\boldsymbol{\eta}}_T \quad \text{and} \quad \widetilde{\boldsymbol{\beta}}_T^- := \widetilde{\boldsymbol{\eta}}_T.$$

Note that these estimators are designed to remove the asymptotic bias as in Phillips and Hansen (1990). To derive the limiting distribution of the FM-OLS estimator, we add the following regularity conditions:

Assumption 2. *(i)* $\boldsymbol{\Sigma}_*$ *is finite and positive definite and* $\widetilde{\boldsymbol{\Sigma}}_T \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}_*$*; and*

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

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

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

Lemma 4. *Given Assumptions 1 and 2:*

$$\widetilde{\mathbf{D}}_T^{-1} \left\{ \left(\sum_{t=1}^{T} \boldsymbol{q}_t u_t \right) - \left(\sum_{t=1}^{T} \boldsymbol{q}_t \Delta \boldsymbol{x}_t' \right) \left(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)} \right)^{-1} \widetilde{\boldsymbol{\Sigma}}_T^{(1,2)} - T \mathbf{S}' \widetilde{\boldsymbol{\Lambda}}_T \right\} \Rightarrow \widetilde{\boldsymbol{\mathcal{U}}} := \tau_* \begin{bmatrix} \int_0^1 d\mathcal{W}_u(r) \\ \boldsymbol{\mu}_*^+ \int_0^1 r d\mathcal{W}_u(r) \\ \int_0^1 \boldsymbol{\mathcal{B}}_x(r) d\mathcal{W}_u(r) \end{bmatrix},$$

where $\tau_*^2 := plim_{T\to\infty} \widetilde{\tau}_T^2$ *and* $\widetilde{\tau}_T^2 := \widetilde{\sigma}_T^{(2,2)} - \widetilde{\boldsymbol{\Sigma}}_T^{(2,1)} (\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1} \widetilde{\boldsymbol{\Sigma}}_T^{(1,2)}$. $\square$

Given Lemma 3(i), $\widehat{\mathbf{Q}}_T \Rightarrow \boldsymbol{\mathcal{Q}}$, which is nonsingular with probability 1. The limit distribution of $\widetilde{\boldsymbol{\varrho}}_T$ can therefore be obtained as the product of $\boldsymbol{\mathcal{Q}}^{-1}$ and $\widetilde{\boldsymbol{\mathcal{U}}}$, as stated in the following corollary:

Corollary 2. *Given Assumption 1,* $\widetilde{\mathbf{D}}_T(\widetilde{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) \Rightarrow \boldsymbol{\mathcal{Q}}^{-1}\widetilde{\boldsymbol{\mathcal{U}}}$. $\square$

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

Remarks.

(a) The limit distribution of the FM-OLS estimator is mixed normal. Conditional on $\sigma\{\boldsymbol{\mathcal{B}}_x(r), r \in (0,1]\}$, the limit distribution of $\widetilde{\mathbf{D}}_T(\widetilde{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*)$ is $N(\mathbf{0}, \tau_*^2 \boldsymbol{\mathcal{Q}}^{-1})$. 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: $T(\widetilde{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+) = T(\widetilde{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-) + o_{\mathbb{P}}(1)$, such that the limit distribution of $\widetilde{\boldsymbol{\beta}}_T^+$ is equivalent to that of $\widetilde{\boldsymbol{\beta}}_T^-$. Furthermore, the limit distribution of $\widetilde{\boldsymbol{\beta}}_T^{-1}$ is given by that of $\widetilde{\boldsymbol{\eta}}_T$.

(c) The convergence rates of $\widetilde{\boldsymbol{\beta}}_T^+$ and $\widetilde{\boldsymbol{\beta}}_T^-$ 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 $\widetilde{u}_{t-1} := y_{t-1} - \widetilde{\alpha}_T - \widetilde{\boldsymbol{\beta}}_T^{+\prime} \boldsymbol{x}_{t-1}^+ - \widetilde{\boldsymbol{\beta}}_T^{-\prime} \boldsymbol{x}_{t-1}^-$. □

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

Theorem 3. *Given Assumptions 1 and 2,* $T[(\widetilde{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)', (\widetilde{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)']' \Rightarrow \boldsymbol{\iota}_2 \otimes \mathbf{S}\boldsymbol{\mathcal{Q}}^{-1}\widetilde{\boldsymbol{\mathcal{U}}}$. □

4 Hypotheses 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.

4.1 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:

$$H_0 : \mathbf{R}_s \boldsymbol{\zeta}_* = \mathbf{r} \quad \text{vs.} \quad H_1 : \mathbf{R}_s \boldsymbol{\zeta}_* \neq \mathbf{r},$$

where $\mathbf{R}_s \in \mathbb{R}^{r \times 1+p+2k}$, and $\mathbf{r} \in \mathbb{R}^r$ ($r \in \mathbb{N}$) are selection matrices. If we define $\mathbf{R}_s := [0'_{1+p}, \boldsymbol{\iota}_k, -\boldsymbol{\iota}_k]$ and $\mathbf{r} = \mathbf{0}'$, then we can test the null hypothesis of additive short-run symmetry against the alternative hypothesis of additive short-run asymmetry:[2]

$$H_0 : \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}^+ = \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}^- \quad \text{vs.} \quad H_1 : \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}^+ \neq \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}^-.$$

[2]Studies in the existing NARDL literature test several different forms of short-run symmetry restrictions, including the additive form that we consider here, as well as pairwise symmetry between $\boldsymbol{\pi}_{j*}^+$ and $\boldsymbol{\pi}_{j*}^-$ for $j = 0, \ldots, q-1$ (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 short-run symmetry restrictions by specifying appropriate selection matrices, $\mathbf{R}_s$ and $\mathbf{r}$.

We construct a Wald test statistic to test the above hypotheses as follows:

$$\mathcal{W}_T^{(s)} := T(\mathbf{R}_s\widehat{\boldsymbol{\zeta}}_T - \mathbf{r})'(\mathbf{R}_s\widehat{\boldsymbol{\Gamma}}_T^{-1}\widehat{\boldsymbol{\Omega}}_T\widehat{\boldsymbol{\Gamma}}_T^{-1}\mathbf{R}_s')^{-1}(\mathbf{R}_s\widehat{\boldsymbol{\zeta}}_T - \mathbf{r}),$$

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, as described above. Furthermore, if the condition in Lemma 2(*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 4, we establish that the null and alternative limit distributions of the Wald test statistic are standard.

Theorem 4. *Given Assumption 1, if $\boldsymbol{\Gamma}_*$ and $\boldsymbol{\Omega}_*$ are positive definite, then:*

(i) $\mathcal{W}_T^{(s)} \overset{A}{\sim} \mathcal{X}_r^2$ *under H_0; and*

(ii) for any sequence, c_T, such that $c_T = o(T)$, $\mathbb{P}(\mathcal{W}_T^{(s)} > c_T) \to 1$ under H_1. □

We omit the proof of Theorem 4 as it is straightforward.

4.2 Testing for Symmetry of the Long-Run Parameters

Consider the following hypotheses for the long-run parameters:

$$H_0' : \mathbf{R}_\ell(\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-) = \mathbf{r} \quad \text{vs.} \quad H_1' : \mathbf{R}_\ell(\boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-) \neq \mathbf{r},$$

for some $\mathbf{R}_\ell \in \mathbb{R}^{r\times k}$, $\mathbf{r} \in \mathbb{R}^r$ ($r \in \mathbb{N}$). By setting $\mathbf{R}_\ell = \mathbf{I}_k$ and $\mathbf{r} = \mathbf{0}_k$, we can test whether $\boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$. In models with multiple independent variables, we can also test the partial equality of $\boldsymbol{\beta}_*^+$ an $\boldsymbol{\beta}_*^-$ by selecting $\mathbf{R}_l$ and $\mathbf{r}$ appropriately.

Recall that $\boldsymbol{\lambda}_* := \boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-$ in (11). Consequently, we can restate H_0' as follows:

$$H_0'' : \mathbf{R}_\ell\boldsymbol{\lambda}_* = \mathbf{r} \quad \text{vs.} \quad H_1'' : \mathbf{R}_\ell\boldsymbol{\lambda}_* \neq \mathbf{r}.$$

Consequently, the long-run symmetry restriction, $\boldsymbol{\beta}_*^+ = \boldsymbol{\beta}_*^-$, is equivalent to the restriction that $\boldsymbol{\lambda}_* = \mathbf{0}$. It is straightforward to test this restriction if $\boldsymbol{\lambda}_*$ is estimated by FM-OLS, because the FM-OLS estimators of the long-run parameters are asymptotically mixed-normally distributed, so the Wald test statistic will follow an asymptotic chi-squared distribution. 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 $\widetilde{\boldsymbol{\lambda}}_T$. If we let $\mathbf{S}_\ell := [\mathbf{0}_{k\times 1}, \mathbf{I}_k, \mathbf{0}_{k\times k}]$, then $T^{3/2}(\widetilde{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda}_*) = \mathbf{S}_\ell \widetilde{\mathbf{D}}_T(\widetilde{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) \Rightarrow \mathbf{S}_\ell \boldsymbol{\mathcal{Q}}^{-1}\widetilde{\boldsymbol{\mathcal{U}}}$, implying that $T^{3/2}\mathbf{R}_\ell(\widetilde{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda}_*) \Rightarrow \mathbf{R}_\ell\mathbf{S}_\ell \boldsymbol{\mathcal{Q}}^{-1}\widetilde{\boldsymbol{\mathcal{U}}}$, so that $T^{3/2}(\mathbf{R}_\ell\widehat{\boldsymbol{\lambda}}_T - \mathbf{r}) \Rightarrow \mathbf{R}_\ell\mathbf{S}_\ell \boldsymbol{\mathcal{Q}}^{-1}\widetilde{\boldsymbol{\mathcal{U}}}$ under H_0''. The Wald test statistic is constructed in the usual manner:

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

Note that the Wald statistic above may be inappropriate to test other forms of hypothesis. For example, consider the following hypotheses:

$$H_0''' : \mathbf{R}\boldsymbol{\beta}_* = \mathbf{r} \quad \text{vs.} \quad H_1''' : \mathbf{R}\boldsymbol{\beta}_* \neq \mathbf{r},$$

for some $\mathbf{R} \in \mathbb{R}^{r\times 2k}$ and $\mathbf{r} \in \mathbb{R}^r$, where $\boldsymbol{\beta}_* := (\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$. Define:

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

such that these hypotheses can be rewritten as follows:

$$H_0''' : \widetilde{\mathbf{R}}\boldsymbol{\varrho}_* = \mathbf{r} \quad \text{vs.} \quad H_1''' : \widetilde{\mathbf{R}}\boldsymbol{\varrho}_* \neq \mathbf{r},$$

where we note that $\widetilde{\mathbf{R}}\boldsymbol{\varrho}_* = \boldsymbol{\beta}_*$ and $\widetilde{\mathbf{R}} := \mathbf{R}\widetilde{\mathbf{R}}_\ell$. In this case, we 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 5. *Given Assumptions 1 and 2:*

(i) $\mathcal{W}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_r^2$ *under* H_0'' *and* $\widetilde{\mathcal{W}}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_{2k}^2$ *under* H_0'''*; and*

(ii) 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'''. □

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

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. First, we study the finite sample bias and mean squared error (MSE) of the parameters estimated in two steps as described in Section 3, where the first step estimator is either OLS or FM-OLS and the second step estimator is OLS in both cases. We then examine the properties of the Wald test statistics developed in Section 4.

5.1 Finite Sample Performance of the Two-step Estimators

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

$$\Delta y_t = \gamma_* + \rho_* u_{t-1} + \varphi_* \Delta y_{t-1} + \pi_*^+ \Delta x_t^+ + \pi_*^- \Delta x_t^- + e_t, \tag{14}$$

where:

$$u_{t-1} := y_{t-1} - \alpha_* - \beta_*^+ x_{t-1}^+ - \beta_*^- x_{t-1}^-, \quad \Delta x_t := \kappa_* \Delta x_{t-1} + \sqrt{1-\kappa_*^2} v_t, \quad \text{and} \quad (e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2).$$

We set $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 2, 1, 0, -2/3, \varphi_*, 1, 1/2, 1/2)$. We allow the sample size, T, to vary over 100, 250, 500, 750 and 1,000 and we allow φ_* to vary over -0.50, -0.25, 0.00, 0.25 and 0.5 to introduce different degrees of serial correlation. Note that Δx_t is generated according to an AR(1) process with normally distributed disturbances and that u_t is both serially correlated and contemporaneously correlated with Δx_t.

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

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

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. Recall that the convergence rates of the long-run and short-run parameter estimators are T and $\sqrt{T}$, respectively. It is necessary to adjust for these different convergence rates in order to facilitate reliable comparisons of the finite sample bias and MSE of the long-run and short-run parameter estimators across different sample sizes. We therefore compute the asymptotic bias of $\widehat{\beta}_T^+$ and $\widehat{\varphi}_T$ as follows:

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

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 take account of the different convergence rates when calculating the finite sample MSE of $\widehat{\beta}_T^+$ and $\widehat{\varphi}_T$ as follows:

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

The finite sample bias and MSE of the estimated parameters based on $R = 5,000$ replications of the simulation experiments are recorded in Tables 1 and 2, respectively. To conserve space, we do not report the finite sample bias and MSE for either of the intercepts, α or γ; these results are available 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 first step FM-OLS estimator is typically substantially smaller than that of the first step OLS estimator. Recall that the first-stage 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 first step FM-OLS estimator is close to zero, because $T(\widehat{\beta}_T^+ - \beta_*^+)$ and $T(\widehat{\beta}_T^- - \beta_*^-)$ are asymptotically mixed-normally distributed around zero. By contrast, the first step OLS estimator is not asymptotically distributed around zero and exhibits non-negligible bias. In addition, our simulation results indicate that the first step FM-OLS estimator is often more efficient than the first step OLS estimator, resulting in a smaller MSE as the sample size increases. This tendency is particularly apparent for small and/or negative values of φ_*. Taken as a whole, these results strongly favor the use of the FM-OLS estimator in the first step.

Now consider the short-run parameter estimators obtained by OLS in the second step. We find that the finite sample bias and MSE of the second step estimators are similar irrespective of the choice to use either OLS or FM-OLS in the first step. In each case, the bias becomes negligible as the sample size increases. Even for a small sample of just 100 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.

5.2 Finite Sample Performance of the Wald Statistics

This section examines the finite sample performance of the Wald test statistics derived in Section 4.

5.2.1 Testing Restrictions on the Short-Run Parameters

To examine the empirical level properties of the Wald test statistic, we generate data using (14), with $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 2, 1, 0, -2/3, \varphi_*, 1/2, 1/2, 1/2)$ and allowing φ_* to vary over -0.50, -0.25, 0.00, 0.25 and 0.50, as in Section 5.1. 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:

$$H_0^{(s)} : \pi_*^+ - \pi_*^- = 0 \quad \text{versus} \quad H_1^{(s)} : \pi_*^+ - \pi_*^- \neq 0,$$

using $\mathcal{W}_T^{(s)}$ with the heteroskedasticity consistent covariance estimator $\widehat{\mathbf{\Omega}}_T$. The computed 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 3 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 level properties display little sensitivity to the value of φ_* even for moderate sample sizes.

— Insert Table 3 Here —

Next, we examine the empirical power properties of the Wald test statistic. For this exercise, we maintain the same hypotheses but we update the parameters of the DGP, 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, 0.00, 0.25 and 0.50 and we compute $\mathcal{W}_T^{(s)}$ using the heteroskedasticity and autocorrelation consistent covariance matrix estimator. The simulation results are reported in Table 4. 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 the parameter φ_*.[3]

— Insert Table 4 Here —

5.2.2 Testing Restrictions on the Long-Run Parameters

To evaluate the finite sample properties of the Wald test of long-run parameter restrictions, we confine our attention to the case where the FM-OLS estimator is used in the first step. We generate data using (14) 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. We test the following hypotheses:

$$H_0^{(\ell)} : \beta_*^+ - \beta_*^- = 0 \quad \text{versus} \quad H_1^{(\ell)} : \beta_*^+ - \beta_*^- \neq 0.$$

The simulation results reported in Table 5 reveal some mis-sizing in small samples, particularly for negative values of φ_*. However, 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. For sample sizes larger than 500, the approximation is relatively good. 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 5 Here —

To examine the empirical power properties of the Wald test statistic, we generate data from (14) 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, 0.00, 0.25 and 0.50, as before. The simulation results for $\mathcal{W}_T^{(\ell)}$ are reported in Table 6. We

[3] Although we do not report the results here, we also evaluated the empirical level and power properties of the Wald test statistic using OLS as the first step estimator. For small sample sizes, the Wald-test statistic obtained using FM-OLS in the first step is slightly better approximated by the chi-square distribution than its counterpart obtained using OLS in the first step, particularly when φ_* is small. The empirical power of the Wald test statistic is similar regardless of the choice of first step estimator. Detailed results are available from the authors on request.

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_*, indicating that the empirical power of the Wald test statistic is not adversely affected by serial correlation.

— Insert Table 6 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 per share, 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:

$$\Delta D_t = a_* - \zeta_* \left(D_t^* - D_{t-1} \right) + \epsilon_t, \tag{15}$$

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 may re-write Lintner's partial adjustment model in the following form:

$$\Delta D_t = a_* + \zeta_* D_{t-1} + \theta_* E_t + \epsilon_t, \tag{16}$$

where $\theta_* = -\zeta_* \beta_*$ by construction. As a linear and symmetric partial adjustment process, (16) 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:

$$E_t = E_0 + E_t^+ + E_t^-, \tag{17}$$

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 earnings news and negative earnings news, respectively. We may now re-write (16) as follows:

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

where $\theta_*^+ = -\zeta_* \beta_*^+$ and $\theta_*^- = -\zeta_* \beta_*^-$. Unit root testing reveals that D_t, E_t^+ and E_t^- are all first difference stationary time series, implying that there is an asymmetric cointegrating relation between these variables provided that their linear combination is stationary.[4] To account for serial correlation in ϵ_t, (18) may be embedded within a NARDL(p,q) model as follows:

$$\Delta D_t = \alpha_* + \zeta_*(D_{t-1} - \beta_*^+ E_{t-1}^+ - \beta_*^- E_{t-1}^-) + \sum_{j=1}^{p-1} \lambda_{j*} \Delta D_{t-j} + \sum_{j=0}^{q-1} d_{j*}^+ \Delta E_{t-j}^+ + \sum_{j=0}^{q-1} d_{j*}^- \Delta E_{t-j}^- + \varepsilon_t, \tag{19}$$

where the use of a sufficiently rich lag structure will ensure that ε_t is serially uncorrelated. Equation (19) can be estimated either by the single-step procedure advanced by SYG or by the two-step procedure that we propose above. 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.[5] 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 7, we report descriptive statistics for both the level and first difference of real earnings and real dividends. 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 displays 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 in this case. 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 7 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, using

[4] Unit root testing results are available from the authors on request.

[5] Shiller's dataset is available from http://www.econ.yale.edu/~shiller/data/ie_data.xls.

FM-OLS in the first step. In Table 8, 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 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 in this case. 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 8 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.[6]

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 over our sample period.

— 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 9 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 9 is striking.

— Insert Table 9 Here —

In practice, neither the single-step nor the two-step estimation results provide any 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.

[6]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, the NARDL model 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 has 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 the first step, the parameters of the transformed asymmetric long-run relationship are estimated 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 the FM-OLS estimator of Phillips and Hansen (1990) in the first step, because it accounts for serial correlation and potential endogeneity of the explanatory variables and it facilitates standard inference by virtue of its asymptotic mixed normality. 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 continue to hold to an acceptable degree 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.

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*) We note that:

$$\frac{1}{T^3}\sum_{t=1}^{T}\boldsymbol{z}_{1t}\boldsymbol{z}_{1t}' = \frac{1}{T^3}\sum_{t=1}^{T}\begin{bmatrix} y_{t-1}^2 & y_{t-1}\boldsymbol{x}_{t-1}^{+\prime} & y_{t-1}\boldsymbol{x}_{t-1}^{-\prime} \\ \boldsymbol{x}_{t-1}^{+}y_{t-1} & \boldsymbol{x}_{t-1}^{+}\boldsymbol{x}_{t-1}^{+\prime} & \boldsymbol{x}_{t-1}^{+}\boldsymbol{x}_{t-1}^{-\prime} \\ \boldsymbol{x}_{t-1}^{-}y_{t-1} & \boldsymbol{x}_{t-1}^{-}\boldsymbol{x}_{t-1}^{+\prime} & \boldsymbol{x}_{t-1}^{-}\boldsymbol{x}_{t-1}^{-\prime}\end{bmatrix}.$$

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*) We note that:

$$\frac{1}{T^2}\sum_{t=1}^{T}\boldsymbol{z}_{1t}\boldsymbol{z}_{2t}' \stackrel{\mathbb{P}}{\to} M_{12} = \frac{1}{T^2}\sum_{t=1}^{T}\begin{bmatrix} y_{t-1} & y_{t-1}\boldsymbol{w}_{1t}' & y_{t-1}\boldsymbol{w}_{2t}' & y_{t-1}\boldsymbol{w}_{3t}' \\ \boldsymbol{x}_{t-1}^{+} & \boldsymbol{x}_{t-1}^{+}\boldsymbol{w}_{1t}' & \boldsymbol{x}_{t-1}^{+}\boldsymbol{w}_{2t}' & \boldsymbol{x}_{t-1}^{+}\boldsymbol{w}_{3t}' \\ \boldsymbol{x}_{t-1}^{-} & \boldsymbol{x}_{t-1}^{-}\boldsymbol{w}_{1t}' & \boldsymbol{x}_{t-1}^{-}\boldsymbol{w}_{2t}' & \boldsymbol{x}_{t-1}^{-}\boldsymbol{w}_{3t}'\end{bmatrix}.$$

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:

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{2t}\boldsymbol{z}_{2t}' = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c|ccc} 1 & \boldsymbol{w}_{1t}' & \boldsymbol{w}_{2t}' & \boldsymbol{w}_{3t}' \\ \hline \boldsymbol{w}_{1t} & & & \\ \boldsymbol{w}_{2t} & & \boldsymbol{w}_t\boldsymbol{w}_t' & \\ \boldsymbol{w}_{3t} & & & \end{array}\right].$$

Given this, we note that:

- $T^{-1}\sum_{t=1}^{T} \boldsymbol{w}_{1t}' = \mathbb{E}[\Delta \boldsymbol{y}_{t-1}]' + o_{\mathbb{P}}(1) = \delta_*\boldsymbol{\iota}_{p-1}' + o_{\mathbb{P}}$ by the ergodic theorem;

- $T^{-1}\sum_{t=1}^{T} \boldsymbol{w}_{2t}' = [\mathbb{E}[\Delta\boldsymbol{x}_t^{+\prime}], \mathbb{E}[\Delta\boldsymbol{x}_{t-1}^{+\prime}], \ldots, \mathbb{E}[\Delta\boldsymbol{x}_{t-q+1}^{+\prime}]] + o_{\mathbb{P}}(1) = [\boldsymbol{\mu}_*^{+\prime}, \boldsymbol{\mu}_*^{+\prime}, \ldots, \boldsymbol{\mu}_*^{+\prime}] + o_{\mathbb{P}} = \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^{+}$ by the ergodic theorem;

- $T^{-1}\sum_{t=1}^{T} \boldsymbol{w}_{3t}' = [\mathbb{E}[\Delta\boldsymbol{x}_t^{-\prime}], \mathbb{E}[\Delta\boldsymbol{x}_{t-1}^{-\prime}], \ldots, \mathbb{E}[\Delta\boldsymbol{x}_{t-q+1}^{-\prime}]] + o_{\mathbb{P}}(1) = [\boldsymbol{\mu}_*^{-\prime}, \boldsymbol{\mu}_*^{-\prime}, \ldots, \boldsymbol{\mu}_*^{-\prime}] + o_{\mathbb{P}} = \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^{-}$ by the ergodic theorem; and

- $T^{-1}\sum_{t=1}^{T} \boldsymbol{w}_t\boldsymbol{w}_t' = \mathbb{E}[\boldsymbol{w}_t\boldsymbol{w}_t'] + o_{\mathbb{P}}(1)$ by the ergodic theorem.

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.

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

Proof of Theorem 1.

(i) Given (9), we can combine Lemmas 2 (i and ii) to obtain the desired result.

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

Proof of Lemma 3.

(i) We note that:

$$\widehat{\mathbf{Q}}_T = \begin{bmatrix} 1 & T^{-2}\sum_{t=1}^T \boldsymbol{x}_t^{+\prime} & T^{-3/2}\sum_{t=1}^T \boldsymbol{x}_t' \\ T^{-2}\sum_{t=1}^T \boldsymbol{x}_t^{+} & T^{-3}\sum_{t=1}^T \boldsymbol{x}_t^{+}\boldsymbol{x}_t^{+\prime} & T^{-5/2}\sum_{t=1}^T \boldsymbol{x}_t^{+}\boldsymbol{x}_t' \\ T^{-3/2}\sum_{t=1}^T \boldsymbol{x}_t & T^{-5/2}\sum_{t=1}^T \boldsymbol{x}_t\boldsymbol{x}_t^{+\prime} & T^{-2}\sum_{t=1}^T \boldsymbol{x}_t\boldsymbol{x}_t' \end{bmatrix}.$$

In addition, we note that:

- $T^{-2}\sum_{t=1}^T \boldsymbol{x}_t^{+} = T^{-1}\sum_{t=1}^T \boldsymbol{\mu}_*^{+}(t/T) + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} \frac{1}{2}\boldsymbol{\mu}_*^{+}$;
- $T^{-3/2}\sum_{t=1}^T \boldsymbol{x}_t = T^{-1}\sum_{t=1}^T (T^{-1/2}\sum_{i=1}^t \Delta\boldsymbol{x}_i) \Rightarrow \int_0^1 \boldsymbol{\mathcal{B}}_x(r)dr$ using that $T^{-1/2}\sum_{i=1}^{[T(\cdot)]}\Delta\boldsymbol{x}_i \Rightarrow \int_0^{(\cdot)} d\boldsymbol{\mathcal{B}}_x(r)$;
- $T^{-3}\sum_{t=1}^T \boldsymbol{x}_t^{+}\boldsymbol{x}_t^{+\prime} = T^{-1}\sum_{t=1}^T \boldsymbol{\mu}_*^{+}\boldsymbol{\mu}_*^{+\prime}(t/T)^2 + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} \frac{1}{3}\boldsymbol{\mu}_*^{+}\boldsymbol{\mu}_*^{+\prime}$;
- $T^{-5/2}\sum_{t=1}^T \boldsymbol{x}_t^{+}\boldsymbol{x}_t' = T^{-1}\sum_{t=1}^T \boldsymbol{\mu}_*^{+}(t/T)(T^{-1/2}\sum_{i=1}^t \Delta\boldsymbol{x}_t') + o_{\mathbb{P}}(1) \Rightarrow \boldsymbol{\mu}_*^{+}\int_0^1 r\boldsymbol{\mathcal{B}}_x(r)'dr$; and
- $T^{-2}\sum_{t=1}^T \boldsymbol{x}_t\boldsymbol{x}_t' = T^{-1}\sum_{t=1}^T (T^{-1/2}\sum_{i=1}^t \Delta\boldsymbol{x}_t)(T^{-1/2}\sum_{i=1}^t \Delta\boldsymbol{x}_t') \Rightarrow \int_0^1 \boldsymbol{\mathcal{B}}_x(r)\boldsymbol{\mathcal{B}}_x(r)'dr$.

Therefore, we obtain the following:

$$\widehat{\mathbf{Q}}_T \Rightarrow \begin{bmatrix} 1 & \frac{1}{2}\boldsymbol{\mu}_*^{+\prime} & \int_0^1 \boldsymbol{\mathcal{B}}_x(r)'dr \\ \frac{1}{2}\boldsymbol{\mu}_*^{+} & \frac{1}{3}\boldsymbol{\mu}_*^{+}\boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\mu}_*^{+}\int_0^1 r\boldsymbol{\mathcal{B}}_x(r)'dr \\ \int_0^1 \boldsymbol{\mathcal{B}}_x(r)dr & \int_0^1 r\boldsymbol{\mathcal{B}}_x(r)dr\boldsymbol{\mu}_*^{+\prime} & \int_0^1 \boldsymbol{\mathcal{B}}_x(r)\boldsymbol{\mathcal{B}}_x(r)'dr \end{bmatrix},$$

as desired.

(*ii*) We note that:

$$\widehat{\mathbf{U}}_T = \begin{bmatrix} T^{-1/2}\sum_{t=1}^T u_t \\ T^{-3/2}\sum_{t=1}^T \boldsymbol{x}_t^+ u_t \\ T^{-1}\sum_{t=1}^T \boldsymbol{x}_t u_t \end{bmatrix}.$$

In addition, we note that:

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

Therefore,

$$\widehat{\mathbf{U}}_T \Rightarrow \begin{bmatrix} \int_0^1 d\mathcal{B}_u(r) \\ \boldsymbol{\mu}_*^+ \int_0^1 r d\mathcal{B}_u(r) \\ \int_0^1 \boldsymbol{\mathcal{B}}_x(r) d\mathcal{B}_u(r) + \boldsymbol{\Lambda}_* \end{bmatrix}.$$

(*iii*) Given the conditions, $\mathcal{B}_u(\cdot) = \sigma_u \mathcal{W}_u(\cdot)$ and $\boldsymbol{\mathcal{B}}_x(\cdot) = \boldsymbol{\Sigma}_{xx}^{1/2} \boldsymbol{\mathcal{W}}_x(\cdot)$. Furthermore, it follows from the given condition that $\mathbb{E}[\Delta \boldsymbol{x}_t u_s] = \mathbf{0}$. The desired result follows in light of these additional properties. This completes the proof. ■

Proof of Corollary 1.

Given (12), the desired result follows from Lemma 2. ■

Proof of Theorem 2.

By (13), the weak limit of $T(\widehat{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)$ is equivalent to that of $T(\widehat{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)$. Furthermore, Corollary 1 implies that $T(\widehat{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-) \Rightarrow \mathbf{S}\boldsymbol{\mathcal{Q}}^{-1}\boldsymbol{\mathcal{U}}$, leading to the desired result. ■

Proof of Lemma 4.

Given Assumption 2, we note that $\widetilde{\boldsymbol{\Lambda}}_T \overset{\mathbb{P}}{\to} \boldsymbol{\Lambda}_*$ and $(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\widetilde{\boldsymbol{\Sigma}}_T^{(1,2)} \overset{\mathbb{P}}{\to} \boldsymbol{\nu}_* := (\widetilde{\boldsymbol{\Sigma}}_*^{(1,1)})^{-1}\widetilde{\boldsymbol{\Sigma}}_*^{(1,2)}$, where:

$$\boldsymbol{\Sigma}_* := \begin{bmatrix} \boldsymbol{\Sigma}_*^{(1,1)} & \boldsymbol{\Sigma}_*^{(1,2)} \\ \boldsymbol{\Sigma}_*^{(2,1)} & \sigma_*^{(2,2)} \end{bmatrix} := \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T \sum_{s=1}^T \begin{bmatrix} \mathbb{E}[\Delta \boldsymbol{x}_t \Delta \boldsymbol{x}_s'] & \mathbb{E}[\Delta \boldsymbol{x}_t u_s] \\ \mathbb{E}[u_t \Delta \boldsymbol{x}_s'] & \mathbb{E}[u_t u_s] \end{bmatrix}.$$

Therefore, if we let $\ddot{u}_t := u_t - \Delta\boldsymbol{x}_t'\boldsymbol{\nu}_*$:

$$\begin{aligned}
&\widetilde{\mathbf{D}}_T^{-1}\left\{\left(\sum_{t=1}^T \boldsymbol{q}_t u_t\right) - \left(\sum_{t=1}^T \boldsymbol{q}_t\Delta\boldsymbol{x}_t'\right)(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\widetilde{\boldsymbol{\Sigma}}_T^{(1,2)} - \mathbf{S}'\widetilde{\boldsymbol{\Lambda}}_T\right\} \\
&\quad = \widetilde{\mathbf{D}}_T^{-1}\left\{\left(\sum_{t=1}^T \boldsymbol{q}_t u_t\right) - \left(\sum_{t=1}^T \boldsymbol{q}_t\Delta\boldsymbol{x}_t'\right)\boldsymbol{\nu}_* - \mathbf{S}'\boldsymbol{\Lambda}_*\right\} + o_{\mathbb{P}}(1) = \widetilde{\mathbf{D}}_T^{-1}\sum_{t=1}^T\left\{\left(\boldsymbol{q}_t\ddot{u}_t - \mathbf{S}'\boldsymbol{\Lambda}_*\right)\right\} + o_{\mathbb{P}}(1),
\end{aligned}$$

implying that:

$$\widetilde{\mathbf{D}}_T^{-1}\sum_{t=1}^T\left\{\left(\boldsymbol{q}_t\ddot{u}_t - \mathbf{S}'\boldsymbol{\Lambda}_*\right)\right\} \Rightarrow \begin{bmatrix} \int_0^1 d\mathcal{B}_{\ddot{u}}(r) \\ \boldsymbol{\mu}_*^+\int_0^1 r d\mathcal{B}_{\ddot{u}}(r) \\ \int_0^1 \boldsymbol{\mathcal{B}}_x(r) d\mathcal{B}_{\ddot{u}}(r) \end{bmatrix},$$

where $\mathcal{B}_{\ddot{u}}(\cdot) := \tau_*\mathcal{W}_u(\cdot)$. Therefore:

$$\widetilde{\mathbf{D}}_T^{-1}\left\{\left(\sum_{t=1}^T \boldsymbol{q}_t u_t\right) - \left(\sum_{t=1}^T \boldsymbol{q}_t\Delta\boldsymbol{x}_t'\right)(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\widetilde{\boldsymbol{\Sigma}}_T^{(1,2)} - \mathbf{S}'\widetilde{\boldsymbol{\Lambda}}_T\right\} \Rightarrow \widetilde{\boldsymbol{\mathcal{U}}}.$$

This completes the proof. ■

Proof of Corollary 2.

Given that:

$$\begin{aligned}
&\widetilde{\mathbf{D}}_T(\widetilde{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) \\
&\quad = \left[\widetilde{\mathbf{D}}_T^{-1}\left(\sum_{t=1}^T \boldsymbol{q}_t\boldsymbol{q}_t\right)\widetilde{\mathbf{D}}_T^{-1}\right]^{-1}\widetilde{\mathbf{D}}_T^{-1}\left[\left(\sum_{t=1}^T \boldsymbol{q}_t u_t\right) - \left(\sum_{t=1}^T \boldsymbol{q}_t\Delta\boldsymbol{x}_t'\right)\left(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)}\right)^{-1}\widetilde{\boldsymbol{\Sigma}}_T^{(1,2)} - T\mathbf{S}'\widetilde{\boldsymbol{\Lambda}}_T\right],
\end{aligned}$$

the desired result follows from Lemmas 3(i) and 4. ■

Proof of Theorem 3.

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

Proof of Lemma 5.

We omit the proof due to its similarity to the proof of Lemma 1. ■

Proof of Theorem 4.

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

Proof of Theorem 5.

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

In addition, we note that $\widetilde{\mathcal{W}}_T^{(\ell)} = (\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r})'\widetilde{\mathbf{D}}_T(\widehat{\tau}_T^2\widetilde{\mathbf{R}}\widehat{\mathbf{Q}}_T^{-1}\widetilde{\mathbf{R}}')^{-1}\widetilde{\mathbf{D}}_T(\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r})$. Furthermore, Theorem 3 implies that $\widetilde{\mathbf{D}}_T(\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r}) \overset{\mathrm{A}}{\sim} N(\mathbf{0}, \tau_*^2\widetilde{\mathbf{R}}\boldsymbol{\mathcal{Q}}^{-1}\widetilde{\mathbf{R}}')$ conditional on $\sigma\{\boldsymbol{\mathcal{B}}_x(r) : r \in (0,1]\}$ under H_0'''. Given the condition that $\widehat{\mathbf{Q}}_T \Rightarrow \boldsymbol{\mathcal{Q}}$ and $\widehat{\tau}_T^2 \overset{\mathbb{P}}{\to} \tau_*^2$, it now follows that $\widetilde{\mathcal{W}}_T^{(\ell)} \overset{\mathrm{A}}{\sim} \mathcal{X}_{2k}^2$ under H_0'''.

(ii) 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, $(\widetilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_*) = O_{\mathbb{P}}(T^{-1})$, 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. ■

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 (11) 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 (7) with (3) will encounter a further singularity problem. Using the definitions of $\boldsymbol{\lambda}_* := \boldsymbol{\beta}_*^+ - \boldsymbol{\beta}_*^-$ and $\boldsymbol{\eta}_* := \boldsymbol{\beta}_*^-$, it follows that $u_{t-1} = y_{t-1} - \boldsymbol{\lambda}_*'\boldsymbol{x}_{t-1}^+ - \boldsymbol{\beta}_*'\boldsymbol{x}_{t-1}$, such that:

$$\Delta y_t = \rho_* y_{t-1} + (\boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-)'\boldsymbol{x}_{t-1}^+ + \boldsymbol{\theta}_*^{-\prime}\boldsymbol{x}_{t-1} + \gamma_* + \sum_{j=1}^{p-1}\varphi_{j*}\Delta y_{t-j} + \sum_{j=0}^{q-1}\left(\boldsymbol{\pi}_{j*}^{+\prime}\Delta\boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_{j*}^{-\prime}\Delta\boldsymbol{x}_{t-j}^-\right) + e_t, \tag{20}$$

where $\boldsymbol{\beta}_*^+ := -\boldsymbol{\theta}_*^+/\rho_*$ and $\boldsymbol{\beta}_*^- := -\boldsymbol{\theta}_*^-/\rho_*$. Let:

$$\boldsymbol{\xi}_* := \left[\begin{array}{c:c} \boldsymbol{\xi}_{1*}' & \boldsymbol{\xi}_{2*}' \end{array}\right]' := \left[\begin{array}{ccc:c} \rho_* & \boldsymbol{\theta}_*' & \boldsymbol{\theta}_*^{-\prime} & \boldsymbol{\alpha}_{2*}' \end{array}\right]' \quad \text{and}$$
$$\boldsymbol{p}_t := \left[\begin{array}{c:c} \boldsymbol{p}_{1t}' & \boldsymbol{p}_{2t}' \end{array}\right]' := \left[\begin{array}{ccc:c} y_{t-1} & \boldsymbol{x}_{t-1}^{+\prime} & \boldsymbol{x}_{t-1}' & \boldsymbol{z}_{2t}' \end{array}\right]'.$$

Note that $\boldsymbol{\xi}_{2*}$ and $\boldsymbol{p}_{2t}$ are identical to $\boldsymbol{\alpha}_{2*}$ and $\boldsymbol{z}_{2t}$, respectively, where $\boldsymbol{\theta}_* := \boldsymbol{\theta}_*^+ - \boldsymbol{\theta}_*^-$. If we attempt to estimate the vector of unknown parameters, $\boldsymbol{\xi}_*$, in (20) 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 5. *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*$[T^{3/2}\mathbf{I}_{1+k}, T\mathbf{I}_k]$ *and:*

$$\boldsymbol{\mathcal{P}}_{11} := \begin{bmatrix} \frac{1}{3}\delta_*^2 & \frac{1}{3}\delta_*\boldsymbol{\mu}_*^{+\prime} & \delta_*\int_0^1 r\boldsymbol{\mathcal{B}}_x(r)'dr \\ \frac{1}{3}\delta_*\boldsymbol{\mu}_*^{+} & \frac{1}{3}\boldsymbol{\mu}_*^{+}\boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\mu}_*^{+}\int_0^1 r\boldsymbol{\mathcal{B}}_x(r)'dr \\ \delta_*\int_0^1 r\boldsymbol{\mathcal{B}}_x(r)dr & \int_0^1 r\boldsymbol{\mathcal{B}}_x(r)dr\boldsymbol{\mu}_*^{+\prime} & \int_0^1 \boldsymbol{\mathcal{B}}_x(r)\boldsymbol{\mathcal{B}}_x(r)'dr \end{bmatrix};$$

(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*$[T^{1/2}\mathbf{I}_{1+p+2qk}]$ *and:*

$$\boldsymbol{\mathcal{P}}_{12} := \begin{bmatrix} \frac{1}{2}\delta_* & \frac{1}{2}\delta_*^2\boldsymbol{\iota}_{p-1}' & \frac{1}{2}\delta_*\boldsymbol{\iota}_q'\otimes\boldsymbol{\mu}_*^{+\prime} & \frac{1}{2}\delta_*\boldsymbol{\iota}_q'\otimes\boldsymbol{\mu}_*^{-\prime} \\ \frac{1}{2}\boldsymbol{\mu}_*^{+} & \frac{1}{2}\delta_*\boldsymbol{\mu}_*^{+}\boldsymbol{\iota}_{p-1}' & \frac{1}{2}\boldsymbol{\iota}_q'\otimes\boldsymbol{\mu}_*^{+}\boldsymbol{\mu}_*^{+\prime} & \frac{1}{2}\boldsymbol{\iota}_q'\otimes\boldsymbol{\mu}_*^{+}\boldsymbol{\mu}_*^{-\prime} \\ \int_0^1 \boldsymbol{\mathcal{B}}_x(r)dr & \delta_*\int_0^1 \boldsymbol{\mathcal{B}}_x(r)dr\boldsymbol{\iota}_{p-1}' & \boldsymbol{\iota}_q'\otimes\int_0^1 \boldsymbol{\mathcal{B}}_x(r)dr\boldsymbol{\mu}_*^{+\prime} & \boldsymbol{\iota}_q'\otimes\int_0^1 \boldsymbol{\mathcal{B}}_x(r)dr\boldsymbol{\mu}_*^{-\prime} \end{bmatrix}; \quad \textit{and}$$

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

We omit the proof of Lemma 5, as it can be easily derived from the proof of Lemma 1. Let $\ddot{\mathbf{D}}_T :=$ diag$[T^{3/2}\mathbf{I}_{1+k}, T\mathbf{I}_k, T^{1/2}\mathbf{I}_{1+p+2qk}]$, then:

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

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 (11).

	Sample Size	100		250		500		750		1,000	
	First Step	OLS	FM	OLS	FM	OLS	FM	OLS	FM	OLS	FM
φ_*	Second Step	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
-0.50	β_*^+	-13.95	-3.76	-14.54	-1.51	-14.74	0.68	-14.62	1.95	-14.67	2.19
	β_*^-	-13.97	-0.89	-14.55	-0.17	-14.76	1.45	-14.65	2.53	-14.65	2.62
	ρ_*	-0.36	-0.29	-0.19	-0.15	-0.11	-0.04	-0.08	-0.01	-0.08	-0.02
	φ_*	0.55	0.28	0.36	0.15	0.25	0.06	0.20	0.03	0.18	0.03
	π_*^+	-0.26	0.04	-0.16	0.10	-0.10	0.12	-0.09	0.12	-0.10	0.08
	π_*^-	-0.44	-0.08	-0.31	-0.07	-0.21	0.00	-0.21	-0.03	-0.14	0.03
-0.25	β_*^+	-9.84	-2.03	-10.11	-0.35	-10.20	1.05	-9.99	1.57	-10.22	1.65
	β_*^-	-9.89	0.18	-10.16	0.65	-10.18	1.59	-9.98	1.95	-10.18	1.96
	ρ_*	-0.30	-0.26	-0.17	-0.16	-0.10	-0.09	-0.07	-0.05	-0.07	-0.06
	φ_*	0.48	0.16	0.32	0.06	0.24	0.03	0.20	0.02	0.17	0.01
	π_*^+	-0.24	0.02	-0.14	0.07	-0.12	0.05	-0.11	0.04	-0.12	0.02
	π_*^-	-0.32	-0.05	-0.23	-0.04	-0.14	0.02	-0.11	0.02	-0.11	0.00
0.00	β_*^+	-5.68	-1.03	-5.42	0.40	-5.60	0.61	-5.39	0.82	-5.55	0.78
	β_*^-	-5.71	0.36	-5.45	1.07	-5.61	0.93	-5.40	1.05	-5.55	0.95
	ρ_*	-0.35	-0.34	-0.20	-0.19	-0.13	-0.13	-0.12	-0.12	-0.11	-0.11
	φ_*	0.28	0.07	0.17	0.00	0.13	0.00	0.12	0.00	0.08	-0.02
	π_*^+	-0.18	-0.06	-0.09	0.03	-0.07	0.00	-0.02	0.04	-0.02	0.04
	π_*^-	-0.26	-0.11	-0.14	-0.05	-0.11	-0.04	-0.14	-0.08	-0.10	-0.06
0.25	β_*^+	-0.94	-0.58	-0.88	-0.31	-0.88	-0.59	-0.77	-0.55	-0.80	-0.54
	β_*^-	-1.02	-0.30	-0.92	-0.16	-0.89	-0.52	-0.77	-0.49	-0.81	-0.51
	ρ_*	-0.35	-0.36	-0.22	-0.23	-0.16	-0.17	-0.11	-0.12	-0.11	-0.11
	φ_*	0.06	0.02	0.04	0.00	0.02	0.00	0.01	-0.01	0.01	0.00
	π_*^+	-0.04	-0.11	-0.02	-0.07	-0.05	-0.09	0.01	-0.03	0.01	-0.02
	π_*^-	-0.07	-0.15	-0.08	-0.13	0.01	-0.03	-0.02	-0.06	-0.01	-0.04
0.50	β_*^+	3.45	-2.31	3.58	-2.65	3.52	-3.40	3.62	-2.37	3.60	-2.42
	β_*^-	3.43	-3.47	3.53	-3.11	3.51	-3.65	3.61	-2.50	3.58	-2.52
	ρ_*	-0.22	-0.19	-0.13	-0.12	-0.10	-0.08	-0.07	-0.06	-0.07	-0.07
	φ_*	-0.09	0.02	-0.05	0.02	-0.04	0.02	-0.03	0.02	-0.04	0.00
	π_*^+	0.11	-0.21	0.08	-0.14	0.00	-0.16	0.06	-0.06	0.06	-0.04
	π_*^-	0.13	-0.21	0.07	-0.16	0.08	-0.08	0.03	-0.08	0.02	-0.07

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.

	Sample Size	100		250		500		750		1,000	
	First Step	OLS	FM	OLS	FM	OLS	FM	OLS	FM	OLS	FM
φ_*	Second Step	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
-0.50	β_*^+	290.07	92.10	308.75	47.24	316.53	33.89	309.60	34.38	317.14	32.37
	β_*^-	294.96	103.96	314.89	49.31	323.88	38.72	314.44	41.63	317.41	36.63
	ρ_*	0.64	0.62	0.42	0.40	0.35	0.34	0.34	0.33	0.33	0.32
	φ_*	0.63	0.38	0.39	0.26	0.30	0.23	0.27	0.22	0.27	0.23
	π_*^+	6.16	5.34	4.82	4.30	4.42	4.16	4.33	4.14	4.11	3.98
	π_*^-	6.17	5.32	4.93	4.31	4.53	4.11	4.33	4.04	4.18	3.95
-0.25	β_*^+	157.44	56.03	162.25	33.08	162.94	27.45	155.82	25.23	162.62	25.17
	β_*^-	162.01	72.54	164.97	37.75	163.31	30.82	154.90	27.96	161.96	27.41
	ρ_*	0.68	0.68	0.51	0.50	0.47	0.47	0.45	0.45	0.45	0.45
	φ_*	0.65	0.44	0.49	0.38	0.42	0.36	0.39	0.34	0.39	0.35
	π_*^+	5.48	5.08	4.48	4.26	4.20	4.09	4.26	4.15	4.09	4.00
	π_*^-	5.58	5.04	4.58	4.32	4.09	3.92	4.11	3.99	4.06	3.96
0.00	β_*^+	71.88	40.88	61.20	24.89	63.08	22.42	58.96	19.88	62.24	20.32
	β_*^-	72.55	46.21	62.50	29.14	63.37	23.75	58.57	21.07	61.75	21.10
	ρ_*	0.79	0.79	0.64	0.64	0.57	0.57	0.57	0.57	0.57	0.57
	φ_*	0.48	0.44	0.42	0.40	0.40	0.39	0.39	0.38	0.40	0.40
	π_*^+	4.96	4.81	4.40	4.31	4.01	3.96	4.23	4.22	4.10	4.08
	π_*^-	4.96	4.78	4.41	4.32	4.03	3.96	4.07	4.01	3.98	3.94
0.25	β_*^+	24.60	28.47	19.84	20.19	18.52	18.60	18.42	18.62	17.39	17.18
	β_*^-	25.14	36.22	20.41	21.39	18.71	18.75	18.18	18.48	17.24	17.01
	ρ_*	0.72	0.73	0.59	0.60	0.56	0.56	0.57	0.57	0.57	0.57
	φ_*	0.35	0.39	0.34	0.36	0.35	0.36	0.34	0.35	0.35	0.35
	π_*^+	4.65	4.64	4.24	4.24	4.14	4.15	4.02	4.02	4.12	4.12
	π_*^-	4.65	4.70	4.18	4.18	4.03	4.04	3.90	3.90	3.89	3.89
0.50	β_*^+	38.92	39.84	35.34	31.54	32.78	34.45	33.36	25.86	32.66	25.45
	β_*^-	41.36	71.65	34.80	38.13	33.06	38.17	33.71	27.18	32.68	26.25
	ρ_*	0.45	0.44	0.41	0.41	0.39	0.39	0.39	0.39	0.38	0.38
	φ_*	0.29	0.31	0.28	0.29	0.27	0.27	0.27	0.27	0.28	0.28
	π_*^+	4.54	4.58	4.18	4.19	4.18	4.20	3.94	3.93	3.93	3.93
	π_*^-	4.64	4.71	4.17	4.20	3.98	3.98	3.89	3.90	4.07	4.07

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 $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$. The simulation results are obtained using $R = 5,000$ replications.

φ_*	sample size	100	250	500	750	1,000
-0.50	1%	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 3: 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 $(e_t, v_t)' \sim$ 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 4: 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 $(e_t, v_t)' \sim$ 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%	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 5: 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 6: 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.

	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 7: 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 8: 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 9: 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.

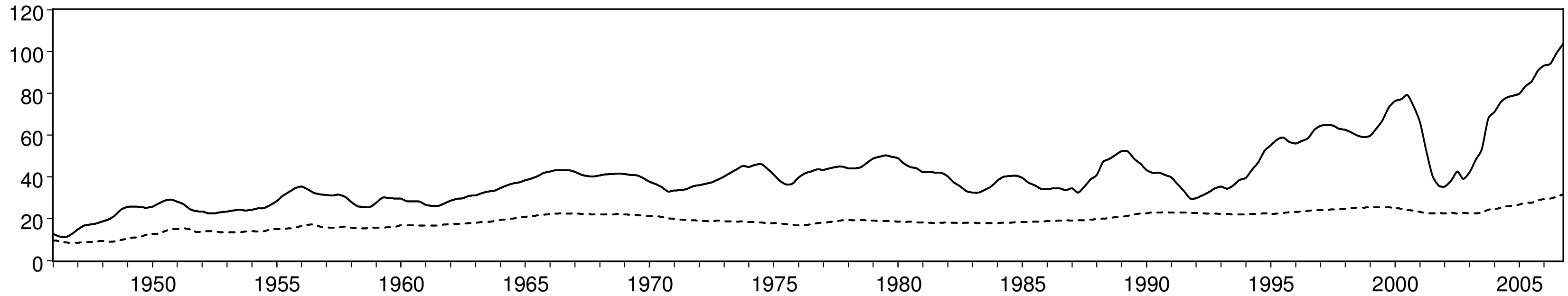

Figure 1: REAL EARNINGS VS. REAL DIVIDENDS. The solid line represents real earnings and the dashed line real dividends. Both series are measured in US Dollars at January 2000 prices. We convert from the original monthly sampling frequency to quarterly frequency by taking the end-of-period value.

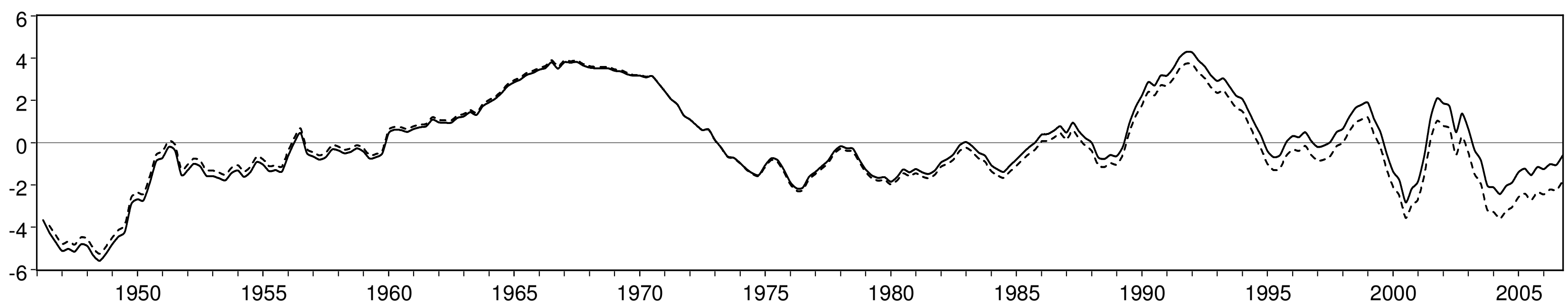

Figure 2: LONG-RUN DISEQUILIBRIUM ERROR. The long-run disequilibrium error obtained from the single-step estimation procedure is shown as a solid line. The values are obtained as $\xi_t = D_t - \beta^+ E_t^+ - \beta^- E_t^-$. In the figure, we de-mean ξ_t before we plot it. The long-run disequilibrium 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.