Two-Step Nonlinear ARDL Estimation: Theory and Application*

JIN SEO CHO
School of Economics, Yonsei University, Seoul, Korea
jinseocho@yonsei.ac.kr

MATTHEW GREENWOOD-NIMMO
Faculty of Business and Economics, University of Melbourne, Carlton, Australia
Centre for Applied Macroeconomic Analysis, Australian National University, Canberra, Australia
Codera Analytics, Johannesburg, GP 2193, South Africa
matthew.greenwood@unimelb.edu.au

YONGCHEOL SHIN
Department of Economics and Related Studies, University of York, York, U.K.
yongcheol.shin@york.ac.uk

March 2026

Abstract

The nonlinear autoregressive distributed lag (NARDL) model is a single-equation error correction model that captures asymmetric long-run and short-run relationships through partial sum decompositions of the explanatory variable(s). The NARDL model exhibits an asymptotic singularity issue that frustrates efforts to derive the asymptotic properties of the established single-step estimator. We propose a tractable two-step estimation procedure, and show that the two-step NARDL estimators follow a limiting normal distribution, the validity of which is confirmed by Monte Carlo simulations. We demonstrate the utility of our approach with an application to the asymmetric relationship between R&D intensity and investment in the U.S.

Key Words: Two-step Estimation of the NARDL Model, Asymptotic Singularity, Fully-Modified Estimator, Asymmetric Relationship between R&D Intensity and Investment.
JEL Classifications: C22, E22, O32.

*We are deeply grateful to the editor, Yuya Sasaki, an anonymous associate editor, and two anonymous referees for their insightful comments. We also benefited from valuable discussions with In Choi, Tae-Hwan Kim, Rui Lin, Viet Nguyen, Barbara Rossi, Michael Thornton and Ting Xie, as well as delegates at the 2025 World Congress of the Econometric Society (Seoul, 2025) and seminar participants at the Universities of Melbourne, Yonsei, and York. Cho is grateful for financial support from the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (Grant number NRF-2019S1A5A2A01035568). Greenwood-Nimmo and Shin acknowledge financial support from the Economic and Social Research Council (Grant number ES/T01573X/1). The views expressed herein are those of the authors and should not be attributed to Codera Analytics. The usual disclaimer applies.

Disclosure statement: The authors declare no conflicts of interest.

Data availability statement: The dataset used in our empirical analysis is publicly available at: http://www.greenwoodeconomics.com/RnD_Data.xlsx.

1 Introduction

The nonlinear autoregressive distributed lag (NARDL) model proposed by Shin et al. (2014) (hereafter SYG) is an asymmetric extension of the ARDL framework of Pesaran and Shin (1998) and Pesaran et al. (2001). It is a single-equation error-correction model designed to capture asymmetry in long-run equilibrium relationships and short-run dynamics through partial sum decomposition of the explanatory variables. The NARDL approach has been widely applied across diverse fields, including criminology (Box et al., 2019), economic growth (Eberhardt and Presbitero, 2015), energy economics (Hammoudeh et al., 2015), exchange rates and trade (Brun-Aguerre et al., 2017), financial economics (He and Zhou, 2018), health economics (Barati and Fariditavana, 2020), and political science (Ferris et al., 2020). For an extensive review, see Cho et al. (2023b). Despite its popularity, the theoretical foundations for estimation and inference in the NARDL model remain underdeveloped. 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). However, the theoretical properties of the single-step estimator remain unresolved. The challenge arises because the positive and negative partial sums of the regressors are dominated by deterministic trends that are asymptotically perfectly collinear. This collinearity creates an asymptotic singularity that represents a barrier to the development of asymptotic theory for the single-step estimator, frustrating efforts to derive its limit distribution.As a result, existing work relies on Monte Carlo simulations to assess finite-sample performance rather than providing a rigorous asymptotic framework. Developing the theoretical foundations for estimation of and inference on NARDL models is overdue given their widespread use in applied research.

To address this important issue, we first consider a bivariate model with a scalar dependent variable, y_t, and a scalar explanatory variable, x_t. In this case, the asymmetric long-run relationship is expressed among the level of the dependent variable and the positive and negative cumulative partial sums of the regressor, denoted x_t^+ and x_t^-, respectively, the latter of which share asymptotically collinear trends. The long-run relationship can be expressed equivalently via a one-to-one transformation as a relationship between y_t, x_t and x_t^+. By excluding one partial sum process, the asymptotic singularity in the long-run relationship is resolved. It is important to realize, however, that this reparameterization is insufficient to resolve the singularity problem associated with the single-step NARDL estimator; in fact, we show that it introduces a further asymptotic singularity problem, once again frustrating efforts to obtain the necessary limit theory.

To overcome the asymptotic singularity problem, we propose a two-step estimation framework, which we refer to as the two-step NARDL or 2SNARDL procedure. In the first step, we estimate the parameters

of the transformed long-run relationship using any consistent estimator that converges faster than the square root of the sample size, $\sqrt{T}$. We advocate the use of the fully modified (FM) estimator of Phillips and Hansen (1990), which follows an asymptotic mixed normal distribution that facilitates standard inference on the long-run parameters. Importantly, the FM estimator is robust to residual serial correlation, making it well-suited to dynamic settings. Furthermore, given the super-consistency of the FM estimator, the estimated error-correction term can be treated as known in the second step, where OLS provides consistent and asymptotically normal estimates of the short-run dynamic parameters. This two-step approach not only resolves the singularity issue but also allows us to establish a rigorous asymptotic framework for inference in NARDL models.

However, the two-step estimator described above cannot be directly applied to models with multiple explanatory variables. In such cases, a further singularity problem arises when estimating the reparameterized long-run equation due to collinearity among the trends in $\mathbf{x}_t^+$ and $\mathbf{x}_t$. To resolve this issue, we propose a modified two-step procedure: first, detrend $\mathbf{x}_t^+$ by OLS and use the residuals as regressors alongside a time trend and $\mathbf{x}_t$. This adjustment eliminates the singularity issue and enables consistent estimation of the long-run relationship in models with multiple regressors. The short-run parameters can then be estimated by OLS in the next step.

Because the NARDL model accommodates asymmetry in both long-run equilibrium and short-run dynamics, testing for symmetry is a key aspect of inference. To this end, we develop Wald tests for symmetry in the long-run parameters and in the short-run dynamic coefficients. We show that, under the null hypothesis, the Wald statistics converge weakly to a chi-squared distribution. Monte Carlo simulations confirm that these tests are well-sized and exhibit high power even in small samples, providing a reliable basis for empirical applications.

To demonstrate the utility of the two-step NARDL estimation framework, we study the asymmetric relationship between research and development (R&D) intensity and physical investment using quarterly U.S. data covering the period 1960Q1 to 2019Q4. Despite the extensive literature on innovation and growth (e.g., Romer, 1990; Agarwal and Audretsch, 2001; Aghion et al., 2009; Chung and Shin, 2020), the possibility of asymmetry in the relationship between R&D expenditure and investment has received little attention. This is surprising given that different types of R&D expenditure (e.g., establishing new products vs. streamlining production processes) are likely to have contrasting effects on physical investment decisions.

In the Online Supplement, we develop a theory that distinguishes between two stages of R&D activity. Early-stage *innovative* R&D creates new products or technologies and expands the scope of production. This type of R&D is expected to complement physical investment, as new technologies often require addi-

tional capital to scale production. Later-stage *managerial* R&D focuses on improving efficiency in existing production processes. This type of R&D may act as a substitute for physical investment, as efficiency gains reduce the need for additional capital. This framework gives rise to the theoretical prediction that innovative R&D expenditure is positively related to investment, while managerial R&D may exhibit a negative or weaker positive relationship.

Using our two-step NARDL approach, we estimate the asymmetric long-run and short-run effects of R&D intensity on investment. We find that, in the long run, investment responds positively to R&D expenditures when they grow faster than GDP, consistent with complementarity during periods of strong innovation. Conversely, when R&D grows more slowly than GDP, the long-run effect on investment is negative, reflecting substitution effects associated with managerial R&D. Furthermore, we find that investment is more sensitive to changes in R&D intensity when managerial R&D activity prevails.

These findings support the theoretical prediction that innovative R&D complements investment, while managerial R&D acts as a substitute. This has clear policy relevance. Measures that incentivize innovative R&D, such as tax credits for innovative technologies, are likely to stimulate capital formation. By contrast, policies that promote efficiency-oriented R&D may reduce investment. Understanding these asymmetries is essential to the design of effective innovation policies that balance technological progress with sustained physical investment.

This paper proceeds in 7 sections. In Section 2, we introduce the NARDL model and analyze the asymptotic singularity problem. Sections 3 and 4 introduce the two-step estimation framework and develop Wald tests for the null hypotheses of the short-run and long-run symmetry. In Section 5, we scrutinize the finite sample properties of the tests by simulation. Section 6 is devoted to our empirical application, and Section 7 concludes. Proofs of the main claims, further singularity issues associated with the one-step NARDL estimator, the theory relating early-stage innovative and later-stage managerial R&D expenditures to investment, and additional simulation/empirical results are relegated to the Online Supplement.

2 The NARDL Model

Consider the NARDL(p, q) process: for some $(\nu_*, \phi_{1*}, \ldots, \phi_{p*})' \in \mathbb{R}^{p+1}$ and $(\boldsymbol{\theta}_{j*}^{+\prime}, \boldsymbol{\theta}_{j*}^{-\prime})' \in \mathbb{R}^{2k}$ for $j = 1, 2, \ldots, q$, $(p, q, k \in \mathbb{N})$,

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

where $\boldsymbol{x}_t \in \mathbb{R}^k$, $\boldsymbol{x}_t^+ := \sum_{j=1}^t \Delta\boldsymbol{x}_j^+$, $\boldsymbol{x}_t^- := \sum_{j=1}^t \Delta\boldsymbol{x}_j^-$, $\Delta\boldsymbol{x}_t^+ := \max[\boldsymbol{0}, \Delta\boldsymbol{x}_t]$, and $\Delta\boldsymbol{x}_t^- := \min[\boldsymbol{0}, \Delta\boldsymbol{x}_t]$, such that $\Delta\boldsymbol{x}_t := \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$ is a stationary process. The corresponding error-correction model is given by

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

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

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

where $u_{t-1} := y_{t-1} - \alpha_* - \boldsymbol{\beta}_*^{+\prime}\boldsymbol{x}_{t-1}^+ - \boldsymbol{\beta}_*^{-\prime}\boldsymbol{x}_{t-1}^-$ is the cointegrating error, $\boldsymbol{\beta}_*^+ := -(\boldsymbol{\theta}_*^+/\rho_*)$ and $\boldsymbol{\beta}_*^- := -(\boldsymbol{\theta}_*^-/\rho_*)$. Here, u_{t-1} is a stationary process that may be correlated with $\Delta\boldsymbol{x}_t$, and α_* is introduced so that $\mathbb{E}[u_{t-1}] = 0$. Therefore, $\gamma_* - \rho_*\alpha_* = \nu_*$. We call $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$ the long-run parameters and the parameters in (2) the short-run parameters.

The NARDL process can capture a cointegrating relationship between a deterministic time trend process driven by a unit-root process and other unit-root processes, possibly associated with a time trend. Suppose that $\mathbb{E}[\Delta\boldsymbol{x}_t] = \boldsymbol{0}$ and define: $\boldsymbol{\mu}_*^+ := \mathbb{E}[\Delta\boldsymbol{x}_t^+]$ and $\boldsymbol{\mu}_*^- := \mathbb{E}[\Delta\boldsymbol{x}_t^-]$. By construction $\boldsymbol{\mu}_*^+ + \boldsymbol{\mu}_*^- \equiv \boldsymbol{0}$. If we further let $\boldsymbol{s}_t^+ := \Delta\boldsymbol{x}_t^+ - \boldsymbol{\mu}_*^+$ and $\boldsymbol{s}_t^- := \Delta\boldsymbol{x}_t^- - \boldsymbol{\mu}_*^-$, then $\boldsymbol{x}_t^+ = \boldsymbol{\mu}_*^+ t + \sum_{j=1}^t \boldsymbol{s}_j^+$ and $\boldsymbol{x}_t^- = \boldsymbol{\mu}_*^- t + \sum_{j=1}^t \boldsymbol{s}_j^-$. It is clear that $\boldsymbol{x}_t^+$ and $\boldsymbol{x}_t^-$ are deterministic time-trend processes driven by unit-root processes. Consequently, 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 the NARDL$(p, q)$ process, 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 $y_t = \delta_* t + \sum_{j=1}^t d_j$, which shows that y_t is a deterministic time-trend process driven by a unit-root process if $\delta_* \neq 0$. This has an important implication—the NARDL model can analyze an asymmetric cointegrating relationship between two integrated variables even with mismatched drifts without including a deterministic time trend in the model.

SYG propose to estimate the parameters of the NARDL model (1) using a single-step OLS procedure. However, as shown in Lemma 1, the single-step estimator suffers from the asymptotic singularity of the

inverse matrix associated with the NARDL parameter estimates.

We make the following assumptions:

Assumption 1. *(i)* $\{(\Delta \boldsymbol{x}_t', u_t)'\}$ *is the* $(k+1)\times 1$ *vector of globally covariance stationary* ϕ*-mixing processes of size* $-r/(2(r-1))$ *or* α*-mixing processes 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 (PD); and (iv) for some* $(\rho_*, \boldsymbol{\theta}_*^{+\prime}, \boldsymbol{\theta}_*^{-\prime}, \nu_*, \varphi_{1*}, \ldots, \varphi_{p-1*}, \boldsymbol{\pi}_{0*}^{+\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{-\prime})'$*,* Δy_t *is generated by* (1) *such that* $\{e_t, \mathcal{F}_t\}$ *is a martingale difference sequence and* $\mathcal{F}_t$ *is the smallest* σ*-algebra driven by* $\{y_{t-1}, \boldsymbol{x}_t^+, \boldsymbol{x}_t^-, y_{t-2}, \boldsymbol{x}_{t-1}^+, \boldsymbol{x}_{t-1}^-, \ldots\}$. □

Let $\boldsymbol{z}_t := [\boldsymbol{z}_{1t}' \vdots \boldsymbol{z}_{2t}']' := [y_{t-1}, \boldsymbol{x}_{t-1}^{+\prime}, \boldsymbol{x}_{t-1}^{-\prime} \vdots 1, \Delta \boldsymbol{y}_{t-1}', \Delta \boldsymbol{x}_t^{+\prime}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{+\prime}, \Delta \boldsymbol{x}_t^{-\prime}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{-\prime}]'$, 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. $\boldsymbol{z}_{2t}$ is further partitioned as $\boldsymbol{z}_{2t} := [1 \vdots \boldsymbol{w}_t']' := [1 \vdots \boldsymbol{w}_{1t}' \vdots \boldsymbol{w}_{2t}' \vdots \boldsymbol{w}_{3t}']' := [1 \vdots \Delta \boldsymbol{y}_{t-1}' \vdots \Delta \boldsymbol{x}_t^{+\prime}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{+\prime} \vdots \Delta \boldsymbol{x}_t^{-\prime}, \ldots, \Delta \boldsymbol{x}_{t-q+1}^{-\prime}]'$. Next, we define $\boldsymbol{\alpha}_* := [\boldsymbol{\alpha}_{1*}' \vdots \boldsymbol{\alpha}_{2*}']' := [\rho_*, \boldsymbol{\theta}_*^{+\prime}, \boldsymbol{\theta}_*^{-\prime} \vdots \nu_*, \boldsymbol{\varphi}_*', \boldsymbol{\pi}_{0*}^{+\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{-\prime}]'$, where $\boldsymbol{\varphi}_* := [\varphi_{1*}, \varphi_{2*}, \ldots, \varphi_{p-1*}]'$. Then, the OLS estimator is:

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

Inference using $\widehat{\boldsymbol{\alpha}}_T$ is challenging because $\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t'$ is asymptotically singular, as shown in Lemma 1.

Lemma 1. *Under Assumption 1, (i)* $T^{-3}\sum_{t=1}^{T} \boldsymbol{z}_{t1}\boldsymbol{z}_{t1}' \xrightarrow{\mathbb{P}} \mathbf{M}_{11} := \frac{1}{3}\boldsymbol{n}_1\boldsymbol{n}_1'$*; (ii)* $T^{-2}\sum_{t=1}^{T} \boldsymbol{z}_{1t}\ \boldsymbol{z}_{2t}' \xrightarrow{\mathbb{P}} \mathbf{M}_{12} := \frac{1}{2}\boldsymbol{n}_1\boldsymbol{n}_2'$*; and (iii)* $T^{-1}\sum_{t=1}^{T} \boldsymbol{z}_{2t}\boldsymbol{z}_{2t}' \xrightarrow{\mathbb{P}} \mathbf{M}_{22}$*, where*

$$\mathbf{M}_{22} := \begin{bmatrix} 1 & \boldsymbol{n}_2^{(1)\prime} \\ \boldsymbol{n}_2^{(1)} & \mathbb{E}[\boldsymbol{w}_t\boldsymbol{w}_t'] \end{bmatrix},$$

$\boldsymbol{n}_1 := [\delta_*, \boldsymbol{\mu}_*^{+\prime}, \boldsymbol{\mu}_*^{-\prime}]'$*,* $\boldsymbol{n}_2 := [1, \boldsymbol{n}_2^{(1)}]$ *and* $\boldsymbol{n}_2^{(1)} := [\delta_*\boldsymbol{\iota}_{p-1}', \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^{+\prime}, \boldsymbol{\iota}_q' \otimes \boldsymbol{\mu}_*^{-\prime}]'$ *with* $\boldsymbol{\iota}_a$ *being an* $a \times 1$ *vector of ones.* □

Lemma 1 implies that, if we define the scaling matrix $\mathbf{D}_T := \text{diag}[T^{3/2}\mathbf{I}_{2+2k}, T^{1/2}\mathbf{I}_{p+2qk}]$, then $\mathbf{D}_T^{-1}(\sum_{t=1}^{T} \boldsymbol{z}_t\boldsymbol{z}_t')\ \mathbf{D}_T^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_*$, where $\mathbf{M}_*$ is 2×2 block matrix, whose i-th row and j-th column block is $\mathbf{M}_{ij}$ and $\mathbf{M}_{21} = \mathbf{M}_{12}'$. Therefore, $[\mathbf{M}_{11}, \mathbf{M}_{12}] = \boldsymbol{n}_1[\frac{1}{3}\boldsymbol{n}_1', \frac{1}{2}\boldsymbol{n}_2']$. This shows that the first-row block sub-matrices can be expressed as the product of the column and row vectors. implying that $\mathbf{M}$ is singular. Due to this singularity, it is difficult to derive the limit distribution of $\widehat{\boldsymbol{\alpha}}_T$ directly, unless we can derive the limit distribution of the determinant of $\sum_{t=1}^{T} \boldsymbol{z}_t\boldsymbol{z}_t'$, which is analytically challenging.

3 Asymptotic Theory for the Two-step NARDL Estimator

We propose an analytically tractable two-step estimation procedure, building on the approaches of Engle and Granger (1987) and Phillips and Hansen (1990), and derive the corresponding limit distributions. For clarity, we present separate treatments for the cases $k = 1$ and $k > 1$.

3.1 The Two-step NARDL Estimation with $k = 1$

3.1.1 Estimation of the Long-Run Parameters

First Step Estimation by OLS. Recall that the long-run relationship is written as $y_t = \alpha_* + \beta_*^+ x_t^+ + \beta_*^- x_t^- + u_t$. Following the two-step estimation framework of Engle and Granger (1987), we may estimate the long-run parameters by OLS. Define: $\bar{\mathbf{D}}_T := \text{diag}[T^{1/2}, T^{3/2}\mathbf{I}_2]$ and $\boldsymbol{v}_t := (1, x_t^+, x_t^-)'$ such that $\bar{\mathbf{D}}_T^{-1}\left(\sum_{t=1}^T \boldsymbol{v}_t\boldsymbol{v}_t'\right)\bar{\mathbf{D}}_T^{-1} \xrightarrow{\mathbb{P}} \mathbf{M}_{11}$. By Lemma 1(i), this limit is singular due to collinear trends in x_t^+ and x_t^-, which complicates the derivation of the OLS limit distribution.

To address this, we reparameterize the long-run relationship as: $y_t = \alpha_* + \lambda_* x_t^+ + \eta_* x_t + u_t$, where $x_t \equiv x_t^+ + x_t^-$, $\lambda_* = \beta_*^+ - \beta_*^-$ and $\eta_* = \beta_*^-$. It follows that $\beta_*^+ = \lambda_* + \eta_*$ and $\beta_*^- = \eta_*$. The OLS estimator of $\boldsymbol{\varrho}_* := (\alpha_*, \lambda_*, \eta_*)'$ is: $\widehat{\boldsymbol{\varrho}}_T := (\widehat{\alpha}_T, \widehat{\lambda}_T, \widehat{\eta}_T)' := \arg\min_{\alpha,\,\lambda,\,\eta} \sum_{t=1}^T (y_t - \alpha - \lambda x_t^+ - \eta x_t)^2$ from which we recover $\widehat{\beta}_T^+ := \widehat{\lambda}_T + \widehat{\eta}_T$ and $\widehat{\beta}_T^- = \widehat{\eta}_T$.

Let $\boldsymbol{q}_t := (1, x_t^+, x_t)'$. Then, $\widehat{\boldsymbol{\varrho}}_T = \boldsymbol{\varrho}_* + (\sum_{t=1}^T \boldsymbol{q}_t\boldsymbol{q}_t')^{-1}(\sum_{t=1}^T \boldsymbol{q}_t u_t)$. Define

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

where $\boldsymbol{g}_t := [\Delta x_t, u_t]'$ and $[\mathcal{W}_x(\cdot), \mathcal{W}_u(\cdot)]'$ is a 2×1 vector of independent Wiener processes. Let $\sigma_*^{(i,j)}$ be the i-th row and j-th column element of $\boldsymbol{\Sigma}_*$. If $\{u_t\}$ is serially uncorrelated and independent of $\{\Delta x_t\}$, then $\boldsymbol{\Sigma}_*$ and $[\mathcal{B}_x(\cdot), \mathcal{B}_u(\cdot)]'$ are simplified to diag $\left[\sigma_x^2, \sigma_u^2\right]$ and $[\sigma_x\mathcal{W}_x(\cdot), \sigma_u\mathcal{W}_u(\cdot)]'$, respectively, where $\sigma_x^2 := \lim_{T\to\infty} T^{-1}\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}[\Delta x_t \Delta x_s]$ and $\sigma_u^2 := \mathbb{E}[u_t^2]$.

Lemma 2 provides the limit behaviors of the components of the OLS estimator.

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

(i) Define $\widehat{\mathbf{Q}}_T := \widetilde{\mathbf{D}}_T^{-1}\left(\sum_{t=1}^T \boldsymbol{q}_t\boldsymbol{q}_t'\right)\widetilde{\mathbf{D}}_T^{-1}$ *and* $\widetilde{\mathbf{D}}_T := diag[T^{1/2}, T^{3/2}, T]$. *Then,*

$$\widehat{\mathbf{Q}}_T \Rightarrow \boldsymbol{\mathcal{Q}} := \begin{bmatrix} 1 & \frac{1}{2}\mu_*^+ & \int \mathcal{B}_x \\ \frac{1}{2}\mu_*^+ & \frac{1}{3}\mu_*^{+2} & \mu_*^+ \int r\mathcal{B}_x \\ \int \mathcal{B}_x & \mu_*^+ \int r\mathcal{B}_x & \int \mathcal{B}_x\mathcal{B}_x \end{bmatrix}.$$

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

$$\widehat{\mathbf{U}}_T := \widetilde{\mathbf{D}}_T^{-1}\left(\sum_{t=1}^T \boldsymbol{q}_t u_t\right) \Rightarrow \boldsymbol{\mathcal{U}} := \left[\int d\mathcal{B}_u,\ \mu_*^+ \int r\, d\mathcal{B}_u,\ \int \mathcal{B}_x\, d\mathcal{B}_u + \upsilon_*\right].$$

Lemma 2 can be compared with the case where y_t is regressed directly on (x_t^+, x_t^-). Under Assumption 1, we have: $x_t^+ = \mu_*^+ t + \sum_{j=1}^t s_j^+$ and $x_t^- = \mu_*^- t + \sum_{j=1}^t s_j^-$, where $\sum_{j=1}^t s_j^+ = O_{\mathbb{P}}(\sqrt{n})$ and $\sum_{j=1}^t s_j^- = O_{\mathbb{P}}(\sqrt{n})$, while $t = O(n)$. Thus, the dominant term in x_t^+ and x_t^- is linear in t. Consequently

$$\frac{1}{n^3}\sum_{t=1}^n \begin{bmatrix} x_t^{+2} & x_t^+x_t^- \\ x_t^+x_t^- & x_t^{-2} \end{bmatrix} = \frac{1}{3}\begin{bmatrix} \mu_*^{+2} & \mu_*^+\mu_*^- \\ \mu_*^+\mu_*^- & \mu_*^{-2} \end{bmatrix} + o_{\mathbb{P}}(1) = \frac{1}{3}\begin{bmatrix} \mu_*^+ \\ \mu_*^- \end{bmatrix}\begin{bmatrix} \mu_*^+ & \mu_*^- \end{bmatrix} + o_{\mathbb{P}}(1),$$

which is asymptotically singular. In contrast, note that $x_t^+ + x_t^- = x_t = \sum_{j=1}^t \Delta x_j$, which is $O_{\mathbb{P}}(\sqrt{n})$ and therefore has a different convergence rate than x_t^+ and x_t^-. By the functional central limit theorem, $n^{-1/2}x_{\lfloor n(\cdot)\rfloor} \Rightarrow \mathcal{B}_x(\cdot)$, so that

$$\begin{bmatrix} n^{-3}\sum_{t=1}^n x_t^{+2} & n^{-5/2}\sum_{t=1}^n x_t^+x_t \\ n^{-5/2}\sum_{t=1}^n x_t^+x_t & n^{-2}\sum_{t=1}^n x_t^2 \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{1}{3}\mu_*^{+2} & \mu_*^+\int_0^1 r\mathcal{B}_x(r)dr \\ \mu_*^+\int_0^1 r\mathcal{B}_x(r)dr & \int_0^1 \mathcal{B}_x^2(r)dr \end{bmatrix}.$$

Unlike the previous case, this limit cannot be expressed as the outer product of two vectors. This difference highlights that the proposed reparameterization avoids the asymptotic singularity inherent in the original specification.

As $\boldsymbol{\mathcal{Q}}$ is nonsingular with probability one, the limit distribution of $\widehat{\boldsymbol{\varrho}}_T$ is obtained as the product of $\boldsymbol{\mathcal{Q}}^{-1}$ and $\boldsymbol{\mathcal{U}}$, as stated in Corollary 1.

[1] All integrals are taken over $r \in [0,1]$. For example, $\int r\mathcal{B}_x$ denotes $\int_0^1 r\mathcal{B}_x(r)\, dr$.

Corollary 1. *Under Assumption 1, if $k = 1$ and $\boldsymbol{\Sigma}_*$ is positive definite, then, $\widetilde{\mathbf{D}}_T(\widehat{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) \Rightarrow \boldsymbol{\mathcal{Q}}^{-1}\boldsymbol{\mathcal{U}}$.*[2] □

Corollary 1 has important implications. First, by reparameterization, the collinearity between x_t^+ and x_t^- is avoided because $\sum_{t=1}^T x_t = O_{\mathbb{P}}(T^{3/2})$ and $\sum_{t=1}^T x_t^+ = O_{\mathbb{P}}(T^2)$ yield different convergence rates for $\widehat{\lambda}_T$ and $\widehat{\eta}_T$, namely, $\widehat{\lambda}_T - \lambda_* = O_{\mathbb{P}}(T^{-3/2})$ and $\widehat{\eta}_T - \eta_* = O_{\mathbb{P}}(T^{-1})$. Second, to derive the limit distribution of $\widehat{\boldsymbol{\varrho}}_T$, we apply the functional central limit theorem only to $\sum_{t=1}^{T(\cdot)} \boldsymbol{g}_t$, where $\boldsymbol{g}_t := [\Delta x_t, u_t]'$, rather than to $\sum_{t=1}^{[T(\cdot)]}(x_t^+ - \mu_*^+)$. Third, using the definition of $\widehat{\lambda}_T$, we have: $T\{(\widehat{\beta}_T^+ - \widehat{\beta}_T^-) - (\beta_*^+ - \beta_*^-)\} = O_{\mathbb{P}}(T^{-1/2})$, implying that the limit distributions of $T(\widehat{\beta}_T^+ - \beta_*^+)$ and $T(\widehat{\beta}_T^- - \beta_*^-)$ are asymptotically equivalent. Finally, as the long-run parameter estimator is super-consistent (converging faster than $T^{1/2}$), $\widehat{\beta}_T^+$ and $\widehat{\beta}_T^-$ can be treated as given when estimating short-run dynamic parameters.

Theorem 1 presents the limit distribution of the OLS estimator of the long-run parameters.

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

The limit distribution in Theorem 1 is non-normal due to its dependence on the nuisance parameters $\boldsymbol{\Sigma}_*$ and υ_*. Consequently, the OLS estimator of the long-run parameters is not conducive to standard inference. Except in the special case where $\{u_t\}$ is independent of $\{\Delta x_t\}$ and/or serially uncorrelated, the OLS estimator exhibits an asymptotic bias determined by υ_*.

First Step Estimation by FM. Phillips and Hansen (1990) propose the fully modified (FM) estimator, which is free from asymptotic bias even in the presence of serial correlation, and follows an asymptotic mixed normal distribution. Hence, we advocate the use of FM to estimate the long-run parameters in the first step.

Suppose that $\boldsymbol{\Sigma}_*$ is consistently estimated by a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator, $\widetilde{\boldsymbol{\Sigma}}_T$, with $\widetilde{\sigma}_T^{(i,j)}$ being the (i,j)-th element. For example, following Newey and

[2] Alternatively, the same limit distribution can be obtained using a rotation matrix:

$$\mathbf{A} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad \text{so that} \quad (\mathbf{A}')^{-1} \begin{pmatrix} \widehat{\alpha}_T - \alpha_* \\ \widehat{\beta}_*^+ - \beta_*^+ \\ \widehat{\beta}_*^- - \beta_*^- \end{pmatrix} = \begin{pmatrix} \widehat{\alpha}_T - \alpha_* \\ (\widehat{\beta}_*^+ - \beta_*^+) - (\widehat{\beta}_*^- - \beta_*^-) \\ \widehat{\beta}_*^- - \beta_*^- \end{pmatrix} = \widehat{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*.$$

Then,

$$\widetilde{\mathbf{D}}_T(\widehat{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) = \widetilde{\mathbf{D}}_T(\mathbf{A}')^{-1} \begin{pmatrix} \widehat{\alpha}_T - \alpha_* \\ \widehat{\beta}_*^+ - \beta_*^+ \\ \widehat{\beta}_*^- - \beta_*^- \end{pmatrix} = \left(\widetilde{\mathbf{D}}_T^{-1} \sum_{t=1}^T \mathbf{A}\boldsymbol{v}_t\boldsymbol{v}_t'\mathbf{A}'\widetilde{\mathbf{D}}_T^{-1}\right)^{-1} \left(\widetilde{\mathbf{D}}_T^{-1} \sum_{t=1}^T \mathbf{A}\boldsymbol{v}_t u_t\right)$$
$$= \left(\widetilde{\mathbf{D}}_T^{-1} \sum_{t=1}^T \boldsymbol{q}_t\boldsymbol{q}_t'\widetilde{\mathbf{D}}_T^{-1}\right)^{-1} \left(\widetilde{\mathbf{D}}_T^{-1} \sum_{t=1}^T \boldsymbol{q}_t u_t\right) \Rightarrow \boldsymbol{\mathcal{Q}}^{-1}\boldsymbol{\mathcal{U}},$$

where the third equality holds by noting that $\mathbf{A}\boldsymbol{v}_t = \boldsymbol{q}_t$.

West (1987), we have

$$\widetilde{\mathbf{\Sigma}}_T := \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{g}}_t\widehat{\boldsymbol{g}}_t' + \frac{1}{T}\sum_{k=1}^{\ell}\omega_{\ell k}\sum_{t=k+1}^{T}\{\widehat{\boldsymbol{g}}_{t-k}\widehat{\boldsymbol{g}}_t' + \widehat{\boldsymbol{g}}_t\widehat{\boldsymbol{g}}_{t-k}'\},$$

where $\widehat{\boldsymbol{g}}_t := [\Delta x_t, \widehat{u}_t]'$, $\omega_{\ell k} := 1 - k/(1+\ell)$, $\ell = O(T^{1/4})$ and $\widehat{u}_t := y_t - \widehat{\alpha}_T - \widehat{\beta}_T^{+}x_t^{+} - \widehat{\beta}_T^{-}x_t^{-}$. Under mild regularity conditions, the asymptotic bias υ_* in $\boldsymbol{\mathcal{U}}$ can be consistently estimated by $\widetilde{\mathbf{\Pi}}_T := \frac{1}{T}\sum_{k=0}^{\ell}\sum_{t=k+1}^{T}\widehat{\boldsymbol{g}}_{t-k}\widehat{\boldsymbol{g}}_t'$, with $\widetilde{\pi}_T^{(i,j)}$ denoting the (i,j)-th element. Let the long-run parameter estimator be

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

where $\widetilde{y}_t := y_t - \Delta x_t(\widetilde{\sigma}_T^{(1,1)})^{-1}\widetilde{\sigma}_T^{(1,2)}$ and $\widetilde{\upsilon}_T := \widetilde{\pi}_T^{(1,2)} - \widetilde{\pi}_T^{(1,1)}(\widetilde{\sigma}_T^{(1,1)})^{-1}\,\widetilde{\sigma}_T^{(1,2)}$. Then, the FM estimators of the long-run parameters are obtained as $\widetilde{\beta}_T^{+} := \widetilde{\lambda}_T + \widetilde{\eta}_T$ and $\widetilde{\beta}_T^{-} := \widetilde{\eta}_T$.

To derive the limit distribution of the FM estimator, we make the following assumptions:

Assumption 2. *(i)* $\mathbf{\Sigma}_*$ *is finite and positive definite and* $\widetilde{\mathbf{\Sigma}}_T \xrightarrow{\mathbb{P}} \mathbf{\Sigma}_*$*; and (ii)* $\mathbf{\Pi}_*$ *is finite and* $\widetilde{\mathbf{\Pi}}_T \xrightarrow{\mathbb{P}} \mathbf{\Pi}_*$*, where* $\mathbf{\Pi}_* := \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\mathbb{E}[\boldsymbol{g}_i\boldsymbol{g}_t']$.

Assumption 2 ensures that the key components required for bias correction and variance estimation in the FM estimator are consistently estimated. Specifically, condition (i) guarantees that the stochastic trends and error terms have well-defined second moments and that the HAC estimator, $\widetilde{\mathbf{\Sigma}}_T$, converges in probability to $\mathbf{\Sigma}_*$. Condition (ii) ensures that the long-run bias term contained in $\mathbf{\Pi}_*$, which reflects the cumulative dependence between Δx_t and u_t over time, is finite and consistently estimated by $\widetilde{\mathbf{\Pi}}_T$. This is crucial because the FM estimator adjusts for asymptotic bias and serial correlation by subtracting a bias correction term based on $\mathbf{\Pi}_*$. Together, these conditions imply that the FM estimator can remove asymptotic bias and achieve mixed normality without imposing restrictive assumptions on the underlying data-generating process. In practice, these high-level conditions are satisfied by many HAC estimators under mild regularity conditions, making the methodology broadly applicable.

Let $\pi_*^{(i,j)}$ be the (i,j)-th element of $\mathbf{\Pi}_*$ such that $\pi_*^{(1,2)} = \upsilon_*$. Lemma 3 provides the limit behavior of the components of the FM estimator.

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

Lemma 3 formalizes the asymptotic behavior of the stochastic component of the FM estimator after bias correction. The term inside braces adjusts the original OLS score using HAC-based corrections. The scalar τ_*^2 represents the long-run variance of the adjusted error term, ensuring that the FM estimator achieves mixed normality, as opposed to the non-normal distribution obtained under OLS.

By Lemma 2(*i*), $\widehat{\mathbf{Q}}_T \Rightarrow \boldsymbol{\mathcal{Q}}$, which is nonsingular with probability 1.

Corollary 2. *Under Assumption 1 and for* $k = 1$, $\widetilde{\mathbf{D}}_T(\widetilde{\boldsymbol{\varrho}}_T - \boldsymbol{\varrho}_*) \Rightarrow \boldsymbol{\mathcal{Q}}^{-1}\widetilde{\boldsymbol{\mathcal{U}}}$. □

Corollary 2 has several implications. First, the limit distribution of the FM estimator is mixed normal. Conditional on $\sigma\{\mathcal{B}_x(r), r \in (0,1]\}$, 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, the null limit distribution of a Wald test constructed using the FM estimator is chi-squared. Second, we have: $T(\widetilde{\beta}_T^+ - \beta_*^+) = T(\widetilde{\beta}_T^- - \beta_*^-) + o_{\mathbb{P}}(1)$, implying that the limit distribution of $\widetilde{\beta}_T^+$ is equivalent to that of $\widetilde{\beta}_T^-$, which is determined by that of $\widetilde{\eta}_T$. Third, the convergence rates of $\widetilde{\beta}_T^+$ and $\widetilde{\beta}_T^-$ are T, enabling estimation of the short-run parameters in the second stage by replacing u_{t-1} with $\widetilde{u}_{t-1} := y_{t-1} - \widetilde{\alpha}_T - \widetilde{\beta}_T^+ x_{t-1}^+ - \widetilde{\beta}_T^- x_{t-1}^-$.

Theorem 2 formally presents the limit distribution of the FM estimator:

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

Because the FM estimator achieves mixed normality, inference based on the FM estimator follows standard chi-squared limits under the null.

3.1.2 Estimation of the Short-Run Parameters

As the long-run coefficients are super-consistent (converging faster than $T^{1/2}$), we can treat them as known when estimating the short-run parameters in (2). The short-run dynamics can be expressed compactly as: $\Delta y_t = \boldsymbol{\zeta}_*' \boldsymbol{h}_t + e_t$, where $\boldsymbol{\zeta}_* := (\rho_*, \boldsymbol{\beta}_{2*}')'$, $\boldsymbol{\beta}_{2*} := (\gamma_*, \varphi_{1*}, \ldots, \varphi_{p-1*}, \pi_{0*}^{+\prime}, \ldots, \pi_{q-1*}^{+\prime}, \pi_{0*}^{-\prime}, \ldots, \pi_{q-1*}^{-\prime},)'$ and $\boldsymbol{h}_t := (u_{t-1}, \boldsymbol{z}_{2t}')'$. Since all variables are stationary, the short-run parameters can be estimated by OLS

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

Lemma 4 shows the limit behaviors of the components of $\widehat{\boldsymbol{\zeta}}_T$.

Lemma 4. *Under Assumption 1, (i)* $\widehat{\boldsymbol{\Gamma}}_T := T^{-1}\sum_{t=1}^T \boldsymbol{h}_t \boldsymbol{h}_t' \xrightarrow{\mathbb{P}} \boldsymbol{\Gamma}_* := \mathbb{E}[\boldsymbol{h}_t \boldsymbol{h}_t']$*; (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)* $\boldsymbol{\Omega}_*$ *simplifies to* $\sigma_*^2 \boldsymbol{\Gamma}_*$ *in the special case where* $\mathbb{E}[e_t^2 | \boldsymbol{h}_t] = \sigma_*^2$.

Using Lemma 4, we derive the limit distribution of $\widehat{\boldsymbol{\zeta}}_T$ in Theorem 3.

Theorem 3. *Suppose that $\boldsymbol{\Gamma}_*$ and $\boldsymbol{\Omega}_*$ are PD. Under Assumption 1, (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) further if $\mathbb{E}[e_t^2|\boldsymbol{h}_t] = \sigma_*^2$, then $\sqrt{T}(\widehat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}_*) \overset{A}{\sim} N(\mathbf{0}, \sigma_*^2\boldsymbol{\Gamma}_*^{-1})$.* □

Note that, because the convergence rate of the short-run parameter estimators is $\sqrt{T}$, standard Wald tests are applicable.

Before moving on to the case with $k > 1$, we justify the use of the two-step approach by demonstrating that the estimation of the short-run and long-run parameters jointly in a single step results in an additional singularity problem. Using $\lambda_* := \beta_*^+ - \beta_*^-$ and $\eta_* := \beta_*^-$, it follows that $u_{t-1} = y_{t-1} - \alpha_* - \lambda_* x_{t-1}^+ - \beta_* x_{t-1}$. Then, the model can be written as

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

Define

$$\boldsymbol{\xi}_* := \left[\begin{array}{c:c} \boldsymbol{\xi}_{1*}' & \boldsymbol{\xi}_{2*}' \end{array}\right]' := \left[\begin{array}{ccc:c} \rho_* & \theta_* & \theta_*^- & \boldsymbol{\alpha}_{2*}' \end{array}\right]', \quad \boldsymbol{p}_t := \left[\begin{array}{c:c} \boldsymbol{p}_{1t}' & \boldsymbol{p}_{2t}' \end{array}\right]' := \left[\begin{array}{ccc:c} y_{t-1} & x_{t-1}^+ & x_{t-1} & \boldsymbol{z}_{2t}' \end{array}\right]',$$

where $\boldsymbol{\xi}_{2*}$ and $\boldsymbol{p}_{2t}$ correspond to $\boldsymbol{\alpha}_{2*}$ and $\boldsymbol{z}_{2t}$, and $\theta_* := \theta_*^+ - \theta_*^-$. If we attempt to estimate $\boldsymbol{\xi}_*$ by OLS, we obtain

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

Lemma 5 shows that the inverse matrix in $\widehat{\boldsymbol{\xi}}_T$ is asymptotically singular.

Lemma 5. *Under Assumption 1:*

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

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

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

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

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

Let $\ddot{\mathbf{D}}_T:=\text{diag}[T^{3/2}\mathbf{I}_2,T,T^{1/2}\mathbf{I}_{1+p+2q}]$. Then

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

where $\boldsymbol{\mathcal{P}}_{21}:=\boldsymbol{\mathcal{P}}_{12}'$. Since $\boldsymbol{\mathcal{P}}$ is singular, it is difficult to obtain the limit distribution of $\widehat{\boldsymbol{\xi}}_T$ using the one-step approach, even after re-parameterizing the long-run levels relationship. This justifies the two-step estimation procedure.

3.2 The Two-step NARDL Estimation with $k>1$

If there are multiple explanatory variables in the NARDL model, the two-step estimation procedure described in Section 3.1 must be modified as follows.

Let $\boldsymbol{x}_t\equiv\boldsymbol{x}_t^{+}+\boldsymbol{x}_t^{-}$, $\boldsymbol{\lambda}_*=\boldsymbol{\beta}_*^{+}-\boldsymbol{\beta}_*^{-}$ and $\boldsymbol{\eta}_*=\boldsymbol{\beta}_*^{-}$ for $k>1$. Then, the long-run relationship can be written as: $y_t=\alpha_*+\boldsymbol{\lambda}_*'\boldsymbol{x}_t^{+}+\boldsymbol{\eta}_*'\boldsymbol{x}_t+u_t$. This representation is obtained from the original long-run specification: $y_t=\alpha_*+\boldsymbol{\beta}_*^{+\prime}\boldsymbol{x}_t^{+}+\boldsymbol{\beta}_*^{-\prime}\boldsymbol{x}_t^{-}+u_t$. By extending Lemma 2, we have:

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

where $\boldsymbol{q}_t:=(1,\boldsymbol{x}_t^{+\prime},\boldsymbol{x}_t')'$, $\widetilde{\mathbf{D}}_T:=\text{diag}[T^{1/2},T^{3/2}\mathbf{I}_k,T\mathbf{I}_k]$, $\boldsymbol{\mathcal{B}}_x(\cdot)$ is a $k\times1$ vector of Brownian motions with $[\boldsymbol{\mathcal{B}}_x(\cdot)',\mathcal{B}_u(\cdot)]':=\boldsymbol{\Sigma}_*^{1/2}[\boldsymbol{\mathcal{W}}_x(\cdot)',\mathcal{W}_u(\cdot)']'$ and $[\boldsymbol{\mathcal{W}}_x(\cdot)',\mathcal{W}_u(\cdot)]'$ being a $(k+1)\times1$ vector of independent Wiener processes, and $\boldsymbol{\Sigma}_*:=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\boldsymbol{g}_t\boldsymbol{g}_s]$ with $\boldsymbol{g}_t:=[\Delta\boldsymbol{x}_t,u_t]'$. The second row block sub-matrices of $\boldsymbol{\mathcal{Q}}$, namely $[\frac{1}{2}\boldsymbol{\mu}_*^{+},\frac{1}{3}\boldsymbol{\mu}_*^{+}\boldsymbol{\mu}_*^{+\prime},\boldsymbol{\mu}_*^{+}\int r\boldsymbol{\mathcal{B}}_x']$, form a sub-matrix with rank one because it can be written as: $\boldsymbol{\mu}_*^{+}[\frac{1}{2},\frac{1}{3}\boldsymbol{\mu}_*^{+\prime},\int r\boldsymbol{\mathcal{B}}_x']$, which is the outer product of a column vector and a row vector. This implies that $\boldsymbol{\mathcal{Q}}$ is singular with probability 1. Consequently, the two-step procedure in Section 3.1 cannot be applied to estimate the long-run parameters when $k>1$.

3.2.1 Estimation of the Long-Run Parameters

First-Step Transformed OLS Estimator. To address the above singularity issue,[3] consider the following modification. Let $\boldsymbol{m}_t:=\sum_{j=1}^{t}\boldsymbol{s}_j^{+}$, which is a unit-root process with zero-mean increments. If we regress

[3] Cho et al. (2023a) allow $E(\Delta\boldsymbol{x}_t)\neq0$ and propose a substantial modification to the estimation and inference procedure.

$\boldsymbol{x}_t^+$ against t, then we estimate $\boldsymbol{\mu}_*^+$ by $\widehat{\boldsymbol{\mu}}_T^+ := (\sum_{t=1}^T t^2)^{-1}\sum_{t=1}^T t\boldsymbol{x}_t^+ = \boldsymbol{\mu}_*^+ + (\sum_{t=1}^T t^2)^{-1}\sum_{t=1}^T t\boldsymbol{m}_t$ and obtain $\widehat{\boldsymbol{m}}_t := \boldsymbol{x}_t^+ - \widehat{\boldsymbol{\mu}}_T^+ t$. Then, it follows that $\boldsymbol{m}_t = \widehat{\boldsymbol{m}}_t + t\boldsymbol{d}_T$, where $\boldsymbol{d}_T := (\sum_{t=1}^T t^2)^{-1}\sum_{t=1}^T t\boldsymbol{m}_t$. Thus, $\boldsymbol{x}_t^+ = \widehat{\boldsymbol{m}}_t + \boldsymbol{\delta}_{*T} t$, where $\boldsymbol{\delta}_{*T} := \boldsymbol{\mu}_*^+ + \boldsymbol{d}_T$. Under regularity conditions, $\boldsymbol{d}_T = O_{\mathbb{P}}(T^{-1/2})$ and thus $\boldsymbol{\delta}_{*T} = \boldsymbol{\mu}_* + O_{\mathbb{P}}(T^{-1/2})$. Rewriting the long-run relationship as $y_t = \alpha_* + \xi_{*T} t + \boldsymbol{\lambda}_*'\widehat{\boldsymbol{m}}_t + \boldsymbol{\eta}_*'\boldsymbol{x}_t + u_t$, where $\xi_{*T} := \boldsymbol{\lambda}_*'\boldsymbol{\delta}_{*T}$, we can estimate $\boldsymbol{\lambda}_*$ and $\boldsymbol{\eta}_*$ by regressing y_t on $\boldsymbol{r}_t := (1, t, \widehat{\boldsymbol{m}}_t', \boldsymbol{x}_t')'$. Let $\ddot{\boldsymbol{\varpi}}_T := (\ddot{\alpha}_T, \ddot{\xi}_{*T}, \ddot{\boldsymbol{\lambda}}_T', \ddot{\boldsymbol{\eta}}_T')'$ be the OLS estimator of $\varpi_{*T} := (\alpha_*, \xi_{*T}, \boldsymbol{\lambda}_*', \boldsymbol{\eta}_*')'$. The long-run estimators of $\boldsymbol{\beta}_*^+$ and $\boldsymbol{\beta}_*^-$ can be obtained as $\ddot{\boldsymbol{\beta}}_T^+ := \ddot{\boldsymbol{\lambda}}_T + \ddot{\boldsymbol{\eta}}_T$ and $\ddot{\boldsymbol{\beta}}_T^- := \ddot{\boldsymbol{\eta}}_T$. Define

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

We refer to this as the *first-step transformed OLS (TOLS) estimator*. The collinear trend in $\boldsymbol{x}_t^+$ causes the singularity of $\boldsymbol{\mathcal{Q}}$, rendering first-step estimation by OLS or FM inoperable. By detrending $\boldsymbol{x}_t^+$ prior to estimation, the transformed OLS (TOLS) estimator restores full rank and enables consistent estimation of long-run parameters.

The limit distribution of the first-step TOLS estimator is obtained similarly to that of the first-step OLS estimator. Note that: $\ddot{\boldsymbol{\varpi}}_T = \boldsymbol{\varpi}_{*T} + (\sum_{t=1}^T \boldsymbol{r}_t\boldsymbol{r}_t')^{-1}\sum_{t=1}^T \boldsymbol{r}_t u_t$. To characterize the limit behavior of the components, define $\ddot{\boldsymbol{\Sigma}}_* := \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}[\ddot{\boldsymbol{g}}_t\ddot{\boldsymbol{g}}_t']$ and $[\boldsymbol{\mathcal{B}}_m(\cdot)', \boldsymbol{\mathcal{B}}_x(\cdot)', \mathcal{B}_u(\cdot)]' := \ddot{\boldsymbol{\Sigma}}_*^{1/2}[\boldsymbol{\mathcal{W}}_m(\cdot)', \boldsymbol{\mathcal{W}}_x(\cdot)', \mathcal{W}_u(\cdot)]'$, where $\ddot{\boldsymbol{g}}_t := [\Delta\boldsymbol{m}_t', \Delta\boldsymbol{x}_t, u_t]'$ and $[\boldsymbol{\mathcal{W}}_m(\cdot)', \boldsymbol{\mathcal{W}}_x(\cdot)', \mathcal{W}_u(\cdot)]'$ is a $(2k+1)\times 1$ vector of independent Wiener processes.

Lemma 6. *Suppose that $\ddot{\boldsymbol{\Sigma}}_*$ is positive definite. Under Assumption 1, (i) $\ddot{\mathbf{R}}_T := \ddot{\mathbf{D}}_T^{-1}(\sum_{t=1}^T \boldsymbol{r}_t\boldsymbol{r}_t')\ddot{\mathbf{D}}_T^{-1}$ $\Rightarrow \boldsymbol{\mathcal{R}}$, where $\boldsymbol{\mathcal{R}}$ is defined as*

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

and $\ddot{\mathbf{D}}_T := diag[T^{1/2}, T^{3/2}, T\mathbf{I}_{2k}]$; and (ii) if $\boldsymbol{\upsilon}_{x*} := \lim_{T\to\infty} T^{-1}\sum_{t=1}^T\sum_{i=1}^t \mathbb{E}[\Delta\boldsymbol{x}_i u_t]$ and $\boldsymbol{\upsilon}_{m*} := \lim_{T\to\infty} T^{-1}\sum_{t=1}^T\sum_{i=1}^t \mathbb{E}[\Delta\boldsymbol{m}_i u_t]$ are finite, then $\ddot{\mathbf{U}}_T := \ddot{\mathbf{D}}_T^{-1}(\sum_{t=1}^T \boldsymbol{r}_t u_t) \Rightarrow \ddot{\boldsymbol{\mathcal{U}}} := [\int d\mathcal{B}_u, \int r d\mathcal{B}_u,$ $\int \boldsymbol{\mathcal{B}}_m' d\mathcal{B}_u - 3\int r d\mathcal{B}_u \int r\boldsymbol{\mathcal{B}}_m' + \boldsymbol{\upsilon}_{m*}', \int \boldsymbol{\mathcal{B}}_x' d\mathcal{B}_u + \boldsymbol{\upsilon}_{x*}']'$. □

The transformed OLS estimator uses detrended $\boldsymbol{x}_t^+$ (via $\widehat{\boldsymbol{m}}_t$) to eliminate collinear trends. Lemma 6 characterizes the asymptotic behavior of the design matrix and score vector after this transformation. The

limiting matrix $\boldsymbol{\mathcal{R}}$ now involves Brownian motion functionals for both $\boldsymbol{\mathcal{B}}_m$ (detrended component) and $\boldsymbol{\mathcal{B}}_x$ (original regressors), with correction terms to account for trend coefficient estimation error. It ensures that the transformed estimator avoids singularity and has a well-defined asymptotic distribution, enabling inference even when $k > 1$. $\boldsymbol{\mathcal{R}}$ is no longer singular because $\sum_{t=1}^{[T(\cdot)]} \ddot{\boldsymbol{g}}_t$ obeys the FCLT using partially correlated increments.

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

The first part of Corollary 3 follows from the structure of the first-step TOLS estimator. For the second part, notice that $\widehat{\xi}_T$ is not of primary interest. While $(\widehat{\xi}_T - \xi_{*T})$ converges at rate $T^{3/2}$, $\xi_{*T} := \boldsymbol{\lambda}_*'\boldsymbol{\delta}_{*T}$ converges to $\boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+$ at rate $\sqrt{T}$. Thus, $T^{1/2}(\widehat{\xi}_T - \boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+)$ is asymptotically bounded in probability.

Theorem 4 provides the limit distribution of the TOLS estimator.

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

Theorem 4 is analogous to Theorem 2. The TOLS estimator is consistent for $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$ and its limit distribution depends on the asymptotic bias terms $(\boldsymbol{\upsilon}_{m*}', \boldsymbol{\upsilon}_{x*}')'$. By detrending $\boldsymbol{x}_t^+$, the singularity problem is resolved, and inference becomes feasible for $k > 1$.

First-Step Transformed FM Estimator. The limit distribution of the TOLS estimator in Theorem 4 does not provide a basis for standard inference, as it exhibits asymptotic biases driven by $\boldsymbol{\upsilon}_{m*}$ and $\boldsymbol{\upsilon}_{x*}$. To address this issue, we propose the first-step transformed fully modified (TFM) estimator, which combines the detrending approach of TOLS with the bias correction mechanism of the FM estimator.

We make the following assumptions:

Assumption 3. *(i) For finite and PD* $\ddot{\boldsymbol{\Sigma}}_*$ *there exists a consistent estimator of* $\ddot{\boldsymbol{\Sigma}}_*$

$$\bar{\boldsymbol{\Sigma}}_T := \begin{bmatrix} \bar{\boldsymbol{\Sigma}}_T^{(1,1)} & \bar{\boldsymbol{\sigma}}_T^{(1,2)} \\ \bar{\boldsymbol{\sigma}}_T^{(2,1)} & \bar{\sigma}_T^{(2,2)} \end{bmatrix} \xrightarrow{\mathbb{P}} \ddot{\boldsymbol{\Sigma}}_* := \begin{bmatrix} \ddot{\boldsymbol{\Sigma}}_*^{(1,1)} & \ddot{\boldsymbol{\sigma}}_*^{(1,2)} \\ \ddot{\boldsymbol{\sigma}}_*^{(2,1)} & \sigma_*^{(2,2)} \end{bmatrix};$$

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

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

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

Assumption 3 corresponds to Assumption 2 for $k > 1$. Formally, the TFM estimator is defined as

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

where $\bar{y}_t := y_t - \boldsymbol{\ell}_t'(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)}$, $\boldsymbol{\ell}_t := (\Delta\widehat{\boldsymbol{m}}_t', \Delta\boldsymbol{x}_t')'$, $\bar{\boldsymbol{\upsilon}}_T := \bar{\boldsymbol{\pi}}_T^{(1,2)} - \bar{\boldsymbol{\Pi}}_T^{(1,1)}(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)}$ and $\bar{\mathbf{S}} := [\mathbf{0}_{2k\times 2}, \mathbf{I}_{2k}]$. The TFM estimator corrects for collinear deterministic trends via detrending and for asymptotic bias/serial correlation via FM-style bias adjustments. This ensures a mixed normal asymptotic distribution for inference, even when $k > 1$. The following lemma corresponds to Lemma 3 when $k > 1$:

Lemma 7. *Under Assumptions 1 and 3,* $\bar{\mathbf{U}}_T := \ddot{\mathbf{D}}_T^{-1}\{\sum_{t=1}^{T} \boldsymbol{r}_t(u_t - \boldsymbol{\ell}_t'(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)}) - T\bar{\mathbf{S}}'\bar{\boldsymbol{\upsilon}}_T\} \Rightarrow \bar{\boldsymbol{\mathcal{U}}} := \ddot{\tau}[\int d\mathcal{W}_u, \int r d\mathcal{W}_u, \int \boldsymbol{\mathcal{B}}_m' d\mathcal{W}_u - 3\int r d\mathcal{W}_u \int r\boldsymbol{\mathcal{B}}_m', \int \boldsymbol{\mathcal{B}}_x' d\mathcal{W}_u]'$, *where* $\ddot{\tau}^2 := plim_{T\to\infty}\bar{\tau}_T^2$ *and* $\bar{\tau}_T^2 := \bar{\sigma}_T^{(2,2)} - \bar{\boldsymbol{\sigma}}_T^{(2,1)}(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)}$. □

By Lemmas 6 and 7, the limit distribution of the TFM estimator is obtained as the product of $\boldsymbol{\mathcal{R}}^{-1}$ and $\bar{\boldsymbol{\mathcal{U}}}$. Letting $[\bar{\boldsymbol{\beta}}_T^{+\prime}, \bar{\boldsymbol{\beta}}_T^{-\prime}]' := \ddot{\mathbf{S}}\bar{\boldsymbol{\varpi}}_T$, we obtain its weak limit as follows:

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

Theorem 5 extends Theorem 4 by establishing that the TFM estimator is consistent for $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$ and eliminates the asymptotic bias terms $(\boldsymbol{\upsilon}_{m*}', \boldsymbol{\upsilon}_{x*}')'$. Moreover, its limit distribution is mixed normal. Specifically, conditional on the sigma-field generated by $\sigma\{(\boldsymbol{\mathcal{B}}_m(r)', \boldsymbol{\mathcal{B}}_x(r)')', r \in (0,1]\}$, we have: $\bar{\mathbf{D}}_T(\bar{\boldsymbol{\varpi}}_T - \bar{\boldsymbol{\varpi}}_{*T}) \Rightarrow N(\mathbf{0}, \ddot{\tau}^2\boldsymbol{\mathcal{R}}^{-1})$. This result provides the foundation on which to develop hypothesis testing and inference procedures for the unknown parameters.

3.2.2 Estimation of the Short-Run Parameters

Because the TFM estimator for the long-run parameters converges at a rate faster than $\sqrt{T}$, we can treat these estimates as known when estimating the short-run parameters in the second step. Define $u_{t-1} := y_{t-1} - \alpha_* - \boldsymbol{\beta}_*^{+\prime}\boldsymbol{x}_{t-1}^{+} - \boldsymbol{\beta}_*^{-\prime}\boldsymbol{x}_{t-1}^{-} = y_{t-1} - \alpha_* - \boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^{+}(t-1) - \boldsymbol{\lambda}_*'\boldsymbol{m}_{t-1} - \boldsymbol{\eta}_*'\boldsymbol{x}_{t-1}$, where $\boldsymbol{\lambda}_*$, $\boldsymbol{\mu}_*^{+}$, and $\boldsymbol{\eta}_*$ are assumed to be known. This enables us to apply (2) in Section 2, and estimate the short-run parameters using the OLS estimator. Then, Theorem 3 can be applied.

4 Hypothesis Testing

We develop a formal testing procedure to detect the presence of asymmetries in the model, both in the long-run relationship and in the short-run dynamics.

4.1 Hypothesis Testing with $k = 1$

4.1.1 Testing Symmetry of the Long-Run Parameters

Consider the hypotheses $H_0' : (\beta_*^+ - \beta_*^-) = r$ vs. $H_1' : (\beta_*^+ - \beta_*^-) \neq r$ for some $r \in \mathbb{R}$. Setting $r = 0$ allows us to test whether $\beta_*^+ = \beta_*^-$. As $\lambda_* := \beta_*^+ - \beta_*^-$, the hypotheses can be restated as $H_0'' : \lambda_* = r$ vs. $H_1'' : \lambda_* \neq r$, where the long-run symmetry restriction $\beta_*^+ = \beta_*^-$ corresponds to $\lambda_* = 0$. Testing this restriction is straightforward when λ_* is estimated by FM, as the FM estimator is asymptotically mixed-normal. Thus, the Wald test follows an asymptotic chi-squared distribution under the null, which represents a key advantage of FM over OLS.

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

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

However, this Wald statistic is specific to testing λ_* and may not be suitable for more general hypotheses. For example, consider: $H_0''' : \mathbf{R}\boldsymbol{\beta}_* = \mathbf{r}$ vs. $H_1''' : \mathbf{R}\boldsymbol{\beta}_* \neq \mathbf{r}$ for some $\mathbf{R} \in \mathbb{R}^{r \times 2}$ and $\mathbf{r} \in \mathbb{R}^r$ ($r \in \{1, 2\}$), where $\boldsymbol{\beta}_* := (\beta_*^+, \beta_*^-)'$. Define

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

These hypotheses can be rewritten as $H_0''' : \widetilde{\mathbf{R}}\boldsymbol{\varrho}_* = \mathbf{r}$ vs. $H_1''' : \widetilde{\mathbf{R}}\boldsymbol{\varrho}_* \neq \mathbf{r}$, where $\widetilde{\mathbf{R}}\boldsymbol{\varrho}_* = \boldsymbol{\beta}_*$ and $\widetilde{\mathbf{R}} := \mathbf{R}\widetilde{\mathbf{R}}_\ell$. The corresponding Wald statistic is

$$\widetilde{\mathcal{W}}_T^{(\ell)} := (\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r})'(\widetilde{\tau}_T^2 \widetilde{\mathbf{R}} \mathbf{Q}_T^{-1} \widetilde{\mathbf{R}}')^{-1}(\widetilde{\mathbf{R}}\widetilde{\boldsymbol{\varrho}}_T - \mathbf{r}),$$

where $\mathbf{Q}_T := \sum_{t=1}^T \boldsymbol{q}_t \boldsymbol{q}_t'$.

Theorem 6 describes the limit behavior of $\widetilde{\mathcal{W}}_T^{(\ell)}$:

Theorem 6. *Under Assumptions 1 and 2, $\mathcal{W}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_1^2$ under H_0'' and $\widetilde{\mathcal{W}}_T^{(\ell)} \overset{A}{\sim} \mathcal{X}_2^2$ under H_0'''. 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'''. □

4.1.2 Testing Symmetry of the Short-Run Parameters

In the NARDL literature, it is common to test for additive symmetry in the short-run dynamic parameters.[4] Consider the general hypothesis: $H_0 : \mathbf{R}_s\boldsymbol{\zeta}_* = \mathbf{r}$ vs. $H_1 : \mathbf{R}_s\boldsymbol{\zeta}_* \neq \mathbf{r}$, where $\mathbf{R}_s \in \mathbb{R}^{r\times(1+p+2q)}$ and $\mathbf{r} \in \mathbb{R}^r$ ($r \in \mathbb{N}$) are selection matrices. For additive short-run symmetry, the null and alternative hypotheses are $H_0 : \sum_{j=0}^{q-1}\pi_{j*}^{+} = \sum_{j=0}^{q-1}\pi_{j*}^{-}$ and $H_1 : \sum_{j=0}^{q-1}\pi_{j*}^{+} \neq \sum_{j=0}^{q-1}\pi_{j*}^{-}$. Let $\mathbf{R}_s := \left[\mathbf{0}'_{1+p}, \boldsymbol{\iota}'_q, -\boldsymbol{\iota}'_q\right]$ and $\mathbf{r} = 0$. Then the Wald test is constructed as

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

where $\widehat{\boldsymbol{\Omega}}_T := T^{-1}\sum_{t=1}^{T}\widehat{e}_t^2\boldsymbol{h}_t\boldsymbol{h}_t'$ is a consistent estimator of $\boldsymbol{\Omega}_*$. If Lemma 4(*iii*) holds, then the Wald test 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 $\widehat{\sigma}_{e,T}^2 := T^{-1}\sum_{t=1}^{T}\widehat{e}_t^2$ and $\widehat{e}_t := \Delta y_t - \widehat{\boldsymbol{\zeta}}_T'\boldsymbol{h}_t$.

Theorem 7 establishes that the null and alternative limit distributions of the Wald test are standard.

Theorem 7. *Suppose that* $\boldsymbol{\Gamma}_*$ *and* $\boldsymbol{\Omega}_*$ *are PD. Under Assumption 1,* $\mathcal{W}_T^{(s)} \overset{A}{\sim} \mathcal{X}_r^2$ *under* H_0. *For any sequence,* c_T *such that* $c_T = o(T)$, $\mathbb{P}(\mathcal{W}_T^{(s)} > c_T) \to 1$ *under* H_1. □

4.2 Hypotheses Testing with $k > 1$

4.2.1 Testing for Symmetry of the Long-Run Parameters

Define $\boldsymbol{\beta}_* := (\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$ and consider the general hypotheses $H_0^{(4)} : \mathbf{R}\boldsymbol{\beta}_* = \boldsymbol{r}$ vs. $H_1^{(4)} : \mathbf{R}\boldsymbol{\beta}_* \neq \boldsymbol{r}$ for some $\mathbf{R} \in \mathbb{R}^{r\times 2k}$ and $\boldsymbol{r} \in \mathbb{R}^r$ ($r \leq 2k$). We then construct the Wald test as $\ddot{\mathcal{W}}_T := T^2(\mathbf{R}\bar{\boldsymbol{\beta}}_T - \boldsymbol{r})'\{\bar{\tau}_T^2\mathbf{R}\ddot{\mathbf{S}}\ddot{\mathbf{R}}_T^{-1}\ddot{\mathbf{S}}'\mathbf{R}'\}^{-1}(\mathbf{R}\bar{\boldsymbol{\beta}}_T - \boldsymbol{r})$, where $\bar{\boldsymbol{\beta}}_T := (\bar{\boldsymbol{\beta}}_T^{+\prime}, \bar{\boldsymbol{\beta}}_T^{-\prime})'$.

Theorem 8 describes the limit behavior of $\ddot{\mathcal{W}}_T$.

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

[4] Alternative short-run symmetry restrictions include pairwise symmetry between $\boldsymbol{\pi}_{j*}^{+}$ and $\boldsymbol{\pi}_{j*}^{-}$ for $j = 0, \ldots, q-1$ (SYG) and impact symmetry defined by $\boldsymbol{\pi}_{0*}^{+} = \boldsymbol{\pi}_{0*}^{-}$ (Greenwood-Nimmo and Shin, 2013). These can be tested by appropriately specifying selection matrices, $\mathbf{R}_s$ and $\mathbf{r}$.

4.2.2 Testing for Symmetry of the Short-Run Parameters

To test for additive symmetry of the short-run dynamic parameters when $k > 1$, consider the hypotheses $H_0 : \mathbf{R}_s\boldsymbol{\zeta}_* = \mathbf{r}$ vs. $H_1 : \mathbf{R}_s\boldsymbol{\zeta}_* \neq \mathbf{r}$, where $\mathbf{R}_s \in \mathbb{R}^{r\times(1+p+2qk)}$ and $\mathbf{r} \in \mathbb{R}^r$ are selection matrices, and $\boldsymbol{\zeta}_* := (\rho_*, \gamma_*, \varphi_{1*}, \ldots, \varphi_{p-1*}, \boldsymbol{\pi}_{0*}^{+\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime}, \ldots, \boldsymbol{\pi}_{q-1*}^{-\prime})$, which generalizes the definition of $\boldsymbol{\zeta}_*$ for $k = 1$ in Section 4.1.2. Let $\mathbf{R}_s := \left[\mathbf{0}_{k\times(1+p)}, \boldsymbol{\iota}_q' \otimes \mathbf{I}_k, -\boldsymbol{\iota}_q \otimes \mathbf{I}_k\right]$ and $\mathbf{r} = \mathbf{0}$, then we can test the null hypothesis of additive short-run symmetry as $H_0 : \sum_{j=0}^{q-1}\boldsymbol{\pi}_{j*}^{+} = \sum_{j=0}^{q-1}\boldsymbol{\pi}_{j*}^{-}$ vs. $H_1 : \sum_{j=0}^{q-1}\boldsymbol{\pi}_{j*}^{+} \neq \sum_{j=0}^{q-1}\boldsymbol{\pi}_{j*}^{-}$. If the cointegration residuals are obtained as in Section 3.2.2, the same Wald test described in Section 4.1.2 can be applied, as both the TOLS and TFM estimators are super-consistent.

5 Monte Carlo Simulations

We conduct stochastic simulations to evaluate the finite-sample properties of the Wald tests introduced in Section 4.[5]

5.1 Simulations Results for $k = 1$

Consider the NARDL(2,1) data generating process (DGP): $\Delta y_t = \gamma_* + \rho_* u_{t-1} + \varphi_*\Delta y_{t-1} + \pi_*^{+}\Delta x_t^{+} + \pi_*^{-}\Delta x_t^{-} + e_t$, where $u_{t-1} := y_{t-1} - \alpha_* - \beta_*^{+}x_{t-1}^{+} - \beta_*^{-}x_{t-1}^{-}$, $\Delta x_t := \kappa_*\Delta x_{t-1} + \sqrt{1-\kappa_*^2}v_t$, and $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$.

Comparison of 2SNARDL with One-Step Estimation. We set the parameter vector $(\alpha_*, \beta_*^{+}, \beta_*^{-}, \gamma_*, \rho_*, \varphi_*, \pi_*^{+}, \pi_*^{-}, \kappa_*) = (0, 2, 1, 0, -2/3, \varphi_*, 1, 1/2, 1/2)$ and allow both the sample size T and the parameter φ_* to vary. Note that Δx_t is generated by an AR(1) process with normally distributed disturbances and that the equilibrium error, u_t, is serially correlated and contemporaneously correlated with Δx_t. This design enables us to assess the relative performance of the two-step NARDL estimator against the single-step OLS estimator.

Next, we consider the following specifications for the long-run and short-run models

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

where $\widehat{u}_t := y_t - \widehat{\alpha}_T - \widehat{\lambda}_T x_t^{+} - \widehat{\eta}_T x_t$. In the first step, we estimate the parameters of the long-run relationship using either OLS or FM estimator. In the second step, we estimate the short-run parameters by OLS. To

[5] Section D of the Online Supplement examines the finite-sample bias and mean squared error of the two-step NARDL estimators and demonstrates their consistency.

assess the performance of the estimators, we compute the finite-sample bias and mean squared error (MSE) for the key parameters. Simulation results using $R = 5,000$ replications are reported in Tables 1 and 2.[6]

— Insert Tables 1 and 2 Here —

First, consider the long-run parameter estimators obtained in the first step. The finite-sample bias of the FM estimator is substantially smaller than that of the OLS estimator. This is because the FM estimator of the long-run parameters is asymptotically mixed-normally distributed around zero, so $T(\widetilde{\beta}_T^+ - \beta_*^+)$ and $T(\widetilde{\beta}_T^- - \beta_*^-)$ converge to zero in expectation. By contrast, the OLS estimator is not centered asymptotically at zero, leading to non-negligible bias. In addition, the FM estimator is generally more efficient than OLS, producing smaller MSE values across sample sizes. The efficiency gain is particularly pronounced for small and/or negative values of φ_*, where dynamic feedback effects are stronger. Overall, these findings strongly support the use of the FM estimator in the first step of the two-step NARDL procedure.

Next, consider the short-run parameter estimators obtained by OLS in the second step. The finite-sample biases of these estimators are generally negligible, even for small samples such as $T = 50$, regardless of whether the first-step estimator is OLS or FM. However, the smallest biases are typically observed when the FM estimator is employed in the first step. In terms of efficiency, the MSEs of the second-step OLS estimators are very similar across both first-step methods, indicating that the choice of long-run estimator has little impact on short-run precision. This evidence is encouraging, as many applications of the NARDL model rely on relatively small samples due to low data frequency and the limited historical coverage of many macroeconomic datasets.

Testing Restrictions on the Long-Run Parameters. We focus on the case where the FM estimator is used in the first step and set the parameter vector as $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 1, 1, 0, -2/3, \varphi_*, 1/3, 1/2, 1/2)$. We test the null hypothesis of long-run symmetry, $H_0^{(\ell)} : \beta_*^+ - \beta_*^- = 0$ vs. $H_1^{(\ell)} : \beta_*^+ - \beta_*^- \neq 0$, allowing φ_* to vary over $-1/2$, $-1/4$, 0, $1/4$ and $1/2$.

The simulation results in Table 3 indicate some size distortion in small samples for negative values of φ_*. As T increases, the distribution of the Wald test becomes well-approximated by the $\mathcal{X}_1^2$ distribution. To assess power, we generate data under the alternative hypothesis 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 allow φ_* to vary as before. The results in Table 4 confirm that the $\mathcal{W}_T^{(\ell)}$ test is consistent under the alternative and that its power is largely insensitive to the value of φ_*.

— Insert Tables 3 and 4 Here —

[6] To conserve space, we omit results for the intercepts α and γ, which are available upon request.

Testing Restrictions on the Short-Run Parameters. To evaluate the empirical size of the Wald test, we set the parameter vector as $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 2, 1, 0, -2/3, \varphi_*, 1/2, 1/2, 1/2)$ and allow φ_* to vary as before. We first estimate the long-run parameters using the FM estimator and compute $\widehat{u}_t$, then estimate the short-run parameters by OLS. We test the hypotheses $H_0^{(s)} : \pi_*^+ - \pi_*^- = 0$ vs. $H_1^{(s)} : \pi_*^+ - \pi_*^- \neq 0$ using the Wald statistic $\mathcal{W}_T^{(s)}$ with a heteroskedasticity-consistent covariance estimator, $\widehat{\mathbf{\Omega}}_T$. The simulation results in Table 5 show that the finite-sample distribution of $\mathcal{W}_T^{(s)}$ is well approximated by the $\mathcal{X}_1^2$ distribution. Its empirical size approaches the nominal level for $T \geq 500$ and size distortion is minimal even for small samples. Moreover, the empirical size is largely insensitive to the value of φ_*.

To assess the power of the test, we maintain the same hypotheses but update the parameters to $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_*^+, \pi_*^-, \kappa_*) = (0, 2, 1, 0, -2/3, \varphi_*, 1, 1/2, 1/2)$. Two points regarding the simulation results in Table 6 are noteworthy. First, the empirical power of the Wald test rises with T, indicating that the test is consistent. Second, the power of the Wald test exhibits little sensitivity to the degree of autocorrelation captured by the value of φ_*.

— Insert Tables 5 and 6 Here —

5.2 Simulations Results for *k*=2

We generate data from the NARDL(2,1) process $\Delta y_t = \gamma_* + \rho_* u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \pi_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_{t-1} := y_{t-1} - \alpha_* - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_{t-1}^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_{t-1}^-$, $\Delta \boldsymbol{x}_t := \kappa_* \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - \kappa_*^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t)' \sim \text{IID} N(\mathbf{0}_3, \mathbf{I}_3)$.

Comparison of 2SNARDL with One-Step Estimation. We set $(\alpha_*, \gamma_*, \rho_*, \varphi_*, \kappa_*) = (0, 0, -1, \varphi_*, 0.5)$, $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})' = (-1, 0.5, 0.75, -1.5)'$, and $(\boldsymbol{\pi}_{0*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime})' = (0.5, -0.5, -1, 1)'$. As before, we allow φ_* to vary to examine the effects of serial correlation.

The long-run and short-run models are specified as: $y_t = \alpha + \boldsymbol{\lambda}' \boldsymbol{x}_t^+ + \boldsymbol{\eta} \boldsymbol{x}_t + u_t$ and $\Delta y_t = \gamma + \rho \widehat{u}_{t-1} + \varphi_1 \Delta y_{t-1} + \boldsymbol{\pi}_0^+ \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_0^- \Delta \boldsymbol{x}_t^- + e_t$, where $\widehat{u}_t$ is the residual from the first-step estimation. In the first step, we estimate the long-run parameters using either TOLS or TFM, and then estimate the short-run parameters by OLS. We assess finite-sample performance in terms of bias and MSE based on $R = 5{,}000$ replications. The results are reported in Tables 7 and 8.

— Insert Tables 7 and 8 Here —

As the sample size, T, increases, the finite-sample bias of the TFM estimator for the long-run parameters becomes substantially smaller than that of TOLS. Because TFM yields asymptotically mixed-normal

estimators for the long-run parameters, its bias approaches zero much faster than TOLS. Moreover, the TFM estimator is generally more efficient, producing smaller MSE values across nearly all sample sizes. These results strongly support the use of TFM in the first step of the two-step NARDL procedure. For the short-run dynamic parameters, the finite-sample bias of the OLS estimator becomes negligible as T grows, regardless of whether TOLS or TFM is used in the first step. Similarly, the MSEs of the short-run estimator is broadly comparable across both approaches, especially for large samples. This suggests that the efficiency gains from TFM primarily accrue to the estimation of the long-run parameters, while short-run estimates are robust to the choice of first-step estimator.

Testing Restrictions on the Long-Run Parameters. We confine our attention to the case where the TFM estimator is used in the first step. We set the parameter configuration as $(\alpha_*, \gamma_*, \rho_*, \varphi_*, \kappa_*) = (0, 0, -1, \varphi_*, 0.5)$, $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})' = (-1, 0.5, 0.75, -1.5)'$, and $(\boldsymbol{\pi}_{0*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime})' = (0.5, -0.5, -1, 1)'$. We test the null hypothesis $\dot{H}_0^{(\ell)} : \boldsymbol{\iota}_2'\boldsymbol{\beta}_*^{+} = -0.50$ and $\boldsymbol{\iota}_2'\boldsymbol{\beta}_*^{-} = -0.75$ against the alternative $\dot{H}_1^{(\ell)} : \boldsymbol{\iota}_2'\boldsymbol{\beta}_*^{+} \neq -0.50$ or $\boldsymbol{\iota}_2'\boldsymbol{\beta}_*^{-} \neq -0.75$ while allowing φ_* to vary over -0.3, -0.1, 0, 0.1 and 0.3.

The simulation results reported in Table 9 show that, as T increases, the distribution of the Wald statistic is well approximated by the $\mathcal{X}_2^2$ distribution. For $\varphi_* = 0.3$, larger sample sizes are required to achieve a satisfactory approximation. To examine the empirical power, we test $\dot{H}_0^{(\ell)} : \boldsymbol{\iota}_2'\boldsymbol{\beta}_*^{+} = -0.40$ and $\boldsymbol{\iota}_2'\boldsymbol{\beta}_*^{-} = -0.65$ vs. $\dot{H}_1^{(\ell)} : \boldsymbol{\iota}_2'\boldsymbol{\beta}_*^{+} \neq -0.40$ or $\boldsymbol{\iota}_2'\boldsymbol{\beta}_*^{-} \neq -0.65$. The results in Table 10 confirm that the Wald test is consistent, with empirical rejection rates converging to 100% across all values of φ_*.

— Insert Tables 9 and 10 Here —

Testing Restrictions on the Short-Run Parameters. We generate data using the DGP parameters $(\alpha_*, \gamma_*, \rho_*, \varphi_*, \kappa_*) = (0, 0, -1, \varphi_*, 0.5)$, $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})' = (-1, 0.5, 0.75, -1.5)'$, and $(\boldsymbol{\pi}_{0*}^{+\prime}, \boldsymbol{\pi}_{0*}^{-\prime})' = (0.5, 0.2, 0.5, 0.2)'$. As before, we allow φ_* to vary to capture different degrees of autocorrelation. We again focus on the case where the long-run parameters are estimated by TFM and construct $\widehat{u}_t$ prior to estimating the short-run parameters by OLS. We test the null hypothesis of short-run symmetry $\dot{H}_0^{(s)} : \boldsymbol{\pi}_{0*}^{+} - \boldsymbol{\pi}_{0*}^{-} = \mathbf{0}$ against the alternative hypothesis $\dot{H}_1^{(s)} : \boldsymbol{\pi}_{0*}^{+} - \boldsymbol{\pi}_{0*}^{-} \neq \mathbf{0}$ using the Wald statistic with a heteroskedasticity-consistent covariance estimator $\widehat{\boldsymbol{\Omega}}_T$.

The simulation results in Table 11 show that the finite-sample distribution of the Wald statistic is well approximated by the $\mathcal{X}_2^2$ distribution. Its empirical size is close to the nominal level for large T and exhibits little sensitivity to φ_*, even in small samples. To examine the empirical power of the test, we retain the same DGP and test $\ddot{H}_0^{(s)} : \boldsymbol{\pi}_{0*}^{+} - \boldsymbol{\pi}_{0*}^{-} = 0.3\boldsymbol{\iota}$ vs. $\ddot{H}_1^{(s)} : \boldsymbol{\pi}_{0*}^{+} - \boldsymbol{\pi}_{0*}^{-} \neq 0.3\boldsymbol{\iota}$. The results shown in Table 12 indicate

that the Wald test is consistent, with empirical rejection rates converging to 100% as T increases. Moreover, the power of the test is largely unaffected by the degree of autocorrelation captured by φ_*.

— Insert Tables 11 and 12 Here —

6 Asymmetric Relationship between R&D Intensity and Investment

Following Schumpeter's seminal 1942 work on creative destruction, a substantial literature on R&D activities has emerged. However, the potentially asymmetric relationship between R&D expenditure and physical investment has received limited attention. The product life cycle literature distinguishes between early-stage R&D—often referred to as innovative R&D—that is typically associated with new product development, and later-stage R&D—commonly termed managerial R&D—that focuses on scaling production and improving efficiency (e.g., Gort and Wall, 1986; Audretsch, 1987). Early-stage (innovative) R&D focused on the development of new products or technologies, often leading to large-scale investment. By contrast, later-stage (managerial) R&D emphasizes production efficiency and cost reduction, such that this type of R&D expenditure does not exceed the expected increase in output and generally resulting in smaller-scale investments.[7] Overall, R&D expenditure tends to rise sharply during the early stage before leveling off or declining in the later stage.

Our empirical specification builds on two stylized features of innovative and managerial R&D expenditures highlighted in the theoretical framework presented in Section D.2 of the Online Supplement. First, innovative R&D, which focuses on product innovation, typically involves large-scale expenditures relative to output. This implies that R&D spending is expected to grow faster than output during the early stage, when start-up costs are high and production scale is small. Second, managerial R&D emphasizes production efficiency, so its scale is generally smaller than output.

Let r_t denote R&D intensity in period t, defined as the ratio of aggregate R&D expenditure to GDP. Because aggregate R&D expenditure reflects the full spectrum of R&D activities in the economy, the sign of Δr_t may indicate the relative prevalence of innovative versus managerial R&D. If $\Delta r_t > 0$, R&D expenditure grows at least as fast as output, suggesting the prevalence of innovative R&D. Conversely, if $\Delta r_t < 0$, output grows faster than R&D expenditure, indicating the prevalence of managerial R&D. Given the different characteristics of innovative and managerial R&D, we derive a theoretical prediction that innovative R&D acts as a complement to physical investment, while managerial R&D serves as a substitute. This

[7] Several studies differentiate between innovative and managerial R&D and examine their distinct effects on economic outcomes (e.g., Klepper, 1996; Zif and McCarthy, 1997; Agarwal and Audretsch, 2001; Comin and Philippon, 2005; Aghion et al., 2009; Chung and Shin, 2020).

gives rise to the testable hypothesis that innovative R&D expenditure is positively related to investment due to complementarity, while the relationship between managerial R&D expenditure and investment may be negative due to their nature as substitutes (see Section D.2 of the Online Supplement).

We examine the asymmetric relationship between R&D intensity and investment using quarterly U.S. data from 1960Q1 to 2019Q4, obtained from the Federal Reserve Economic Data (FRED) service at the Federal Reserve Bank of St. Louis. R&D intensity (r_t) is defined as 100 times the ratio of seasonally adjusted nominal R&D expenditure to seasonally adjusted nominal GDP. Investment (i_t) is measured as the logarithm of seasonally adjusted real gross private domestic investment (GPDI) in 2012 prices. The National Income and Product Accounts (NIPAs) separately aggregate R&D expenditure and GPDI, ensuring no double-counting of R&D expenditure.[8]

The Phillips and Perron (1988) unit root test results indicate that both r_t and i_t are nonstationary. Consequently, we report descriptive statistics for the first differences of both series (see Section E in the Online Supplement). While the growth rate of R&D intensity is approximately normally distributed around zero, the growth rate of GPDI exhibits a non-zero mean, negative skewness, and excess kurtosis. This suggests that i_t follows a unit-root process with a deterministic drift, whereas r_t is a driftless unit-root process.[9] As discussed in Section D.3 of the Online Supplement, the NARDL model can accommodate an asymmetric cointegrating relationship between two integrated variables with different drift characteristics without explicitly including a deterministic trend in the specification.

We estimate the NARDL model using the two-step procedure outlined in Section 3, applying the FM estimator in the first step. Based on the Akaike Information Criterion (AIC), we select a lag order of two and specify the following NARDL(2,2) error-correction model:

$$\Delta i_t = \gamma_* + \rho_* u_{t-1} + \varphi_* \Delta i_{t-1} + \pi_{0*}^{+} \Delta r_t^{+} + \pi_{0*}^{-} \Delta r_t^{-} + \pi_{1*}^{+} \Delta r_{t-1}^{+} + \pi_{1*}^{-} \Delta r_{t-1}^{-} + e_t, \tag{3}$$

where the long-run relationship between i_t and r_t is given by

$$i_t = \alpha_* + \beta_*^{+} r_t^{+} + \beta_*^{-} r_t^{-} + u_t.$$

Applying the Phillips and Perron (1988) unit root test to the residuals, we reject the null hypothesis of no cointegration at the 1% significance level, confirming that there exists an asymmetric long-run relationship

[8] The Bureau of Economic Analysis constructs a partial R&D satellite account and revises the NIPAs by treating R&D expenditure as part of investment (Fraumeni and Okubo, 2005).

[9] Univariate time-series regressions confirm that the time trend coefficient is statistically significant for i_t, but not for r_t.

between i_t, r_t^+ and r_t^-.

The long-run coefficients reported in Table 13 are not only different in magnitude, but also in sign: $\widetilde{\beta}_*^+$ is positive and $\widetilde{\beta}_*^-$ is negative and both are highly significant. Moreover, the null hypothesis of long-run symmetry is strongly rejected.[10] These findings indicate that an increase in R&D spending equivalent to 1% of GDP is associated with a 2.7% increase in real investment in the long run when R&D growth exceeds GDP growth (i.e., when innovative R&D dominates). Conversely, when R&D grows more slowly than GDP (i.e., when managerial R&D dominates), an increase in R&D spending of 1% of GDP reduces real investment by 6.4%.[11] This is consistent with our theoretical prediction that $\beta_*^+ > 0$ due to the complementarity between innovative R&D expenditure and investment, while $\beta_*^- > 0$ is consistent with managerial R&D and investment acting as substitutes. Notably, the substitution effect appears stronger than the complementarity effect in the long run.

From the short-run dynamic estimation results reported in Table 14, we find no evidence of residual autocorrelation up to order 4, as indicated by the Breusch–Godfrey LM test ($p = 0.39$). However, the Breusch–Pagan LM test rejects the null of conditional homoskedasticity ($p = 0.035$). Thus, we report robust standard errors based on the HAC covariance matrix estimator. The results suggest that disequilibrium errors are corrected significantly at a rate of 6.8% per quarter, while no evidence of short-run asymmetry is found.[12]

— Insert Tables 13 and 14 Here —

In Figure 1, we report the cumulative dynamic multipliers associated with a 1% increase in R&D intensity (as a share of GDP) under two regimes: when R&D grows faster than GDP (innovative R&D) and when it grows more slowly (managerial R&D). We also report the difference between these multipliers as a measure of asymmetry at each horizon, along with 95% confidence intervals obtained from 5,000 iterations of a moving block bootstrap procedure with a block length of $T^{1/3}$.[13] Panel (a) shows that, under the innovative R&D regime, a 1% increase in R&D intensity initially reduces real investment, with a peak decline of 6.7% after one quarter. This short-run contraction is consistent with the concept of creative destruction—

[10] Due to the reparamerization, the Wald test of $H_0 : \beta_*^+ = \beta_*^-$ versus $H_1 : \beta_*^+ \neq \beta_*^-$ is equivalent to a t-test of $H_0 : \lambda_* = 0$ versus $H_1 : \lambda_* \neq 0$, which yields a p-value very close to zero.

[11] The average R&D intensity in the U.S. over the sample period is 2.69% of GDP with a standard deviation of 0.2%. Thus, a shock equivalent to 1% of GDP is relatively large by historical standards.

[12] The Wald test for the of impact symmetry, $H_0 : \pi_{0*}^+ = \pi_{0*}^-$ vs. $H_1 : \pi_{0*}^+ \neq \pi_{0*}^-$, returns a p-value of 0.677. Moreover, the null hypothesis of additive short-run symmetry, $H_0 : \pi_{0*}^+ + \pi_{1*}^+ = \pi_{0*}^- + \pi_{1*}^-$, is not rejected against $H_1 : \pi_{0*}^+ + \pi_{1*}^+ \neq \pi_{0*}^- + \pi_{1*}^-$, with a p-value of 0.146.

[13] The moving block bootstrap is well-suited to time series applications, as it preserves local dependence by resampling blocks of adjacent observations. We tested a range of alternative block lengths to verify that the empirical confidence intervals for the cumulative dynamic multipliers do not depend critically on the chosen block length. Further details are available from the authors on request.

large innovative R&D expenditures focused on new product development create some obsolescence in the existing capital stock and reduce incentives to invest in these technologies. The dynamic multiplier then rises steadily, turning positive after eight quarters as new technologies mature and scale. In the long run, the impact of innovative R&D expenditure on real investment reaches the long-run multiplier value of 2.7%. Panel (b) indicates that, when managerial R&D dominates, a 1% increase in R&D intensity leads to an immediate decline of 3.6% in real investment, reflecting a short-term substitution effect. Thereafter, the dynamic multiplier remains statistically indistinguishable from zero until horizon 11, after which it converges to the long-run multiplier value of –6.4%. Thus, an environment favoring managerial R&D is associated with reduced real investment over time, as efficiency gains increase returns on existing capital rather than incentivizing new investment.

— Insert Figure 1 Here —

In Figure 1(c), we illustrate the asymmetry between the two regimes by plotting the difference between the cumulative dynamic multipliers associated with innovative and managerial R&D expenditures along with 95% bootstrap confidence intervals. In the short run, we observe substantial negative asymmetry that is significant at the 10% level, indicating that real investment responds more strongly to innovative R&D than to managerial R&D. In the long run, however, the asymmetry becomes positive as the substitution effect of managerial R&D outweighs the complementary effect of innovative R&D. This pattern aligns with the product life cycle. Large innovative R&D expenditures often occur early, disrupting existing products, while managerial R&D occurs later and focuses on scaling production and delivering efficiency gains. Collectively, these results have important implications for the endogenous growth literature, where linear functional forms are routinely imposed to characterize the relationship between the stock of knowledge and R&D intensity (e.g. Romer, 1990). Our findings suggest that the imposition of a linear functional form may be overly restrictive and potentially misleading, given the dynamic and asymmetric nature of R&D effects across regimes and horizons.

Next, we present additional estimation results obtained via the single-step OLS procedure popularized by SYG, which we summarize using cumulative dynamic multipliers in Figure 2. This exercise provides a direct comparison between the single-step approach and our two-step framework. We find that both procedures yield similar estimation and testing results, suggesting that they may be used interchangeably in practice. However, the two-step framework offers greater precision in estimating the long-run parameters, because it is not influenced by nuisance parameters. This improved precision enhances its ability to detect long-run asymmetry, particularly in small samples. This represents an important practical benefit of our two-step

estimation framework, 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.

— Insert Figure 2 Here —

The contrasting effects of innovative and managerial R&D on real investment have important policy implications. Policies that incentivize innovative R&D—such as innovation subsidies or tax benefits—can foster long-term capital formation and technological progress. However, our results indicate that such policies may generate a short-run contraction in real investment, due to resource reallocation by firms and the obsolescence of existing capital. Policies to ease these transitional costs include enhanced depreciation allowances for obsolete capital assets, incentives to invest in complementary assets and retraining programs for affected workers. Meanwhile, policies that incentivize managerial R&D may achieve efficiency gains but risk weakening physical investment in the long-run, as efficiency improvements reduce the need for new capital. This may drag on overall economic growth, as private investment is an important component of GDP. Collectively, our results suggest that policymakers should seek a balanced policy setting that encourages innovative R&D while ensuring that managerial R&D does not crowd out investment in new technologies.

7 Concluding Remarks

In this paper, we analyze the potentially asymmetric relationship between R&D intensity and physical investment. Building on documented differences between innovative and managerial R&D, we develop a theoretical model that predicts an asymmetric relationship whereby innovative R&D expenditure is complementary to investment, while managerial R&D serves as a substitute.

To test this hypothesis, we specify a NARDL model in which real investment is regressed on the positive and negative partial sums of R&D intensity. While NARDL models have been widely applied to study asymmetries (see Cho et al., 2023b, for a recent survey), the asymptotic theory for the single-step estimator remains incomplete due to asymptotic singularity issues. To address this, we introduce a tractable two-step estimation procedure. In the first step, we estimate a reparameterized long-run relationship using the FM estimator, which eliminates asymptotic bias, accounts for serial correlation and facilitates standard inference. In the second step, the short-run dynamic parameters are estimated by OLS, treating the error-correction term as given. We derive the asymptotic distributions of the two-step NARDL estimators and develop Wald tests for inference on the short- and long-run parameters. A suite of Monte Carlo simulations demonstrates that our asymptotic results offer good approximations in finite samples.

Employing quarterly U.S. data covering the period from 1960q1 to 2019q4, we find compelling evidence of asymmetry in the relationship between R&D intensity and physical investment. In the long run, investment responds positively to R&D expenditures when their growth rate exceeds that of GDP but negatively when they grow more slowly than GDP. This supports our theoretical prediction that innovative R&D complements investment, while managerial R&D substitutes for it. Furthermore, we find that investment is more sensitive to changes in R&D intensity when managerial R&D is prevalent, highlighting the importance of distinguishing between R&D types in policy design.

Our work opens several avenues for future research. Methodologically, there is scope to extend our approach to accommodate trending regressors or to estimate unknown threshold parameters in the construction of the partial sum processes. Empirically, our findings motivate the study of asymmetries in other areas of the endogenous growth literature, such as the knowledge production function, where linear specifications are routinely imposed. Relaxing these assumptions may yield richer insights into the dynamics of innovation and growth.

References

AGARWAL, R. AND D. B. AUDRETSCH (2001): "Does Entry Size Matter? The Impact of the Life Cycle and Technology on Firm Survival," *Journal of Industrial Economics*, 49, 21–43. 2, 22, 7

AGHION, P., R. BLUNDELL, R. GRIFFTH, P. HOWITT, AND S. PRANTL (2009): "The Effects of Entry on Incumbent Innovation and Productivity," *Review of Economics and Statistics*, 91, 20–32. 2, 22, 7

AUDRETSCH, D. B. (1987): "An Empirical Test of the Industry Life Cycle," *Weltwirtschaftliches Archiv*, 123, 297–308. 22, 6

BARATI, M. AND H. FARIDITAVANA (2020): "Asymmetric Effect of Income on the US Healthcare Expenditure: Evidence from the Nonlinear Autoregressive Distributed Lag (ARDL) Approach," *Empirical Economics*, 58, 1979–2008. 1

BAUSSOLA, M. (2000): "The Causality Between R&D and Investment," *Economics of Innovation and New Technology*, 9, 385–399. 7

BOX, M., K. GRATZER, AND X. LIN (2019): "The Asymmetric Effect of Bankruptcy Fraud in Sweden: A Long-Term Perspective," *Journal of Quantitative Criminology*, 35, 287–312. 1

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. 1

CHIAO, C. (2001): "The Relationship between R&D and Physical Investment of Firms in Science-Based Industries," *Applied Economics*, 33, 23–35. 7

CHO, J., M. GREENWOOD-NIMMO, AND Y. SHIN (2023a): "Nonlinear Autoregressive Distributed Lag Model Estimation Using Data with Drifts," Working paper, School of Economics, Yonsei University, Seoul. 12

——— (2023b): "Recent Developments of the Autoregressive Distributed Lag Modelling Framework," *Journal of Economic Surveys*, 37, 7–32. 1, 26

CHUNG, D. AND D. SHIN (2020): "When do Firms Invest in R&D? Two Types of Performance Feedback and Organizational Search in the Korean Shipbuilding Industry," *Asian Business & Management*, 20, 583–617. 2, 22, 7

COMIN, D. AND T. PHILIPPON (2005): "The Rise in Firm-level Volatility: Causes and Consequences," *NBER/Macroeconomics Annual*, 20, 167–201. 22, 7

EBERHARDT, M. AND A. F. PRESBITERO (2015): "Public Debt and Growth: Heterogeneity and Non-Linearity," *Journal of International Economics*, 97, 45–58. 1

ENGLE, R. F. AND C. W. GRANGER (1987): "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276. 6

FERRIS, J. S., S. L. WINER, AND D. E. OLMSTEAD (2020): "A Dynamic Model of Political Party Equilibrium: The Evolution of ENP in Canada, 1870 – 2015," Working Paper 8387, CESifo, Munich. 1

FRAUMENI, B. M. AND S. OKUBO (2005): "R&D in the National Income and Product Accounts: A First Look at Its Effect on GDP," in *Measuring Capital in the New Economy*, ed. by J. H. C. Corrado and D. Sichel, Chicago (IL): The University of Chicago Press, 275–322. 23

FUDENBERG, D., R. GILBERT, J. STIGLITZ, AND J. TIROLE (1983): "Preemption, Leapfrogging and Competition in Patent Races," *European Economic Review*, 22, 3–31. 7

GORT, M. AND R. A. WALL (1986): "The Evolution of Technologies and Investment in Innovations," *Economic Journal*, 96, 741–757. 22, 6

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. 17

GROSSMAN, G. M. AND C. SHAPIRO (1987): "Dynamic R&D Competition," *Economic Journal*, 97, 372–387. 7

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. 1

HARRIS, C. J. AND J. S. VICKERS (1987): "Racing with Uncertainty," *Review of Economic Studies*, 54, 1–21. 7

HE, Z. AND F. ZHOU (2018): "Time-Varying and Asymmetric Eeffects of the Oil-Specific Demand Shock on Investor Sentiment," *PLOS ONE*, 13, e0200734. 1

KAMIEN, M. I. AND N. L. SCHWARTZ (1972): "Timing of Innovations under Rivalry," *Econometrica*, 40, 43–60. 7

KLEPPER, S. (1996): "Entry, Exit, Growth, and Innovation over the Product Life Cycle," *American Economic Review*, 86, 562–583. 22, 7

——— (1997): "Industry Life Cycles," *Industrial and Corporate Change*, 6, 145–182. 7

LACH, S. AND R. ROB (1996): "R&D Investment and Industry Dynamics," *Journal of Economics and Management Strategy*, 5, 217–249. 7

LACH, S. AND M. SCHANKERMAN (1989): "Dynamics of R&D and Physical Investment in the Scientific Sector," *Journal of Political Economy*, 97, 880–904. 7

NEWEY, W. K. AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. 8

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. 1

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. 1

PHILLIPS, P. AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346. 23, 12, 15

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. 2, 6, 8

REINGANUM, J. F. (1982): "A Dynamic Game of R&D: Patent Protection and Competitive Behavior," *Econometrica*, 50, 671–688. 7

ROMER, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98, 71–102. 2, 25

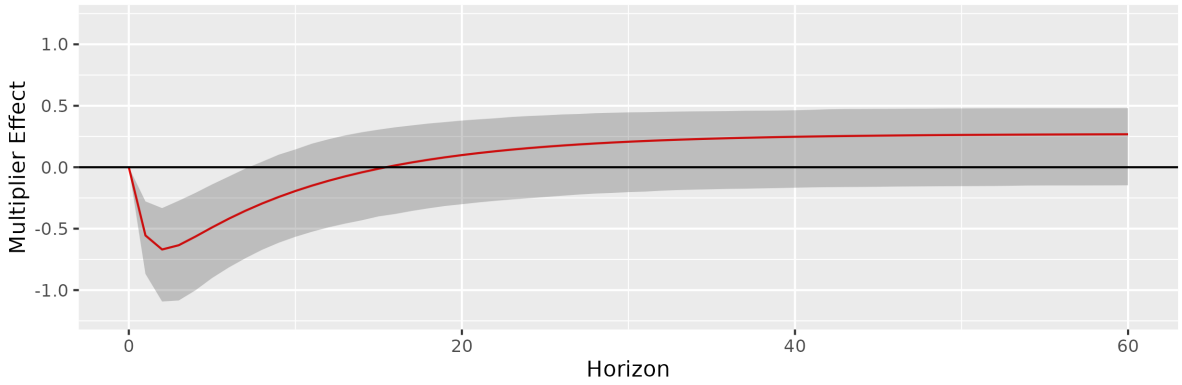
SCHMOOKLER, J. (1966): *Innovation and Economic Growth*, Cambridge (MA): Harvard University Press. 7

SCHUMPETER, J. A. (1942): *Capitalism, Socialism, and Democracy*, New York (NY): Harper & Row. 22, 6

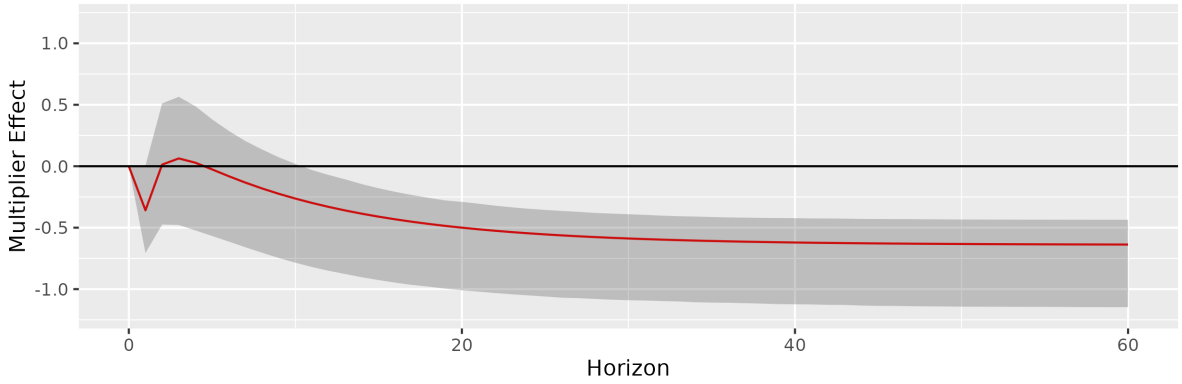
SHIN, Y., B. YU, AND M. J. GREENWOOD-NIMMO (2014): "Modelling Asymmetric Cointegration and Dynamic Multipliers in a Nonlinear ARDL Framework," in *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*, ed. by W. Horrace and R. Sickles, New York (NY): Springer Science & Business Media, 281–314. 1, 4, 17, 25, 12

UTTERBACK, J. AND W. Y. ABERNATHY (1975): "A Dynamic Model of Process and Product Innovation," *Omega*, 3, 639–659. 6

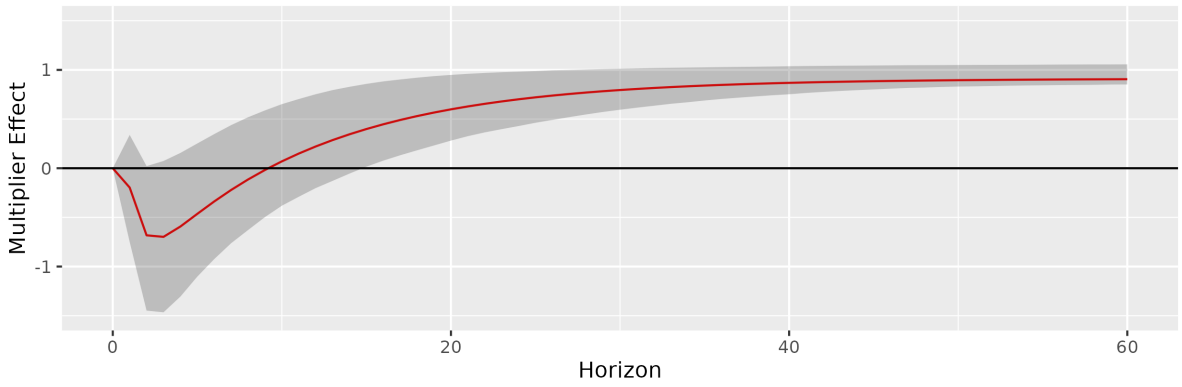
ZIF, J. AND D. J. MCCARTHY (1997): "The R&D Cycle: The Influence of Product and Process R&D on Short-Term ROI," *IEEE Transactions on Engineering Management*, 44, 114–123. 22, 7



(a) Cumulative response of i_{t+h} to a +1 unit shock to r_t^+ in period 1

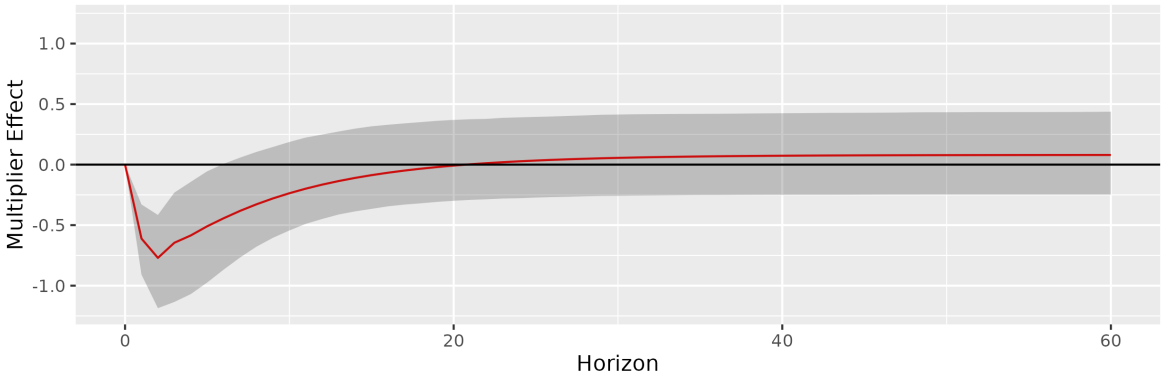


(b) Cumulative response of i_{t+h} to a +1 unit shock to r_t^- in period 1

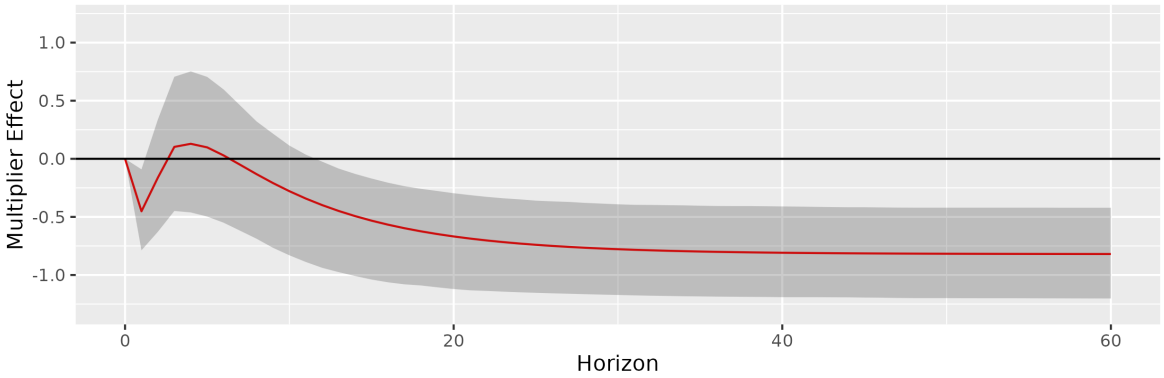


(c) Asymmetry across horizons

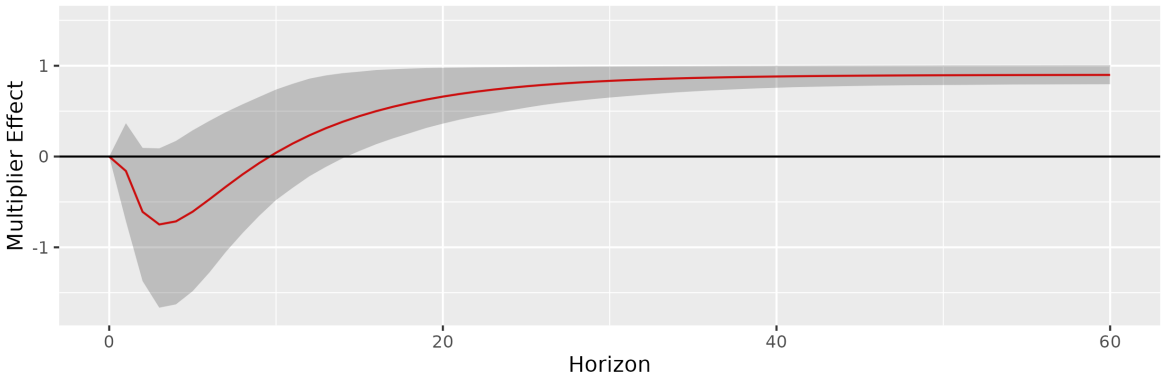
Figure 1: CUMULATIVE DYNAMIC MULTIPLIERS BASED ON THE TWO-STEP ESTIMATOR. Panels (a) and (b) present the cumulative dynamic multipliers with respect to unit shocks to r_t^+ and r_t^-, respectively. Panel (c) shows the asymmetry at each horizon, i.e., the difference between the cumulative dynamic multipliers in panel (a) and those in panel (b). We also report empirical 95% confidence intervals that are obtained using a moving block bootstrap procedure with 5,000 replications.



(a) Cumulative response of i_{t+h} to a +1 unit shock to r_t^+ in period 1



(b) Cumulative response of i_{t+h} to a +1 unit shock to r_t^- in period 1



(c) Asymmetry across horizons

Figure 2: CUMULATIVE DYNAMIC MULTIPLIERS BASED ON THE ONE-STEP ESTIMATOR. Panels (a) and (b) present the cumulative dynamic multiplier effects with respect to unit shocks to r_t^+ and r_t^-, respectively, occurring in period 1. Panel (c) shows the asymmetry at each horizon, i.e., the difference between the cumulative dynamic multipliers in panel (a) and those in panel (b). We also report empirical 95% confidence intervals that are obtained using a moving block bootstrap procedure with 5,000 replications.

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

Table 1: FINITE SAMPLE BIAS OF THE TWO-STEP ESTIMATORS FOR $k = 1$. This table reports the finite sample biases when OLS/FM is used in the first step and OLS is used in the second step. The data is generated as $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + \Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, 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 $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$.

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

Table 2: FINITE SAMPLE MEAN SQUARED ERROR (MSE) OF THE TWO-STEP ESTIMATORS FOR $k = 1$. This table reports the finite sample MSEs when OLS/FM is used in the first step and OLS is used in the second step. The data is generated as $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + \Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, 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 $(e_t, v_t)' \sim \text{IID}N(\mathbf{0}_2, \mathbf{I}_2)$.

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

Table 3: EMPIRICAL LEVELS OF THE WALD TEST FOR LONG-RUN SYMMETRY. This table reports the empirical level (in %) of the Wald test for the symmetry of the long-run parameters estimated by FM in the first step. The data is generated by $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + (1/3)\Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, 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. $H_1^{(\ell)} : \beta_*^+ - \beta_*^- \neq 0$.

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

Table 4: EMPIRICAL POWER OF THE WALD TEST FOR LONG-RUN SYMMETRY. This table shows the empirical power (in %) of the Wald test for the symmetry of the long-run parameters estimated by FM in the first step. The data is generated by $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + (1/3)\Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, 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. $H_1^{(\ell)} : \beta_*^+ - \beta_*^- \neq 0$.

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

Table 5: EMPIRICAL LEVELS OF THE WALD TEST FOR SHORT-RUN SYMMETRY. This table reports the empirical levels (in %) of the Wald test for the symmetry of the short-run parameters, where FM is used in the first step and OLS used in the second step. The data is generated as $\Delta y_t = -(2/3)u_{t-1} + \varphi_* \Delta y_{t-1} + (1/2)\Delta x_t^+ + (1/2)\Delta x_t^- + e_t$, 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. $H_1^{(s)} : H_0 : \pi_*^+ - \pi_*^- \neq 0$.

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

Table 6: EMPIRICAL POWER OF THE WALD TEST FOR SHORT-RUN SYMMETRY. This table reports the empirical rejection rates (in %) of the Wald test for the symmetry of the short-run parameters, where FM 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. $H_1^{(s)} : H_0 : \pi_*^+ - \pi_*^- \neq 0$.

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

Table 7: FINITE SAMPLE BIAS OF THE TWO-STEP ESTIMATORS FOR $k = 2$. This table reports the finite sample biases when OLS/FM is used in the first step and OLS is used in the second step. The data is generated as $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}^{+\prime}_{0*} \Delta \boldsymbol{x}^+_t + \boldsymbol{\pi}^{-\prime}_{0*} \Delta \boldsymbol{x}^-_t + e_t$, where $u_t := y_t - \boldsymbol{\beta}^{+\prime}_* \boldsymbol{x}^+_t - \boldsymbol{\beta}^{-\prime}_* \boldsymbol{x}^-_t$, $\Delta \boldsymbol{x}_t = 0.5 \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}'_t)' \sim \text{IID} N(\mathbf{0}_3, \mathbf{I}_3)$.

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

Table 8: FINITE SAMPLE MEAN SQUARED ERROR (MSE) OF THE TWO-STEP ESTIMATORS FOR $k = 2$. This table reports the finite sample MSEs when OLS/FM is used in the first step and OLS is used in the second step. The data is generated as $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}^{+\prime}_{0*} \Delta \boldsymbol{x}^+_t + \boldsymbol{\pi}^{-\prime}_{0*} \Delta \boldsymbol{x}^-_t + e_t$, where $u_t := y_t - \boldsymbol{\beta}^{+\prime}_* \boldsymbol{x}^+_t - \boldsymbol{\beta}^{-\prime}_* \boldsymbol{x}^-_t$, $\Delta \boldsymbol{x}_t = 0.5 \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}'_t)' \sim \text{IID} N(\mathbf{0}_3, \mathbf{I}_3)$.

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

Table 9: EMPIRICAL LEVELS OF THE WALD TEST FOR LONG-RUN SYMMETRY. This table reports the empirical level (in %) of the Wald test for the symmetry of the long-run parameters estimated by the TFM estimator in the first step. The data is generated by $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5 \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \text{IID} N(\mathbf{0}_3, \mathbf{I}_3)$. $\dot{H}_0^{(\ell)} : \boldsymbol{\iota}_2' \boldsymbol{\beta}_*^+ = -0.5$ and $\boldsymbol{\iota}_2' \boldsymbol{\beta}_*^- = -0.75$ vs. $\dot{H}_1^{(\ell)} : \boldsymbol{\iota}_2' \boldsymbol{\beta}_*^+ \neq -0.5$ or $\boldsymbol{\iota}_2' \boldsymbol{\beta}_*^- \neq -0.75$.

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

Table 10: EMPIRICAL POWER OF THE WALD TEST FOR LONG-RUN SYMMETRY This table shows the empirical power (in %) of the Wald test for the symmetry of the long-run parameter estimated by the TFM estimator in the first step. The data is generated by $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5 \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \text{IID} N(\mathbf{0}_3, \mathbf{I}_3)$. $\ddot{H}_0^{(\ell)} : \boldsymbol{\iota}_2' \boldsymbol{\beta}_*^+ = -0.4$ and $\boldsymbol{\iota}_2' \boldsymbol{\beta}_*^- = -0.65$ vs. $\ddot{H}_1^{(\ell)} : \boldsymbol{\iota}_2' \boldsymbol{\beta}_*^+ \neq -0.4$ or $\boldsymbol{\iota}_2' \boldsymbol{\beta}_*^- \neq -0.65$.

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

Table 11: EMPIRICAL LEVELS OF THE WALD TEST FOR SHORT-RUN SYMMETRY. This table reports the empirical levels (in %) of the Wald test for the symmetry of the short-run parameters, where the TFM estimator is used in the first step and OLS in the second step. The data is generated as $\Delta y_t = -u_{t-1} + \varphi_* \Delta y_{t-1} + \boldsymbol{\pi}_{0*}^{+\prime} \Delta \boldsymbol{x}_t^+ + \boldsymbol{\pi}_{0*}^{-\prime} \Delta \boldsymbol{x}_t^- + e_t$, where $u_t := y_t - \boldsymbol{\beta}_*^{+\prime} \boldsymbol{x}_t^+ - \boldsymbol{\beta}_*^{-\prime} \boldsymbol{x}_t^-$, $\Delta \boldsymbol{x}_t = 0.5 \Delta \boldsymbol{x}_{t-1} + \sqrt{1 - 0.5^2} \boldsymbol{v}_t$, and $(e_t, \boldsymbol{v}_t')' \sim \text{IID}N(\mathbf{0}_3, \mathbf{I}_3)$. $\dot{H}_0^{(s)} : \boldsymbol{\pi}_{0*}^+ - \boldsymbol{\pi}_{0*}^- = \mathbf{0}_2$ vs. $\dot{H}_1^{(s)} : \boldsymbol{\pi}_{0*}^+ - \boldsymbol{\pi}_{0*}^- \neq \mathbf{0}_2$.

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

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

	Estimate	S.E.
Intercept	7.453	0.381
β_*^+	0.271	0.133
β_*^-	-0.640	0.154

Table 13: THE LONG-RUN PARAMETER ESTIMATES. This table reports the long-run parameter estimates obtained by our two-step estimation procedure applied to quarterly observations from 1960q1 to 2019q4, where FM is used in the first step.

	Estimate	S.E.
γ_*	0.015	0.005
ρ_*	-0.068	0.016
φ_*	0.255	0.066
π_{0*}^+	-0.555	0.179
π_{1*}^+	-0.029	0.150
π_{0*}^-	-0.359	0.344
π_{1*}^-	0.482	0.176
Adjusted R^2	0.199	
$\mathcal{X}^2_{\text{S.Corr.}}$	0.385	
$\mathcal{X}^2_{\text{Hetero.}}$	0.035	

Table 14: THE SHORT-RUN DYNAMIC PARAMETER ESTIMATES. This table reports parameter estimates for the NARDL(2,2) ECM model obtained using the two-step procedure applied to quarterly observations from 1960q1 to 2019q4, where FM is used in the first step and OLS in the second step. The lag order is selected by AIC. The standard errors are evaluated using HAC covariance matrix estimation. $\mathcal{X}^2_{\text{S.Corr.}}$ and $\mathcal{X}^2_{\text{Hetero.}}$ denote the Breusch–Godfrey LM test for serial correlation (up to order four) and the Breusch–Pagan–Godfrey LM test for residual heteroskedasticity, respectively. We report asymptotic p-values for these two tests.

Online Supplement for

"Two-Step Nonlinear ARDL Estimation: Theory and Application"

by Jin Seo Cho, Matthew Greenwood-Nimmo and Yongcheol Shin

This Online Supplement is organized into five sections. In Sections A and B we provide the proofs of the main claims in the manuscript. In Section C, we provide a range of additional simulation results. In Section D, we develop the theory relating early-stage innovative and later-stage managerial R&D expenditures to physical investment and conduct a comparative static analysis to confirm our theoretical predictions. In Section E, we report additional estimation results.

A Preliminary Equations

We first provide some equations for an efficient exposition of our proofs. As they are already explained in the manuscript, we provide them without reiterating their motivation and derivations.

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

$$y_t = \delta_* t + \sum_{j=1}^{t} d_j; \tag{A.2}$$

$$\begin{aligned}\Delta y_t = {} & \rho_* y_{t-1} + (\theta_*^+ - \theta_*^-) x_{t-1}^+ + \theta_*^- x_{t-1} \\ & + \nu_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\pi_{j*}^+ \Delta x_{t-j}^+ + \pi_{j*}^- \Delta x_{t-j}^- \right) + e_t;\end{aligned} \tag{A.3}$$

$$y_t = \alpha_* + \lambda_* x_t^+ + \eta_* x_t + u_t; \tag{A.4}$$

$$u_{t-1} := y_{t-1} - \alpha_* - \beta_*^+ x_{t-1}^+ - \beta_*^- x_{t-1}^-; \tag{A.5}$$

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^{+\prime} \Delta \boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_{j*}^{-\prime} \Delta \boldsymbol{x}_{t-j}^- \right) + e_t; \tag{A.6}$$

$$\widehat{\boldsymbol{\varrho}}_T = \boldsymbol{\varrho}_* + \left(\sum_{t=1}^{T} \boldsymbol{q}_t \boldsymbol{q}_t' \right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{q}_t u_t \right); \tag{A.7}$$

$$\widehat{\boldsymbol{\zeta}}_T := \left(\sum_{t=1}^{T} \boldsymbol{h}_t \boldsymbol{h}_t'\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{h}_t \Delta y_t\right) = \boldsymbol{\zeta}_* + \left(\sum_{t=1}^{T} \boldsymbol{h}_t \boldsymbol{h}_t'\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{h}_t e_t\right). \tag{A.8}$$

B Proofs

Proof of Lemma 1. (*i*) By (A.1) and (A.2), we obtain the following results:

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

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

(*ii*) By (A.1) and (A.2), we note that:

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

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

(*iii*) We note that:

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

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

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

Proof of Lemma 2. (*i*) We note that:

- $T^{-2}\sum_{t=1}^{T}x_t^+ = T^{-1}\sum_{t=1}^{T}\mu_*^+(t/T) + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \frac{1}{2}\mu_*^+$;
- $T^{-3/2}\sum_{t=1}^{T}x_t = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta x_i) \Rightarrow \int \mathcal{B}_x$ using that $T^{-1/2}\sum_{i=1}^{[T(\cdot)]}\Delta x_i \Rightarrow \int_0^{(\cdot)} d\mathcal{B}_x$;
- $T^{-3}\sum_{t=1}^{T}x_t^+x_t^+ = T^{-1}\sum_{t=1}^{T}\mu_*^+\mu_*^+(t/T)^2 + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \frac{1}{3}\mu_*^+\mu_*^+$;
- $T^{-5/2}\sum_{t=1}^{T}x_t^+x_t = T^{-1}\sum_{t=1}^{T}\mu_*^+(t/T)(T^{-1/2}\sum_{i=1}^{t}\Delta x_t) + o_{\mathbb{P}}(1) \Rightarrow \mu_*^+\int r\mathcal{B}_x$; and
- $T^{-2}\sum_{t=1}^{T}x_tx_t = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta x_t)(T^{-1/2}\sum_{i=1}^{t}\Delta x_t) \Rightarrow \int \mathcal{B}_x^2$.

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

(*ii*) We note that:

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

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

Proof of Corollary 1. Given (A.7), the desired result follows from Lemma 4. ■

Proof of Theorem 1. Using the definition of $\widehat{\lambda}_T$, we have: $T\{(\widehat{\beta}_T^+ - \widehat{\beta}_T^-) - (\beta_*^+ - \beta_*^-)\} = O_{\mathbb{P}}(T^{-1/2})$, implying that the weak limit of $T(\widehat{\beta}_T^+ - \beta_*^+)$ is equivalent to that of $T(\widehat{\beta}_T^- - \beta_*^-)$. Furthermore, by Corollary 1, we obtain the desired result, $T(\widehat{\beta}_T^- - \beta_*^-) \Rightarrow \mathbf{S}\boldsymbol{\mathcal{Q}}^{-1}\boldsymbol{\mathcal{U}}$. ■

Proof of Lemma 3. Under Assumption 2, $\widetilde{\upsilon}_T \xrightarrow{\mathbb{P}} \upsilon_*$ and $(\widetilde{\sigma}_T^{(1,1)})^{-1}\widetilde{\sigma}_T^{(1,2)} \xrightarrow{\mathbb{P}} \boldsymbol{\nu}_* := (\sigma_*^{(1,1)})^{-1}\sigma_*^{(1,2)}$. Next, let $\dot{u}_t := u_t - \Delta x_t\boldsymbol{\nu}_*$, $\widetilde{\mathbf{U}}_T = \widetilde{\mathbf{D}}_T^{-1}\sum_{t=1}^{T}\{\boldsymbol{q}_t\dot{u}_t - \mathbf{S}'\upsilon_*\} + o_{\mathbb{P}}(1)$, then $\widetilde{\mathbf{U}}_T \Rightarrow [\int d\mathcal{B}_{\dot{u}}, \mu_*^+\int r d\mathcal{B}_{\dot{u}}, \int \mathcal{B}_x d\mathcal{B}_{\dot{u}}]'$, where $\mathcal{B}_{\dot{u}}(\cdot) := \tau_*\mathcal{W}_u(\cdot)$. Therefore, $\widetilde{\mathbf{U}}_T \Rightarrow \widetilde{\boldsymbol{\mathcal{U}}}$. ■

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

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

■

Proof of Lemma 4. The result is established by the ergodic theorem and the multivariate central limit theorem. ■

Proof of Lemma 5. The desired result is easily obtained from the proof of Lemma 1.

Proof of Theorem 3. (*i*) Given (A.8), we can combine Lemmas 4 (*i* and *ii*) and obtain the desired result.

(*ii*) Further, if it holds that $\mathbb{E}[e_t^2|\boldsymbol{h}_t] = \sigma_*^2$, then we have: $\boldsymbol{\Omega}_* = \sigma_*^2\boldsymbol{\Gamma}_*$ by Lemma 4(*iii*). Thus, Theorem 3(*i*) 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 6. (*i*) We note that:

- $T^{-2}\sum_{t=1}^{T} t = \frac{1}{2} + o(1)$;
- $T^{-3/2}\sum_{t=1}^{T}\widehat{\boldsymbol{m}}_t = T^{-3/2}\sum_{t=1}^{T}\boldsymbol{m}_t - T^{-2}\sum_{t=1}^{T} t(T^{-3}\sum_{t=1}^{T} t^2)^{-1}T^{-5/2}\sum_{t=1}^{T} t\boldsymbol{m}_t = T^{-1}\sum_{t=1}^{T} T^{-1/2}(\sum_{i=1}^{t}\Delta\boldsymbol{m}_i) - (T^{-2}\sum_{t=1}^{T} t)(T^{-3}\sum_{t=1}^{T} t^2)^{-1}(T^{-1}\sum_{t=1}^{T}(t/T)T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{m}_i) \Rightarrow \int \boldsymbol{\mathcal{B}}_m - \frac{3}{2}\int r\boldsymbol{\mathcal{B}}_m$ using the fact that $T^{-1/2}\sum_{i=1}^{[T(\cdot)]}\Delta\boldsymbol{m}_i \Rightarrow \int_0^{(\cdot)} d\boldsymbol{\mathcal{B}}_m$;
- $T^{-3/2}\sum_{t=1}^{T}\boldsymbol{x}_t = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{x}_i) \Rightarrow \int \boldsymbol{\mathcal{B}}_x$ using the fact that $T^{-1/2}\sum_{i=1}^{[T(\cdot)]}\Delta\boldsymbol{x}_i \Rightarrow \int_0^{(\cdot)} d\boldsymbol{\mathcal{B}}_x$;
- $T^{-3}\sum_{t=1}^{T} t^2 = \frac{1}{3} + o(1)$;
- $T^{-5/2}\sum_{t=1}^{T} t\widehat{\boldsymbol{m}}_t = T^{-5/2}\sum_{t=1}^{T} t\boldsymbol{m}_t - T^{-5/2}(\sum_{t=1}^{T} t^2)(\sum_{t=1}^{T} t^2)^{-1}(\sum_{t=1}^{T} t\boldsymbol{m}_t) = \mathbf{0}$;
- $T^{-5/2}\sum_{t=1}^{T} t\boldsymbol{x}_t = T^{-1}\sum_{t=1}^{T}(t/T)(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{x}_i) \Rightarrow \int r\boldsymbol{\mathcal{B}}_x$;
- $T^{-2}\sum_{t=1}^{T}\widehat{\boldsymbol{m}}_t\widehat{\boldsymbol{m}}_t' = T^{-2}\sum_{t=1}^{T}\boldsymbol{m}_t\boldsymbol{m}_t' - T^{-2}\sum_{t=1}^{T} t\,\boldsymbol{m}_t\,(\sum_{t=1}^{T} t^2)^{-1}\sum_{t=1}^{T} t\,\boldsymbol{m}_t' = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{m}_i)(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{m}_i)' - T^{-1}\sum_{t=1}^{T}((t/T)\,T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{m}_i)(T^{-3}\sum_{t=1}^{T} t^2)^{-1} T^{-1}\sum_{t=1}^{T}((t/T)T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{m}_i)' \Rightarrow \int \boldsymbol{\mathcal{B}}_m\boldsymbol{\mathcal{B}}_m' - 3\int r\boldsymbol{\mathcal{B}}_m\int r\boldsymbol{\mathcal{B}}_m'$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_t\widehat{\boldsymbol{m}}_t' = T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{m}_t' - T^{-2}\sum_{i=1}^{t} t\boldsymbol{x}_t(\sum_{i=1}^{t} t^2)^{-1}\sum_{t=1}^{T} t\boldsymbol{m}_t' = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{x}_i)(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{m}_i) - T^{-1}\sum_{t=1}^{T}(t/T)T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{x}_i(T^{-3}\sum_{t=1}^{T} t^2)^{-1}((t/T)T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{m}_i') \Rightarrow \int \boldsymbol{\mathcal{B}}_x\boldsymbol{\mathcal{B}}_m' - 3\int r\boldsymbol{\mathcal{B}}_x\int r\boldsymbol{\mathcal{B}}_m'$;
- $T^{-2}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t' = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{x}_t)(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{x}_t) \Rightarrow \int \boldsymbol{\mathcal{B}}_x\boldsymbol{\mathcal{B}}_x'$.

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

(*ii*) We note that:

- $T^{-1/2}\sum_{t=1}^{T} u_t \Rightarrow \int d\mathcal{B}_u$ using the fact that $T^{-1/2}\sum_{t=1}^{[T(\cdot)]} u_t \Rightarrow \int_0^{(\cdot)} d\mathcal{B}_u$;
- $T^{-3/2}\sum_{t=1}^{T} tu_t = T^{-1/2}\sum_{t=1}^{T}(t/T)u_t + o_{\mathbb{P}}(1) \Rightarrow \int r d\mathcal{B}_u$;
- $T^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{m}}_t u_t = T^{-1}\sum_{t=1}^{T} u_t(\boldsymbol{m}_t - t(\sum_{t=1}^{T} t^2)^{-1}\sum_{t=1}^{T} t\boldsymbol{m}_t) = T^{-1}\sum_{t=1}^{T} u_t\boldsymbol{m}_t - T^{-3/2}\sum_{t=1}^{T} u_t t(T^{-3}\sum_{t=1}^{T} t^2)^{-1}T^{-5/2}\sum_{t=1}^{T} t\boldsymbol{m}_t \Rightarrow \int \boldsymbol{\mathcal{B}}_m d\mathcal{B}_u + \boldsymbol{\upsilon}_{m*} - 3\int r d\mathcal{B}_u\int r\boldsymbol{\mathcal{B}}_m$ using the fact that $\boldsymbol{\upsilon}_{m*} := \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t}\mathbb{E}[\Delta\boldsymbol{m}_i u_t]$ is finite.

- $T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t u_t = T^{-1/2}\sum_{t=1}^{T}(T^{-1/2}\sum_{i=1}^{t}\Delta\boldsymbol{x}_i)u_t \Rightarrow \int \boldsymbol{\mathcal{B}}_x d\mathcal{B}_u + \boldsymbol{\upsilon}_{x*}$ using the fact that $\boldsymbol{\upsilon}_{x*} := \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t}\mathbb{E}[\Delta\boldsymbol{x}_i u_t]$ is finite.

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

Proof of Corollary 3. As it is straightforward to show that the first claim follows from Lemma 6, we focus on the second claim. As $T^{3/2}(\widehat{\xi}_T - \xi_{*T}) = O_{\mathbb{P}}(1)$ and $\xi_{*T} = \boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+ + \boldsymbol{\lambda}_*'\sum t\boldsymbol{m}_t(\sum t^2)^{-1}$, $T^{3/2}(\widehat{\xi}_T - \boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+ - \boldsymbol{\lambda}_*'\sum t\boldsymbol{m}_t(\sum t^2)^{-1}) = O_{\mathbb{P}}(1)$, where $T^{1/2}\sum t\boldsymbol{m}_t(\sum t^2)^{-1} \Rightarrow \frac{1}{3}\int r\boldsymbol{\mathcal{B}}_m$. Thus, it follows that $T^{1/2}(\widehat{\xi}_T - \boldsymbol{\lambda}_*'\boldsymbol{\mu}_*^+) = T^{1/2}\boldsymbol{\lambda}_*'\sum t\boldsymbol{m}_t\,(\sum t^2)^{-1}) + O_{\mathbb{P}}(T^{-1}) \Rightarrow 3\boldsymbol{\lambda}_*'\int r\boldsymbol{\mathcal{B}}_m$. ■

Proof of Theorem 4. This result is easily obtained from Corollary 3. ■

Proof of Lemma 7. Under Assumption 3, notice that $\bar{\boldsymbol{\upsilon}}_T \xrightarrow{\mathbb{P}} \ddot{\boldsymbol{\upsilon}}_* := [\boldsymbol{\upsilon}_{m*}', \boldsymbol{\upsilon}_{x*}']'$ and $(\bar{\boldsymbol{\Sigma}}_T^{(1,1)})^{-1}\bar{\boldsymbol{\sigma}}_T^{(1,2)} \xrightarrow{\mathbb{P}} \ddot{\boldsymbol{\nu}}_* := (\boldsymbol{\Sigma}_*^{(1,1)})^{-1}\boldsymbol{\sigma}_*^{(1,2)}$. Let $\mathring{u}_t := u_t - \boldsymbol{\ell}_t'\ddot{\boldsymbol{\nu}}_*$, $\bar{\mathbf{U}}_T = \ddot{\mathbf{D}}_T^{-1}\sum_{t=1}^{T}\{\boldsymbol{r}_t\mathring{u}_t - \bar{\mathbf{S}}'\ddot{\boldsymbol{\upsilon}}_*\} + o_{\mathbb{P}}(1)$, then $\bar{\mathbf{U}}_T \Rightarrow [\int d\mathcal{B}_{\mathring{u}}, \int r d\mathcal{B}_{\mathring{u}}, \int \boldsymbol{\mathcal{B}}_m' d\mathcal{B}_{\mathring{u}} - 3\int r d\mathcal{B}_{\mathring{u}}\int r\boldsymbol{\mathcal{B}}_m', \int \boldsymbol{\mathcal{B}}_x' d\mathcal{B}_{\mathring{u}}]'$, where $\mathcal{B}_{\mathring{u}}(\cdot) := \ddot{\tau}\mathcal{W}_u(\cdot)$. Therefore, $\bar{\mathbf{U}}_T \Rightarrow \bar{\boldsymbol{\mathcal{U}}}$. ■

Proof of Theorem 5. Given that $\ddot{\mathbf{D}}_T(\bar{\boldsymbol{\varpi}}_T - \bar{\boldsymbol{\varpi}}_{*T}) = (\ddot{\mathbf{D}}_T^{-1}(\sum_{t=1}^{T}\boldsymbol{r}_t\boldsymbol{r}_t')\ddot{\mathbf{D}}_T^{-1})^{-1}\bar{\mathbf{U}}_T$, Lemma 7 establishes the first claim. Second, $T[(\bar{\boldsymbol{\beta}}_T^+ - \boldsymbol{\beta}_*^+)', (\bar{\boldsymbol{\beta}}_T^- - \boldsymbol{\beta}_*^-)']' = \ddot{\mathbf{S}}\ddot{\mathbf{D}}_T(\bar{\boldsymbol{\varpi}}_T - \bar{\boldsymbol{\varpi}}_{*T}) \Rightarrow \ddot{\mathbf{S}}\boldsymbol{\mathcal{R}}^{-1}\bar{\boldsymbol{\mathcal{U}}}$, as desired. ■

Proof of Theorem 6. By Corollary 2 we have: $T^{3/2}(\widetilde{\boldsymbol{\lambda}}_T - r) \Rightarrow \mathbf{S}\boldsymbol{\mathcal{Q}}^{-1}\widetilde{\boldsymbol{\mathcal{U}}}$ under H_0''. By Lemma 2(*i*) we have: $\widehat{\mathbf{Q}}_T := \widetilde{\mathbf{D}}_T^{-1}(\sum_{t=1}^{T}\boldsymbol{q}_t\,\boldsymbol{q}_t')\widetilde{\mathbf{D}}_T^{-1} \Rightarrow \boldsymbol{\mathcal{Q}}$. Further, under Assumption 2(*i*), we have: $\widetilde{\tau}_T^2 = \tau_*^2 + o_{\mathbb{P}}(1)$. Given the mixed normal distribution of the FM estimator of the long-run parameter in Corollary 2, it follows that $\mathcal{W}_T^{(\ell)} \overset{\text{A}}{\sim} \mathcal{X}_1^2$ under $\mathcal{H}_0''$.

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

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

Proof of Theorem 7. Due to its similarity to the standard case, we omit the proof. ■

Proof of Theorem 8. Due to its similarity to the standard case, we omit the proof. ■

C Additional Simulation Results

In this section, we provide additional simulation evidence based on the estimated model parameters reported in Section 6. We generate data using the following NARDL(2,2) DGP: $\Delta y_t = \gamma_* + \rho_* u_{t-1} + \varphi_* \Delta y_{t-1} + \pi^+_{0*}\Delta x^+_t + \pi^+_{1*}\Delta x^+_{t-1} + \pi^-_{0*}\Delta x^-_t + \pi^-_{1*}\Delta x^-_{t-1} + e_t$, where $u_{t-1} := y_{t-1} - \alpha_* - \beta^+_* x^+_{t-1} - \beta^-_* x^-_{t-1}$, $\Delta x_t := \kappa_* \Delta x_{t-1} + \sqrt{1-\kappa^2_*} v_t$, and $(e_t, v_t)' \sim$ IID$N(\mathbf{0}_2, \mathbf{I}_2)$. We set $(\alpha_*, \beta^+_*, \beta^-_*, \gamma_*, \rho_*, \varphi_*, \pi^+_{0*}, \pi^+_{1*}, \pi^-_{0*}, \pi^-_{1*}, \kappa_*)$ = (7.453, 0.271, -0.640, 0.015, ρ_*, 0.255, -0.555, -0.029, -0.359, 0.482, 0.5). Except for ρ_* and κ_*, these parameter values correspond to the empirical estimates reported in Tables 13 and 14 in the manuscript. To assess how the estimated parameters respond to the variation in ρ_*, we let $\rho_* = -0.1, -0.2, -0.3, -0.4, -0.5$ and -0.6. Finally, we set $\kappa_* = 1/2$ to generate serial correlation in Δx_t.

— Insert Tables A.1 and A.2 Here —

In parallel to Section 5, we compute the long-run parameters by OLS and FM, and then estimate the short-run parameters by OLS. We evaluate the finite-sample performance of the estimators in terms of their bias and MSE, which are obtained by repeating independent experiments 5,000 times. The results are reported in Tables A.1 and A.2, respectively. This exercise reinforces the findings from the Monte Carlo experiments reported in the manuscript. The finite sample bias and MSE of the FM estimator of β^+_* and β^-_* are smaller than those of OLS. The finite sample bias and MSE of the short-run parameter estimates is largely insensitive to the choice of first-step estimator. In addition, as ρ_* approaches zero, the finite sample bias and MSE of the estimators tend to increase. Overall, this experiment indicates that the parameter estimates reported in Tables 13 and 14 of the manuscript are likely to be relatively precise.

D The Asymmetric Relationship between R&D Intensity and Investment

In this section, we first review the literature on the link between R&D expenditure and physical investment. We then develop a theory that predicts an asymmetric relationship between the two.

D.1 Literature Review

Following Schumpeter's seminal 1942 work on creative destruction, a large literature has emerged on R&D activities. Important contributions include Utterback and Abernathy (1975), who find that R&D activity is conducted differently across the different stages of product life, the product life-cycle theory developed by Gort and Wall (1986) and Audretsch (1987), and the game-theoretic approach to R&D activity associated

with Kamien and Schwartz (1972), Reinganum (1982), Fudenberg et al. (1983), Grossman and Shapiro (1987) and Harris and Vickers (1987).

It is common to distinguish between two different stages of R&D activity. Early-stage (innovative) R&D expenditure focuses on the development of a new product or technology, leading to a subsequent large-scale investment. By contrast, later-stage (managerial) R&D expenditure focuses on improvements to production efficiency. Consequently, managerial R&D expenditure should not exceed the expected increase in output, which results in a smaller-scale investment than innovative R&D. Overall, R&D expenditure tends to increase sharply in the early stage before leveling off or decreasing at the later stage.

A number of theoretical studies distinguish between innovative and managerial R&D activities and their effects on other economic variables, including Klepper (1996, 1997) and Agarwal and Audretsch (2001). Comin and Philippon (2005) and Aghion et al. (2009) empirically examine the relationship between the entry and/or exit rate of firms and innovative R&D expenditure. Similar theories have been developed in other fields including engineering and management—for example, Zif and McCarthy (1997), who classify R&D activity into multiple stages following the product life-cycle theory (see also Chung and Shin, 2020).

However, one area in which the distinction between innovative and managerial R&D activity is yet to be fully investigated is the relationship between R&D expenditure and investment. Early studies (e.g. Schmookler, 1966) focus on the causal relationship between R&D expenditure and investment. Applying vector autoregressions (VARs) to firm- and industry-level data, Lach and Schankerman (1989) and Lach and Rob (1996) find that R&D expenditure Granger causes investment but not *vice versa*. However, using longer time series, Chiao (2001) documents a two-way causal relationship between the growth rates of the R&D expenditure and investment. Employing a vector error-correction (VEC) model, Baussola (2000) documents evidence in favor of unidirectional Granger causality from R&D expenditure to investment. These results should be treated with care because the failure to distinguish between innovative and managerial R&D activity undermines efforts to accurately capture the potentially asymmetric relationship between R&D expenditure and investment.

D.2 Theoretical Predictions

We propose a theoretical model relating innovative and managerial R&D expenditures to investment. Innovative R&D determines the scope of production, as it describes a research activity that creates a new product or technology through the discovery of a novel production function. The more innovative R&D activity, the larger the scale of production, suggesting that the limit of production activity is determined by the amount of innovative R&D activity. By contrast, managerial R&D does not create a new product, but instead produces

an existing product more efficiently, implying that less physical capital is required per unit of output.

Let k and y be the levels of physical capital and output, while we let c and s be the capital levels converted from innovative and managerial R&D expenditures, respectively. We assume that c is complementary to production activity conducted using physical capital while s is a substitute. Consider a production function embodying this mechanism as:

$$y = \min[c, k + s]. \tag{A.9}$$

The complementary relationship with c limits production activity, as output cannot be produced in excess of the level of innovative R&D activity. On the other hand, managerial R&D activity can produce capital s that substitutes for k.

We use a dynamic optimization approach and apply the q-theory of investment to examine how physical investment responds to external shocks to R&D expenditure. First, physical capital k_t is formed by accumulating physical investment i_t through $\dot{k}_t = i_t - \delta k_t$, where δ is the depreciation rate of the physical capital. Similarly, c_t and s_t are accumulated through $\dot{c}_t = r_t - \tau c_t$ and $\dot{s}_t = d_t - \gamma s_t$, where τ and γ denote depreciation rates of c_t and s_t, respectively. Next, consider the cost functions associated with converting R&D expenditures and physical investment into capital. Let $\kappa(r_t)$, $\xi(d_t)$ and $\phi(i_t)$ be the cost levels from innovative and managerial R&D expenditure and physical investment, respectively. We assume that the cost functions are convex with respect to R&D expenditures and physical investment, such that $\kappa'(\cdot) > 0$, $\xi'(\cdot) > 0$, $\phi'(\cdot) > 0$, $\kappa''(\cdot) > 0$, $\xi''(\cdot) > 0$, and $\phi''(\cdot) > 0$.[1]

To generalize the production function (A.9) into a differentiable function, we assume that output is given by $y_t = f(c_t, k_t + s_t)$, where $f(\cdot, \cdot)$ is a differentiable function with $f_c(c, a) > 0$, $f_a(c, a) > 0$, $f_{cc}(c, a) < 0$, $f_{aa}(c, a) < 0$, and $f_{ca}(c, a) > 0$ uniformly on the space of (c, a). Notice that k_t and s_t are strictly substitutes, while c_t and $a_t := k_t + s_t$ are weakly complements.[2]

The representative firm determines the optimal path of capital, investment, and R&D expenditures by

[1] For example, incidental and/or additional costs may be incurred to convert R&D expenditures into capital for production. These include patent fees, monetary or non-monetary incentives for researchers, safety management fees, training costs, and so on. These costs and fees are assumed to form the convex functions.

[2] It is possible to define alternative production functions that exhibit a weakly substitutionary relationship between s_t and k_t. Suppose that the production function (A.9) can be generalized by the constant elasticity of substitution (CES) function, $\ell(x, y; \beta) := (x^{\frac{\beta-1}{\beta}} + y^{\frac{\beta-1}{\beta}})^{\frac{\beta}{\beta-1}}$. Then, we can derive the generalized twofold CES production function: $f(c_t, k_t, s_t; \beta, \sigma) := \ell(c_t, \ell(k_t, s_t; \sigma); \beta)$, from which it follows that $\lim_{\sigma\to\infty} \lim_{\beta\to 0} f(c_t, k_t, s_t; \beta, \sigma) = \min[c_t, k_t + s_t]$. Provided that σ, $\beta > 0$, we can apply optimization theory as the production function is differentiable.

maximizing discounted aggregate profit:

$$\max_{\{c_t,k_t,s_t,r_t,i_t,d_t\}} \int_0^\infty \{f(c_t, k_t + s_t) - \kappa(r_t) - \phi(i_t) - \xi(d_t)\}e^{-\rho t}dt$$
$$\text{subject to } \dot{c}_t = r_t - \tau c_t, \quad \dot{k}_t = i_t - \delta k_t \text{ and } \dot{s}_t = d_t - \gamma s_t,$$

where ρ is the discount rate. This extends the standard q-theory of investment by considering the role of capital converted from R&D expenditures as well as physical capital, with the three different accumulation rules as constraints.[3]

To analyze the long-run relationship between physical investment and R&D expenditures, we set up the dynamic optimization problem. Let $(c_*, k_*, s_*, r_*, i_*, d_*)$ be the steady-state equilibrium, which must satisfy the steady-state conditions given by $f_a(c_*, k_* + s_*) = (\delta + \rho)\phi'(i_*)$, $f_a(c_*, k_* + s_*) = (\gamma + \rho)\xi'(d_*)$, $f_c(c_*, k_* + s_*) = (\tau + \rho)\kappa'(r_*)$, $i_* = \delta k_*$, $d_* = \gamma s_*$, and $r_* = \tau c_*$. To display the steady-state equilibrium, we plot a set of phase diagrams and marginal cost functions in Figure A.1. The panels on the left show the phase diagrams of (r_t, c_t), (i_t, k_t), and (d_t, s_t), while those on the right display the marginal cost functions of innovative R&D expenditure, physical investment, and managerial R&D expenditure, respectively. The phase diagrams also indicate the steady-state equilibrium levels of the variables along with the stable arms denoted by the dotted lines, such that the steady-state equilibrium can be reached by moving toward the equilibrium following the arms. The equilibrium cannot be reached unless the initial levels of (r_0, c_0), (i_0, k_0), and (d_0, s_0) are on the stable arms, simultaneously, as it would violate the transversality conditions.

— Insert Figure A.1 Here —

To examine how the steady-state equilibrium responds to external shocks, we conduct two experiments by changing the marginal cost functions of each type of R&D expenditure. In the first experiment, we let the marginal cost function of innovative R&D expenditure decrease from $\kappa_0'(\cdot)$ to $\kappa_1'(\cdot)$. Denote $(c_*, k_*, s_*, r_*, i_*, d_*)$ and $(c_{**}, k_{**}, s_{**}, r_{**}, i_{**}, d_{**})$ as the initial and new steady-state equilibria. The adjustment processes are displayed in the left panel of Figure A.2. This decline in the marginal cost function shifts the locus of $\dot{c}_t = 0$ to the locus of $\dot{c}_t' = 0$, denoted by the dashed line in the first phase diagram. The steady-state equilibrium (r_{**}, c_{**}) is reached at a level greater than (r_*, c_*). To attain the new steady-state equilibrium, r_* jumps to the new stable arm, denoted by the dotted line. By contrast, c_t is a stock, so it cannot jump to the new stable arm. Thus, (r_t, c_t) gradually tends to (r_{**}, c_{**}). Then, the loci of $\dot{i}_t = 0$ and $\dot{d}_t = 0$ move to

[3]The representative firm is assumed to choose the optimal time paths of r_t and d_t simultaneously, ignoring the fact that innovative R&D activity is conducted earlier than managerial R&D activity. This is not restrictive, as the firm represents a multitude of firms in the economy, where innovative and managerial R&D activities are conducted simultaneously.

$\dot{i}'_t = 0$ and $\dot{d}'_t = 0$, which are denoted by the dashed lines in the second and third diagrams. The steady-state equilibrium is determined at (i_{**}, k_{**}) and (d_{**}, s_{**}), which are greater than (i_*, k_*) and (d_*, s_*). Both i_t and d_t jump to the new stable arms denoted by the dotted lines, and (i_t, k_t) and (d_t, s_t) tend to the new steady-state equilibrium. This reveals that physical investment and innovative R&D expenditure move in the same direction following the change in the cost function of innovative R&D expenditure, implying that they are complements. The economic intuition is straightforward—as innovative R&D activities become relatively cheaper, the firm tends to accumulate more capital from innovative R&D activities, enhancing productivity. This allows the firm to invest more, thereby accumulating more physical capital.

— Insert Figure A.2 Here —

In the second experiment, the marginal cost function of managerial R&D expenditure decreases from $\xi'_0(\cdot)$ to $\xi'_1(\cdot)$. The right panel in Figure A.2 displays the adjustment process to the new equilibrium. The decrease in the marginal cost function shifts the locus of $\dot{d}_t = 0$ to the locus of $\dot{d}'_t = 0$, denoted by the dashed line in the third phase diagram. The new steady-state equilibrium (d_{**}, s_{**}) is reached at a level greater than (d_*, s_*). To attain the new steady-state equilibrium, d_* jumps to the new stable arm, denoted by the dotted line. However, s_t, as a stock, cannot jump to the new stable arm. Consequently, (d_t, s_t) tends to (d_{**}, s_{**}) gradually and the locus of $\dot{i}_t = 0$ shifts to the locus of $\dot{i}'_t = 0$. The steady-state equilibrium level is determined at (i_{**}, k_{**}), where k_t, as a stock, cannot jump to a new level while i_t jumps to the stable arm denoted as the dotted line in the second phase diagram. Overall, following the decrease in the marginal cost of managerial R&D expenditure, d_* rises to d_{**} but i_* falls to i_{**}, revealing a substitutionary relationship between i_t and d_t. This implies that physical capital decreases from k_* to k_{**}, while s increases to s_* from s_{**}.

As s_t and k_t move in opposite directions, $a_{**} := k_{**} + s_{**}$ can be greater or less than $a_* := k_* + s_*$. The sign of the change depends upon the functional shapes of $f_a(\cdot, \cdot)$, $\xi'(\cdot)$, $\phi'(\cdot)$, and the depreciation rates δ and γ. In Figure A.2 under the assumption that $a_{**} > a_*$, the locus of $\dot{c}_t = 0$ is shifted to $\dot{c}'_t = 0$, and a new equilibrium is achieved at (r_{**}, c_{**}), as indicated in the first phase diagram. Then, r_t jumps to the new stable arm denoted as the dotted line, and (r_t, c_t) approaches (r_{**}, c_{**}) gradually. In this case, $r_{**} > r_*$ and $c_{**} > c_*$, which is achieved mainly by virtue of the complementary relationship between c_t and a_t. On the other hand, consider the case with a_{**} less than a_* in which case the locus of $\dot{c}'_t = 0$ is shifted to the left of $\dot{c}_t = 0$. Then, we obtain $r_{**} < r_*$ and $c_{**} < c_*$. The economic intuition of the substitutionary relationship between i_t and d_t is also straightforward. As managerial R&D activity becomes relatively cheaper, the firm tends to accumulate capital by converting managerial R&D expenditures, thereby substituting physical

investment, implying that both physical investment and capital will decrease.

These experiments yield important testable implications—the relationship between physical investment and innovative R&D expenditure is expected to be positive by virtue of their complementarity, while the relationship between managerial R&D expenditure and investment is more likely to be negative due to their nature as substitutes.

D.3 The NARDL Specification

Our empirical specification is grounded in two stylized features of R&D expenditure highlighted in Section D.1 and the theory developed in Section D.2. First, as innovative R&D expenditures tend to focus on product innovation, their scale is often large relative to output. This suggests that R&D expenditure is expected to grow faster than output in the early stage, where start-up costs are large and the scale of production typically small. Second, as managerial R&D expenditures focus on enhancing production efficiency, their scale is typically smaller than output.

Let r_t denote aggregate R&D intensity in the t-th period, defined as a ratio of aggregate R&D expenditure to GDP. Noting that aggregate R&D expenditure incorporates the spectrum of R&D activities conducted throughout the economy, the sign of Δr_t determines the relative prevalence of innovative and managerial R&D activities. If $\Delta r_t \geq 0$, then R&D expenditure grows as fast as output, indicating a prevalence of innovative R&D activity. By contrast, if $\Delta r_t < 0$, then output grows faster than R&D expenditure, indicating a prevalence of managerial R&D activity. Given the different characteristics of innovative and managerial R&D, it is reasonable to expect that the relationship between R&D intensity and physical investment may be asymmetric.

To analyze the potential asymmetric impacts of r_t on the log of investment (i_t) in the short-run and the long-run, we consider the following asymmetric error-correction model:

$$\Delta i_t = \gamma_* + \rho_* u_{t-1} + \sum_{j=1}^{p-1} \varphi_{j*} \Delta i_{t-j} + \sum_{j=0}^{q-1} \pi_{j*}^{+} \Delta r_{t-j}^{+} + \sum_{j=0}^{q-1} \pi_{j*}^{-} \Delta r_{t-j}^{-} + e_t, \tag{A.10}$$

where $u_{t-1} = i_{t-1} - \alpha_* - \beta_*^{+} r_{t-1}^{+} - \beta_*^{-} r_{t-1}^{-}$ is the asymmetric error correction term and e_t is a serially uncorrelated error term, given sufficiently large lag orders, p and q. Here, $\Delta r_t^{+} := \Delta r_t \mathbb{1}_{\{\Delta r_t \geq 0\}}$ and $\Delta r_t^{-} := \Delta r_t \mathbb{1}_{\{\Delta r_t < 0\}}$, where $\mathbb{1}_{\{\cdot\}}$ is an indicator function taking unity if the condition in brace is satisfied, and zero otherwise.

The process in (A.10) is equivalent to the NARDL(p, q) process advanced by SYG,

$$i_t = \nu_* + \sum_{j=1}^{p} \phi_{j*} i_{t-j} + \sum_{j=0}^{q} (\theta_{j*}^{+\prime} r_{t-j}^{+} + \theta_{j*}^{-\prime} r_{t-j}^{-}) + e_t.$$

The NARDL process allows for both the long-run parameters, β_*^+ and β_*^-, and the short-run parameters, π_{j*}^+ and π_{j*}^-, to differ, enabling us to jointly analyze long- and short-run asymmetric relationships between R&D intensity and investment. Furthermore, it is important to notice that the NARDL process can accommodate a cointegrating relationship between integrated time series with mismatched time drifts. As $\Delta r_t^+ \geq 0$ and $\Delta r_t^- \leq 0$ with probability one even if $E(\Delta r_t) = 0$, the partial sum processes, r_t^+ and r_t^-, will be integrated series with positive and negative time drifts, respectively. Thus, if there exists an asymmetric cointegrating relationship between i_t and r_t, then the dependent variable, i_t should be an integrated series with a drift. This has the important implication that the NARDL model can analyze an asymmetric cointegrating relationship between two integrated variables with different drifts without the need to include a deterministic time trend in the model. In Section E, we find that r_t is a unit-root process without a drift while i_t is a unit-root process with a drift, and establish that there exists a cointegrating relationship between them without including a deterministic time trend.

E Additional Empirical Results

Table A.3 reports the descriptive statistics of both R&D intensity and the log of GPDI. R&D intensity has the characteristics of a normal distribution centered at zero. By contrast, the real investment data is centered at a non-zero value and is more widely distributed with a negative skew and excess kurtosis. These characteristics suggest that the relationship between these variables may not be adequately explained by a simple linear model. We apply the Phillips and Perron (1988) unit root test to each series, both including and excluding a time trend. The results reported in Table A.4 show that the unit-root hypothesis cannot be rejected irrespective of the presence of the time trend, implying that both series are unit-root nonstationary. Furthermore, the time trend coefficient in the univariate regression is statistically significant for the log of GPDI, but insignificant for R&D intensity, implying that the log of GPDI is a unit root process with a time drift, while R&D intensity is a unit root process without a time drift, viz., $\mathbb{E}[\Delta r_t] = 0$ but $\mathbb{E}[\Delta i_t] > 0$.

— Insert Tables A.3 and A.4 Here —

	Sample Size	50		100		200		300		400		500	
	First Step	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS
ρ_*	Second Step	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
-0.10	β_*^+	-0.765	-0.547	-0.533	-0.276	-0.337	-0.120	-0.239	-0.039	-0.191	-0.026	-0.159	-0.016
	β_*^-	0.577	0.406	0.420	0.215	0.268	0.092	0.191	0.029	0.157	0.016	0.130	0.013
	ρ_*	-0.198	-0.200	-0.081	-0.081	-0.036	-0.037	-0.022	-0.023	-0.016	-0.016	-0.012	-0.013
	φ_*	0.061	0.043	0.030	0.020	0.015	0.011	0.010	0.007	0.008	0.004	0.008	0.003
	π_{0*}^+	-0.038	-0.004	-0.024	0.000	-0.012	0.004	-0.011	0.002	-0.007	0.003	-0.008	0.003
	π_{1*}^+	0.054	0.021	0.030	-0.008	0.008	-0.009	0.010	-0.007	0.007	-0.009	0.006	-0.007
	π_{0*}^-	0.024	0.008	0.019	0.003	0.009	-0.007	0.008	-0.002	0.006	0.000	0.004	-0.002
	π_{1*}^-	-0.068	-0.058	-0.032	-0.017	-0.013	0.000	-0.008	0.002	-0.006	0.003	-0.003	0.002
-0.20	β_*^+	-0.523	-0.216	-0.327	-0.039	-0.185	0.017	-0.129	0.039	-0.096	0.034	-0.076	0.030
	β_*^-	0.426	0.163	0.262	0.034	0.157	-0.016	0.108	-0.034	0.080	-0.031	0.065	-0.026
	ρ_*	-0.178	-0.172	-0.074	-0.072	-0.033	-0.032	-0.022	-0.019	-0.016	-0.013	-0.013	-0.010
	φ_*	0.066	0.039	0.032	0.018	0.015	0.008	0.011	0.004	0.008	0.001	0.006	0.002
	π_{0*}^+	-0.039	0.019	-0.032	0.018	-0.012	0.014	-0.011	0.012	-0.006	0.006	-0.006	0.006
	π_{1*}^+	0.056	-0.034	0.029	-0.030	0.013	-0.025	0.012	-0.017	0.007	-0.015	0.004	-0.010
	π_{0*}^-	0.038	-0.020	0.016	-0.012	0.010	-0.011	0.010	-0.009	0.006	-0.008	0.004	-0.007
	π_{1*}^-	-0.065	-0.018	-0.025	0.007	-0.015	0.007	-0.007	0.007	-0.007	0.008	-0.005	0.006
-0.30	β_*^+	-0.404	-0.121	-0.231	0.010	-0.122	0.035	-0.084	0.034	-0.062	0.032	-0.051	0.028
	β_*^-	0.339	0.101	0.208	-0.010	0.108	-0.032	0.076	-0.032	0.056	-0.030	0.045	-0.028
	ρ_*	-0.164	-0.159	-0.070	-0.066	-0.032	-0.029	-0.021	-0.018	-0.015	-0.013	-0.012	-0.011
	φ_*	0.065	0.040	0.030	0.015	0.014	0.006	0.010	0.004	0.008	0.003	0.007	0.003
	π_{0*}^+	-0.048	0.005	-0.028	0.019	-0.017	0.016	-0.007	0.009	-0.009	0.010	-0.004	0.008
	π_{1*}^+	0.052	-0.035	0.026	-0.036	0.013	-0.028	0.010	-0.017	0.008	-0.016	0.005	-0.013
	π_{0*}^-	0.040	-0.014	0.029	-0.020	0.015	-0.011	0.008	-0.009	0.009	-0.010	0.005	-0.007
	π_{1*}^-	-0.053	-0.013	-0.030	0.013	-0.016	0.011	-0.009	0.010	-0.008	0.012	-0.006	0.007
-0.40	β_*^+	-0.320	-0.095	-0.173	-0.007	-0.088	0.027	-0.059	0.023	-0.043	0.023	-0.035	0.020
	β_*^-	0.286	0.059	0.159	0.000	0.084	-0.029	0.055	-0.025	0.041	-0.024	0.033	-0.021
	ρ_*	-0.158	-0.151	-0.068	-0.065	-0.031	-0.030	-0.021	-0.018	-0.016	-0.013	-0.012	-0.010
	φ_*	0.065	0.044	0.031	0.017	0.016	0.010	0.009	0.004	0.008	0.003	0.005	0.002
	π_{0*}^+	-0.038	0.004	-0.020	0.010	-0.011	0.012	-0.011	0.009	-0.006	0.007	-0.007	0.008
	π_{1*}^+	0.044	-0.041	0.030	-0.033	0.014	-0.024	0.007	-0.018	0.006	-0.015	0.005	-0.014
	π_{0*}^-	0.051	-0.017	0.025	-0.015	0.015	-0.013	0.007	-0.009	0.007	-0.008	0.006	-0.007
	π_{1*}^-	-0.055	0.000	-0.026	0.011	-0.014	0.017	-0.007	0.012	-0.006	0.009	-0.007	0.010
-0.50	β_*^+	-0.265	-0.103	-0.135	-0.022	-0.069	0.016	-0.046	0.010	-0.034	0.012	-0.027	0.011
	β_*^-	0.242	0.061	0.131	0.007	0.068	-0.020	0.044	-0.014	0.033	-0.014	0.027	-0.014
	ρ_*	-0.148	-0.135	-0.067	-0.062	-0.031	-0.029	-0.020	-0.020	-0.015	-0.014	-0.011	-0.011
	φ_*	0.061	0.041	0.031	0.021	0.016	0.008	0.011	0.006	0.007	0.003	0.006	0.003
	π_{0*}^+	-0.044	-0.006	-0.018	0.007	-0.010	0.009	-0.003	0.005	-0.005	0.002	-0.006	0.005
	π_{1*}^+	0.050	-0.026	0.024	-0.024	0.011	-0.024	0.008	-0.015	0.008	-0.014	0.006	-0.009
	π_{0*}^-	0.048	-0.006	0.029	-0.007	0.012	-0.011	0.009	-0.007	0.005	-0.004	0.006	-0.006
	π_{1*}^-	-0.057	-0.003	-0.026	0.007	-0.014	0.017	-0.008	0.012	-0.008	0.011	-0.006	0.008
-0.60	β_*^+	-0.213	-0.115	-0.107	-0.034	-0.054	0.003	-0.036	0.002	-0.027	0.006	-0.021	0.005
	β_*^-	0.208	0.076	0.106	0.022	0.055	-0.009	0.037	-0.008	0.028	-0.010	0.022	-0.009
	ρ_*	-0.147	-0.132	-0.065	-0.061	-0.031	-0.030	-0.020	-0.019	-0.014	-0.015	-0.012	-0.011
	φ_*	0.061	0.044	0.033	0.022	0.015	0.010	0.009	0.005	0.007	0.004	0.006	0.003
	π_{0*}^+	-0.029	-0.020	-0.012	0.000	-0.008	0.001	-0.005	0.004	-0.005	0.002	-0.002	0.006
	π_{1*}^+	0.045	-0.009	0.024	-0.021	0.011	-0.020	0.006	-0.012	0.008	-0.012	0.006	-0.010
	π_{0*}^-	0.050	-0.011	0.019	-0.003	0.015	-0.002	0.010	-0.006	0.008	-0.007	0.005	-0.006
	π_{1*}^-	-0.050	-0.005	-0.033	0.015	-0.013	0.012	-0.009	0.011	-0.008	0.010	-0.006	0.008

Table A.1: FINITE SAMPLE BIAS OF THE TWO-STEP ESTIMATORS FOR NARDL(2,2). This table reports the finite sample biases when OLS/FM is used in the first step and OLS is used in the second step. DGP: $\Delta y_t = \gamma_* + \rho_* u_{t-1} + \varphi_* \Delta y_{t-1} + \pi_{0*}^+ \Delta x_t^+ + \pi_{1*}^+ \Delta x_{t-1}^+ + \pi_{0*}^- \Delta x_t^- + \pi_{1*}^- \Delta x_{t-1}^- + e_t$, where $u_{t-1} := y_{t-1} - \alpha_* - \beta_*^+ x_{t-1}^+ - \beta_*^- x_{t-1}^-$, $\Delta x_t := \kappa_* \Delta x_{t-1} + \sqrt{1-\kappa_*^2} v_t$, and $(e_t, v_t)' \sim \text{IID} N(\mathbf{0}_2, \mathbf{I}_2)$; $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_{0*}^+, \pi_{1*}^+, \pi_{0*}^-, \pi_{1*}^-, \kappa_*)$ = (7.453, 0.271, -0.640, 0.015, ρ_*, 0.255, -0.555, -0.029, -0.359, 0.482, 0.5).

	Sample Size	50		100		200		300		400		500	
	First Step	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS	OLS	FM-OLS
ρ_*	Second Step	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
-0.10	β_*^+	1.329	2.045	0.571	0.602	0.221	0.157	0.115	0.071	0.070	0.038	0.048	0.024
	β_*^-	1.075	1.415	0.508	0.583	0.181	0.150	0.098	0.070	0.060	0.038	0.041	0.024
	ρ_*	0.055	0.055	0.011	0.010	0.002	0.003	0.001	0.001	0.001	0.001	0.000	0.000
	φ_*	0.022	0.020	0.010	0.009	0.005	0.005	0.003	0.003	0.002	0.002	0.002	0.002
	π_{0*}^+	0.158	0.162	0.060	0.055	0.025	0.025	0.016	0.015	0.011	0.011	0.009	0.009
	π_{1*}^+	0.111	0.139	0.050	0.056	0.023	0.026	0.016	0.017	0.012	0.011	0.009	0.010
	π_{0*}^-	0.152	0.161	0.061	0.061	0.025	0.025	0.016	0.015	0.011	0.011	0.009	0.009
	π_{1*}^-	0.109	0.142	0.048	0.055	0.022	0.024	0.015	0.016	0.011	0.011	0.009	0.009
-0.20	β_*^+	0.714	0.834	0.238	0.210	0.073	0.049	0.034	0.021	0.019	0.012	0.012	0.008
	β_*^-	0.664	0.777	0.211	0.198	0.066	0.047	0.030	0.021	0.017	0.012	0.011	0.007
	ρ_*	0.049	0.045	0.010	0.010	0.003	0.003	0.001	0.001	0.001	0.001	0.001	0.001
	φ_*	0.022	0.019	0.010	0.009	0.004	0.004	0.003	0.003	0.002	0.002	0.002	0.002
	π_{0*}^+	0.155	0.158	0.058	0.057	0.025	0.024	0.016	0.015	0.011	0.011	0.009	0.009
	π_{1*}^+	0.112	0.149	0.050	0.058	0.024	0.027	0.016	0.017	0.011	0.013	0.009	0.010
	π_{0*}^-	0.166	0.158	0.062	0.055	0.026	0.024	0.016	0.015	0.012	0.011	0.009	0.008
	π_{1*}^-	0.115	0.138	0.049	0.054	0.023	0.024	0.015	0.015	0.011	0.012	0.009	0.009
-0.30	β_*^+	0.439	0.502	0.128	0.111	0.035	0.023	0.016	0.011	0.009	0.006	0.006	0.004
	β_*^-	0.425	0.444	0.122	0.106	0.033	0.023	0.015	0.011	0.008	0.006	0.005	0.004
	ρ_*	0.044	0.043	0.011	0.010	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001
	φ_*	0.021	0.020	0.009	0.009	0.004	0.004	0.003	0.003	0.002	0.002	0.002	0.002
	π_{0*}^+	0.159	0.163	0.060	0.056	0.025	0.024	0.016	0.015	0.011	0.011	0.009	0.009
	π_{1*}^+	0.113	0.157	0.051	0.061	0.025	0.027	0.016	0.017	0.012	0.013	0.010	0.010
	π_{0*}^-	0.168	0.157	0.061	0.060	0.025	0.024	0.017	0.015	0.011	0.011	0.009	0.009
	π_{1*}^-	0.114	0.139	0.049	0.059	0.023	0.025	0.015	0.016	0.011	0.012	0.009	0.009
-0.40	β_*^+	0.298	0.328	0.076	0.069	0.019	0.014	0.008	0.006	0.005	0.003	0.003	0.002
	β_*^-	0.292	0.315	0.074	0.066	0.019	0.014	0.008	0.006	0.005	0.003	0.003	0.002
	ρ_*	0.044	0.042	0.012	0.011	0.004	0.004	0.002	0.002	0.002	0.002	0.001	0.001
	φ_*	0.020	0.019	0.009	0.008	0.004	0.004	0.003	0.003	0.002	0.002	0.002	0.002
	π_{0*}^+	0.165	0.161	0.061	0.057	0.025	0.025	0.016	0.016	0.012	0.011	0.009	0.009
	π_{1*}^+	0.111	0.159	0.053	0.062	0.025	0.027	0.017	0.018	0.013	0.013	0.010	0.010
	π_{0*}^-	0.168	0.169	0.061	0.059	0.025	0.024	0.015	0.015	0.011	0.011	0.009	0.009
	π_{1*}^-	0.108	0.141	0.049	0.057	0.024	0.024	0.015	0.017	0.011	0.012	0.009	0.010
-0.50	β_*^+	0.223	0.277	0.050	0.049	0.012	0.009	0.005	0.004	0.003	0.002	0.002	0.001
	β_*^-	0.211	0.243	0.051	0.047	0.012	0.009	0.005	0.004	0.003	0.002	0.002	0.001
	ρ_*	0.043	0.039	0.013	0.012	0.005	0.004	0.003	0.003	0.002	0.002	0.001	0.002
	φ_*	0.021	0.020	0.009	0.009	0.004	0.004	0.003	0.003	0.002	0.002	0.002	0.002
	π_{0*}^+	0.166	0.167	0.059	0.058	0.025	0.023	0.016	0.015	0.012	0.011	0.009	0.009
	π_{1*}^+	0.120	0.165	0.051	0.066	0.025	0.029	0.016	0.017	0.012	0.013	0.010	0.010
	π_{0*}^-	0.154	0.176	0.060	0.059	0.025	0.024	0.016	0.015	0.012	0.011	0.009	0.009
	π_{1*}^-	0.112	0.148	0.052	0.058	0.023	0.026	0.015	0.016	0.011	0.012	0.009	0.009
-0.60	β_*^+	0.155	0.201	0.035	0.036	0.008	0.006	0.003	0.003	0.002	0.001	0.001	0.001
	β_*^-	0.166	0.188	0.036	0.035	0.009	0.006	0.004	0.003	0.002	0.001	0.001	0.001
	ρ_*	0.045	0.041	0.014	0.013	0.005	0.005	0.003	0.003	0.002	0.002	0.002	0.002
	φ_*	0.021	0.019	0.009	0.009	0.004	0.004	0.003	0.003	0.002	0.002	0.002	0.002
	π_{0*}^+	0.158	0.178	0.058	0.060	0.025	0.024	0.016	0.016	0.012	0.012	0.009	0.009
	π_{1*}^+	0.117	0.174	0.052	0.064	0.025	0.028	0.016	0.019	0.013	0.013	0.010	0.010
	π_{0*}^-	0.169	0.181	0.057	0.059	0.025	0.025	0.016	0.016	0.011	0.012	0.009	0.009
	π_{1*}^-	0.117	0.154	0.051	0.059	0.023	0.026	0.015	0.016	0.012	0.012	0.009	0.009

Table A.2: FINITE SAMPLE MEAN SQUARED ERROR (MSE) OF THE TWO-STEP ESTIMATORS FOR NARDL(2,2). This table reports the finite sample biases when OLS/FM is used in the first step and OLS is used in the second step. DGP: $\Delta y_t = \gamma_* + \rho_* u_{t-1} + \varphi_* \Delta y_{t-1} + \pi_{0*}^+ \Delta x_t^+ + \pi_{1*}^+ \Delta x_{t-1}^+ + \pi_{0*}^- \Delta x_t^- + \pi_{1*}^- \Delta x_{t-1}^- + e_t$, where $u_{t-1} := y_{t-1} - \alpha_* - \beta_*^+ x_{t-1}^+ - \beta_*^- x_{t-1}^-$, $\Delta x_t := \kappa_* \Delta x_{t-1} + \sqrt{1-\kappa_*^2} v_t$, and $(e_t, v_t)' \sim$ IID$N(\mathbf{0}_2, \mathbf{I}_2)$; $(\alpha_*, \beta_*^+, \beta_*^-, \gamma_*, \rho_*, \varphi_*, \pi_{0*}^+, \pi_{1*}^+, \pi_{0*}^-, \pi_{1*}^-, \kappa_*)$ = (7.453, 0.271, -0.640, 0.015, ρ_*, 0.255, -0.555, -0.029, -0.359, 0.482, 0.5).

	$\Delta \log$ GPDI	Δ(R&D/GDP)
Mean	0.009	0.004
Median	0.009	0.001
Maximum	0.110	0.082
Minimum	-0.176	-0.082
Standard Deviation	0.039	0.028
Skewness	-0.752	0.190
Excess Kurtosis	2.701	0.256
Sample Size	240	240

Table A.3: DESCRIPTIVE STATISTICS. Descriptive statistics are computed over 240 quarters from 1960q1 to 2019q4. GPDI is measured in US Dollars at 2012 prices and seasonally adjusted. R&D and GDP are seasonally adjusted nominal values.

PP test	$\log$ GPDI	R&D/GDP
PP test w/o trend	-1.138	-1.946
p-value	0.701	0.311
PP test w/ trend	-3.198	-2.255
p-value	0.087	0.457

Table A.4: PHILLIPS AND PERRON'S (1988) UNIT-ROOT TEST STATISTICS. Two Phillips and Perron tests are computed, one including and the other excluding a time trend. When the time trend is included, it is statistically significant for the log of GPDI but not for R&D intensity. The lag lengths of the ADF regressions are selected by the SIC.

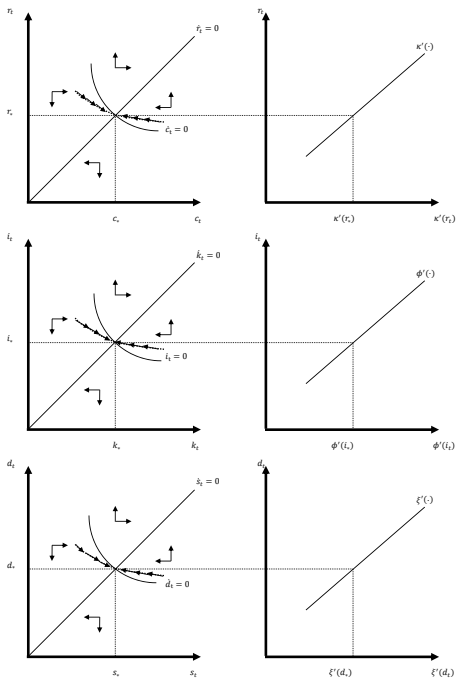


Figure A.1: PHASE DIAGRAMS AND STEADY STATE. This figure shows the phase diagrams of (r_t, c_t), (i_t, k_t), and (d_t, k_t) and their relationships with the marginal cost functions.

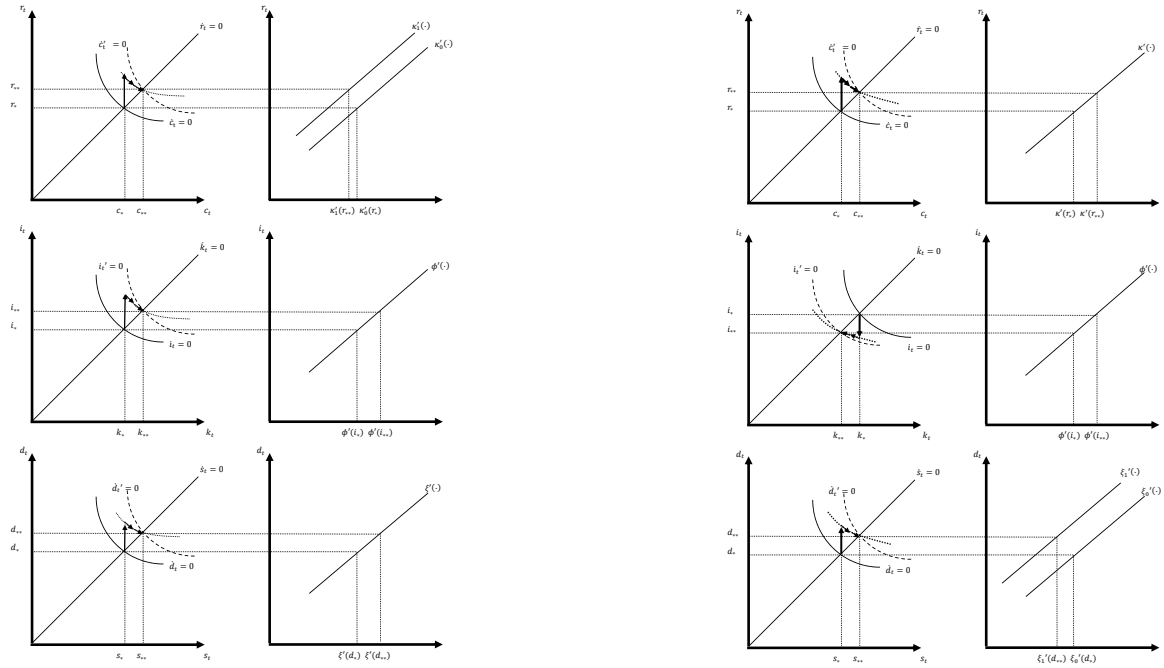


Figure A.2: PHASE DIAGRAMS AND STEADY STATE. The left figure demonstrates how the steady-state levels are adjusted as the marginal cost function $\kappa'(\cdot)$ of innovative R&D expenditure decreases from $\kappa_0'(\cdot)$ to $\kappa_1'(\cdot)$. The right figure demonstrates how the steady-state levels are adjusted as the marginal cost function of managerial R&D expenditure decreases from $\xi_0'(\cdot)$ to $\xi_1'(\cdot)$.