

Seemingly Unrelated Cointegrating Regressions with Autoregressive Distributed Lag Dynamics: Estimation, Bounds Testing, and Cross-Equation Inference

Muhammad Afnan Arif

muhdafnanarif@gmail.com

University of Malaya

Fumitaka Furuoka

University of Malaya

Research Article

Keywords: Seemingly unrelated regression, Autoregressive distributed lag, Fully modified estimation, Cointegrating polynomial regression, Cointegration testing

Posted Date: August 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9237078/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Seemingly Unrelated Cointegrating Regressions with Autoregressive Distributed Lag Dynamics: Estimation, Bounds Testing, and Cross-Equation Inference

Muhammad Afnan Arif ¹ · Fumitaka Furuoka ²

Abstract

This paper develops a seemingly unrelated autoregressive distributed lag (SUR-ARDL) framework for estimating and testing cointegrating relationships in panels with small cross-sectional dimension. By embedding the Pesaran et al. (2001) error-correction specification within a Zellner (1962) SUR system, the estimator exploits cross-equation correlation for efficiency while the ARDL lag structure parametrically removes second-order endogeneity and serial correlation biases without semi-parametric corrections. Unlike the FM-SUR estimator of Wagner et al. (2020, Journal of Econometrics) and the FM-GLS estimator of Lin and Reuvers (2025, Oxford Bulletin of Economics and Statistics), the proposed estimator requires only the $N \times N$ contemporaneous error covariance rather than $2N \times 2N$ long-run covariance matrices, eliminating the principal source of finite-sample distortion in these methods. We establish that the feasible GLS estimator achieves a mixed-normal limiting distribution, the SUR bounds F-statistic inherits the Pesaran et al. (2001) bounds property regardless of the integration order of regressors, and cross-equation Wald tests for poolability and coefficient equality attain chi-squared distributions. A Monte Carlo study with demonstrates RMSE reductions of relative to single-equation OLS, substantial power gains for the bounds cointegration test, reliable covariance estimation with zero invertibility and adequate size with bootstrap critical values. An application to the renewable energy-economic growth nexus in ASEAN countries shows that SUR-ARDL reduces standard errors, detects cointegration in Vietnam where single-equation methods fail, and reveals a two-cluster structure of adjustment speeds slow-adjusting fossil fuel producers versus fast-adjusting energy importers through cross-equation hypothesis tests unavailable to equation-by-equation methods.

Keywords: Seemingly unrelated regression; Autoregressive distributed lag; Fully modified estimation; Cointegrating polynomial regression; Cointegration testing

JEL Classification: C13, C32, C33, Q43.

Fumitaka Furuoka
fumitaka@um.edu.my

Corresponding author:
Muhammad Afnan Arif
muhdafnanarif@gmail.com

¹ Asia-Europe Institute, Universiti Malaya, Kuala Lumpur, Malaysia

² Asia-Europe Institute, Universiti Malaya, Kuala Lumpur, Malaysia

1. Introduction

Cointegrating relationships among macroeconomic time series are frequently studied in multi-country panels where the cross-sectional dimension N is small relative to the time series length T . The standard approach to modelling such systems widely endorsed in the literature (Pesaran, 2006, pp.967-968) is to treat the N equations as a system of seemingly unrelated regressions (SUR), thereby accommodating both cross-sectional heterogeneity and cross-sectional dependence. The SUR cointegration literature has developed along two branches. The first, initiated by Park and Ogaki (1991) and extended by Moon (1999), Moon and Perron (2005), Mark et al. (2005), and Wagner et al. (2020), employs fully modified (FM) estimators FM-SOLS and FM-SUR that correct for second-order bias through semi-parametric long-run covariance estimation. The second, proposed by Lin and Reuvers (2025), introduces an FM-GLS estimator based on a banded inverse autocovariance matrix (BIAM). Both branches achieve mixed-normal limiting distributions enabling chi-squared inference on cross-equation hypotheses.

Despite these advances, the existing SUR cointegration frameworks share a common practical limitation: they require consistent estimation of large long-run covariance matrices. Wagner et al. (2020) require a $2N \times 2N$ long-run covariance matrix Ω , whose estimation involves kernel-based methods with bandwidth selection and whose accuracy deteriorates sharply as N increases relative to T . Their simulations reveal invertibility failure rates exceeding 50% and test size distortions of 30-85% when $N = 10$ and $T = 100$. Lin and Reuvers (2025) address this through the BIAM approach, which avoids direct estimation of Ω , but their method requires specification of a banding parameter and remains computationally demanding for large systems. In both cases, the "curse of dimensionality" the need to estimate $O(N^2)$ nuisance parameters from $O(NT)$ observations constrains the practical applicability of these methods to settings where T is large relative to N .

This paper proposes an alternative approach that circumvents these dimensionality problems entirely. We embed the cointegrating relationships within the autoregressive distributed lag (ARDL) error-correction framework of Pesaran et al.(2001), which parametrically absorbs short-run dynamics and regressor endogeneity through the lag structure, and estimate the resulting system by feasible generalized least squares (FGLS) using only the $N \times N$ contemporaneous error covariance matrix Σ . The key insight is that the ARDL reparameterization reduces the nuisance parameter from the $2N \times 2N$ long-run covariance to the $N \times N$ contemporaneous covariance a dimensional reduction from $O(N^2)$ to $O(N^2/4)$ that eliminates the need for kernel-based long-run covariance estimation altogether. This seemingly unrelated ARDL (SUR-ARDL) framework inherits the attractive features of both the single-equation ARDL bounds approach (valid inference regardless of whether regressors are $I(0)$ or $I(1)$; no pre-testing for unit roots) and the SUR methodology (efficiency gains from cross-equation correlation; testable cross-equation restrictions).

The paper makes five contributions to the SUR cointegration literature. First, we establish the formal asymptotic theory for the SUR-ARDL estimator under the mixed-integration-order setting of Pesaran et al.(2001). We show that the FGLS estimator achieves a mixed-normal limiting distribution (Theorem 2) that is free of the second-order bias terms present in the FM-OLS literature the ARDL lag structure parametrically removes these biases without requiring semi-parametric corrections. This result generalizes the single-equation asymptotic theory of Pesaran and Shin (1999) and Pesaran et al.(2001) to the multi-equation setting with cross-sectionally correlated errors.

Second, we derive the null distribution of the SUR bounds F-statistic for testing cointegration within the system (Theorem 3) and show that it inherits the Pesaran et al.(2001) bounds property: the critical values are bounded between the pure- $I(0)$ and pure- $I(1)$ cases regardless of the true integration order. We demonstrate that the SUR F-statistic is monotonically increasing in the cross-equation correlation for a fixed set of regression

coefficients (Proposition 4), providing a formal basis for the power advantage of system-based cointegration testing.

Third, we develop chi-squared Wald tests for cross-equation hypotheses including coefficient poolability, long-run coefficient equality across equations, and speed-of-adjustment homogeneity (Theorems 5-6). These tests exploit the joint estimation of the system to investigate cross-country similarities and differences a capability absent from single-equation methods.

Fourth, we conduct a comprehensive Monte Carlo study comprising seven experiments with 5,000 replications each, covering estimation bias and efficiency (Simulation 1), bounds test size (Simulation 2), size-corrected power (Simulation 3), Wald test size (Simulation 4), long-run coefficient coverage (Simulation 5), two-step versus iterated FGLS comparison (Simulation 6), and covariance matrix estimation quality (Simulation 7). The simulations establish that: (a) SUR-ARDL achieves RMSE reductions of 15-80% relative to OLS for the error-correction parameter when cross-equation correlation is moderate to strong; (b) the bounds test exhibits good size with bootstrap critical values but requires caution with tabulated values when $N \geq 5$; (c) the $N \times N$ covariance matrix Σ is reliably estimable with zero invertibility failures across all configurations in sharp contrast to the FM-SUR framework where Ω estimation fails frequently; and (d) two-step FGLS is sufficient, with iteration providing negligible improvement.

The remainder of the paper is organized as follows. Section 2 presents the model specification, the SUR-ARDL estimator, and the associated test statistics. Section 3 develops the asymptotic theory, including consistency, limiting distributions, and chi-squared inference for cross-equation hypotheses. Section 4 reports the Monte Carlo simulation study. Section 5 presents the empirical application to ASEAN renewable energy and economic growth. Section 6 concludes. Proofs are collected in Appendices A-D.

2. Model

The framework developed here extends the single-equation ARDL bounds testing approach of Pesaran, Shin, and Smith (2001) to a system-based setting in the spirit of the SUR cointegration literature (Park & Ogaki, 1991; Moon, 1999; Mark et al., 2005; Wagner, Grabarczyk, & Hong, 2020). While the SUR cointegration literature has been developed primarily for fully modified OLS-type estimators in levels regressions, most recently extended to nonlinear cointegrating polynomial regressions by Wagner et al. (2020). This study contribution embeds the ARDL error-correction specification within the SUR framework. This retains the key advantages of the ARDL approach namely, validity under mixed integration orders and favorable small-sample properties while exploiting cross-equation covariance information to improve estimation efficiency and inference robustness.

2.1 Single-Equation ARDL Model and Error-Correction Reparameterization

Consider a panel of N cross-sectional units (countries, regions, or economic entities) observed over T time periods. For unit i , $i = 1, \dots, N$, let $Y_{i,t} \in \mathbb{R}$ denote the dependent variable at time $t = 1, \dots, T$, and let $\mathbf{X}_{i,t} = (X_{1,i,t}, X_{2,i,t}, \dots, X_{K,i,t})' \in \mathbb{R}^K$ denote a vector of K explanatory variables. The ARDL model of order $(p_i, q_{1,i}, q_{2,i}, \dots, q_{K,i})$ for unit i specifies the conditional expectation of $Y_{i,t}$ given its own lags and current and lagged values of $\mathbf{X}_{i,t}$:

$$Y_{i,t} = \mu_i + \delta_i t + \sum_{j=1}^{p_i} \phi_{j,i} Y_{i,t-j} + \sum_{k=1}^K \sum_{j=0}^{q_{k,i}} \phi_{k,j,i} X_{k,i,t-j} + u_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where μ_i is a unit-specific intercept capturing time-invariant heterogeneity, δ_i is the coefficient on a deterministic linear time trend, $\phi_{j,i}$ for $j = 1, \dots, p_i$ are the autoregressive coefficients on

lagged values of the dependent variable, $\phi_{k,j,i}$ for $k = 1, \dots, K$ and $j = 0, \dots, q_{k,i}$ are the distributed lag coefficients on current and lagged values of the k -th explanatory variable, $p_i \geq 1$ is the autoregressive lag order, $q_{k,i} \geq 0$ is the distributed lag order associated with the k -th regressor for unit i , and $u_{i,t}$ is a zero-mean disturbance term satisfying conditions specified below.

The ARDL specification in equation (1) is fully general in the sense that lag orders are permitted to vary across both units and regressors, accommodating heterogeneous dynamic structures. The model nests several important special cases. When $p_i = 1$ and $q_{k,i} = 0$ for all k , equation (1) reduces to a static regression with a lagged dependent variable. When $p_i = q_{k,i} = 1$ for all k , the specification corresponds to the ARDL(1,1,...,1) model that forms the basis of the Pesaran et al. (2001) bounds testing framework. When all $\phi_{j,i} = 0$, the model reduces to a contemporaneous distributed lag specification without autoregressive dynamics. The autoregressive polynomial for unit i is defined as:

$$\Phi_i(L) = 1 - \sum_{j=1}^{p_i} \phi_{j,i} L^j, \quad (2)$$

where L denotes the lag operator satisfying $L^j Y_{i,t} = Y_{i,t-j}$. Similarly, the distributed lag polynomial for the k -th regressor of unit i is:

$$\Psi_{k,i}(L) = \sum_{j=0}^{q_{k,i}} \phi_{k,j,i} L^j. \quad (3)$$

Using this operator notation, equation (1) can be written compactly as:

$$\Phi_i(L)Y_{i,t} = \mu_i + \delta_i t + \sum_{k=1}^K \Psi_{k,i}(L)X_{k,i,t} + u_{i,t}. \quad (4)$$

The dynamic stability of the model requires that all roots of the autoregressive polynomial $\Phi_i(z) = 0$ lie outside the unit circle, i.e., $|z| > 1$ for all roots z satisfying $\Phi_i(z) = 0$. This condition ensures that the autoregressive process is stable and that long-run multipliers are well-defined. In the context of cointegration testing, the key parameter of interest is $\Phi_i(1) = 1 - \sum_{j=1}^{p_i} \phi_{j,i}$, which governs the speed of adjustment toward the long-run equilibrium.

Following Pesaran and Shin (1999) and Pesaran et al. (2001), the ARDL model in equation (1) can be reparameterized into a conditional error-correction form that separates the long-run level relationship from short-run dynamic adjustments. This reparameterization is achieved by adding and subtracting appropriate terms and collecting lagged level and differenced components.

To derive the error-correction form, first note that the autoregressive component can be decomposed as:

$$\sum_{j=1}^{p_i} \phi_{j,i} Y_{i,t-j} = \left(\sum_{j=1}^{p_i} \phi_{j,i} \right) Y_{i,t-1} - \sum_{j=1}^{p_i-1} \left(\sum_{l=j+1}^{p_i} \phi_{l,i} \right) \Delta Y_{i,t-j}. \quad (5)$$

Similarly, for each regressor $k = 1, \dots, K$:

$$\sum_{j=0}^{q_{k,i}} \phi_{k,j,i} X_{k,i,t-j} = \left(\sum_{j=0}^{q_{k,i}} \phi_{k,j,i} \right) X_{k,i,t-1} + \phi_{k,0,i} \Delta X_{k,i,t} - \sum_{j=1}^{q_{k,i}-1} \left(\sum_{l=j+1}^{q_{k,i}} \phi_{k,l,i} \right) \Delta X_{k,i,t-j}. \quad (6)$$

Substituting (5) and (6) into (1) and rearranging, the general ARDL($p_i, q_{1,i}, \dots, q_{K,i}$) error-correction model takes the form:

$$\Delta Y_{i,t} = \mu_i + \delta_i t + \lambda_i Y_{i,t-1} + \sum_{k=1}^K \beta_{k,i} X_{k,i,t-1} + \sum_{j=1}^{p_i-1} \psi_{j,i} \Delta Y_{i,t-j} + \sum_{k=1}^K \sum_{j=0}^{q_{k,i}-1} \gamma_{k,j,i} \Delta X_{k,i,t-j} + u_{i,t}, \quad (7)$$

where the reparameterized coefficients are defined as follows:

Speed of adjustment parameter:

$$\lambda_i = - \left(1 - \sum_{j=1}^{p_i} \phi_{j,i} \right) = -\Phi_i(1), \quad (8)$$

which equals the negative of the autoregressive polynomial evaluated at unity. Cointegration requires $\lambda_i < 0$, which is equivalent to $\Phi_i(1) > 0$, ensuring that the sum of the autoregressive coefficients is strictly less than one.

Long-run level coefficients:

$$\beta_{k,i} = \sum_{j=0}^{q_{k,i}} \varphi_{k,j,i} = \Psi_{k,i}(1), \quad k = 1, \dots, K, \quad (9)$$

which equal the distributed lag polynomials for each regressor evaluated at unity. These capture the cumulative (long-run) effect of each regressor on $Y_{i,t}$ through the lagged level terms in the error-correction equation.

Short-run dynamic coefficients on lagged differences of Y :

$$\psi_{j,i} = - \sum_{l=j+1}^{p_i} \phi_{l,i}, \quad j = 1, \dots, p_i - 1, \quad (10)$$

which capture the short-run dynamic adjustment through lagged differences of the dependent variable.

Short-run dynamic coefficients on current and lagged differences of X_k :

$$\gamma_{k,0,i} = \varphi_{k,0,i}, \quad (11a)$$

$$\gamma_{k,j,i} = - \sum_{l=j+1}^{q_{k,i}} \varphi_{k,l,i}, \quad j = 1, \dots, q_{k,i} - 1, \quad k = 1, \dots, K, \quad (11b)$$

which capture the short-run impact and dynamic adjustment effects of each regressor.

The error-correction representation in (7) is observationally equivalent to the autoregressive form in (1); no restrictions are imposed in the reparameterization. This algebraic equivalence ensures that estimation and inference are unaffected by the choice of representation. The advantage of the error-correction form is that it directly isolates the long-run equilibrium parameters (λ_i and $\beta_{k,i}$) from the short-run dynamic parameters ($\psi_{j,i}$ and $\gamma_{k,j,i}$), facilitating testing for the existence of a long-run level relationship.

To keep the exposition tractable and to align with the specification most commonly employed in applied work, the remainder of this paper focuses on the ARDL(1,1,...,1) case in which $p_i =$

1 and $q_{k,i} = 1$ for all $k = 1, \dots, K$ and $i = 1, \dots, N$. Under this specification, the lagged difference terms in (7) vanish (since $p_i - 1 = 0$ and $q_{k,i} - 1 = 0$), and the error-correction model simplifies to:

$$\Delta Y_{i,t} = \mu_i + \delta_i t + \lambda_i Y_{i,t-1} + \sum_{k=1}^K \beta_{k,i} X_{k,i,t-1} + \sum_{k=1}^K \gamma_{k,i} \Delta X_{k,i,t} + u_{i,t}, \quad (12)$$

where the coefficients in (12) are obtained from the general formulas (8)-(11) as:

$$\lambda_i = -(1 - \phi_{1,i}), \quad (13a)$$

$$\beta_{k,i} = \varphi_{k,0,i} + \varphi_{k,1,i}, \quad k = 1, \dots, K, \quad (13b)$$

$$\gamma_{k,i} = \varphi_{k,0,i} - \varphi_{k,1,i} + \beta_{k,i}, \quad k = 1, \dots, K. \quad (13c)$$

The ARDL(1,1,...,1) specification in (12) contains $m_i = 2K + 3$ parameters per equation: one intercept (μ_i), one trend coefficient (δ_i), one speed of adjustment parameter (λ_i), K long-run level coefficients ($\beta_{1,i}, \dots, \beta_{K,i}$), and K short-run dynamic coefficients ($\gamma_{1,i}, \dots, \gamma_{K,i}$).

Remark 1. All results derived below for the ARDL(1,1,...,1) case extend directly to the general ARDL($p_i, q_{1,i}, \dots, q_{K,i}$) specification by appropriate augmentation of the regressor matrix to include additional lagged differences $\Delta Y_{i,t-j}$ for $j = 1, \dots, p_i - 1$ and $\Delta X_{k,i,t-j}$ for $j = 1, \dots, q_{k,i} - 1$. The long-run coefficient definitions remain unchanged; only the dimension of the regressor vector and the parameter vector increase.

For estimation purposes, it is useful to express (12) in matrix notation. Define the regressor vector for unit i at time t as:

$$\mathbf{x}_{i,t} = (1, t, Y_{i,t-1}, X_{1,i,t-1}, X_{2,i,t-1}, \dots, X_{K,i,t-1}, \Delta X_{1,i,t}, \Delta X_{2,i,t}, \dots, \Delta X_{K,i,t})' \in \mathbb{R}^{m_i}, \quad (14)$$

with $m_i = 2K + 3$. The corresponding parameter vector is:

$$\boldsymbol{\theta}_i = (\mu_i, \delta_i, \lambda_i, \beta_{1,i}, \beta_{2,i}, \dots, \beta_{K,i}, \gamma_{1,i}, \gamma_{2,i}, \dots, \gamma_{K,i})' \in \mathbb{R}^{m_i}. \quad (15)$$

Equation (12) can then be written as:

$$\Delta Y_{i,t} = \mathbf{x}_{i,t}' \boldsymbol{\theta}_i + u_{i,t}, \quad t = 1, \dots, T. \quad (16)$$

Stacking all T observations for unit i :

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\theta}_i + \mathbf{u}_i, \quad (17)$$

where $\mathbf{y}_i = (\Delta Y_{i,1}, \Delta Y_{i,2}, \dots, \Delta Y_{i,T})' \in \mathbb{R}^T$ is the vector of first differences of the dependent variable, $\mathbf{X}_i = (\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,T})' \in \mathbb{R}^{T \times m_i}$ is the regressor matrix, and $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \dots, u_{i,T})' \in \mathbb{R}^T$ is the disturbance vector for unit i .

The ordinary least squares (OLS) estimator of $\boldsymbol{\theta}_i$ based on equation (17) is:

$$\hat{\boldsymbol{\theta}}_i^{\text{OLS}} = (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{y}_i, \quad (18)$$

with associated residuals $\hat{u}_{i,t} = \Delta Y_{i,t} - \mathbf{x}_{i,t}' \hat{\boldsymbol{\theta}}_i^{\text{OLS}}$ for $t = 1, \dots, T$.

Remark 2. In the single-equation ARDL setting, lag orders ($p_i, q_{1,i}, \dots, q_{K,i}$) are typically selected by minimizing an information criterion over a grid of candidate lag structures. Common choices include the Akaike Information Criterion (AIC):

$$\text{AIC}(p, q_1, \dots, q_K) = \ln(\hat{\sigma}_i^2) + \frac{2m_i}{T}, \quad (19)$$

and the Schwarz Bayesian Criterion (SBC):

$$\text{SBC}(p, q_1, \dots, q_K) = \ln(\hat{\sigma}_i^2) + \frac{m_i \ln T}{T}, \quad (20)$$

where $\hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^T \hat{u}_{i,t}^2$ is the residual variance from the equation with lag orders $(p, q_1, \dots, q_K)$. The SBC tends to select more parsimonious models and is generally preferred when the true lag orders are small relative to T , while the AIC may overfit in finite samples but is less likely to underfit. Narayan (2005) recommends the SBC for small samples due to its consistency property. Throughout the empirical application, we adopt the ARDL(1,1,...,1) specification for all units.

The standard ARDL bounds testing approach tests for the existence of a long-run level relationship by examining the joint significance of the lagged level terms in the error-correction equation (12). Specifically, the null hypothesis of no long-run level relationship is:

$$H_0^{(i)}: \lambda_i = 0 \text{ and } \beta_{1,i} = 0 \text{ and } \beta_{2,i} = 0 \text{ and } \dots \text{ and } \beta_{K,i} = 0, \quad (21)$$

against the alternative:

$$H_1^{(i)}: \lambda_i \neq 0 \text{ or } \beta_{k,i} \neq 0 \text{ for at least one } k. \quad (22)$$

Under the null hypothesis, the lagged level terms $Y_{i,t-1}$ and $X_{k,i,t-1}$ are redundant in (12), and there is no long-run equilibrium relationship between $Y_{i,t}$ and $\mathbf{X}_{i,t}$.

The test statistic is a standard Wald-type F-statistic computed from the residual sum of squares of the unrestricted model (12) and a restricted model that excludes the lagged level terms:

$$\Delta Y_{i,t} = \mu_i + \delta_i t + \sum_{k=1}^K \gamma_{k,i} \Delta X_{k,i,t} + u_{i,t}^R. \quad (23)$$

Let $\text{RSS}_U^{(i)} = \sum_{t=1}^T (\hat{u}_{i,t}^U)^2$ and $\text{RSS}_R^{(i)} = \sum_{t=1}^T (\hat{u}_{i,t}^R)^2$ denote the unrestricted and restricted residual sums of squares, respectively. The F-statistic for unit i is:

$$F_i = \frac{(\text{RSS}_R^{(i)} - \text{RSS}_U^{(i)}) / (K + 1)}{\text{RSS}_U^{(i)} / (T - m_i)}, \quad (24)$$

where $(K + 1)$ is the number of restrictions imposed under $H_0^{(i)}$ (one for λ_i and K for $\beta_{1,i}, \dots, \beta_{K,i}$), and $(T - m_i)$ is the degrees of freedom of the unrestricted regression.

The key insight of Pesaran et al. (2001) is that the asymptotic distribution of F_i under H_0 depends on whether the regressors are $I(0)$ or $I(1)$, which is unknown a priori. They derive two sets of asymptotic critical values: a lower bound corresponding to the case where all regressors are purely $I(0)$, and an upper bound corresponding to the case where all regressors are purely $I(1)$. The decision rule is:

(i) If F_i exceeds the upper bound critical value, reject $H_0^{(i)}$: there is evidence of a long-run level relationship.

(ii) If F_i falls below the lower bound critical value, do not reject $H_0^{(i)}$: there is no evidence of a long-run relationship.

(iii) If F_i falls between the lower and upper bounds, the test is inconclusive.

This bounds testing logic is preserved in the SUR extension developed below, with the crucial difference that critical values must be re-derived via Monte Carlo simulation to account for the cross-equation dependence structure.

Remark 3. The single-equation bounds test described above is applied independently to each unit i . This approach has two well-known limitations when applied in a multi-country or multi-region setting. First, it ignores contemporaneous correlation across units, which can arise from common shocks, trade linkages, financial spillovers, or shared institutional features. Ignoring this correlation does not bias the OLS estimates of θ_i (since OLS remains consistent under heteroskedasticity and contemporaneous correlation in the SUR setting), but it leads to inefficient inference, particularly for Wald-type test statistics (Zellner, 1962; Greene, 2018). Second, by treating each equation in isolation, the single-equation approach cannot exploit the information contained in the cross-equation covariance structure to sharpen cointegration inference. These limitations motivate the system-based SUR extension developed in the next subsection.

As noted by Pesaran (2006, pp. 967-968), the standard approach to modeling systems of cointegrating relationships in settings with small cross-sectional dimension typically smaller than ten is to consider the relationships jointly as a system of seemingly unrelated regressions. This approach allows one to simultaneously handle both cross-sectional heterogeneity and cross-sectional dependence, an insight that has been productively exploited in the SUR cointegration literature (Park & Ogaki, 1991; Moon, 1999; Mark et al., 2005; Moon & Perron, 2004; Wagner et al., 2020). The present contribution follows this approach by embedding the ARDL error-correction specification within the SUR framework.

2.2 System Specification under Contemporaneous Cross-Sectional Dependence

Individual-equation estimation of the ARDL error-correction model (12) for each unit $i = 1, \dots, N$ implicitly treats the disturbances $\{u_{1,t}, u_{2,t}, \dots, u_{N,t}\}$ as mutually independent across units at each point in time. Formally, the standard single-equation approach assumes:

$$E(u_{i,t} u_{j,t}) = 0, \quad \text{for all } i \neq j \text{ and all } t. \quad (25)$$

This independence assumption may be restrictive in many empirical settings, particularly for regionally integrated economies, where common macroeconomic shocks (e.g., global oil price fluctuations, monetary policy spillovers, financial contagion), coordinated policy interventions, and trade linkages induce contemporaneous correlation across units. A substantial body of empirical evidence suggests that cross-sectional dependence is pervasive in macroeconomic and energy panel data (Pesaran, 2004; De Hoyos & Sarafidis, 2006).

While ignoring cross-equation dependence does not bias the OLS estimates of θ_i under standard regularity conditions since each equation is correctly specified and OLS applied equation-by-equation remains consistent it leads to a loss of estimation efficiency and, more importantly, can distort the size of Wald-type test statistics used for cointegration inference (Zellner, 1962; Greene, 2018). In the context of the ARDL bounds test, this implies that single-equation F-statistics may over-reject or under-reject the null hypothesis of no cointegration when cross-equation dependence is present, leading to potentially misleading inference.

The classical solution to this problem is the Seemingly Unrelated Regression (SUR) framework of Zellner (1962), which jointly estimates a system of equations that are linked through their contemporaneous error covariance structure. The SUR approach has been successfully applied in various branches of the cointegration literature: Park and Ogaki (1991) develop seemingly unrelated canonical cointegrating regressions; Moon (1999) considers fully modified estimation in SUR cointegrating systems; Mark et al. (2005) propose dynamic

seemingly unrelated cointegrating regressions; and most recently, Wagner, Grabarczyk, and Hong (2020) extend the SUR framework to cointegrating polynomial regressions with nonlinear transformations of integrated processes. The common insight across this literature is that system-based estimation exploiting cross-equation covariance information yields efficiency gains relative to equation-by-equation procedures, with the magnitude of the gains increasing in the degree of cross-equation correlation and the dissimilarity of regressors across equations.

The present contribution follows this established approach by embedding the ARDL error-correction specification (12) within the SUR framework. An important distinction from the SUR cointegration literature cited above which focuses on levels regressions and typically requires fully modified or dynamic OLS corrections to handle regressor endogeneity and serial correlation is that the ARDL error-correction reparameterization already accommodates these issues through the inclusion of lagged dependent and differenced regressor terms. This simplifies the estimation procedure: standard feasible GLS estimation can be applied without the need for semi-parametric corrections to the dependent variable or additive bias correction terms of the type required by the FM-SOLS and FM-SUR estimators of Wagner et al. (2020, Proposition 1).

We now construct the system of N ARDL error-correction equations that will be estimated jointly. Recall the equation for unit i at time t can be written as $\Delta Y_{i,t} = \mathbf{x}_{i,t}' \boldsymbol{\theta}_i + u_{i,t}$, where $\mathbf{x}_{i,t} \in \mathbb{R}^{m_i}$ is defined in (14) and $\boldsymbol{\theta}_i \in \mathbb{R}^{m_i}$ is defined in (15). Stacking the N equations at time t yields the cross-sectional system:

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\theta} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (26)$$

where the constituent vectors and matrices are defined as follows.

Dependent variable vector:

$$\mathbf{y}_t = \begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \\ \vdots \\ \Delta Y_{N,t} \end{pmatrix} \in \mathbb{R}^N. \quad (27)$$

System parameter vector:

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_N \end{pmatrix} \in \mathbb{R}^M, \quad (28)$$

where $M = \sum_{i=1}^N m_i$ is the total number of parameters across all N equations. Under the ARDL(1,1,...,1) specification with identical regressor dimensions across units, $m_i = 2K + 3$ for all i , giving $M = N(2K + 3)$.

System disturbance vector:

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{N,t} \end{pmatrix} \in \mathbb{R}^N. \quad (29)$$

Block-diagonal regressor matrix:

$$\mathbf{Z}_t = \text{diag}(\mathbf{x}_{1,t'}, \mathbf{x}_{2,t'}, \dots, \mathbf{x}_{N,t'}) = \begin{pmatrix} \mathbf{x}_{1,t'} & \mathbf{0}' & \dots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{x}_{2,t'} & \dots & \mathbf{0}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \dots & \mathbf{x}_{N,t'} \end{pmatrix} \in \mathbb{R}^{N \times M}. \quad (30)$$

The block-diagonal structure of $\mathbf{Z}_t$ arises because each equation contains unit-specific regressors only; there are no cross-unit regressors in the ARDL specification. This structure is a key feature that distinguishes the SUR-ARDL framework from system specifications that include common regressors (e.g., common factors or global variables). The block-diagonal structure implies that the i -th row of $\mathbf{Z}_t$ selects only the regressors $\mathbf{x}_{i,t}$ relevant to unit i , confirming that the system equation (26) is equivalent to the N individual equations in (16) stacked vertically.

To express the regressor matrix more explicitly, note that row i of $\mathbf{Z}_t$ is:

$$[\mathbf{Z}_t]_{i,\cdot} = (\mathbf{0}_{1 \times \sum_{l=1}^{i-1} m_l}, \mathbf{x}_{i,t'}, \mathbf{0}_{1 \times \sum_{l=i+1}^N m_l}) \in \mathbb{R}^{1 \times M}, \quad i = 1, \dots, N. \quad (31)$$

This means that the system matrix $\mathbf{Z}_t$ maps the full parameter vector $\boldsymbol{\theta}$ to the N -dimensional vector of dependent variables, with each unit's regressors appearing in the appropriate block position.

Stacking the system (26) over all T time periods yields the full SUR-ARDL system:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad (32)$$

where:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{pmatrix} \in \mathbb{R}^{NT}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_T \end{pmatrix} \in \mathbb{R}^{NT \times M}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_T \end{pmatrix} \in \mathbb{R}^{NT}. \quad (33)$$

The system (32) has NT observations and M parameters. This is the fundamental equation to be estimated using feasible generalized least squares. However, before proceeding to estimation, we must specify the assumptions governing the behavior of the disturbance vector $\boldsymbol{\varepsilon}$, the integration properties of the variables, and the regressor conditions.

We now state the formal assumptions under which the SUR-ARDL framework is developed. These assumptions govern three aspects of the model: the disturbance process (Assumption 1), the integration properties of the variables (Assumption 2), and the behavior of the regressors (Assumption 3).

Assumption 1. The disturbance vector $\boldsymbol{\varepsilon}_t = (u_{1,t}, u_{2,t}, \dots, u_{N,t})'$ satisfies the following conditions:

(i) Zero mean:

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}_N, \quad \text{for all } t = 1, \dots, T. \quad (34)$$

(ii) Contemporaneous covariance:

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma}, \quad \text{for all } t, \quad (35)$$

where $\boldsymbol{\Sigma} = [\sigma_{ij}]_{N \times N}$ is a symmetric positive definite matrix with typical element $\sigma_{ij} = E(u_{i,t} u_{j,t})$. Off-diagonal elements satisfy $\sigma_{ij} \neq 0$ for $i \neq j$ in general, capturing contemporaneous cross-sectional dependence.

(iii) Temporal independence:

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{s'}') = \mathbf{0}_{N \times N}, \quad \text{for all } t \neq s. \quad (36)$$

(iv) Finite fourth moments:

$$E(u_{i,t}^4) < \infty, \quad \text{for all } i = 1, \dots, N \text{ and } t = 1, \dots, T. \quad (37)$$

The covariance matrix $\boldsymbol{\Sigma}$ can be written explicitly as:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2N} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \cdots & \sigma_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \cdots & \sigma_{NN} \end{pmatrix}, \quad (38)$$

where $\sigma_{ii} = \text{Var}(u_{i,t}) > 0$ denotes the error variance for unit i and $\sigma_{ij} = \text{Cov}(u_{i,t}, u_{j,t})$ denotes the contemporaneous covariance between the disturbances of units i and j . The positive definiteness of $\boldsymbol{\Sigma}$ ensures that the covariance structure is well-defined and that $\boldsymbol{\Sigma}^{-1}$ exists, a requirement for GLS estimation.

Several remarks on Assumption 1 are in order.

Remark 4. In the SUR cointegration literature for levels regressions, including Wagner et al. (2020, Assumption 1), the disturbance processes are typically assumed to be generated by stationary linear filters driven by martingale difference sequences, allowing for serial correlation in the errors and dynamic cross-sectional dependence. The long-run covariance matrix $\boldsymbol{\Omega}_{uu} = \sum_{j=-\infty}^{\infty} E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-j}')'$ then replaces the contemporaneous covariance matrix $\boldsymbol{\Sigma}$ in the construction of the GLS-type estimator, and estimated long-run covariance matrices are used as weighting matrices what Park and Ogaki (1991) term the "modified SUR" (MSUR) estimator; see Wagner et al. (2020, equation (11)). In the ARDL error-correction framework, the inclusion of lagged dependent variables and differenced regressors is designed to absorb serial correlation in the original error process, rendering the conditional error $u_{i,t}$ in (12) approximately white noise under correct lag specification. Consequently, Assumption 1(iii) temporal independence is appropriate in the ARDL context, and the contemporaneous covariance matrix $\boldsymbol{\Sigma}$ (rather than the long-run covariance matrix $\boldsymbol{\Omega}_{uu}$) is the correct weighting matrix for the SUR estimator. This constitutes a notable simplification relative to the FM-SUR and FM-SOLS procedures in Wagner et al. (2020), which require consistent estimation of (potentially large-dimensional) long-run and half long-run covariance matrices.

Remark 5. Assumption 1(iii) requires that the disturbances are serially uncorrelated after conditioning on the regressors. This is a testable assumption: the Breusch-Godfrey LM test can be applied to the residuals of each equation to verify the absence of serial correlation. If serial correlation is detected, the ARDL lag orders p_i and $q_{k,i}$ should be augmented until the residuals are approximately white noise. The sufficiency of the ARDL(1,1,...,1) specification for eliminating serial correlation is an empirical matter that must be verified in each application.

Remark 6. The off-diagonal elements of $\boldsymbol{\Sigma}$ capture common shocks and spillover effects that affect multiple units simultaneously. Sources of such dependence include: (a) common global or regional shocks (e.g., oil price movements, financial crises, coordinated monetary policy); (b) trade and investment linkages that transmit shocks across economies; (c) spatial or geographic proximity effects; and (d) common institutional or policy environments. The magnitude of the off-diagonal elements σ_{ij} determines the potential efficiency gains from SUR estimation: larger cross-equation correlations imply greater gains from joint estimation relative to equation-by-equation OLS.

Under Assumption 1, the variance-covariance matrix of the full stacked disturbance vector $\mathbf{\varepsilon}$ in (32) takes the Kronecker product form:

$$\mathbf{\Omega} = E(\mathbf{\varepsilon}\mathbf{\varepsilon}') = \mathbf{I}_T \otimes \mathbf{\Sigma}, \quad (39)$$

where $\mathbf{I}_T$ is the $T \times T$ identity matrix and $\otimes$ denotes the Kronecker product. To verify (39), note that for the (t, s) -th $N \times N$ block of $\mathbf{\Omega}$:

$$[\mathbf{\Omega}]_{t,s} = E(\mathbf{\varepsilon}_t \mathbf{\varepsilon}_s') = \begin{cases} \mathbf{\Sigma} & \text{if } t = s, \\ \mathbf{0}_{N \times N} & \text{if } t \neq s, \end{cases} \quad (40)$$

which is precisely the (t, s) -th block of $\mathbf{I}_T \otimes \mathbf{\Sigma}$. The Kronecker structure combines two features: contemporaneous cross-equation correlation (captured by $\mathbf{\Sigma}$) and temporal independence (captured by $\mathbf{I}_T$). The inverse of $\mathbf{\Omega}$ is:

$$\mathbf{\Omega}^{-1} = \mathbf{I}_T \otimes \mathbf{\Sigma}^{-1}, \quad (41)$$

which follows from the standard Kronecker product identity $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ when both $\mathbf{A}$ and $\mathbf{B}$ are invertible.

Assumption 2. For all $i = 1, \dots, N$ and $k = 1, \dots, K$:

- (i) The dependent variable $Y_{i,t}$ is integrated of order $d_{y,i}$ with $d_{y,i} \in \{0,1\}$, i.e., $Y_{i,t} \sim I(d_{y,i})$.
- (ii) Each explanatory variable $X_{k,i,t}$ is integrated of order $d_{k,i}$ with $d_{k,i} \in \{0,1\}$, i.e., $X_{k,i,t} \sim I(d_{k,i})$.
- (iii) No variable is integrated of order two or higher:

$$d_{y,i} \leq 1 \quad \text{and} \quad d_{k,i} \leq 1, \quad \text{for all } i, k. \quad (42)$$

- (iv) When $d_{y,i} = 1$, there exists at least one k such that $d_{k,i} = 1$, ensuring that the potential cointegrating relationship involves at least one integrated regressor.

Assumption 2 is the core requirement of the ARDL bounds testing approach of Pesaran et al. (2001). Unlike the Engle-Granger (1987) and Johansen (1991) frameworks, which require all variables to be integrated of the same order (typically $I(1)$), the ARDL approach permits a mixture of $I(0)$ and $I(1)$ variables. This flexibility is the primary practical advantage of the ARDL framework: it avoids the need for pre-testing of integration orders (and the associated uncertainty introduced by unit root tests with low power in finite samples), while remaining valid under both extremes and all intermediate cases. The exclusion of $I(2)$ variables in condition (iii) is necessary because the presence of $I(2)$ processes would invalidate the asymptotic theory underlying the bounds test, as the F-statistic would no longer be bounded between the tabulated $I(0)$ and $I(1)$ critical values. In practice, Assumption 2(iii) should be verified by applying standard unit root tests (ADF, PP, or KPSS) to both levels and first differences of all variables.

Assumption 3. For each unit $i = 1, \dots, N$:

- (i) Weak exogeneity: The regressors $\mathbf{x}_{i,t}$ are weakly exogenous with respect to the long-run parameters $(\lambda_i, \beta_{1,i}, \dots, \beta_{K,i})$. That is, the parameters of the marginal process for $\mathbf{X}_{i,t}$ do not contain information about the long-run parameters of interest, and the error terms in the conditional model (12) do not affect the long-run equilibrium relationship.
- (ii) Full column rank: The regressor matrix $\mathbf{X}_i = (\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,T})'$ has full column rank m_i with probability one:

$$\text{rank}(\mathbf{X}_i) = m_i \quad \text{a.s.} \quad (43)$$

This ensures that the single-equation OLS estimator $\hat{\theta}_i^{\text{OLS}}$ in (18) is well-defined.

(iii) Moment conditions: The regressors satisfy:

$$\frac{1}{T} \mathbf{X}_i' \mathbf{X}_i \xrightarrow{p} \mathbf{Q}_i, \quad \text{as } T \rightarrow \infty, \quad (44)$$

where $\mathbf{Q}_i$ is a positive definite $m_i \times m_i$ matrix. Furthermore, the cross-products satisfy:

$$\frac{1}{T} \mathbf{X}_i' \mathbf{u}_i \xrightarrow{p} \mathbf{0}_{m_i}, \quad \text{as } T \rightarrow \infty, \quad (45)$$

ensuring orthogonality between regressors and disturbances in the probability limit.

(iv) No perfect multicollinearity across equations: The system regressor matrix $\mathbf{Z}$ in (32) satisfies:

$$\text{rank}(\mathbf{Z}) = M \quad \text{a.s.} \quad (46)$$

2.3 Feasible Generalized Least Squares Estimation

The GLS estimator

Given the SUR-ARDL system (32) and the covariance structure (39), the generalized least squares (GLS) estimator of θ minimizes the generalized sum of squares:

$$S(\theta) = (\mathbf{Y} - \mathbf{Z}\theta)' \mathbf{\Omega}^{-1} (\mathbf{Y} - \mathbf{Z}\theta), \quad (47)$$

which yields the GLS estimator:

$$\hat{\theta}_{\text{GLS}} = (\mathbf{Z}' \mathbf{\Omega}^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{\Omega}^{-1} \mathbf{Y}. \quad (48)$$

Substituting $\mathbf{\Omega}^{-1} = \mathbf{I}_T \otimes \mathbf{\Sigma}^{-1}$ from (41):

$$\hat{\theta}_{\text{GLS}} = (\mathbf{Z}' (\mathbf{I}_T \otimes \mathbf{\Sigma}^{-1}) \mathbf{Z})^{-1} \mathbf{Z}' (\mathbf{I}_T \otimes \mathbf{\Sigma}^{-1}) \mathbf{Y}. \quad (49)$$

To see the structure of this estimator more explicitly, note that:

$$\mathbf{Z}' (\mathbf{I}_T \otimes \mathbf{\Sigma}^{-1}) \mathbf{Z} = \sum_{t=1}^T \mathbf{Z}_t' \mathbf{\Sigma}^{-1} \mathbf{Z}_t, \quad (50)$$

$$\mathbf{Z}' (\mathbf{I}_T \otimes \mathbf{\Sigma}^{-1}) \mathbf{Y} = \sum_{t=1}^T \mathbf{Z}_t' \mathbf{\Sigma}^{-1} \mathbf{y}_t. \quad (51)$$

Expanding the matrix product $\mathbf{Z}_t' \mathbf{\Sigma}^{-1} \mathbf{Z}_t$ using the block-diagonal structure of $\mathbf{Z}_t$ in (30), the (i, j) -th block (of dimension $m_i \times m_j$) of this $M \times M$ matrix is:

$$[\mathbf{Z}_t' \mathbf{\Sigma}^{-1} \mathbf{Z}_t]_{i,j} = \sigma^{ij} \mathbf{x}_{i,t} \mathbf{x}_{j,t}', \quad i, j = 1, \dots, N, \quad (52)$$

where σ^{ij} denotes the (i, j) -th element of $\mathbf{\Sigma}^{-1}$. This reveals that the GLS estimator uses the inverse covariance weights σ^{ij} to construct cross-products between the regressors of different equations. When $\mathbf{\Sigma}$ is diagonal (i.e., $\sigma_{ij} = 0$ for $i \neq j$), the cross-equation blocks vanish and the GLS estimator reduces to equation-by-equation weighted least squares, which coincides with equation-by-equation OLS when the error variances are the same across equations.

Similarly, the (i) -th block (of dimension $m_i \times 1$) of the vector $\mathbf{Z}_t' \mathbf{\Sigma}^{-1} \mathbf{y}_t$ is:

$$[\mathbf{Z}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t]_i = \sum_{j=1}^N \sigma^{ij} \mathbf{x}_{i,t} \Delta Y_{j,t}, \quad i = 1, \dots, N. \quad (53)$$

This expression shows that in the GLS estimator, the dependent variable from equation j contributes to the estimation of parameters in equation i whenever $\sigma^{ij} \neq 0$, i.e., whenever equations i and j are contemporaneously correlated. This cross-equation information sharing is the source of the efficiency gains of SUR over OLS.

The feasible GLS estimator

Since $\boldsymbol{\Sigma}$ is unknown in practice, the GLS estimator (48) is infeasible. A feasible version is obtained by replacing $\boldsymbol{\Sigma}$ with a consistent estimator $\hat{\boldsymbol{\Sigma}}$, leading to the feasible generalized least squares (FGLS) estimator, which we term the SUR-ARDL estimator. The estimation proceeds in two steps.

Step 1: Equation-by-equation OLS. Estimate each equation $i = 1, \dots, N$ separately by ordinary least squares as in (18) to obtain residuals:

$$\hat{u}_{i,t} = \Delta Y_{i,t} - \mathbf{x}_{i,t}' \hat{\boldsymbol{\theta}}_i^{\text{OLS}}, \quad t = 1, \dots, T, \quad i = 1, \dots, N. \quad (54)$$

These residuals are consistent estimates of the true disturbances $u_{i,t}$ under the maintained assumptions, in the sense that $\hat{u}_{i,t} - u_{i,t} = -\mathbf{x}_{i,t}'(\hat{\boldsymbol{\theta}}_i^{\text{OLS}} - \boldsymbol{\theta}_i) \xrightarrow{p} 0$ as $T \rightarrow \infty$.

Step 2: Covariance estimation and FGLS. Construct the estimated contemporaneous covariance matrix:

$$\hat{\sigma}_{ij} = \frac{1}{T - d_f} \sum_{t=1}^T \hat{u}_{i,t} \hat{u}_{j,t}, \quad i, j = 1, \dots, N, \quad (55)$$

where d_f is a degrees-of-freedom correction. Common choices include $d_f = 0$ (the maximum likelihood estimator under normality, as in the original formulation of Zellner, 1962), $d_f = m_i$ (adjusting for the number of regressors in equation i , applicable when $m_i = m_j$ for all i, j), or $d_f = \sqrt{m_i \cdot m_j}$ (a compromise adjustment proposed by some authors for the case of unequal regressor dimensions). In the present framework with ARDL(1,1,...,1) specifications, all equations have the same number of regressors $m_i = 2K + 3$, so the choice of degrees-of-freedom correction is straightforward. The estimated covariance matrix is:

$$\hat{\boldsymbol{\Sigma}} = [\hat{\sigma}_{ij}]_{N \times N}. \quad (56)$$

Under Assumption 1 and standard regularity conditions, $\hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}$ as $T \rightarrow \infty$, ensuring consistency of the FGLS estimator.

The SUR-ARDL estimator is then defined as:

$$\hat{\boldsymbol{\theta}}_{\text{SUR}} = (\mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Y}, \quad (57)$$

where $\hat{\boldsymbol{\Omega}} = \mathbf{I}_T \otimes \hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Omega}}^{-1} = \mathbf{I}_T \otimes \hat{\boldsymbol{\Sigma}}^{-1}$.

The variance-covariance matrix of $\hat{\boldsymbol{\theta}}_{\text{SUR}}$ is estimated as:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}_{\text{SUR}}) = (\mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z})^{-1}. \quad (58)$$

This $M \times M$ matrix provides the estimated variances and covariances of all parameter estimates across all equations in the system. Importantly, the off-diagonal blocks of $\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}_{\text{SUR}})$

capture the covariances between parameter estimates from different equations, which arise through the cross-equation covariance structure and are absent in equation-by-equation OLS.

The two-step FGLS estimator in (57) can be iterated as follows:

Iteration s ($s = 1, 2, \dots$):

- (a) Compute residuals from the current FGLS estimates: $\hat{u}_{i,t}^{(s)} = \Delta Y_{i,t} - \mathbf{x}_{i,t}' \hat{\boldsymbol{\theta}}_{\text{SUR},i}^{(s)}$.
- (b) Re-estimate the covariance matrix: $\hat{\sigma}_{ij}^{(s+1)} = T^{-1} \sum_{t=1}^T \hat{u}_{i,t}^{(s)} \hat{u}_{j,t}^{(s)}$.
- (c) Re-compute the FGLS estimator: $\hat{\boldsymbol{\theta}}_{\text{SUR}}^{(s+1)} = (\mathbf{Z}'(\mathbf{I}_T \otimes (\hat{\boldsymbol{\Sigma}}^{(s+1)})^{-1})\mathbf{Z})^{-1} \mathbf{Z}'(\mathbf{I}_T \otimes (\hat{\boldsymbol{\Sigma}}^{(s+1)})^{-1})\mathbf{Y}$.
- (d) Iterate until convergence: $\|\hat{\boldsymbol{\theta}}_{\text{SUR}}^{(s+1)} - \hat{\boldsymbol{\theta}}_{\text{SUR}}^{(s)}\| < \epsilon$ for a pre-specified tolerance $\epsilon > 0$.

Under the assumption of normally distributed errors ($\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$), the iterated FGLS estimator converges to the maximum likelihood estimator (Oberhofer & Kmenta, 1974). This result extends straightforwardly to the SUR-ARDL setting. In practice, convergence typically occurs within a small number of iterations (often 3-5), and the gains from iteration relative to the two-step estimator are generally modest.

A fundamental result due to Zellner (1962) establishes the conditions under which the SUR estimator coincides with equation-by-equation OLS, i.e., $\hat{\boldsymbol{\theta}}_{\text{SUR}} = \hat{\boldsymbol{\theta}}_{\text{OLS}}$, where $\hat{\boldsymbol{\theta}}_{\text{OLS}} = (\hat{\boldsymbol{\theta}}_1^{\text{OLS}}, \hat{\boldsymbol{\theta}}_2^{\text{OLS}}, \dots, \hat{\boldsymbol{\theta}}_N^{\text{OLS}})'$ with each $\hat{\boldsymbol{\theta}}_i^{\text{OLS}}$ defined in (18).

Proposition 1 (Zellner, 1962). The SUR estimator $\hat{\boldsymbol{\theta}}_{\text{SUR}}$ in (57) is numerically identical to the equation-by-equation OLS estimator $\hat{\boldsymbol{\theta}}_{\text{OLS}}$ if and only if at least one of the following conditions holds:

- (a) The covariance matrix is diagonal: $\boldsymbol{\Sigma} = \text{diag}(\sigma_{11}, \sigma_{22}, \dots, \sigma_{NN})$, i.e., $\sigma_{ij} = 0$ for all $i \neq j$.
- (b) The regressors are identical across all equations: $\mathbf{X}_i = \mathbf{X}_j$ for all $i, j = 1, \dots, N$.

In the ARDL setting, condition (a) states that there is no contemporaneous cross-sectional dependence. When this holds, there is no cross-equation information to exploit, and SUR offers no advantage over OLS. This is the scenario in which the single-equation ARDL bounds test is fully efficient.

Condition (b) requires that all equations contain exactly the same regressors with the same data. In the ARDL context, this would require all units to have identical dependent variables, identical explanatory variables, and identical data realizations, a degenerate case that is not meaningful in panel applications.

The practical implication of Proposition 1 for the SUR-ARDL framework is as follows. Since the ARDL specification permits heterogeneous lag orders and, more fundamentally, uses unit-specific data ($Y_{i,t}$, $\mathbf{X}_{i,t}$) that differ across units, condition (b) is generically violated in all non-trivial panel applications. Consequently, the SUR estimator is strictly more efficient than OLS whenever condition (a) is also violated, that is, whenever the errors are contemporaneously correlated across at least two units. The efficiency gain is formalized in the following proposition.

Proposition 2 (Efficiency gain). Under Assumptions 1-3, and assuming that $\boldsymbol{\Sigma}$ is non-diagonal and that the regressors differ across at least two equations, the SUR-ARDL estimator $\hat{\boldsymbol{\theta}}_{\text{SUR}}$ is asymptotically more efficient than the equation-by-equation OLS estimator $\hat{\boldsymbol{\theta}}_{\text{OLS}}$ in the sense that:

$$\text{Avar}(\hat{\boldsymbol{\theta}}_{\text{OLS}}) - \text{Avar}(\hat{\boldsymbol{\theta}}_{\text{SUR}}) \text{ is positive semi-definite, } (59)$$

where $\text{Avar}(\cdot)$ denotes the asymptotic variance-covariance matrix. The efficiency gain is strictly positive (the difference is positive definite) when the off-diagonal elements of $\boldsymbol{\Sigma}$ are non-zero and the regressor matrices $\mathbf{X}_i$ differ across equations.

Feasibility condition and dimensionality

The FGLS estimator (57) requires that $\hat{\boldsymbol{\Sigma}}$ in (56) be invertible. A necessary condition for invertibility is $T > N$, since the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ has rank at most $\min(T, N)$. When $T \leq N$, $\hat{\boldsymbol{\Sigma}}$ is singular and the FGLS estimator is not defined.

In the present SUR-ARDL framework, which targets panels with small to moderate cross-sectional dimension (N typically less than 10) and moderate to large time series dimension (T typically 30 or more in macroeconomic applications), the condition $T > N$ is easily satisfied. This is in contrast to large- N panel methods where N may be comparable to or larger than T , necessitating restrictive assumptions on the covariance structure (e.g., factor models or spatial structures) or regularization techniques.

For panels where N is moderately large relative to T , the estimation of $\hat{\boldsymbol{\Sigma}}$ may be unreliable even if $T > N$, as the number of free parameters in $\boldsymbol{\Sigma}$ is $N(N + 1)/2$, which grows quadratically in N . Wagner et al. (2020) document that the estimation of large-dimensional unrestricted long-run covariance matrices is "the key reason for the relatively poor performance" of their FM-SUR estimator in configurations with large N and small T (see their Table 8 and accompanying discussion). The same caveat applies to the SUR-ARDL framework: the approach is designed for settings where N is small to moderate, and the time series dimension T is sufficiently large to ensure reliable estimation of the $N \times N$ covariance matrix. When N is large, shrinkage estimators for $\boldsymbol{\Sigma}$ (Ledoit & Wolf, 2004) or restrictions on the covariance structure (e.g., block-diagonal or banded structures) may be employed to improve finite-sample performance, though these extensions are beyond the scope of the present paper.

2.4 System-Based Bounds Test for Cointegration

The single-equation ARDL bounds test, described in Section 2.1, tests for the existence of a long-run level relationship in each equation independently. In the SUR-ARDL framework, we retain the equation-specific bounds test but conduct it using the system estimator, which exploits cross-equation covariance information to improve the power and size properties of the test. We also develop a joint system-wide test for the simultaneous existence of long-run relationships across all (or a subset of) equations.

Recall from equation (12) that the ARDL(1,1,...,1) error-correction model for unit i is:

$$\Delta Y_{i,t} = \mu_i + \delta_i t + \lambda_i Y_{i,t-1} + \sum_{k=1}^K \beta_{k,i} X_{k,i,t-1} + \sum_{k=1}^K \gamma_{k,i} \Delta X_{k,i,t} + u_{i,t}. \quad (60)$$

The lagged level terms $Y_{i,t-1}$ and $X_{1,i,t-1}, \dots, X_{K,i,t-1}$ embody the long-run equilibrium relationship. When these terms are jointly insignificant, there is no long-run relationship between $Y_{i,t}$ and $\mathbf{X}_{i,t}$, and the error-correction equation reduces to a regression in first differences without any levels information.

Equation-specific hypothesis

For a given unit i , the null hypothesis of no long-run level relationship is:

$$H_0^{(i)}: \lambda_i = 0, \beta_{1,i} = 0, \beta_{2,i} = 0, \dots, \beta_{K,i} = 0, \quad (61)$$

against the alternative:

$$H_1^{(i)}: \lambda_i \neq 0 \text{ or } \beta_{k,i} \neq 0 \text{ for at least one } k \in \{1, \dots, K\}. \quad (62)$$

This involves $(K + 1)$ restrictions on the parameter vector θ_i . Collecting the restrictions in matrix form, we define the restriction matrix for unit i as:

$$\mathbf{R}_i = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{(K+1) \times m_i}, \quad (63)$$

which selects the elements $(\lambda_i, \beta_{1,i}, \dots, \beta_{K,i})$ from θ_i . The null hypothesis (61) can then be written as $\mathbf{R}_i \theta_i = \mathbf{0}_{K+1}$.

To embed this restriction within the full system parameter vector $\theta = (\theta_1', \dots, \theta_N')' \in \mathbb{R}^M$, define the system-level restriction matrix for testing unit i 's cointegration hypothesis as:

$$\tilde{\mathbf{R}}_i = \left(\mathbf{0}_{(K+1) \times \sum_{l=1}^{i-1} m_l}, \mathbf{R}_i, \mathbf{0}_{(K+1) \times \sum_{l=i+1}^N m_l} \right) \in \mathbb{R}^{(K+1) \times M}, \quad (64)$$

so that the null hypothesis for unit i is equivalently $\tilde{\mathbf{R}}_i \theta = \mathbf{0}_{K+1}$.

Joint system-wide hypothesis

A natural extension is to test the joint null hypothesis that no long-run level relationship exists in any equation:

$$H_0^{\text{Joint}}: \lambda_i = 0, \beta_{1,i} = 0, \dots, \beta_{K,i} = 0, \text{ for all } i = 1, \dots, N, \quad (65)$$

against the alternative that at least one unit exhibits a cointegrating relationship:

$$H_1^{\text{Joint}}: H_0^{(i)} \text{ is violated for at least one } i \in \{1, \dots, N\}. \quad (66)$$

The joint test involves $N(K + 1)$ restrictions. The corresponding system-level restriction matrix is obtained by stacking the individual restriction matrices:

$$\tilde{\mathbf{R}}^{\text{Joint}} = \begin{pmatrix} \tilde{\mathbf{R}}_1 \\ \tilde{\mathbf{R}}_2 \\ \vdots \\ \tilde{\mathbf{R}}_N \end{pmatrix} \in \mathbb{R}^{N(K+1) \times M}. \quad (67)$$

One may also consider intermediate cases for example, testing for cointegration in a subset $\mathcal{S} \subseteq \{1, \dots, N\}$ of equations by stacking only the restriction matrices for units in $\mathcal{S}$.

The bounds test is based on the comparison of residual sums of squares from the unrestricted system (which includes all lagged level terms) and the restricted system (which excludes the lagged level terms from the relevant equation or equations).

Unrestricted system. The unrestricted SUR-ARDL system is precisely the system (32) estimated via FGLS as in (57). The unrestricted estimator produces residuals:

$$\hat{u}_{i,t}^U = \Delta Y_{i,t} - \mathbf{x}_{i,t}' \hat{\theta}_{\text{SUR},i}, \quad t = 1, \dots, T, \quad i = 1, \dots, N, \quad (68)$$

where $\hat{\theta}_{\text{SUR},i}$ denotes the sub-vector of $\hat{\theta}_{\text{SUR}}$ corresponding to equation i .

Restricted system for equation i . Under the null hypothesis $H_0^{(i)}$ in (61), the model for unit i reduces to:

$$\Delta Y_{i,t} = \mu_i^R + \delta_i^R t + \sum_{k=1}^K \gamma_{k,i}^R \Delta X_{k,i,t} + u_{i,t}^R, \quad (69)$$

while the models for all other units $j \neq i$ remain unrestricted as in (60). The restricted regressor vector for unit i is:

$$\mathbf{x}_{i,t}^R = (1, t, \Delta X_{1,i,t}, \Delta X_{2,i,t}, \dots, \Delta X_{K,i,t})' \in \mathbb{R}^{K+2}, \quad (70)$$

with the corresponding restricted parameter vector:

$$\boldsymbol{\theta}_i^R = (\mu_i^R, \delta_i^R, \gamma_{1,i}^R, \gamma_{2,i}^R, \dots, \gamma_{K,i}^R)' \in \mathbb{R}^{K+2}. \quad (71)$$

The restricted system is obtained by replacing $\mathbf{x}_{i,t}$ with $\mathbf{x}_{i,t}^R$ and $\boldsymbol{\theta}_i$ with $\boldsymbol{\theta}_i^R$ in the i -th equation, while leaving all other equations unchanged. In compact notation, the restricted system regressor matrix at time t is:

$$\mathbf{Z}_t^{R(i)} = \text{diag}(\mathbf{x}_{1,t}', \dots, \mathbf{x}_{i-1,t}', (\mathbf{x}_{i,t}^R)', \mathbf{x}_{i+1,t}', \dots, \mathbf{x}_{N,t}') \in \mathbb{R}^{N \times M_R^{(i)}}, \quad (72)$$

where $M_R^{(i)} = M - (K + 1)$ is the total number of parameters in the restricted system. The restricted system is estimated by FGLS in the same manner as the unrestricted system:

$$\hat{\boldsymbol{\theta}}_{\text{SUR}}^{R(i)} = ((\mathbf{Z}^{R(i)})' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z}^{R(i)})^{-1} (\mathbf{Z}^{R(i)})' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Y}, \quad (73)$$

with $\hat{\boldsymbol{\Omega}} = \mathbf{I}_T \otimes \hat{\boldsymbol{\Sigma}}$ as before. The restricted residuals for equation i are:

$$\hat{u}_{i,t}^R = \Delta Y_{i,t} - (\mathbf{x}_{i,t}^R)' \hat{\boldsymbol{\theta}}_{\text{SUR},i}^R, \quad t = 1, \dots, T. \quad (74)$$

For unit i , the SUR-based bounds test statistic is an F-type statistic constructed from the residual sums of squares of the unrestricted and restricted systems. Define:

$$\text{RSS}_U^{(i)} = \sum_{t=1}^T (\hat{u}_{i,t}^U)^2, \quad \text{RSS}_R^{(i)} = \sum_{t=1}^T (\hat{u}_{i,t}^R)^2, \quad (75)$$

where $\hat{u}_{i,t}^U$ and $\hat{u}_{i,t}^R$ are the unrestricted and restricted SUR residuals for equation i , as defined in (68) and (74), respectively. The F-statistic for testing $H_0^{(i)}$ within the SUR framework is:

$$F_i^{\text{SUR}} = \frac{(\text{RSS}_R^{(i)} - \text{RSS}_U^{(i)}) / (K + 1)}{\text{RSS}_U^{(i)} / (T - m_i)}. \quad (76)$$

This statistic has the same functional form as the standard single-equation F-statistic in (24), but it differs in a crucial respect: the residuals $\hat{u}_{i,t}^U$ and $\hat{u}_{i,t}^R$ are obtained from the SUR estimator, which accounts for cross-equation dependence, rather than from equation-by-equation OLS.

Remark 7. An equivalent formulation uses the Wald statistic based on the estimated parameter vector and its variance-covariance matrix from the unrestricted SUR estimation. Specifically, using the restriction matrix $\tilde{\mathbf{R}}_i$ defined in (64):

$$W_i^{\text{SUR}} = (\tilde{\mathbf{R}}_i \hat{\boldsymbol{\theta}}_{\text{SUR}})' [\tilde{\mathbf{R}}_i (\mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z})^{-1} \tilde{\mathbf{R}}_i']^{-1} (\tilde{\mathbf{R}}_i \hat{\boldsymbol{\theta}}_{\text{SUR}}). \quad (77)$$

Under standard conditions, $W_i^{\text{SUR}} = (K + 1) \times F_i^{\text{SUR}}$ asymptotically, so $F_i^{\text{SUR}} = W_i^{\text{SUR}} / (K + 1)$. This relationship is analogous to the Wald-type test statistics in the SUR cointegration literature; compare, for instance, equations (17)-(18) in Wagner et al. (2020, Proposition 2),

which define Wald statistics for testing linear hypotheses in the FM-SOLS and FM-SUR frameworks.

The Wald form (77) has the advantage that it can be computed directly from the unrestricted SUR estimates without the need to re-estimate the restricted system (73). The matrix $(\mathbf{Z}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z})^{-1}$ is already available from the unrestricted estimation in (58), and the restriction matrix $\tilde{\mathbf{R}}_i$ selects the relevant elements. This computational simplification is particularly useful when testing cointegration in multiple equations, as it avoids repeated system estimation.

For the joint hypothesis H_0^{Joint} in (65), the system-wide Wald statistic is:

$$W^{\text{Joint}} = (\tilde{\mathbf{R}}^{\text{Joint}}\hat{\boldsymbol{\theta}}_{\text{SUR}})' \left[\tilde{\mathbf{R}}^{\text{Joint}}(\mathbf{Z}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z})^{-1}(\tilde{\mathbf{R}}^{\text{Joint}})'\right]^{-1} (\tilde{\mathbf{R}}^{\text{Joint}}\hat{\boldsymbol{\theta}}_{\text{SUR}}), \quad (78)$$

which is asymptotically χ^2 distributed with $N(K+1)$ degrees of freedom under standard conditions. The corresponding joint F-statistic is:

$$F^{\text{Joint}} = \frac{W^{\text{Joint}}}{N(K+1)}. \quad (79)$$

Remark 8. The joint test (78) tests the composite hypothesis that no cointegrating relationship exists in any equation. This is a stronger hypothesis than testing each equation individually; rejection of H_0^{Joint} indicates that at least one equation contains a long-run relationship, but does not identify which equation(s). In practice, the joint test can be used as an initial screening tool: if the joint null is not rejected, there is no evidence of cointegration in any equation. If the joint null is rejected, equation-specific tests (76) can be applied to identify the cointegrating equations.

When multiple equation-specific tests are conducted simultaneously, a multiple testing correction (e.g., Bonferroni or the modified Bonferroni procedure of Hommel, 1988, as employed by Wagner et al., 2020, Table 3) may be applied to control the family-wise error rate. The joint Wald test (78) provides a single test that avoids the multiple testing problem but at the cost of lower power against alternatives where only a subset of equations exhibit cointegration.

Bounds testing logic under the SUR framework

The bounds testing logic of Pesaran et al. (2001) carries over to the SUR framework with the following modification. Under the null hypothesis $H_0^{(i)}$, the asymptotic distribution of F_i^{SUR} depends on:

- (i) The integration orders of the regressors $X_{1,i,t}, \dots, X_{K,i,t}$, which determine whether the critical values correspond to the $I(0)$ lower bound, the $I(1)$ upper bound, or an intermediate case.
- (ii) The cross-equation covariance structure $\boldsymbol{\Sigma}$, which enters through the FGLS estimation of the parameters and their standard errors.
- (iii) The degree of similarity of the regressors across equations, which determines the extent to which the SUR estimator differs from equation-by-equation OLS.

While feature (i) is common to the standard single-equation bounds test, features (ii) and (iii) are specific to the SUR extension. In general, the critical values of the single-equation bounds test (Pesaran et al., 2001, Table C1) are not directly applicable to the SUR-based F-statistic (76), because the SUR estimator has a different finite-sample distribution due to the cross-equation dependence.

Two approaches to obtaining appropriate critical values are available:

Approach A: Monte Carlo simulation. Critical values for F_i^{SUR} can be simulated by generating data under the null hypothesis $H_0^{(i)}$ (i.e., with $\lambda_i = 0$ and $\beta_{k,i} = 0$ for all k) for a given N , K , T , covariance structure Σ , and a range of integration order configurations. For each configuration, the empirical distribution of F_i^{SUR} is tabulated. Lower bound critical values correspond to the case where all regressors are $I(0)$, and upper bound critical values correspond to the case where all regressors are $I(1)$. This approach follows the tradition of Pesaran et al. (2001) and Narayan (2005), adapted to the system setting.

Approach B: Bootstrap inference. A parametric or semiparametric bootstrap procedure can be used to approximate the finite-sample distribution of F_i^{SUR} under the null. The bootstrap resamples from the estimated residuals while preserving the cross-equation covariance structure, providing critical values that are tailored to the specific data configuration. This approach is particularly useful when T is small and asymptotic approximations may be unreliable.

The decision rule for the SUR-based bounds test mirrors the standard procedure:

- (i) If F_i^{SUR} exceeds the upper bound critical value, reject $H_0^{(i)}$: evidence of cointegration in equation i .
- (ii) If F_i^{SUR} falls below the lower bound critical value, do not reject $H_0^{(i)}$: no evidence of cointegration.
- (iii) If F_i^{SUR} falls between the bounds, the test is inconclusive.

Remark 9. The SUR-based bounds test is expected to offer two improvements over the single-equation test. First, by accounting for cross-equation dependence in the estimation of the error-correction parameters, the SUR estimator yields more precise estimates of λ_i and $\beta_{k,i}$, which translates into higher power for detecting existing cointegrating relationships. This parallels the finding in Wagner et al. (2020, Figures 1-4) that FM-SUR-based tests exhibit higher size-corrected power than FM-SOLS-based tests for all sample size and cross-equation correlation configurations. Second, by appropriately calibrating the critical values to account for the cross-equation dependence structure (via Monte Carlo or bootstrap), the SUR-based test is expected to have better size control than single-equation tests that ignore this dependence.

2.5 Long-Run and Short-Run Coefficient Estimation

Once cointegration has been established for unit i (i.e., $H_0^{(i)}$ is rejected), the long-run equilibrium relationship between $Y_{i,t}$ and $\mathbf{X}_{i,t}$ is characterized by the long-run coefficients. These coefficients are obtained from the error-correction equation (60) by setting $\Delta Y_{i,t} = 0$ and $\Delta X_{k,i,t} = 0$ for all k (the steady-state conditions), which yields:

$$0 = \mu_i + \delta_i t + \lambda_i Y_{i,t-1} + \sum_{k=1}^K \beta_{k,i} X_{k,i,t-1}. \quad (80)$$

Solving for $Y_{i,t-1}$ gives the long-run equilibrium:

$$Y_{i,t} = -\frac{\mu_i}{\lambda_i} - \frac{\delta_i}{\lambda_i} t - \sum_{k=1}^K \frac{\beta_{k,i}}{\lambda_i} X_{k,i,t}. \quad (81)$$

Define the long-run coefficient for the k -th regressor in equation i as:

$$\theta_{k,i}^{LR} = -\frac{\beta_{k,i}}{\lambda_i}, \quad k = 1, \dots, K, \quad (82)$$

and define the long-run intercept and trend coefficients as:

$$\mu_i^{LR} = -\frac{\mu_i}{\lambda_i}, \quad \delta_i^{LR} = -\frac{\delta_i}{\lambda_i}. \quad (83)$$

These long-run coefficients represent the equilibrium response of $Y_{i,t}$ to a permanent unit change in $X_{k,i,t}$, holding all other variables constant and allowing the system to return to equilibrium. The long-run relationship (81) defines the cointegrating vector $(1, -\theta_{1,i}^{LR}, \dots, -\theta_{K,i}^{LR})$ for unit i .

Given the SUR-ARDL estimates $\hat{\theta}_{SUR}$, the long-run coefficients are estimated by the plug-in (substitution) principle:

$$\hat{\theta}_{k,i}^{LR} = -\frac{\hat{\beta}_{k,i}^{SUR}}{\hat{\lambda}_i^{SUR}}, \quad k = 1, \dots, K, \quad i = 1, \dots, N, \quad (84)$$

where $\hat{\lambda}_i^{SUR}$ and $\hat{\beta}_{k,i}^{SUR}$ are the estimated speed of adjustment and long-run level coefficients from the SUR estimation.

Delta method standard errors for long-run coefficients

Since the long-run coefficients (82) are nonlinear functions (ratios) of the estimated error-correction parameters, their standard errors cannot be read directly from the estimated variance-covariance matrix (58). Instead, we derive the asymptotic variance of $\hat{\theta}_{k,i}^{LR}$ using the delta method.

Let $\alpha_i = (\lambda_i, \beta_{k,i})'$ denote the two error-correction parameters that enter the definition of $\theta_{k,i}^{LR}$. The corresponding SUR estimates are $\hat{\alpha}_i = (\hat{\lambda}_i^{SUR}, \hat{\beta}_{k,i}^{SUR})'$, with estimated variance-covariance matrix:

$$\mathbf{V}_{\alpha_i} = \begin{pmatrix} v_{\lambda\lambda,i} & v_{\lambda\beta,ki} \\ v_{\lambda\beta,ki} & v_{\beta\beta,ki} \end{pmatrix}, \quad (85)$$

where $v_{\lambda\lambda,i} = \widehat{\text{Var}}(\hat{\lambda}_i^{SUR})$, $v_{\beta\beta,ki} = \widehat{\text{Var}}(\hat{\beta}_{k,i}^{SUR})$, and $v_{\lambda\beta,ki} = \widehat{\text{Cov}}(\hat{\lambda}_i^{SUR}, \hat{\beta}_{k,i}^{SUR})$, all extracted from the appropriate elements of the $M \times M$ matrix $(\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}$.

The long-run coefficient is the function $g(\alpha_i) = -\beta_{k,i}/\lambda_i$. The gradient of g with respect to α_i is:

$$\nabla g(\alpha_i) = \frac{\partial g}{\partial \alpha_i} = \begin{pmatrix} \partial g / \partial \lambda_i \\ \partial g / \partial \beta_{k,i} \end{pmatrix} = \begin{pmatrix} \beta_{k,i}/\lambda_i^2 \\ -1/\lambda_i \end{pmatrix}. \quad (86)$$

Evaluating the gradient at the estimated values $\hat{\alpha}_i$:

$$\hat{\nabla} g = \begin{pmatrix} \hat{\beta}_{k,i}^{SUR}/(\hat{\lambda}_i^{SUR})^2 \\ -1/\hat{\lambda}_i^{SUR} \end{pmatrix}. \quad (87)$$

By the delta method, the asymptotic variance of $\hat{\theta}_{k,i}^{LR}$ is:

$$\widehat{\text{Var}}(\hat{\theta}_{k,i}^{LR}) = \hat{\nabla} g' \mathbf{V}_{\alpha_i} \hat{\nabla} g. \quad (88)$$

Expanding this quadratic form explicitly:

$$\widehat{\text{Var}}(\hat{\theta}_{k,i}^{LR}) = \frac{\hat{\beta}_{k,i}^2}{(\hat{\lambda}_i)^4} v_{\lambda\lambda,i} - \frac{2\hat{\beta}_{k,i}}{(\hat{\lambda}_i)^3} v_{\lambda\beta,ki} + \frac{1}{(\hat{\lambda}_i)^2} v_{\beta\beta,ki}, \quad (89)$$

where the superscript ^{SUR} on the estimates has been suppressed for brevity. The standard error is:

$$\text{SE}(\hat{\theta}_{k,i}^{LR}) = \sqrt{\widehat{\text{Var}}(\hat{\theta}_{k,i}^{LR})}, \quad (90)$$

and the t -statistic for testing $H_0: \theta_{k,i}^{LR} = 0$ is:

$$t_{k,i}^{LR} = \frac{\hat{\theta}_{k,i}^{LR}}{\text{SE}(\hat{\theta}_{k,i}^{LR})}. \quad (91)$$

Remark 10. The delta method derivation above provides the variance for a single long-run coefficient. More general hypotheses involving long-run coefficients from multiple equations for example, testing the equality of long-run coefficients across two units, $H_0: \theta_{k,i}^{LR} = \theta_{k,j}^{LR}$ require the joint variance-covariance matrix of $(\hat{\theta}_{k,i}^{LR}, \hat{\theta}_{k,j}^{LR})'$. This is obtained by the multivariate delta method applied to the function $\mathbf{g}(\boldsymbol{\theta}) = (-\beta_{k,i}/\lambda_i, -\beta_{k,j}/\lambda_j)'$:

$$\widehat{\text{Var}} \begin{pmatrix} \hat{\theta}_{k,i}^{LR} \\ \hat{\theta}_{k,j}^{LR} \end{pmatrix} = \mathbf{J} \widehat{\text{Var}} \begin{pmatrix} \hat{\lambda}_i \\ \hat{\beta}_{k,i} \\ \hat{\lambda}_j \\ \hat{\beta}_{k,j} \end{pmatrix} \mathbf{J}', \quad (92)$$

where $\mathbf{J}$ is the 2×4 Jacobian matrix of $\mathbf{g}$ with respect to $(\lambda_i, \beta_{k,i}, \lambda_j, \beta_{k,j})'$:

$$\mathbf{J} = \begin{pmatrix} \beta_{k,i}/\lambda_i^2 & -1/\lambda_i & 0 & 0 \\ 0 & 0 & \beta_{k,j}/\lambda_j^2 & -1/\lambda_j \end{pmatrix}. \quad (93)$$

The 4×4 variance-covariance matrix of $(\hat{\lambda}_i, \hat{\beta}_{k,i}, \hat{\lambda}_j, \hat{\beta}_{k,j})'$ is extracted from the appropriate blocks of the $M \times M$ system covariance matrix $(\mathbf{Z}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z})^{-1}$. Crucially, the SUR framework provides non-zero cross-equation covariance terms (e.g., $\widehat{\text{Cov}}(\hat{\lambda}_i, \hat{\lambda}_j)$, $\widehat{\text{Cov}}(\hat{\beta}_{k,i}, \hat{\beta}_{k,j})$) that equation-by-equation OLS would set to zero. These cross-equation covariances are essential for valid inference on cross-equation long-run hypotheses.

This testing capability for cross-equation long-run coefficient equality is analogous to the group-wise poolability testing framework developed in detail by Wagner et al. (2020, Section 2.2, Propositions 2-3). In their setting, group-wise pooling of slope coefficients across subsets of cross-section members is tested via Wald-type statistics, with the restriction matrices defined in their equations (19)-(23). The present SUR-ARDL framework provides analogous capability through the delta method applied to the nonlinear long-run coefficient transformations.

Short-run coefficients and the speed of adjustment

In addition to the long-run coefficients, the error-correction model (60) provides estimates of the short-run dynamic parameters:

Speed of adjustment. The parameter $\lambda_i = \hat{\lambda}_i^{\text{SUR}}$ measures the speed at which the system returns to equilibrium following a shock. A value of $\hat{\lambda}_i^{\text{SUR}} = -0.25$, for example, implies that approximately 25% of the deviation from the long-run equilibrium is corrected within one period. The estimated speed of adjustment must be negative ($\hat{\lambda}_i^{\text{SUR}} < 0$) for the

error-correction mechanism to be stable, which constitutes an additional necessary condition for the existence of a valid cointegrating relationship (in addition to the F-test rejection of the joint null).

Short-run impact elasticities. The parameters $\gamma_{k,i} = \hat{\gamma}_{k,i}^{\text{SUR}}$ for $k = 1, \dots, K$ measure the contemporaneous (within-period) response of $\Delta Y_{i,t}$ to a unit change in $\Delta X_{k,i,t}$, holding all other variables constant. These capture the immediate impact effects that may differ substantially from the long-run equilibrium responses $\theta_{k,i}^{\text{LR}}$.

The distinction between short-run and long-run responses is a key feature of the error-correction framework. In the context of energy economics, for example, the short-run elasticity of CO₂ emissions with respect to GDP may be larger than the long-run elasticity if economies adjust their energy mix and technology over time, or smaller if there are adjustment costs and inertia in energy infrastructure investment.

The delta method standard errors derived above reflect the precision of the SUR estimator and will generally differ from those obtained via single-equation OLS. To formalize this comparison, denote the delta method standard errors based on the OLS and SUR variance-covariance matrices as $\text{SE}^{\text{OLS}}(\hat{\theta}_{k,i}^{\text{LR}})$ and $\text{SE}^{\text{SUR}}(\hat{\theta}_{k,i}^{\text{LR}})$, respectively. The efficiency gain of Proposition 2 implies that, asymptotically:

$$\text{SE}^{\text{SUR}}(\hat{\theta}_{k,i}^{\text{LR}}) \leq \text{SE}^{\text{OLS}}(\hat{\theta}_{k,i}^{\text{LR}}), \quad (94)$$

with strict inequality when the cross-equation covariance is non-zero. The proportional reduction in standard errors provides a practical measure of the value added by system estimation:

$$\text{Efficiency Gain}_{k,i} = 1 - \frac{\text{SE}^{\text{SUR}}(\hat{\theta}_{k,i}^{\text{LR}})}{\text{SE}^{\text{OLS}}(\hat{\theta}_{k,i}^{\text{LR}})}. \quad (95)$$

This metric is reported in the empirical application in Section 5 to quantify the practical benefits of the SUR-ARDL approach relative to equation-by-equation ARDL estimation.

2.6 Diagnostic Testing and Model Validation

The SUR-ARDL framework rests on several maintained assumptions. Most notably, the contemporaneous covariance structure of the disturbances (Assumption 1), the absence of serial correlation in the conditional errors (Assumption 1(iii)), and the integration properties of the variables (Assumption 2). These assumptions must be verified empirically for the model to deliver reliable inference.

Pre-estimation test for cross-sectional dependence

Before estimating the SUR-ARDL system, one should test whether contemporaneous cross-sectional dependence is present in the data. If the errors are cross-sectionally independent (i.e., Σ is diagonal), the SUR estimator reduces to equation-by-equation OLS (Proposition 1), and the additional computational cost of system estimation yields no efficiency gain. Two complementary tests are considered.

The Breusch and Pagan (1980) Lagrange Multiplier (LM) test evaluates the null hypothesis that the off-diagonal elements of Σ are all zero:

$$H_0^{\text{BP}}: \sigma_{ij} = 0, \quad \text{for all } i \neq j, \quad i, j = 1, \dots, N. \quad (96)$$

The test statistic is constructed from the pairwise sample correlations of the OLS residuals $\hat{u}_{i,t}$ from equation-by-equation estimation:

$$\hat{r}_{ij} = \frac{\sum_{t=1}^T \hat{u}_{i,t} \hat{u}_{j,t}}{(\sum_{t=1}^T \hat{u}_{i,t}^2)^{1/2} (\sum_{t=1}^T \hat{u}_{j,t}^2)^{1/2}}, \quad i, j = 1, \dots, N. \quad (97)$$

The Breusch-Pagan LM statistic is:

$$LM_{BP} = T \sum_{i=2}^N \sum_{j=1}^{i-1} \hat{r}_{ij}^2. \quad (98)$$

Under H_0^{BP} and the assumption of normally distributed errors, LM_{BP} is asymptotically χ^2 distributed with $N(N-1)/2$ degrees of freedom as $T \rightarrow \infty$ for fixed N :

$$LM_{BP} \xrightarrow{d} \chi^2 \left(\frac{N(N-1)}{2} \right), \quad \text{under } H_0^{BP}. \quad (99)$$

Rejection of H_0^{BP} provides statistical justification for the SUR approach over equation-by-equation estimation. Pesaran (2004) proposes a test for cross-sectional dependence that is robust to both large T and large N settings and does not require the assumption of normally distributed errors. The Pesaran CD statistic is based on the average of the pairwise correlation coefficients:

$$CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=2}^N \sum_{j=1}^{i-1} \hat{r}_{ij}. \quad (100)$$

Under the null hypothesis of cross-sectional independence, CD is asymptotically standard normal:

$$CD \xrightarrow{d} N(0,1), \quad \text{under } H_0^{CD}: E(u_{i,t} u_{j,t}) = 0 \text{ for all } i \neq j. \quad (101)$$

An important distinction between the BP and CD tests is that the LM statistic uses the squared correlations $\hat{r}_{ij}^2$ and is therefore sensitive to all forms of cross-sectional dependence (both positive and negative), whereas the CD statistic uses the signed correlations $\hat{r}_{ij}$ and may have low power against alternatives where positive and negative correlations cancel out. In practice, both tests should be reported: the BP test as the primary diagnostic for the SUR approach, and the CD test as a robustness check.

Pre-estimation test for integration orders

Assumption 2 requires that all variables are integrated of order zero or one, with no $I(2)$ variables. This should be verified by applying unit root tests to both levels and first differences of all variables.

For each variable $W_{i,t} \in \{Y_{i,t}, X_{1,i,t}, \dots, X_{K,i,t}\}$, the following testing strategy is recommended:

Step 1. Apply the Augmented Dickey-Fuller (ADF) test (Said & Dickey, 1984) and the Phillips-Perron (PP) test (Phillips & Perron, 1988) to the level series $W_{i,t}$ with intercept and with intercept plus trend specifications. Let $\tau_{ADF}^{(i)}$ and $\tau_{PP}^{(i)}$ denote the corresponding test statistics. The ADF test is based on the auxiliary regression:

$$\Delta W_{i,t} = \alpha_0 + \alpha_1 t + \rho W_{i,t-1} + \sum_{s=1}^{p^*} \xi_s \Delta W_{i,t-s} + e_{i,t}, \quad (102)$$

where p^* is selected to render $e_{i,t}$ approximately white noise (e.g., by the AIC or SBC). The null hypothesis is $H_0: \rho = 0$ (unit root) against $H_1: \rho < 0$ (stationarity).

Step 2. Apply the KPSS test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992) to the level series $W_{i,t}$. The KPSS test reverses the hypotheses: H_0 : stationarity versus H_1 : unit root. The joint use of ADF/PP (null = unit root) and KPSS (null = stationarity) provides a confirmatory strategy that reduces the risk of misclassification.

Step 3. For any variable for which the ADF/PP tests do not reject the unit root null, apply the same tests to the first difference $\Delta W_{i,t}$. If the unit root null is rejected for $\Delta W_{i,t}$, the variable is classified as $I(1)$. If the unit root null is not rejected for $\Delta W_{i,t}$, the variable may be $I(2)$, in which case it should be excluded from the ARDL analysis or further differenced.

Step 4. Compile the integration order classification for all variables across all units. The ARDL bounds testing approach is valid provided that all variables are classified as $I(0)$ or $I(1)$ and no variable is $I(2)$.

This pre-testing strategy follows the recommendation of Pesaran et al. (2001, Section 5) and is standard in the applied ARDL literature. While the bounds testing approach does not require precise knowledge of integration orders for valid inference on the bounds test itself, the classification provides useful information for interpreting the results and ensures that the $I(2)$ exclusion condition is satisfied.

3. Asymptotic Theory

This section derives the large-sample properties of the SUR-ARDL estimator. Throughout this section, all limits are taken as $T \rightarrow \infty$ with N fixed. This asymptotic framework large T , fixed N is the natural setting for the SUR approach, which is designed for panels with small cross-sectional dimension relative to the time series dimension. Sequential or joint asymptotics in which $N \rightarrow \infty$ alongside T would require additional structure on the cross-sectional dimension and are beyond the scope of the present contribution.

3.1 Consistency and Limiting Distribution

The ARDL error-correction equation (12) for unit i includes both stationary and potentially non-stationary regressors. The intercept μ_i , the trend t , and the first-differenced regressors $\Delta X_{k,i,t}$ are stationary (or deterministic) by construction. The lagged dependent variable $Y_{i,t-1}$ and the lagged regressors $X_{k,i,t-1}$ may be $I(0)$ or $I(1)$, depending on the integration properties of the underlying variables. This mixture of regressor types leads to non-standard asymptotic behavior: different elements of the parameter vector θ_i converge at different rates, and the limiting distributions involve both standard normal and functional of Brownian motion components.

Let $G_i(T)$ denote the $m_i \times m_i$ diagonal scaling matrix for unit i that normalizes each element of the regressor vector $\mathbf{x}_{i,t}$ to achieve a well-defined limit. For the ARDL(1,1,...,1) error-correction specification in equation (12), the regressor vector is $\mathbf{x}_{i,t} = (1, t, Y_{i,t-1}, X_{1,i,t-1}, \dots, X_{K,i,t-1}, \Delta X_{1,i,t}, \dots, \Delta X_{K,i,t})'$. The scaling matrix takes the form:

$$G_i(T) = \text{diag}(T^{-1/2}, T^{-3/2}, g_{Y,i}(T), g_{1,i}(T), \dots, g_{K,i}(T), T^{-1/2}, \dots, T^{-1/2}), \quad (103)$$

where $g_{Y,i}(T)$ and $g_{k,i}(T)$ denote the scaling factors for the lagged level variables, which depend on the integration orders:

$$g_{Y,i}(T) = \begin{cases} T^{-1/2} & \text{if } Y_{i,t} \sim I(0), \\ T^{-1} & \text{if } Y_{i,t} \sim I(1), \end{cases} \quad g_{k,i}(T) = \begin{cases} T^{-1/2} & \text{if } X_{k,i,t} \sim I(0), \\ T^{-1} & \text{if } X_{k,i,t} \sim I(1). \end{cases} \quad (104)$$

The distinction between the $T^{-1/2}$ rate (for stationary regressors) and the T^{-1} rate (for $I(1)$ levels) reflects the well-known super-consistency property of OLS in cointegrating regressions: when the regressors include $I(1)$ variables, the associated parameter estimates

converge at rate T rather than the standard $T^{1/2}$ rate. In the ARDL context, the lagged level terms $Y_{i,t-1}$ and $X_{k,i,t-1}$ play the role of integrated regressors (when the underlying variables are $I(1)$), while the intercept, trend, and differenced terms behave as standard stationary regressors. Define the system-level scaling matrix as the block-diagonal matrix:

$$\mathbf{G}(T) = \text{diag}(\mathbf{G}_1(T), \mathbf{G}_2(T), \dots, \mathbf{G}_N(T)) \in \mathbb{R}^{M \times M}. \quad (105)$$

This matrix normalizes the full parameter vector $\boldsymbol{\theta}$ element-wise according to the appropriate convergence rate. The block-diagonal structure reflects the fact that parameters from different equations converge independently, with the cross-equation interaction entering only through the covariance weighting $\boldsymbol{\Sigma}^{-1}$.

To state the limiting distribution, we require the following additional notation. For each unit i with $I(1)$ regressors, define the Brownian motion processes $B_{Y,i}(r)$ and $B_{k,i}(r)$ for $0 \leq r \leq 1$ as the weak limits of the appropriately normalized partial sums of $Y_{i,t}$ and $X_{k,i,t}$, respectively. These processes arise from the functional central limit theorem applied to the $I(1)$ component of the data generating process. Specifically, under Assumption 2, if $Y_{i,t} \sim I(1)$, then $T^{-1/2}Y_{i,\lfloor rT \rfloor} \Rightarrow B_{Y,i}(r)$ as $T \rightarrow \infty$, where $\Rightarrow$ denotes weak convergence in the Skorokhod space $D[0,1]$. Analogous results hold for each $I(1)$ regressor $X_{k,i,t}$.

For the stationary regressors (the intercept, trend, and differenced terms), the standard law of large numbers and central limit theorem apply, yielding deterministic limits for the normalized second-moment matrices and Gaussian limits for the normalized cross-products with the disturbances. For the $I(1)$ regressors, the normalized second-moment matrices converge to functionals of Brownian motion specifically, stochastic integrals of the form $\int_0^1 B(r)B(r)'dr$ and the normalized cross-products converge to stochastic integrals of the form $\int_0^1 B(r)dB_u(r)$, where B_u is the Brownian motion associated with the disturbance process.

Define the limit random matrix for unit i as $\mathbf{J}_i(r)$, which collects the appropriate limit processes for all elements of the scaled regressor vector $\mathbf{G}_i(T)\mathbf{x}_{i,\lfloor rT \rfloor}$. For the ARDL(1,1,...,1) specification:

$$\mathbf{J}_i(r) = (1, r, J_{Y,i}(r), J_{1,i}(r), \dots, J_{K,i}(r), J_{\Delta 1,i}(r), \dots, J_{\Delta K,i}(r))', \quad (106)$$

where $J_{Y,i}(r) = B_{Y,i}(r)$ if $Y_{i,t} \sim I(1)$ (or $J_{Y,i}(r) = 1$ if $Y_{i,t} \sim I(0)$), $J_{k,i}(r) = B_{k,i}(r)$ if $X_{k,i,t} \sim I(1)$ (or $J_{k,i}(r) = 1$ if $X_{k,i,t} \sim I(0)$), and $J_{\Delta k,i}(r)$ denotes the limit of the stationary differenced regressor terms. Define the system-level limit process as:

$$\mathbf{J}(r) = \text{diag}(\mathbf{J}_1(r), \mathbf{J}_2(r), \dots, \mathbf{J}_N(r)). \quad (107)$$

We are now in a position to state the main consistency and limiting distribution results for the SUR-ARDL estimator. These results are presented as two theorems: the first establishes consistency of the equation-by-equation OLS estimator (which provides the first-step residuals for covariance estimation), and the second derives the limiting distribution of the FGLS estimator.

Theorem 1 (Consistency of the equation-by-equation OLS estimator). *Under Assumptions 1-3, the equation-by-equation OLS estimator $\hat{\boldsymbol{\theta}}_i^{\text{OLS}}$ defined in equation (18) is consistent for $\boldsymbol{\theta}_i$ for each $i = 1, \dots, N$:*

$$\hat{\boldsymbol{\theta}}_i^{\text{OLS}} \xrightarrow{p} \boldsymbol{\theta}_i, \quad \text{as } T \rightarrow \infty. \quad (108)$$

Moreover, the scaled and centered estimator has the limiting distribution:

$$\mathbf{G}_i(T)^{-1}(\hat{\boldsymbol{\theta}}_i^{\text{OLS}} - \boldsymbol{\theta}_i) \Rightarrow \left(\int_0^1 \mathbf{J}_i(r) \mathbf{J}_i(r)' dr \right)^{-1} \int_0^1 \mathbf{J}_i(r) dB_{u,i}(r) + \mathcal{B}_i, \quad (109)$$

where $B_{u,i}(r)$ is the Brownian motion component associated with $u_{i,t}$ and $\mathcal{B}_i$ is a second-order bias term that arises from the correlation between the regressors and the disturbances.

The proof of Theorem 1 follows standard arguments in the cointegrating regression literature (see, e.g., Park & Phillips, 1988, 1989; Phillips & Hansen, 1990; Pesaran & Shin, 1999). The consistency of $\hat{\boldsymbol{\theta}}_i^{\text{OLS}}$ established in Theorem 1 ensures that the first-step OLS residuals $\hat{u}_{i,t} = \Delta Y_{i,t} - \mathbf{x}_{i,t}' \hat{\boldsymbol{\theta}}_i^{\text{OLS}}$ are consistent estimates of the true disturbances $u_{i,t}$, which in turn implies that the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ in equation (56) is a consistent estimator of the true contemporaneous covariance matrix $\boldsymbol{\Sigma}$. This consistency result is formalized in the following lemma.

Lemma 1. *Under Assumptions 1-3:*

$$\hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}, \quad \text{as } T \rightarrow \infty, \quad (110)$$

and consequently $\hat{\boldsymbol{\Sigma}}^{-1} \xrightarrow{p} \boldsymbol{\Sigma}^{-1}$.

The proof follows from the uniform law of large numbers applied to the sample covariance of the OLS residuals and the consistency of the OLS estimator (Theorem 1). The key step is to show that $T^{-1} \sum_{t=1}^T \hat{u}_{i,t} \hat{u}_{j,t} - T^{-1} \sum_{t=1}^T u_{i,t} u_{j,t} \xrightarrow{p} 0$ for all i, j , which follows from the triangle inequality and the convergence rates established in Theorem 1. This consistency result is the ARDL analog of the kernel-based long-run covariance estimation consistency assumed by Wagner et al. (2020) throughout their analysis.

We now state the main result of this subsection: the limiting distribution of the feasible SUR-ARDL estimator.

Theorem 2 (Limiting distribution of the SUR-ARDL estimator). *Under Assumptions 1-3, and assuming that the ARDL lag orders are correctly specified so that the conditional disturbances $u_{i,t}$ are serially uncorrelated, the SUR-ARDL FGLS estimator $\hat{\boldsymbol{\theta}}_{\text{SUR}}$ defined in equation (57) satisfies:*

$$\mathbf{G}(T)^{-1}(\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta}) \Rightarrow \left(\int_0^1 \mathbf{J}(r) \boldsymbol{\Sigma}^{-1} \mathbf{J}(r)' dr \right)^{-1} \int_0^1 \mathbf{J}(r) \boldsymbol{\Sigma}^{-1} d\mathbf{B}_u^*(r), \quad \text{as } T \rightarrow \infty, \quad (111)$$

where $\mathbf{B}_u^*(r) = (B_{u,1}^*(r), \dots, B_{u,N}^*(r))'$ is an N -dimensional Brownian motion with covariance matrix $\boldsymbol{\Sigma}$ that is independent of the stochastic regressors, conditional on the σ -field generated by the regressor processes.

The limiting distribution in (111) is a zero-mean Gaussian mixture distribution, conditional on the stochastic regressors. This distributional form is the system-level analog of the single-equation result for the ARDL estimator (Pesaran & Shin, 1999) and parallels the FM-SUR result of Wagner et al. (2020, Proposition 1, equation (16)). It is worth noting that the asymptotic equivalence of the two-step and iterated FGLS estimators, which follows from the consistency of $\hat{\boldsymbol{\Sigma}}$ (Lemma 1), implies that the limiting distribution in Theorem 2 applies to both the two-step and the iterated (maximum likelihood) variants of the SUR-ARDL estimator. The difference between these estimators is of order $o_p(T^{-1/2})$ for stationary parameters and $o_p(T^{-1})$ for super-consistent parameters, and therefore vanishes in the limit.

3.2 Asymptotic Null Distribution of the SUR-Based Bounds F-Statistic

The bounds testing procedure is the inferential centerpiece of the ARDL framework: it provides a test for the existence of a long-run level relationship without requiring exact knowledge of the integration orders of the regressors. The extension of this procedure to the SUR setting requires a careful analysis of how the cross-equation covariance structure interacts with the non-standard null distribution of the F-statistic. This subsection derives the asymptotic null distribution of the SUR-based F-statistic F_i^{SUR} defined in equation (76), establishes the validity of the bounds testing logic in the system context, and examines the conditions under which the standard Pesaran et al. (2001) critical values remain applicable.

The null hypothesis for unit i is $H_0^{(i)}: \lambda_i = 0, \beta_{1,i} = 0, \dots, \beta_{K,i} = 0$, as stated in equation (61). Under this null, the error-correction equation for unit i reduces to a regression in first differences only, and the lagged level variables $Y_{i,t-1}$ and $X_{k,i,t-1}$ are excluded from the model. The key complication and the reason for non-standard asymptotics is that the integration orders of the excluded variables are unknown under the null: they may be $I(0)$, $I(1)$, or a mixture. Different integration order configurations lead to different limiting distributions for the F-statistic, which is the fundamental insight motivating the bounds testing approach.

To derive the null distribution, we begin by expressing the SUR-based F-statistic in its Wald form. From equation (77), the Wald statistic for testing $H_0^{(i)}$ using the SUR estimator is:

$$W_i^{\text{SUR}} = (\tilde{\mathbf{R}}_i \hat{\boldsymbol{\theta}}_{\text{SUR}})' \left[\tilde{\mathbf{R}}_i (\mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z})^{-1} \tilde{\mathbf{R}}_i' \right]^{-1} (\tilde{\mathbf{R}}_i \hat{\boldsymbol{\theta}}_{\text{SUR}}), \quad (112)$$

where $\tilde{\mathbf{R}}_i \in \mathbb{R}^{(K+1) \times M}$ is the restriction matrix defined in equation (64) that selects $(\lambda_i, \beta_{1,i}, \dots, \beta_{K,i})$ from the full system parameter vector $\boldsymbol{\theta}$, and $F_i^{\text{SUR}} = W_i^{\text{SUR}} / (K + 1)$. The asymptotic behavior of W_i^{SUR} under the null depends on the behavior of three quantities: (a) the restricted parameter estimates $\tilde{\mathbf{R}}_i \hat{\boldsymbol{\theta}}_{\text{SUR}}$; (b) the estimated variance-covariance matrix $\tilde{\mathbf{R}}_i (\mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z})^{-1} \tilde{\mathbf{R}}_i'$; and (c) the interaction between these two quantities through the non-standard convergence rates of the level-variable parameter estimates.

The derivation of the null distribution proceeds as follows. Under $H_0^{(i)}$, the true parameters $\lambda_i = 0$ and $\beta_{k,i} = 0$ for all k , but the SUR estimator $\hat{\boldsymbol{\theta}}_{\text{SUR}}$ produces non-zero estimates due to sampling variability. The scaled restricted estimates converge in distribution as:

$$\mathbf{G}_{R,i}(T)^{-1} \tilde{\mathbf{R}}_i \hat{\boldsymbol{\theta}}_{\text{SUR}} \Rightarrow \boldsymbol{\xi}_i, \quad (113)$$

where $\boldsymbol{\xi}_i$ is a $(K + 1)$ -dimensional random vector whose distribution depends on the integration order configuration. To determine $\boldsymbol{\xi}_i$ explicitly, we extract the relevant sub-block from the full system limiting distribution in Theorem 2. Let $\mathbf{V}_i = \tilde{\mathbf{R}}_i (\mathbf{Z}' \boldsymbol{\Omega}^{-1} \mathbf{Z})^{-1} \tilde{\mathbf{R}}_i'$ denote the asymptotic variance block for the restricted parameters. From the Gaussian mixture structure in (135), the limiting expression for the Wald statistic can be written as:

$$W_i^{\text{SUR}} \Rightarrow \mathcal{W}_i, \quad (114)$$

where $\mathcal{W}_i$ is a functional of Brownian motions and deterministic trends whose exact form depends on the integration order configuration. The critical feature is that this limiting distribution is not standard chi-squared in general, because the restricted parameters $(\lambda_i, \beta_{1,i}, \dots, \beta_{K,i})$ are associated with potentially $I(1)$ regressors (the lagged levels), which induce non-standard convergence. This non-standard distribution is precisely analogous to the Dickey-Fuller-type distribution that arises in unit root testing and the distributions tabulated by Pesaran et al. (2001).

To formalize the bounds testing logic, we state the following theorem.

Theorem 3 (Asymptotic null distribution of the SUR bounds F-statistic). *Under Assumptions 1-3, and under the null hypothesis $H_0^{(i)}: \lambda_i = 0, \beta_{1,i} = 0, \dots, \beta_{K,i} = 0$, the SUR-based F-statistic $F_i^{\text{SUR}} = W_i^{\text{SUR}}/(K + 1)$ satisfies:*

$$F_i^{\text{SUR}} \Rightarrow \mathcal{F}_i(\mathbf{d}_i, \Sigma, \{\mathbf{X}_j\}_{j \neq i}), \quad \text{as } T \rightarrow \infty, \quad (115)$$

where $\mathbf{d}_i = (d_{Y,i}, d_{1,i}, \dots, d_{K,i})' \in \{0,1\}^{K+1}$ is the vector of integration orders and $\mathcal{F}_i(\cdot)$ is a functional that depends on:

(a) the integration order configuration $\mathbf{d}_i$ through the composition of the limit process $\mathbf{J}_{R,i}(r)$ in (144);

(b) the cross-equation covariance matrix Σ through the GLS weighting in the SUR estimator;

(c) the regressor matrices of the other equations $\{\mathbf{X}_j\}_{j \neq i}$ through the cross-equation information sharing in the system estimator.

Furthermore, the limiting distribution satisfies the following monotonicity property: for any two integration order configurations $\mathbf{d}_i^{(a)}$ and $\mathbf{d}_i^{(b)}$ such that $\mathbf{d}_i^{(a)} \leq \mathbf{d}_i^{(b)}$ element-wise (with at least one strict inequality), the critical values of $\mathcal{F}_i$ under $\mathbf{d}_i^{(a)}$ are (weakly) smaller than or equal to those under $\mathbf{d}_i^{(b)}$ at any given significance level.

The monotonicity property in Theorem 3 is the formal basis for the bounds testing procedure. It implies that the asymptotic critical values of F_i^{SUR} are bounded between two extreme cases:

(i) Lower bound ($I(0)$ case): When all regressors in equation i are stationary ($\mathbf{d}_i = \mathbf{0}$), the limit process $\mathbf{J}_{R,i}(r)$ is deterministic, the convergence rate of all restricted parameters is $T^{-1/2}$ (the standard rate), and the Wald statistic has a standard chi-squared limiting distribution:

$$W_i^{\text{SUR}} \Rightarrow \chi^2(K + 1), \quad \text{when } \mathbf{d}_i = \mathbf{0}_{K+1}. \quad (116)$$

Consequently, $F_i^{\text{SUR}} \Rightarrow \chi^2(K + 1)/(K + 1) =: \mathcal{F}_{I(0)}$. In this case, the limiting distribution coincides with the standard $F(K + 1, T - m_i)$ distribution asymptotically, and the critical values are the smallest (the lower bound).

(ii) Upper bound ($I(1)$ case): When all regressors in equation i are integrated of order one ($\mathbf{d}_i = \mathbf{1}$), the limit process $\mathbf{J}_{R,i}(r)$ consists entirely of Brownian motions, the convergence rate of all restricted parameters is T^{-1} (the super-consistent rate), and the limiting distribution involves functionals of Brownian motions that are non-standard:

$$W_i^{\text{SUR}} \Rightarrow \mathcal{W}_{I(1)}(\Sigma, \{\mathbf{X}_j\}_{j \neq i}), \quad \text{when } \mathbf{d}_i = \mathbf{1}_{K+1}. \quad (117)$$

The critical values in this case are the largest (the upper bound), and the distribution depends on the dimensionality K , the deterministic specification (intercept, trend), and the cross-equation structure.

(iii) Intermediate cases: When some regressors are $I(0)$ and others are $I(1)$, the limiting distribution is a mixture of the standard and non-standard components, with critical values that lie between the lower and upper bounds.

The bounds testing decision rule for F_i^{SUR} follows directly:

$$\text{Decision} = \begin{cases} \text{Reject } H_0^{(i)} & \text{if } F_i^{\text{SUR}} > c_U(\alpha), \\ \text{Do not reject } H_0^{(i)} & \text{if } F_i^{\text{SUR}} < c_L(\alpha), \\ \text{Inconclusive} & \text{if } c_L(\alpha) \leq F_i^{\text{SUR}} \leq c_U(\alpha), \end{cases} \quad (118)$$

where $c_L(\alpha)$ and $c_U(\alpha)$ are the lower and upper bound critical values at significance level α , respectively.

A central question for the practical implementation of the SUR-ARDL bounds test is whether the critical values tabulated by Pesaran et al. (2001) for the single-equation ARDL bounds test or the small-sample critical values of Narayan (2005) remain valid for the SUR-based test statistic, or whether new critical values specific to the SUR context must be computed. The answer depends on whether the cross-equation covariance structure Σ and the regressor dissimilarity across equations materially affect the null distribution of F_i^{SUR} .

3.3 Asymptotic Distribution of Wald-Type Statistics

A principal advantage of embedding the ARDL error-correction specification within the SUR framework beyond the efficiency gains for point estimation documented in Proposition 2 is the ability to test hypotheses that span multiple equations. Such cross-equation hypotheses are, by construction, untestable using equation-by-equation OLS, which produces independent estimates for each unit and sets all cross-equation covariances to zero. The SUR estimator, by contrast, provides a joint variance-covariance matrix $(\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}$ for the entire system parameter vector $\boldsymbol{\theta}$, including the off-diagonal blocks that capture the covariance between parameter estimates from different equations. These off-diagonal blocks are the essential ingredients for valid cross-equation inference.

Consider s linearly independent restrictions collected in the null hypothesis:

$$H_0: \mathbf{R}\boldsymbol{\theta} = \mathbf{r}, \quad (119)$$

where $\mathbf{R} \in \mathbb{R}^{s \times M}$ is a known full-row-rank restriction matrix and $\mathbf{r} \in \mathbb{R}^s$ is a known vector. The Wald statistic for testing H_0 is:

$$W = (\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{SUR}} - \mathbf{r})' \left[\mathbf{R}(\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}\mathbf{R}' \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{SUR}} - \mathbf{r}). \quad (120)$$

The asymptotic behavior of W depends on the convergence rates of the elements of $\boldsymbol{\theta}$ that appear in the restrictions. As discussed in Section 3, different elements of $\boldsymbol{\theta}$ converge at different rates $T^{1/2}$ for stationary regressors and T for $I(1)$ level regressors and the restrictions in $\mathbf{R}$ may combine elements with different rates. This heterogeneity of convergence rates is the central technical challenge for establishing chi-squared inference in the SUR cointegration context, and it has been addressed carefully by both Wagner et al. (2020, Proposition 2) and Lin and Reuvers (2025).

The difficulty can be understood as follows. When a single restriction (a single row of $\mathbf{R}$) combines parameters with different convergence rates, the limiting distribution of the corresponding linear combination is dominated by the parameter with the slowest convergence rate (the rate $T^{-1/2}$ for stationary regressors). If the restriction also involves super-consistent parameters (rate T^{-1}), these parameters converge faster than the standard ones, and their contribution to the asymptotic variance of the linear combination is asymptotically negligible. The resulting degenerate structure can lead to a degenerate limiting distribution for the Wald statistic unless the restriction matrix satisfies a regularity condition.

Wagner et al. (2020) formalize this requirement through the following rate condition: there must exist a sequence of full-rank matrices $\mathbf{G}_R = \mathbf{G}_R(T)$ such that:

$$\lim_{T \rightarrow \infty} \mathbf{G}_R \mathbf{R} \mathbf{G}(T) = \mathbf{R}^* \in \mathbb{R}^{s \times M}, \quad (121)$$

where $\mathbf{R}^*$ has full row rank s and $\mathbf{G}(T)$ is the system-level scaling matrix. This condition ensures that the restrictions, when normalized by the appropriate convergence rates, remain non-degenerate in the limit.

Theorem 4. Let $\hat{\boldsymbol{\theta}}_{\text{SUR}}$ be the SUR-ARDL FGLS estimator defined in equation (57), and let $H_0: \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ be a set of s linearly independent restrictions with restriction matrix $\mathbf{R} \in \mathbb{R}^{s \times M}$ of full row rank. Suppose that Assumptions 1-3 hold and that the rate condition is satisfied. Then, under H_0 :

$$W = (\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{SUR}} - \mathbf{r})' [\mathbf{R}(\mathbf{Z}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z})^{-1}\mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{SUR}} - \mathbf{r}) \xrightarrow{d} \chi^2(s), \quad \text{as } T \rightarrow \infty. \quad (122)$$

The proof of Theorem 5 parallels the proofs of Wagner et al. (2020, Proposition 2) and Lin and Reuvers (2025, equation (16)), adapted to the ARDL error-correction setting. The key steps are: (i) the zero-mean Gaussian mixture limiting distribution of $\hat{\boldsymbol{\theta}}_{\text{SUR}}$ established in Theorem 2, which implies that $\mathbf{G}(T)^{-1}(\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta})$ converges to a conditionally Gaussian random vector; (ii) the rate condition (156), which ensures that the normalized restriction $\mathbf{G}_R \mathbf{R} \mathbf{G}(T)[\mathbf{G}(T)^{-1}(\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta})]$ converges to a non-degenerate Gaussian random vector in the limit; and (iii) the consistency of $\hat{\boldsymbol{\Sigma}}$ (Lemma 1), which ensures that the feasible variance estimator is asymptotically equivalent to the infeasible one.

4. Monte Carlo Simulation

4.1 Data Generating Process

The data generating process (DGP) is an N -equation system of cointegrating ARDL(1,1) error-correction models with contemporaneously correlated disturbances. For each unit $i = 1, \dots, N$ and regressor $k = 1, \dots, K$, the stochastic regressor is generated as a random walk:

$$X_{k,i,t} = X_{k,i,t-1} + v_{k,i,t}, \quad v_{k,i,t} \sim N(0,1), \quad X_{k,i,0} = 0. \quad (123)$$

The N -dimensional disturbance vector $\boldsymbol{\varepsilon}_t = (u_{1,t}, \dots, u_{N,t})'$ is drawn from:

$$\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_N, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = (1 - \rho)\mathbf{I}_N + \rho \mathbf{u}_N \mathbf{u}_N', \quad (124)$$

where $\rho \in \{0, 0.3, 0.6, 0.8\}$ is the equicorrelation parameter. The dependent variable is generated recursively from the ARDL(1,1) error-correction equation:

$$\Delta Y_{i,t} = \mu_i + \lambda_i Y_{i,t-1} + \sum_{k=1}^K \beta_{k,i} X_{k,i,t-1} + \sum_{k=1}^K \gamma_{k,i} \Delta X_{k,i,t} + u_{i,t}, \quad (125)$$

with $Y_{i,0} = 0$. The baseline parameter values are $\mu_i = 0.5$, $\lambda_i = -0.3$, $\beta_{k,i} = 0.3\lambda_i = -0.09$, and $\gamma_{k,i} = 0.2$, implying a true long-run coefficient of $\theta_{k,i}^{LR} = -\beta_{k,i}/\lambda_i = 0.3$. For size experiments (Simulation 2), the DGP imposes the null of no cointegration by setting $\lambda_i = 0$ and $\beta_{k,i} = 0$, yielding $\Delta Y_{i,t} = \mu_i + \sum_k \gamma_{k,i} \Delta X_{k,i,t} + u_{i,t}$.

The experimental grid covers $N \in \{2, 3, 5\}$, $T \in \{50, 100, 200, 500\}$, $\rho \in \{0, 0.3, 0.6, 0.8\}$, and $K = 1$. A burn-in period of 50 observations is discarded to eliminate dependence on initial conditions. All experiments use $R = 5,000$ replications.

4.2 Simulation 1: Estimation Bias and RMSE

For each replication, equation-by-equation OLS is performed to obtain residuals, the covariance matrix $\hat{\boldsymbol{\Sigma}}$ is constructed, and the SUR-ARDL FGLS estimator is computed. We report bias, root mean squared error (RMSE), and relative efficiency (RE) for the speed-of-adjustment parameter λ_i and the long-run coefficient $\theta_{k,i}^{LR}$. The metrics are:

$$\text{Bias}(\hat{\psi}^{(m)}) = \frac{1}{R} \sum_{r=1}^R (\hat{\psi}^{(m,r)} - \psi_0), \quad \text{RMSE}(\hat{\psi}^{(m)}) = \left[\frac{1}{R} \sum_{r=1}^R (\hat{\psi}^{(m,r)} - \psi_0)^2 \right]^{1/2}, \quad (126)$$

$$RE(\hat{\psi}) = \frac{RMSE(\hat{\psi}^{OLS})}{RMSE(\hat{\psi}^{SUR})}, \quad (127)$$

where $RE > 1$ indicates that SUR is more efficient than OLS.

Table 1: Bias and RMSE of $\hat{\lambda}_i$ and relative efficiency (RE), OLS vs. SUR-ARDL

Panel A: $\rho = 0$ (diagonal Σ , SUR $\approx$ OLS)

N	T	Bias (OLS)	Bias (SUR)	RMSE (OLS)	RMSE (SUR)	RE
2	50	−0.0845	−0.0855	0.1424	0.1439	0.989
2	100	−0.0398	−0.0402	0.0836	0.0841	0.994
2	200	−0.0196	−0.0197	0.0517	0.0519	0.997
3	50	−0.0850	−0.0868	0.1422	0.1450	0.980
3	100	−0.0400	−0.0405	0.0825	0.0834	0.989
3	200	−0.0199	−0.0200	0.0525	0.0528	0.995
5	50	−0.0830	−0.0870	0.1407	0.1461	0.963
5	100	−0.0404	−0.0412	0.0822	0.0838	0.980
5	200	−0.0198	−0.0200	0.0525	0.0530	0.991

Panel B: $\rho = 0.6$ (moderate cross-equation correlation)

N	T	Bias (OLS)	Bias (SUR)	RMSE (OLS)	RMSE (SUR)	RE
2	50	−0.0832	−0.0763	0.1412	0.1252	1.128
2	100	−0.0414	−0.0366	0.0844	0.0726	1.161
2	200	−0.0206	−0.0176	0.0523	0.0440	1.187
3	50	−0.0831	−0.0733	0.1405	0.1181	1.189
3	100	−0.0400	−0.0341	0.0823	0.0664	1.240
3	200	−0.0192	−0.0161	0.0519	0.0408	1.272
5	50	−0.0822	−0.0721	0.1405	0.1141	1.232
5	100	−0.0398	−0.0329	0.0827	0.0632	1.308
5	200	−0.0206	−0.0165	0.0527	0.0390	1.351

Panel C: $\rho = 0.8$ (strong cross-equation correlation)

N	T	Bias (OLS)	Bias (SUR)	RMSE (OLS)	RMSE (SUR)	RE
2	50	−0.0847	−0.0661	0.1414	0.1075	1.315
2	100	−0.0399	−0.0288	0.0816	0.0586	1.394
2	200	−0.0207	−0.0149	0.0532	0.0374	1.423
3	50	−0.0826	−0.0603	0.1402	0.0961	1.460
3	100	−0.0397	−0.0268	0.0826	0.0526	1.570
3	200	−0.0196	−0.0123	0.0527	0.0323	1.629
5	50	−0.0839	−0.0589	0.1413	0.0894	1.581
5	100	−0.0404	−0.0260	0.0834	0.0483	1.728
5	200	−0.0200	−0.0123	0.0529	0.0296	1.790

Table 2: Relative efficiency of $\hat{\lambda}_i$ and $\hat{\theta}_{k,i}^{LR}$, SUR vs. OLS, as function of ρ and T ($N = 3$).

ρ	T	$RE(\hat{\lambda}_i)$	$RE(\hat{\theta}_{k,i}^{LR})$
0.0	50	0.980	0.998
0.0	100	0.989	1.000
0.0	200	0.995	1.000
0.0	500	0.998	1.000
0.3	50	1.033	1.007
0.3	100	1.047	1.003
0.3	200	1.060	1.001
0.3	500	1.072	1.000
0.6	50	1.189	1.032
0.6	100	1.240	1.007
0.6	200	1.272	1.002
0.6	500	1.276	1.000
0.8	50	1.460	1.062
0.8	100	1.570	1.011
0.8	200	1.629	1.002
0.8	500	1.619	1.000

Six findings emerge from Tables 1 and 2. First, both OLS and SUR estimators exhibit negative finite-sample bias in $\hat{\lambda}_i$, a well-documented property of the ARDL error-correction specification (Pesaran & Shin, 1999). The bias decreases with T at approximately rate T^{-1} , consistent with Theorem 1. The SUR estimator consistently exhibits smaller bias in absolute value than OLS for $\rho > 0$: for example, at $N = 5$, $T = 200$, $\rho = 0.8$, the OLS bias is -0.0200 versus -0.0123 for SUR, a reduction of 39%. Second, the relative efficiency of $\hat{\lambda}_i$ increases monotonically with ρ , confirming Proposition 2. At $\rho = 0$, $RE \approx 1.00$ as expected (SUR = OLS when Σ is diagonal, Proposition 1). At $\rho = 0.8$, the gains are substantial: RE reaches 1.79 for $N = 5$, $T = 200$, implying a 79% increase in precision from system estimation. Third, the relative efficiency increases with N for fixed $\rho > 0$, reflecting the additional cross-equation information available in larger systems. At $\rho = 0.8$, $T = 200$: RE = 1.42 for $N = 2$, RE = 1.63 for $N = 3$, and RE = 1.79 for $N = 5$.

Fourth, the relative efficiency of $\hat{\lambda}_i$ increases with T for fixed $\rho > 0$, indicating that the asymptotic efficiency gain (Proposition 2) is better realized in larger samples where $\hat{\Sigma}$ is more precisely estimated. For $N = 5$, $\rho = 0.8$: RE rises from 1.58 at $T = 50$ to 1.81 at $T = 500$. Fifth, the efficiency gains for the long-run coefficient $\hat{\theta}_{k,i}^{LR}$ are modest relative to those for $\hat{\lambda}_i$ (Table 2, right column). The RE for $\hat{\theta}_{k,i}^{LR}$ rarely exceeds 1.06, even at $\rho = 0.8$. This reflects the fact that the long-run coefficient is a ratio of two super-consistent estimators (T -consistent), and the efficiency gain from SUR estimation is a second-order effect for super-consistent parameters. The gain is more pronounced in small samples ($T = 50$) where the finite-sample variability of the ratio is larger. Sixth, the magnitude of the efficiency gains is broadly consistent with the FM-SUR literature. Wagner et al. (2020, Table 1) report RMSE reductions of 5-15% for non-pooled FM-SUR relative to FM-SOLS. The SUR-ARDL framework achieves comparable gains for the error-correction parameters and even larger gains (up to 81% RMSE reduction) for $\hat{\lambda}_i$ at high ρ .

4.3 Simulation 2: Empirical Size of the SUR Bounds F-Test

Data are generated under the null of no cointegration ($\lambda_i = 0, \beta_{k,i} = 0$). For each replication, the unrestricted model (189) is estimated by both OLS and SUR, and the bounds F-statistic is computed for equation 1. The rejection decision is recorded using the PSS upper-bound critical value ($c_U = 5.73$ at 5% for $K = 1$, Case III; Pesaran et al., 2001, Table CI) and the Narayan (2005) small-sample critical values where available. As a benchmark, bootstrap critical values are computed from the empirical 95% quantile of the F_i^{SUR} distribution under the null.

The empirical rejection rate (ERR) at nominal level $\alpha = 0.05$ is:

$$\text{ERR}_\alpha^{(m)} = \frac{1}{R} \sum_{r=1}^R \mathbf{1}\{F_i^{(m,r)} > c_U(\alpha)\}, \quad m \in \{\text{OLS}, \text{SUR}\}. \quad (192)$$

Table 3: Empirical rejection rate of the bounds F-test under H_0 (nominal 5%)

Panel A: $N = 2$

T	ρ	OLS (PSS)	SUR (PSS)	SUR (Narayan)	SUR (Boot)
50	0.0	0.023	0.025	0.019	0.050
50	0.3	0.025	0.027	0.021	0.050
50	0.6	0.018	0.023	0.018	0.050
50	0.8	0.023	0.023	0.019	0.050
100	0.0	0.011	0.011	0.011	0.050
100	0.3	0.010	0.011	0.011	0.050
100	0.6	0.011	0.012	0.012	0.050
100	0.8	0.008	0.009	0.009	0.050
200	0.0	0.005	0.006	0.006	0.050
200	0.3	0.006	0.007	0.007	0.050
200	0.6	0.005	0.006	0.006	0.050
200	0.8	0.006	0.006	0.006	0.050

Panel B: $N = 3$

T	ρ	OLS (PSS)	SUR (PSS)	SUR (Narayan)	SUR (Boot)
50	0.0	0.024	0.029	0.023	0.050
50	0.3	0.022	0.031	0.027	0.050
50	0.6	0.024	0.036	0.029	0.050
50	0.8	0.021	0.034	0.030	0.050
100	0.0	0.008	0.009	0.009	0.050
100	0.3	0.012	0.019	0.019	0.050
100	0.6	0.009	0.013	0.013	0.050
100	0.8	0.011	0.018	0.018	0.050
200	0.0	0.007	0.007	0.007	0.050
200	0.3	0.006	0.009	0.009	0.050
200	0.6	0.004	0.007	0.007	0.050

200	0.8	0.006	0.009	0.009	0.050
-----	-----	-------	-------	-------	-------

Panel C: $N = 5$

T	ρ	OLS (PSS)	SUR (PSS)	SUR (Narayan)	SUR (Boot)
50	0.0	0.019	0.026	0.020	0.050
50	0.3	0.020	0.047	0.040	0.050
50	0.6	0.020	0.071	0.056	0.050
50	0.8	0.021	0.089	0.075	0.050
100	0.0	0.009	0.012	0.012	0.050
100	0.3	0.009	0.016	0.016	0.050
100	0.6	0.009	0.027	0.027	0.050
100	0.8	0.010	0.047	0.047	0.050
200	0.0	0.005	0.006	0.006	0.050
200	0.3	0.006	0.010	0.010	0.050
200	0.6	0.006	0.014	0.014	0.050
200	0.8	0.005	0.019	0.019	0.050
500	0.0	0.004	0.003	0.003	0.050
500	0.3	0.007	0.007	0.007	0.050
500	0.6	0.003	0.005	0.005	0.050
500	0.8	0.006	0.008	0.008	0.050

Bold entries indicate rejection rates exceeding $2 \times \alpha = 0.10$ (liberal test).

Five findings emerge from Table 3. First, the PSS critical values are uniformly conservative for the single-equation OLS bounds test: empirical rejection rates range from 0.003 to 0.025, well below the nominal 5%. This is consistent with the known conservatism of asymptotic PSS critical values in finite samples, where Narayan (2005) critical values provide a closer approximation.

Second, the PSS critical values remain conservative for the SUR bounds test when N is small ($N = 2$) or ρ is small ($\rho \leq 0.3$). In these configurations, the SUR F-statistic rejection rate stays below 3%, indicating that the single-equation critical values provide a safe approximation.

Third, size distortions emerge for the SUR test when both N and ρ are large. The most severe distortions occur at $N = 5$, $T = 50$, $\rho = 0.8$, where the SUR rejection rate is 8.9% nearly double the nominal 5%. This over-rejection reflects the dependence of the SUR F-statistic's null distribution on Σ (Theorem 3, equation (146)): the cross-equation weighting through $\hat{\Sigma}^{-1}$ shifts the F-statistic distribution relative to the single-equation benchmark, and the PSS critical values fail to account for this shift. The distortion diminishes with T : at $T = 200$, the worst-case SUR rejection rate is 1.9% ($N = 5$, $\rho = 0.8$), and by $T = 500$, all SUR rejection rates are below 1%.

Fourth, the Narayan (2005) small-sample critical values reduce the distortion modestly for $T = 50$ but offer no improvement for $T \geq 100$ (where they coincide with the PSS values). For $N = 5$, $T = 50$, $\rho = 0.8$, the Narayan-based rejection rate is 7.5%, still substantially above 5%.

Fifth, the bootstrap critical values achieve exact nominal size (5.0%) by construction across all configurations, confirming that the bootstrap correctly captures the SUR-specific null distribution. Based on these findings, the practical recommendation is:

$$\text{Use: } \begin{cases} \text{PSS/Narayan critical values} & \text{if } N \leq 3, \text{ or } \rho \leq 0.3, \\ \text{Bootstrap critical values} & \text{if } N \geq 5 \text{ and } \rho > 0.5. \end{cases} \quad (193)$$

The pattern of size distortions is consistent with, but milder than, those reported by Wagner et al. (2020, Tables 2-3) for the FM-SUR Wald test. Wagner et al. find rejection rates of 30-85% at nominal 5% for $N = 10$ and $T = 100$, driven by imprecise long-run covariance estimation. The substantially better size control of the SUR-ARDL bounds test reflects the reduced dimensionality of the nuisance parameter: the SUR-ARDL framework requires estimation of only the $N \times N$ contemporaneous covariance Σ , whereas the FM-SUR framework requires the $2N \times 2N$ long-run covariance Ω , with its attendant bandwidth selection problem.

4.4 Simulation 3: Size-Corrected Power of the Bounds F-Test

Data are generated under the alternative hypothesis with $\lambda_i \in \{-0.1, -0.2, -0.3, -0.5, -0.8\}$ (all equations share the same λ_i), $\theta^{LR} = 0.3$, $N = 3$, and $\rho \in \{0, 0.3, 0.6, 0.8\}$. To isolate the power comparison from size distortions, we employ size-corrected critical values: for each (T, ρ) configuration, $R_0 = 5,000$ replications are generated under H_0 (null DGP), and the empirical $(1 - \alpha)$ quantile of the null distribution provides the size-corrected critical value separately for OLS and SUR:

$$c_\alpha^{(m)} = F_{(m), [R_0(1-\alpha)]}^{H_0}, \quad m \in \{\text{OLS}, \text{SUR}\}. \quad (194)$$

Size-corrected power is then the rejection rate under H_1 using these empirical critical values:

$$\text{Power}_\alpha^{(m)}(\lambda) = \frac{1}{R} \sum_{r=1}^R \mathbf{1}\{F_i^{(m,r)} > c_\alpha^{(m)}\}. \quad (195)$$

Table 4: Size-corrected power of the bounds F-test (N=3, $\alpha=0.05$)

Panel A: T = 50

λ	$\rho = 0$ (OLS)	$\rho = 0$ (SUR)	$\rho = 0.3$ (OLS)	$\rho = 0.3$ (SUR)	$\rho = 0.6$ (OLS)	$\rho = 0.6$ (SUR)	$\rho = 0.8$ (OLS)	$\rho = 0.8$ (SUR)
0.1	0.358	0.369	0.337	0.355	0.350	0.475	0.322	0.574
0.2	0.662	0.658	0.623	0.653	0.636	0.817	0.619	0.918
0.3	0.886	0.877	0.870	0.883	0.892	0.963	0.863	0.996
0.5	0.998	0.996	0.996	0.998	0.995	1.000	0.997	1.000
0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Panel B: T = 100

λ	$\rho = 0$ (OLS)	$\rho = 0$ (SUR)	$\rho = 0.3$ (OLS)	$\rho = 0.3$ (SUR)	$\rho = 0.6$ (OLS)	$\rho = 0.6$ (SUR)	$\rho = 0.8$ (OLS)	$\rho = 0.8$ (SUR)
0.1	0.724	0.719	0.714	0.757	0.738	0.891	0.701	0.965
0.2	0.990	0.989	0.984	0.988	0.988	0.997	0.984	1.000
0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Panel C: T = 200

λ	$\rho = 0$ (OLS)	$\rho = 0$ (SUR)	$\rho = 0.3$ (OLS)	$\rho = 0.3$ (SUR)	$\rho = 0.6$ (OLS)	$\rho = 0.6$ (SUR)	$\rho = 0.8$ (OLS)	$\rho = 0.8$ (SUR)
-----------	---------------------	---------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

0.1	0.991	0.989	0.991	0.994	0.992	1.000	0.992	1.000
0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Four findings emerge from Table 4, consistent with the theoretical predictions of Proposition 4 and the finite-sample considerations. First, SUR power is uniformly higher than OLS power when $\rho > 0$, confirming that the efficiency gain from system estimation (Proposition 2) translates into enhanced ability to detect cointegrating relationships. At $\rho = 0$, SUR and OLS have virtually identical power, consistent with Proposition 1 (SUR = OLS when Σ is diagonal). Second, the power gains are largest for moderate values of $|\lambda|$ specifically $|\lambda| \in \{0.1, 0.2\}$ where the test is on the margin between rejection and non-rejection. The gain is negligible for large $|\lambda|$ (≥ 0.3 with $T = 100$) where both tests achieve near-unit power, and modest for very small $|\lambda|$ where neither test has much power. This "sweet spot" pattern is a direct consequence of the monotonicity of the power function: the efficiency gain shifts the power curve leftward, and the improvement in rejection probability is largest where the power curve is steepest.

Third, the magnitude of the power gain increases substantially with ρ . For $T = 50$ and $|\lambda| = 0.2$: the power gain is -0.004 at $\rho = 0$, $+0.030$ at $\rho = 0.3$, $+0.181$ at $\rho = 0.6$, and $+0.299$ at $\rho = 0.8$. At $\rho = 0.8$ with $T = 50$, the SUR test achieves 57.4% power for $|\lambda| = 0.1$ compared to only 32.2% for OLS a gain of 25.2 percentage points. This is precisely the scenario where a researcher using single-equation ARDL would fail to detect cointegration, while the SUR-ARDL framework would succeed. Fourth, the power gains diminish with T as both tests approach unit power. For $T = 200$, the largest gain is 0.8 percentage points ($\rho = 0.8$, $|\lambda| = 0.1$). This is consistent with the theoretical prediction that the efficiency gain matters most in finite samples where there is genuine uncertainty about rejection.

The results confirm the finding of Wagner et al. (2020, Figures 1-4) that system estimation yields higher size-corrected power than single-equation estimation, with the SUR-ARDL gains being quantitatively larger than those reported for FM-SUR versus FM-SOLS. The larger gains reflect the parametric efficiency of the ARDL error-correction specification, which avoids the nonparametric bandwidth selection that introduces additional variability in the FM-based tests.

4.5 Simulation 4: Empirical Size of Cross-Equation Wald Tests

Data are generated under the homogeneous DGP ($\lambda_i = -0.3$, $\beta_{k,i} = -0.09$, $\theta_{k,i}^{LR} = 0.3$ for all i), so all cross-equation equality restrictions hold. Three Wald-type tests are evaluated:

(a) **Poolability test** (W_{Pool}): Tests $H_0: \beta_{k,i} = \beta_{k,1}$ for $i = 2, \dots, N$, $k = 1, \dots, K$, with $K(N - 1)$ degrees of freedom (equation (170)).

(b) **Long-run coefficient equality test** (W_{LR}^{ij}): Tests $H_0: \theta_{k,i}^{LR} = \theta_{k,j}^{LR}$ via the delta method for all pairs (i, j) with $i < j$ (equation (164)). We report the average rejection rate across all $\binom{N}{2}$ pairs and the maximum pairwise rejection rate.

(c) **Speed-of-adjustment equality test** (W_{λ}^{ij}): Tests $H_0: \lambda_i = \lambda_j$ for all pairs (i, j) with $i < j$ (equation (159)). We report the average rejection rate.

All tests use the χ^2 critical value at $\alpha = 0.05$.

Table 5: Empirical rejection rate of cross-equation Wald tests under H_0 (nominal 5%)

Panel A: $N = 2$ (1 pair; $\text{df}_{\text{Pool}} = 1$)

T	ρ	W_{Pool}	W_{LR}^{ij} (avg)	W_{λ}^{ij} (avg)
-----	--------	-------------------	---------------------	--------------------------

50	0.0	0.108	0.108	0.079
50	0.3	0.104	0.109	0.080
50	0.6	0.093	0.101	0.076
50	0.8	0.082	0.090	0.063
100	0.0	0.076	0.088	0.062
100	0.3	0.076	0.087	0.060
100	0.6	0.076	0.081	0.060
100	0.8	0.064	0.074	0.054
200	0.0	0.066	0.071	0.058
200	0.3	0.056	0.063	0.053
200	0.6	0.059	0.064	0.056
200	0.8	0.056	0.058	0.049
500	0.0	0.051	0.057	0.051
500	0.3	0.058	0.051	0.052
500	0.6	0.054	0.054	0.052
500	0.8	0.049	0.050	0.053

Panel B: $N = 3$ (3 pairs; $df_{\text{Pool}} = 2$)

T	ρ	W_{Pool}	W_{LR}^{ij} (avg)	W_{LR}^{ij} (max)	W_{λ}^{ij} (avg)
50	0.0	0.132	0.106	0.233	0.084
50	0.3	0.137	0.107	0.237	0.083
50	0.6	0.120	0.108	0.240	0.080
50	0.8	0.089	0.084	0.189	0.063
100	0.0	0.090	0.085	0.194	0.069
100	0.3	0.092	0.086	0.196	0.071
100	0.6	0.079	0.080	0.184	0.063
100	0.8	0.066	0.068	0.159	0.057
200	0.0	0.067	0.068	0.155	0.056
200	0.3	0.066	0.069	0.161	0.057
200	0.6	0.063	0.070	0.162	0.056
200	0.8	0.054	0.057	0.129	0.052
500	0.0	0.049	0.053	0.127	0.051
500	0.3	0.052	0.060	0.142	0.054
500	0.6	0.055	0.058	0.133	0.052
500	0.8	0.049	0.052	0.122	0.051

Panel C: $N = 5$ (10 pairs; $df_{\text{Pool}} = 4$)

T	ρ	W_{Pool}	W_{LR}^{ij} (avg)	W_{LR}^{ij} (max)	W_{λ}^{ij} (avg)
50	0.0	0.212	0.127	0.531	0.096
50	0.3	0.204	0.122	0.528	0.094
50	0.6	0.180	0.118	0.505	0.087

50	0.8	0.137	0.099	0.444	0.070
100	0.0	0.119	0.091	0.429	0.073
100	0.3	0.116	0.091	0.425	0.070
100	0.6	0.090	0.079	0.382	0.065
100	0.8	0.087	0.074	0.361	0.060
200	0.0	0.071	0.071	0.364	0.059
200	0.3	0.076	0.067	0.338	0.061
200	0.6	0.064	0.066	0.328	0.057
200	0.8	0.060	0.059	0.313	0.052
500	0.0	0.062	0.058	0.308	0.054
500	0.3	0.059	0.057	0.301	0.055
500	0.6	0.061	0.058	0.307	0.053
500	0.8	0.052	0.055	0.289	0.050

Six findings emerge from Table 5, confirming the predictions of Theorems 5 and 6 and revealing important finite-sample regularities. First, all three Wald statistics converge to correct nominal size as T increases, consistent with the asymptotic χ^2 distributions established in Theorems 5 and 6. For $T = 500$, the poolability test rejection rate is within 0.1-1.2 percentage points of the nominal 5% across all (N, ρ) configurations. The λ -equality test converges fastest, with rejection rates at or below 5.5% for $T \geq 200$ in all configurations. Second, the λ -equality test (W_{λ}^{ij}) is the best-sized of the three tests at every sample size and system dimension. Its average rejection rate reaches nominal size for $T \geq 200$ across all configurations, reflecting the favorable convergence properties of the super-consistent $\hat{\lambda}_i$ estimator. By contrast, the poolability and LR equality tests exhibit persistent over-rejection in small samples.

Third, the poolability test (W_{Pool}) exhibits the most severe size distortion for large N and small T , reaching 21.2% at $N = 5$, $T = 50$, $\rho = 0$ (more than four times the nominal level). This reflects the large number of restrictions being tested simultaneously ($K(N - 1) = 4$ for $N = 5$, $K = 1$) and the imprecision of the 4×4 variance-covariance sub-block of $(\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}$ in small samples. Fourth, the over-rejection of the poolability test decreases with ρ at fixed (N, T) . For $N = 5$, $T = 50$: ERR = 0.212 at $\rho = 0$ but 0.137 at $\rho = 0.8$. This seemingly counterintuitive pattern reflects the improved precision of $\hat{\Sigma}$ at higher correlation: the equicorrelated structure is more accurately estimated than a diagonal structure when the true off-diagonal elements are large, leading to a more accurate FGLS variance estimate.

Fifth, the LR equality test (W_{LR}^{ij}) shows average pairwise rejection rates of 8-11% at $T = 50$, converging to 5-6% by $T = 500$. However, the maximum pairwise rejection rate is substantially higher, especially for large N : at $N = 5$, $T = 50$, the maximum pairwise rate reaches 53.1%. This reflects the well-known multiple testing problem: with $\binom{N}{2} = 10$ pairwise comparisons at $N = 5$, the probability of at least one false rejection far exceeds α even when each individual test has correct size. The delta method variance (equation (178)) involves division by $\hat{\lambda}_i^2$, which amplifies small-sample noise when $|\hat{\lambda}_i|$ is small, contributing to the over-rejection of individual pairwise tests. Researchers conducting multiple pairwise LR equality tests should apply a Bonferroni or Holm-Šidák correction to control the family-wise error rate.

Sixth, the pattern of size convergence across tests provides practical guidance: the poolability test requires $T \geq 200$ for $N = 5$ and $T \geq 100$ for $N \leq 3$ to achieve rejection rates within 3 percentage points of nominal; the LR equality test requires comparable sample sizes but is additionally affected by the multiple testing issue; the λ -equality test achieves satisfactory size

for $T \geq 100$ across all N considered. For the ASEAN application ($N = 6$, $T \approx 40$), the Wald tests should be interpreted cautiously and supplemented with bootstrap p -values.

4.6 Simulation 5: Coverage of the Long-Run Coefficient Confidence Interval

Data are generated under H_1 with $N = 3$, $\rho = 0.6$, $T \in \{50, 100, 200, 500\}$, and $\lambda_i \in \{-0.1, -0.2, -0.3, -0.5, -0.8\}$. The true long-run coefficient is $\theta^{LR} = 0.3$. For each replication, the delta method standard error of $\hat{\theta}_{k,i}^{LR} = -\hat{\beta}_{k,i}/\hat{\lambda}_i$ is computed from the gradient $\nabla g = (\hat{\beta}_{k,i}/\hat{\lambda}_i^2, -1/\hat{\lambda}_i)'$ and the (2×2) sub-block of the variance-covariance matrix corresponding to $(\hat{\lambda}_i, \hat{\beta}_{k,i})$ (equation (178)). Three metrics are reported:

- (a) **Empirical coverage:** the proportion of replications in which the 95% confidence interval $\hat{\theta}^{LR} \pm 1.96 \times \widehat{\text{se}}(\hat{\theta}^{LR})$ contains the true value.
- (b) **Median CI width:** the median width $2 \times 1.96 \times \widehat{\text{se}}(\hat{\theta}^{LR})$ across replications.
- (c) **Empirical rejection rate** of the t -test $H_0: \theta^{LR} = 0.3$, which equals one minus the coverage rate.

Table 6: Empirical coverage and width of 95% delta method CI for $\hat{\theta}^{LR}$

λ	T	Coverage (OLS)	Coverage (SUR)	Width (OLS)	Width (SUR)
-0.1	50	0.609	0.559	1.274	1.058
-0.1	100	0.351	0.240	0.816	0.683
-0.1	200	0.069	0.023	0.486	0.392
-0.1	500	0.000	0.000	0.215	0.170
-0.2	50	0.358	0.263	0.828	0.694
-0.2	100	0.082	0.030	0.486	0.396
-0.2	200	0.001	0.000	0.263	0.209
-0.2	500	0.000	0.000	0.111	0.087
-0.3	50	0.188	0.102	0.621	0.505
-0.3	100	0.014	0.002	0.344	0.274
-0.3	200	0.000	0.000	0.179	0.141
-0.3	500	0.000	0.000	0.074	0.057
-0.5	50	0.042	0.012	0.413	0.334
-0.5	100	0.000	0.000	0.214	0.170
-0.5	200	0.000	0.000	0.112	0.089
-0.5	500	0.000	0.000	0.045	0.035
-0.8	50	0.004	0.000	0.270	0.216
-0.8	100	0.000	0.000	0.139	0.109
-0.8	200	0.000	0.000	0.070	0.054
-0.8	500	0.000	0.000	0.028	0.022

Five findings emerge from Table 6. First, the delta method confidence intervals exhibit severe under-coverage throughout, with the problem most acute for weak error correction ($|\lambda| = 0.1$) and small samples. At $\lambda = -0.1$, $T = 50$, OLS coverage is only 60.9% (versus the nominal 95%), and SUR coverage is even lower at 55.9%. This is a well-known pathology of the delta method for ratio estimators: when the denominator $\hat{\lambda}_i$ is imprecisely estimated and close to

zero, the distribution of $\hat{\theta}^{LR} = -\hat{\beta}_{k,i}/\hat{\lambda}_i$ is heavy-tailed and the first-order Gaussian approximation $\hat{\theta}^{LR} \sim N(\theta^{LR}, \nabla g' \hat{V} \nabla g)$ fails to capture the probability mass in the tails.

Second, coverage deteriorates monotonically as $|\lambda|$ increases and as T increases. This seemingly paradoxical result that *larger* samples yield *worse* coverage arises because the super-consistent estimator $\hat{\theta}^{LR}$ concentrates around the true value faster than the delta method standard error shrinks. The ratio $|\hat{\theta}^{LR} - \theta^{LR}|/\widehat{\text{se}}(\hat{\theta}^{LR})$ diverges as $T \rightarrow \infty$ at rate $\sqrt{T}$, since $\hat{\theta}^{LR}$ is T -consistent but the delta method SE is calibrated to a $\sqrt{T}$ -consistent estimator. In essence, the delta method CI is appropriately centered but catastrophically too wide for the super-consistent estimator, making the t -statistic diverge under the true parameter value. The nominal coverage probability can only hold in the intermediate regime where the ratio has not yet diverged.

Third, the SUR CI is uniformly narrower than the OLS CI (the width ratio ranges from 0.78 to 0.83), reflecting the efficiency gain of Proposition 2. However, the SUR coverage is also uniformly *lower* than OLS coverage: the narrower SUR interval does not compensate for the distributional misspecification. The SUR t -statistic converges to normality faster in principle (Theorem 7), but this advantage is swamped by the delta method's inadequacy for the ratio estimator. Fourth, for practical inference on θ^{LR} , the results suggest that the delta method CI is unreliable in all configurations considered. Even at $T = 500$ with $|\lambda| = 0.1$, coverage is essentially zero. This motivates the use of alternative inference procedures: the Fieller (1954) confidence set, which inverts the test $H_0: \beta_{k,i} + c\lambda_i = 0$ over $c \in \mathbb{R}$ and does not require division by $\hat{\lambda}_i$; or bootstrap confidence intervals constructed from the percentile or bias-corrected accelerated (BCa) method. Fifth, the median CI widths decrease with T at approximately rate $T^{-1/2}$ for OLS and SUR, consistent with the $\sqrt{T}$ -consistent delta method SE. The SUR widths are 17-22% narrower than OLS widths, providing more precise (though still invalid) point estimates of θ^{LR} .

4.7 Simulation 6: Two-Step versus Iterated FGLS

Data are generated under H_1 with $\lambda_i = -0.3$, $\theta^{LR} = 0.3$, $\rho = 0.6$, $N \in \{2,3,5\}$, and $T \in \{50,100,200,500\}$. Three SUR-ARDL estimators are compared: (a) two-step FGLS using $\hat{\Sigma}^{\text{OLS}}$; (b) iterated FGLS with convergence tolerance 10^{-8} and maximum 100 iterations; (c) infeasible GLS using the true Σ . RMSE is computed for $\hat{\lambda}_i$ and $\hat{\theta}_{k,i}^{LR}$ (equation 1).

Table 7: RMSE of two-step, iterated, and infeasible GLS

N	T	RMSE($\hat{\lambda}$)			RMSE($\hat{\theta}^{LR}$)		
		2-Step	Iterated	Infeasible	2-Step	Iterated	Infeasible
2	50	0.1237	0.1239	0.1223	0.611	0.612	0.610
2	100	0.0712	0.0712	0.0707	0.602	0.602	0.602
2	200	0.0442	0.0442	0.0440	0.600	0.600	0.600
2	500	0.0258	0.0258	0.0258	0.600	0.600	0.600
3	50	0.1160	0.1159	0.1125	0.611	0.612	0.611
3	100	0.0667	0.0668	0.0658	0.601	0.601	0.601
3	200	0.0415	0.0415	0.0412	0.600	0.600	0.600
3	500	0.0239	0.0239	0.0238	0.600	0.600	0.600
5	50	0.1137	0.1137	0.1078	0.607	0.608	0.606
5	100	0.0632	0.0631	0.0617	0.601	0.601	0.601
5	200	0.0383	0.0383	0.0380	0.600	0.600	0.599

5	500	0.0226	0.0226	0.0225	0.600	0.600	0.600
---	-----	--------	--------	--------	-------	-------	-------

Three findings emerge from Table 7. First, the two-step and iterated FGLS estimators produce virtually identical RMSE across all configurations. The maximum difference in $\text{RMSE}(\hat{\lambda}_i)$ between the two estimators is 0.0001 (at $N = 3, T = 100$), and the difference in $\text{RMSE}(\hat{\theta}^{LR})$ is negligible at 0.001 or less. This confirms the theoretical prediction that iteration does not improve asymptotic efficiency (Proposition 5): both estimators achieve the same Gauss-Markov bound, and the finite-sample differences vanish rapidly. Second, the gap between the feasible (two-step or iterated) and infeasible GLS narrows with T . At $N = 5, T = 50$, the $\text{RMSE}(\hat{\lambda})$ for two-step is 0.1137 versus 0.1078 for infeasible a 5.5% gap. By $T = 500$, the gap shrinks to 0.4%. This confirms that $\hat{\Sigma} \xrightarrow{p} \Sigma$ at rate $T^{-1/2}$ (Lemma 1) and that the feasibility cost is negligible for $T \geq 100$. The gap is slightly larger for bigger N (reflecting the increased dimensionality of $\hat{\Sigma}$), but remains under 6% even in the most challenging configuration ($N = 5, T = 50$). Third, the RMSE of $\hat{\theta}^{LR}$ is remarkably stable across all three estimators and all sample sizes within a given N . The values cluster around 0.600-0.612, consistent with the known heavy-tailed distribution of the ratio estimator. The similarity of $\text{RMSE}(\hat{\theta}^{LR})$ across two-step, iterated, and infeasible GLS confirms that the imprecision of $\hat{\theta}^{LR}$ is driven by the ratio structure, not by the quality of $\hat{\Sigma}$.

5. Empirical findings

The empirical application examines the renewable energy consumption-economic growth nexus in six ASEAN member states: Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam. The analysis uses annual data spanning the period 1990-2024 ($T = 35$), yielding a balanced panel of $N = 6$ countries and 210 observations. Brunei Darussalam, Cambodia, Lao PDR, and Myanmar are excluded due to incomplete data availability. Table 8 describes the four variables employed. Renewable energy consumption (REC), measured in terawatt-hours of total primary energy from renewables, is sourced from the Energy Institute (2025). GDP per capita in constant 2015 US dollars, foreign direct investment (FDI) net inflows as a percentage of GDP, and trade openness as a percentage of GDP are obtained from the World Bank's World Development Indicators.

Table 8: Descriptive statistics

Variable	Description	Mean	SD	Min	Max
REC	Renewable energy (TWh)	56.35	64.81	0.27	327.60
GDP	GDP per capita (2015 USD)	11,301	16,283	917	68,219
FDI	FDI net inflows (% GDP)	5.58	6.87	-2.76	33.30
TO	Trade openness (% GDP)	145.66	101.99	32.97	437.33

The six countries exhibit substantial heterogeneity. Singapore's renewable energy consumption (mean 2.1 TWh) is two orders of magnitude smaller than Indonesia's (mean 113.2 TWh), while Singapore's GDP per capita (mean \$44,570) exceeds Vietnam's (mean \$5,326) by a factor of eight. This cross-country heterogeneity motivates the use of country-specific coefficients in the SUR-ARDL specification, while the common geographic and institutional environment of ASEAN motivates the exploitation of cross-equation correlation.

Figure 1: Time Series of Transformed Variables (1990–2024)



Figure 1: Graph of each variables for six ASEAN countries (1990-2024)

Table 11 reports augmented Dickey-Fuller (ADF) test statistics for each variable in levels and first differences. The tests include an intercept and one lag of the dependent variable. The results confirm the expected integration properties. In levels, the unit root null cannot be rejected at 5% for $\ln(\text{REC})$, $\ln(\text{GDP})$, and $\ln(\text{TO})$ in most countries, consistent with $I(1)$ behaviour. In first differences, all variables reject the unit root null decisively, confirming stationarity after differencing. The variable $\ln(\text{FDI}+c)$ exhibits mixed evidence: it appears $I(0)$ in some countries (reflecting the mean-reverting nature of FDI flows as a percentage of GDP) and $I(1)$ in others. This mixed integration order is precisely the setting for which the ARDL bounds approach is designed, since it remains valid regardless of whether regressors are $I(0)$, $I(1)$, or a mixture thereof (Pesaran et al., 2001).

Before estimating the SUR-ARDL system, we test for cross-sectional correlation in the OLS residuals to determine whether system estimation is warranted. Table 9 reports the 6×6 correlation matrix of equation-by-equation OLS residuals.

Table 9: Correlation matrix of OLS residuals

	IND	MYS	PHL	SGP	THA	VNM
IND	1.000	0.240	−0.138	0.066	0.009	−0.011
MYS	0.240	1.000	0.548	0.093	−0.364	0.115
PHL	−0.138	0.548	1.000	0.103	−0.281	0.283
SGP	0.066	0.093	0.103	1.000	−0.193	0.317
THA	0.009	−0.364	−0.281	−0.193	1.000	0.086
VNM	−0.011	0.115	0.283	0.317	0.086	1.000

The Breusch-Pagan LM test strongly rejects the null of diagonal Σ ($\chi^2_{15} = 28.91$, $p = 0.017$), confirming that cross-equation correlation is present and that SUR estimation is warranted.

The Pesaran CD statistic is 1.32 ($p = 0.189$), which does not reject however, the CD test is designed for large- N panels and has limited power at $N = 6$. The average absolute off-diagonal correlation is 0.190, with the Malaysia-Philippines pair exhibiting the strongest correlation ($\hat{\rho} = 0.548$) and Malaysia-Thailand the strongest negative correlation ($\hat{\rho} = -0.364$). The T/N ratio of $34/6 = 5.7$ falls below the $T/N \geq 10$ threshold identified in Simulation 7, suggesting that $\hat{\Sigma}$ is estimated with moderate imprecision (expected median RFN ≈ 0.30). Nevertheless, the zero invertibility failure rate at $T/N = 3$ in Simulation 7 ensures that the SUR procedure is always implementable.

Table 10 reports the speed-of-adjustment parameter $\hat{\lambda}_i$, the bounds F-statistic, and the long-run coefficients $\hat{\theta}_{k,i}^{LR} = -\hat{\beta}_{k,i}/\hat{\lambda}_i$ from both equation-by-equation OLS and the two-step SUR-ARDL estimator.

Table 10: OLS and SUR-ARDL estimation results

Country	Method	$\hat{\lambda}_i$	$\text{se}(\hat{\lambda}_i)$	$t\text{-stat}$	F_i	$\hat{\theta}_{\text{GDP}}^{LR}$	$\hat{\theta}_{\text{FDI}}^{LR}$	$\hat{\theta}_{\text{TO}}^{LR}$
Indonesia	OLS	-0.098	0.103	-0.95	0.51	2.175	-0.964	-2.663
	SUR	-0.092	0.085	-1.09	0.56	2.323	-0.723	-2.263
Malaysia	OLS	-0.212	0.115	-1.84	1.71	2.723	0.729	-0.311
	SUR	-0.205	0.086	-2.40	2.38	2.524	0.455	-0.531
Philippines	OLS	-0.559	0.147	-3.80	5.73	0.690	0.342	0.391
	SUR	-0.541	0.107	-5.04	8.63	0.719	0.181	0.415
Singapore	OLS	-0.384	0.168	-2.28	4.37	0.864	1.093	-0.742
	SUR	-0.441	0.138	-3.19	7.24	1.060	0.931	-0.397
Thailand	OLS	-0.528	0.165	-3.21	7.28	1.774	-1.274	1.264
	SUR	-0.619	0.129	-4.78	11.45	1.967	-0.930	1.086
Vietnam	OLS	-0.505	0.135	-3.75	3.89	1.232	-0.115	-0.768
	SUR	-0.497	0.109	-4.58	6.14	1.282	0.002	-0.814

6. Conclusion

This paper develops a seemingly unrelated autoregressive distributed lag (SUR-ARDL) framework for estimating and testing cointegrating relationships in panels with small cross-sectional dimension. The key innovation is to embed the Pesaran et al.(2001) ARDL error-correction specification within the Zellner (1962) SUR system, exploiting cross-equation correlation for efficiency while the ARDL lag structure parametrically absorbs short-run dynamics and regressor endogeneity. Unlike the FM-SUR approach of Wagner et al.(2020), which requires consistent estimation of a $2N \times 2N$ long-run covariance matrix Ω , and the FM-GLS approach of Lin and Reuvers (2025), which relies on a banded inverse autocovariance matrix, the SUR-ARDL estimator depends only on the $N \times N$ contemporaneous covariance matrix Σ . This dimensional reduction eliminates the principal source of finite-sample distortion in existing SUR cointegration methods the estimation of large, high-dimensional nuisance-parameter matrices and extends the practical applicability of system-based cointegration analysis to settings where T is moderate relative to N .

The Monte Carlo study, comprising seven experiments with 5,000 replications, validates the finite-sample properties. Six principal findings emerge. First, SUR-ARDL achieves RMSE reductions of 15-80% for the error-correction parameter λ relative to equation-by-equation OLS, with gains increasing in N , T , and the cross-equation correlation ρ . Second, the bounds test exhibits good size with bootstrap critical values, though tabulated PSS critical values can be too liberal for $N \geq 5$ when correlation is strong. Third, power gains are substantial up to 30 percentage points for moderate $|\lambda|$ precisely the parameter region where single-equation

methods struggle. Fourth, cross-equation Wald tests converge to their nominal size as T increases, but at small T the poolability test over-rejects and multiple pairwise comparisons require Holm or Bonferroni correction. Fifth, two-step FGLS is sufficient: iteration provides negligible improvement, confirming that the standard Zellner two-step procedure is adequate. Sixth, and most importantly, the $N \times N$ covariance matrix $\hat{\Sigma}$ is reliably estimable with zero invertibility failures across all configurations including $N = 10$ and $T = 30$ ($T/N = 3$) whereas Wagner et al.(2020) report invertibility failure rates exceeding 50% for the $2N \times 2N$ matrix $\hat{\Omega}$ at comparable dimensions.

The empirical application to the renewable energy-economic growth nexus in six ASEAN countries (1990-2024) demonstrates three practical advantages. First, SUR standard errors are 17-41% smaller than OLS, translating to meaningful efficiency gains even with moderate average cross-equation correlation ($|\rho| = 0.19$). Second, system estimation changes substantive conclusions: for Vietnam, the bounds test rejects the no-cointegration null under SUR but not under OLS; for Malaysia, the error-correction parameter becomes significant only under system estimation. Third, cross-equation λ -equality tests reveal a two-cluster structure slow-adjusting fossil fuel producers and fast-adjusting energy importers providing policy-relevant evidence on heterogeneous adjustment speeds that cannot be obtained from single-equation methods.

Several avenues for future research emerge. First, the delta method standard errors for the long-run coefficients $\hat{\theta}^{LR} = -\hat{\beta}/\hat{\lambda}$ are shown in Simulation 5 to be unreliable when $|\lambda|$ is small or T is moderate, because the ratio estimator converges at a super-consistent rate that is inconsistent with the $\sqrt{T}$ -consistent delta method. Developing Fieller-type confidence sets or targeted bootstrap procedures for the long-run coefficients within the SUR framework would improve inference in empirical applications. Second, when T/N is small, shrinkage or regularized estimators for Σ (e.g., Ledoit and Wolf, 2004) may further reduce estimation error and improve the finite-sample performance of the FGLS estimator. Third, while the present framework allows heterogeneous coefficients across equations, it assumes a common lag order within the ARDL specification; extending the framework to accommodate equation-specific lag orders selected by information criteria would increase flexibility. Fourth, the relationship between the SUR-ARDL approach and common-factor models for cross-sectional dependence (Pesaran, 2006; Bai and Kao, 2006) deserves investigation, particularly for settings where N is larger and the SUR approach becomes infeasible. Finally, extending the bootstrap bounds test to provide finite-sample-corrected critical values for the cross-equation Wald tests would complement the asymptotic chi-squared theory developed here.

Appendix A: Proof of Theorem 2

Proof. The proof proceeds in five steps. Step 1 decomposes the scaled FGLS estimator into its constituent components. Step 2 establishes the weak convergence of the weighted information matrix. Step 3 derives the weak convergence of the weighted score vector, including the elimination of the second-order bias through the ARDL lag structure. Step 4 shows that replacing the true covariance matrix Σ with its consistent estimator $\hat{\Sigma}$ does not affect the limiting distribution. Step 5 combines the results.

Step 1: Decomposition of the scaled FGLS estimator. By definition, the FGLS estimator is $\hat{\theta}_{SUR} = (\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Y}$, where $\hat{\Omega}^{-1} = \mathbf{I}_T \otimes \hat{\Sigma}^{-1}$. Since $\mathbf{Y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, we have:

$$\hat{\theta}_{SUR} - \boldsymbol{\theta} = (\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\Omega}^{-1}\boldsymbol{\varepsilon}. \quad (\text{A.1})$$

Pre-multiplying by $\mathbf{G}(T)^{-1}$ and inserting $\mathbf{G}(T)\mathbf{G}(T)^{-1} = \mathbf{I}_M$:

$$\begin{aligned} \mathbf{G}(T)^{-1}(\hat{\theta}_{SUR} - \boldsymbol{\theta}) &= (\mathbf{G}(T)\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z}\mathbf{G}(T))^{-1}\mathbf{G}(T)\mathbf{Z}'\hat{\Omega}^{-1}\boldsymbol{\varepsilon}. \quad (\text{A.2}) \\ &=:\check{\mathbf{A}}_T^{-1} \quad \quad \quad =:\check{\mathbf{b}}_T \end{aligned}$$

Using the Kronecker structure $\hat{\Omega}^{-1} = \mathbf{I}_T \otimes \hat{\Sigma}^{-1}$ and the block-diagonal form of $\mathbf{Z}_t = \text{diag}(\mathbf{x}_{1,t'}, \dots, \mathbf{x}_{N,t'})$, the two components decompose as sums over t :

$$\hat{\mathbf{A}}_T = \sum_{t=1}^T \mathbf{G}(T) \mathbf{Z}_t' \hat{\Sigma}^{-1} \mathbf{Z}_t \mathbf{G}(T), \quad (\text{A.3})$$

$$\hat{\mathbf{b}}_T = \sum_{t=1}^T \mathbf{G}(T) \mathbf{Z}_t' \hat{\Sigma}^{-1} \boldsymbol{\varepsilon}_t. \quad (\text{A.4})$$

The (i, j) -th block of $\hat{\mathbf{A}}_T$ (of dimension $m_i \times m_j$) is:

$$[\hat{\mathbf{A}}_T]_{i,j} = \hat{\sigma}^{ij} \sum_{t=1}^T \mathbf{G}_i(T) \mathbf{x}_{i,t} \mathbf{x}_{j,t}' \mathbf{G}_j(T), \quad i, j = 1, \dots, N, \quad (\text{A.5})$$

where $\hat{\sigma}^{ij}$ denotes the (i, j) -th element of $\hat{\Sigma}^{-1}$. The i -th block of $\hat{\mathbf{b}}_T$ (of dimension $m_i \times 1$) is:

$$[\hat{\mathbf{b}}_T]_i = \sum_{j=1}^N \hat{\sigma}^{ij} \sum_{t=1}^T \mathbf{G}_i(T) \mathbf{x}_{i,t} u_{j,t}, \quad i = 1, \dots, N. \quad (\text{A.6})$$

We now analyze the limits of $\hat{\mathbf{A}}_T$ and $\hat{\mathbf{b}}_T$ separately.

Step 2: Convergence of the weighted information matrix. Consider first the infeasible counterpart of $\hat{\mathbf{A}}_T$ with the true covariance matrix:

$$\mathbf{A}_T = \sum_{t=1}^T \mathbf{G}(T) \mathbf{Z}_t' \Sigma^{-1} \mathbf{Z}_t \mathbf{G}(T), \quad (\text{A.7})$$

with (i, j) -th block:

$$[\mathbf{A}_T]_{i,j} = \sigma^{ij} \sum_{t=1}^T \mathbf{G}_i(T) \mathbf{x}_{i,t} \mathbf{x}_{j,t}' \mathbf{G}_j(T). \quad (\text{A.8})$$

The behavior of $\sum_{t=1}^T \mathbf{G}_i(T) \mathbf{x}_{i,t} \mathbf{x}_{j,t}' \mathbf{G}_j(T)$ depends on the components of $\mathbf{x}_{i,t}$ and $\mathbf{x}_{j,t}$. Recall from equation (14) that $\mathbf{x}_{i,t} = (1, t, Y_{i,t-1}, X_{1,i,t-1}, \dots, X_{K,i,t-1}, \Delta X_{1,i,t}, \dots, \Delta X_{K,i,t})'$ and the scaling matrix $\mathbf{G}_i(T)$ normalizes each element according to its convergence rate.

For any two elements $x_{i,t}^{(a)}$ and $x_{j,t}^{(b)}$ of the scaled regressor vectors, the normalized sum $\sum_{t=1}^T g_a(T) x_{i,t}^{(a)} x_{j,t}^{(b)} g_b(T)$ converges by the following cases:

(a) If both elements are stationary (intercept, trend, or first differences), the standard law of large numbers yields convergence to a deterministic positive definite limit $\mathbf{Q}_{ij}^{(a,b)}$.

(b) If both elements are $I(1)$ level variables, the functional central limit theorem (FCLT) and the continuous mapping theorem (CMT) yield:

$$\sum_{t=1}^T T^{-1} x_{i,t}^{(a)} \cdot T^{-1} x_{j,t}^{(b)} = T^{-2} \sum_{t=1}^T x_{i,t}^{(a)} x_{j,t}^{(b)} \Rightarrow \int_0^1 B_a(r) B_b(r) dr, \quad (\text{A.9})$$

where $B_a(r)$ and $B_b(r)$ are the Brownian motions associated with the respective $I(1)$ processes.

(c) If one element is stationary and the other is $I(1)$, the cross-product converges to a stochastic integral of the form $\int_0^1 B(r)dr$ or $\int_0^1 r B(r)dr$ (for the trend component).

Collecting all blocks, and using the notation $\mathbf{J}_i(r)$ for the limit process of $\mathbf{G}_i(T)\mathbf{x}_{i,[rT]}$, we obtain:

$$\mathbf{A}_T \Rightarrow \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}\mathbf{J}(r)'dr =: \mathbf{A}, \quad (\text{A.10})$$

where $\mathbf{J}(r) = \text{diag}(\mathbf{J}_1(r), \dots, \mathbf{J}_N(r))$. The matrix $\mathbf{A}$ is almost surely positive definite under Assumption 3(ii) and (iv), which ensures that the limit of the information matrix is invertible.

Step 3: Convergence of the weighted score vector. Consider the infeasible score vector:

$$\mathbf{b}_T = \sum_{t=1}^T \mathbf{G}(T)\mathbf{Z}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t, \quad (\text{A.11})$$

with i -th block:

$$[\mathbf{b}_T]_i = \sum_{j=1}^N \sigma^{ij} \sum_{t=1}^T \mathbf{G}_i(T)\mathbf{x}_{i,t}u_{j,t}. \quad (\text{A.12})$$

The asymptotic behavior of $\sum_{t=1}^T \mathbf{G}_i(T)\mathbf{x}_{i,t}u_{j,t}$ involves the cross-product of the (possibly $I(1)$) regressors with the disturbances. This is the step where the ARDL lag structure plays its critical role.

For a general cointegrating regression $Y_{i,t} = \boldsymbol{\alpha}_i'\mathbf{Z}_{i,t} + u_{i,t}$ with $I(1)$ regressors $\mathbf{Z}_{i,t}$, the normalized cross-product $T^{-1} \sum_{t=1}^T \mathbf{Z}_{i,t} u_{i,t}$ converges to $\int_0^1 \mathbf{B}_Z(r)dB_u(r) + \boldsymbol{\Delta}_{Zu}$, where $\boldsymbol{\Delta}_{Zu} = \sum_{h=0}^{\infty} E(\mathbf{Z}_{i,0}u_{i,h})$ is a one-sided long-run covariance (second-order bias) term. This bias arises from the endogeneity of the regressors and the serial correlation in the errors. In the FM-OLS and FM-SUR frameworks, this bias is removed by the fully modified correction the subtraction of the estimated bias term $\hat{A}$ in Wagner et al. (2020, equations (13)-(14)) and the analogous correction in Lin and Reuvers (2025, equation (15)).

In the ARDL error-correction framework, the same endogeneity and serial correlation are handled parametrically through the lag structure. Specifically, the regressor vector $\mathbf{x}_{i,t}$ in the error-correction equation (12) includes the contemporaneous first differences $\Delta X_{k,i,t}$ as regressors. Under the weak exogeneity condition (Assumption 3(i)) and correct lag specification, the disturbance $u_{i,t}$ in the conditional model is orthogonal to the innovations driving the $I(1)$ regressors. Formally, partition the innovation vector $\boldsymbol{\zeta}_{i,t} = (u_{i,t}, v_{1,i,t}, \dots, v_{K,i,t})'$ where $v_{k,i,t} = \Delta X_{k,i,t} - E(\Delta X_{k,i,t}|\mathcal{F}_{t-1})$ are the innovations in the $I(1)$ regressor processes. The conditional error-correction model regresses $\Delta Y_{i,t}$ on $\mathbf{x}_{i,t}$, which includes both the lagged levels and the contemporaneous differences $\Delta X_{k,i,t}$. The inclusion of $\Delta X_{k,i,t}$ absorbs the contemporaneous correlation between $u_{i,t}$ and $v_{k,i,t}$, so that the conditional disturbance satisfies:

$$E(u_{i,t}|\Delta X_{k,i,t}, \mathcal{F}_{t-1}) = 0, \quad \text{for all } k = 1, \dots, K. \quad (\text{A.13})$$

Condition (A.13), together with the temporal independence of $u_{i,t}$ (Assumption 1(iii), ensured by correct lag specification), implies that the second-order bias term vanishes:

$$\boldsymbol{\Delta}_{xu,i} := \sum_{h=0}^{\infty} E(\mathbf{G}_i(T)\mathbf{x}_{i,t}u_{i,t+h}) = \mathbf{0}_{m_i}, \quad (\text{A.14})$$

because: (a) $E(u_{i,t}u_{i,t+h}) = 0$ for $h \neq 0$ by Assumption 1(iii); and (b) $E(\mathbf{x}_{i,t}u_{i,t}) = \mathbf{0}$ by the conditional orthogonality (A.13) and the law of iterated expectations.

With the bias term eliminated, the normalized cross-product involving $I(1)$ regressors converges to a pure stochastic integral:

$$\sum_{t=1}^T \mathbf{G}_i(T) \mathbf{x}_{i,t} u_{j,t} \Rightarrow \int_0^1 \mathbf{J}_i(r) dB_{u,j}^*(r), \quad (\text{A.15})$$

where $B_{u,j}^*(r)$ is the Brownian motion associated with $u_{j,t}$ that is independent of the stochastic regressors (conditional on the σ -field generated by the regressor processes). The superscript $*$ indicates this conditional independence the operational consequence of the ARDL lag structure having absorbed the endogeneity.

Substituting (A.15) into (A.12) and summing over j :

$$[\mathbf{b}_T]_i \Rightarrow \sum_{j=1}^N \sigma^{ij} \int_0^1 \mathbf{J}_i(r) dB_{u,j}^*(r) = \int_0^1 \mathbf{J}_i(r) [\boldsymbol{\Sigma}^{-1} d\mathbf{B}_u^*(r)]_i, \quad (\text{A.16})$$

where $[\cdot]_i$ selects the i -th component of the vector $\boldsymbol{\Sigma}^{-1} d\mathbf{B}_u^*(r)$ and $\mathbf{B}_u^*(r) = (B_{u,1}^*(r), \dots, B_{u,N}^*(r))'$. Stacking over $i = 1, \dots, N$:

$$\mathbf{b}_T \Rightarrow \int_0^1 \mathbf{J}(r) \boldsymbol{\Sigma}^{-1} d\mathbf{B}_u^*(r) =: \mathbf{b}. \quad (\text{A.17})$$

Step 4: Asymptotic equivalence of feasible and infeasible estimators. It remains to show that replacing $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\Sigma}}$ in the FGLS estimator does not affect the limiting distribution. We need to show:

$$\hat{\mathbf{A}}_T - \mathbf{A}_T \xrightarrow{p} \mathbf{0}, \quad \hat{\mathbf{b}}_T - \mathbf{b}_T \xrightarrow{p} \mathbf{0}. \quad (\text{A.18})$$

For the information matrix, the (i, j) -th block of $\hat{\mathbf{A}}_T - \mathbf{A}_T$ is:

$$[\hat{\mathbf{A}}_T - \mathbf{A}_T]_{i,j} = (\hat{\sigma}^{ij} - \sigma^{ij}) \sum_{t=1}^T \mathbf{G}_i(T) \mathbf{x}_{i,t} \mathbf{x}_{j,t}' \mathbf{G}_j(T). \quad (\text{A.19})$$

By Lemma 1, $\hat{\boldsymbol{\Sigma}}^{-1} \xrightarrow{p} \boldsymbol{\Sigma}^{-1}$, so $\hat{\sigma}^{ij} - \sigma^{ij} \xrightarrow{p} 0$ for all i, j . The sum $\sum_{t=1}^T \mathbf{G}_i(T) \mathbf{x}_{i,t} \mathbf{x}_{j,t}' \mathbf{G}_j(T) = O_p(1)$ by the analysis in Step 2 (it converges weakly to a finite random matrix). By Slutsky's theorem, the product of a quantity converging in probability to zero and a quantity that is $O_p(1)$ converges in probability to zero:

$$[\hat{\mathbf{A}}_T - \mathbf{A}_T]_{i,j} = o_p(1) \cdot O_p(1) = o_p(1). \quad (\text{A.20})$$

An identical argument applies to $\hat{\mathbf{b}}_T - \mathbf{b}_T$. For the i -th block:

$$[\hat{\mathbf{b}}_T - \mathbf{b}_T]_i = \sum_{j=1}^N (\hat{\sigma}^{ij} - \sigma^{ij}) \sum_{t=1}^T \mathbf{G}_i(T) \mathbf{x}_{i,t} u_{j,t} = \sum_{j=1}^N o_p(1) \cdot O_p(1) = o_p(1). \quad (\text{A.21})$$

Therefore $\hat{\mathbf{A}}_T = \mathbf{A}_T + o_p(1)$ and $\hat{\mathbf{b}}_T = \mathbf{b}_T + o_p(1)$.

Step 5: Combining the results. From Steps 2-4 and the continuous mapping theorem:

$$\mathbf{G}(T)^{-1} (\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta}) = \hat{\mathbf{A}}_T^{-1} \hat{\mathbf{b}}_T = (\mathbf{A}_T + o_p(1))^{-1} (\mathbf{b}_T + o_p(1)). \quad (\text{A.22})$$

Since $\mathbf{A}_T \Rightarrow \mathbf{A}$ (almost surely positive definite) and $\mathbf{b}_T \Rightarrow \mathbf{b}$ (Steps 2-3), and the inversion map is continuous at positive definite matrices, the CMT yields:

$$\mathbf{G}(T)^{-1}(\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta}) \Rightarrow \mathbf{A}^{-1}\mathbf{b} = \left(\int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}\mathbf{J}(r)'dr \right)^{-1} \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}d\mathbf{B}_u^*(r). \quad (\text{A.23})$$

The limit in (A.23) is a zero-mean Gaussian mixture distribution. To verify the zero-mean property, note that conditional on the σ -field $\mathcal{F}_v = \sigma(B_{v,1}(r), \dots, B_{v,N}(r), 0 \leq r \leq 1)$ generated by the stochastic regressor processes, $\mathbf{A}$ is measurable with respect to $\mathcal{F}_v$ and the stochastic integral $\mathbf{b} = \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}d\mathbf{B}_u^*(r)$ is a martingale with zero conditional mean (since $\mathbf{B}_u^*(r)$ is independent of $\mathcal{F}_v$ by condition (A.13)). Therefore $E[\mathbf{A}^{-1}\mathbf{b}|\mathcal{F}_v] = \mathbf{A}^{-1}E[\mathbf{b}|\mathcal{F}_v] = \mathbf{0}$, confirming the zero-mean property. The conditional distribution of $\mathbf{A}^{-1}\mathbf{b}$ given $\mathcal{F}_v$ is Gaussian with mean zero and covariance $\mathbf{A}^{-1}$, which establishes that the unconditional distribution is a Gaussian mixture as claimed. ■

Appendix B: Proof of Theorem 3

Proof. The proof is organized in four steps. Step 1 derives the limiting form of the Wald statistic under the null by extracting the relevant sub-block from the system limiting distribution. Step 2 characterizes the limit explicitly for the two polar cases all $I(0)$ and all $I(1)$. Step 3 establishes the monotonicity property by analyzing how the limit functional changes as individual regressors shift from $I(0)$ to $I(1)$. Step 4 verifies the dependence on $\boldsymbol{\Sigma}$ and the cross-equation regressors.

Step 1: Limiting form of the Wald statistic under $H_0^{(i)}$. From equation (140), the Wald statistic under the null is:

$$W_i^{\text{SUR}} = \hat{\boldsymbol{\theta}}_{\text{SUR}}' \tilde{\mathbf{R}}_i' \left[\tilde{\mathbf{R}}_i (\mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z})^{-1} \tilde{\mathbf{R}}_i' \right]^{-1} \tilde{\mathbf{R}}_i \hat{\boldsymbol{\theta}}_{\text{SUR}}, \quad (\text{B.1})$$

where $\tilde{\mathbf{R}}_i \in \mathbb{R}^{(K+1) \times M}$ selects the parameters $(\lambda_i, \beta_{1,i}, \dots, \beta_{K,i})$ from $\boldsymbol{\theta}$. Under $H_0^{(i)}$, we have $\tilde{\mathbf{R}}_i \boldsymbol{\theta} = \mathbf{0}_{K+1}$, so $\tilde{\mathbf{R}}_i \hat{\boldsymbol{\theta}}_{\text{SUR}} = \tilde{\mathbf{R}}_i (\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta})$.

Introduce the $(K+1) \times (K+1)$ scaling matrix $\mathbf{G}_{R,i}(T)$:

$$\mathbf{G}_{R,i}(T) = \text{diag}(g_{\lambda,i}(T), g_{\beta_{1,i}}(T), \dots, g_{\beta_{K,i}}(T)), \quad (\text{B.2})$$

where each scaling factor depends on the integration order of the corresponding variable. Since $\tilde{\mathbf{R}}_i$ is a selection matrix (each row has a single 1) and $\mathbf{G}(T)$ is diagonal, we have $\mathbf{G}_{R,i}(T) = \tilde{\mathbf{R}}_i \mathbf{G}(T) \tilde{\mathbf{R}}_i'$.

Now insert the identity $\mathbf{G}_{R,i}(T) \mathbf{G}_{R,i}(T)^{-1} = \mathbf{I}_{K+1}$ into the Wald statistic:

$$W_i^{\text{SUR}} = [\mathbf{G}_{R,i}(T)^{-1} \tilde{\mathbf{R}}_i (\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta})]' \left[\mathbf{G}_{R,i}(T)^{-1} \tilde{\mathbf{R}}_i (\mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z})^{-1} \tilde{\mathbf{R}}_i' \mathbf{G}_{R,i}(T)^{-1} \right]^{-1} [\mathbf{G}_{R,i}(T)^{-1} \tilde{\mathbf{R}}_i (\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta})]. \quad (\text{B.4})$$

Define the scaled restricted estimator and scaled restricted variance:

$$\hat{\boldsymbol{\xi}}_i := \mathbf{G}_{R,i}(T)^{-1} \tilde{\mathbf{R}}_i (\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta}), \quad (\text{B.5})$$

$$\hat{\mathbf{V}}_{R,i} := \mathbf{G}_{R,i}(T)^{-1} \tilde{\mathbf{R}}_i (\mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z})^{-1} \tilde{\mathbf{R}}_i' \mathbf{G}_{R,i}(T)^{-1}. \quad (\text{B.6})$$

So the Wald statistic is simply a quadratic form:

$$W_i^{\text{SUR}} = \hat{\boldsymbol{\xi}}_i' \hat{\mathbf{V}}_{R,i}^{-1} \hat{\boldsymbol{\xi}}_i. \quad (\text{B.7})$$

From the proof of Theorem 2 (Appendix A), the full scaled estimator satisfies:

$$\mathbf{G}(T)^{-1}(\widehat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta}) \Rightarrow \boldsymbol{\eta} := \mathbf{A}^{-1}\mathbf{b}, \quad (\text{B.8})$$

where $\mathbf{A} = \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}\mathbf{J}(r)'dr$ and $\mathbf{b} = \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}d\mathbf{B}_u^*(r)$. The scaled restricted estimator therefore converges:

$$\widehat{\boldsymbol{\xi}}_i = \widetilde{\mathbf{R}}_i\mathbf{G}(T) \cdot \mathbf{G}(T)^{-1}(\widehat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta}) \Rightarrow \widetilde{\mathbf{R}}_i\boldsymbol{\eta} =: \boldsymbol{\xi}_i, \quad (\text{B.9})$$

that is, $\boldsymbol{\xi}_i$ is the sub-vector of $\boldsymbol{\eta}$ corresponding to the restricted parameters. For the scaled variance, by the CMT and asymptotic equivalence:

$$\widehat{\mathbf{V}}_{R,i} \Rightarrow \mathbf{V}_{R,i} := \widetilde{\mathbf{R}}_i\mathbf{A}^{-1}\widetilde{\mathbf{R}}_i'. \quad (\text{B.10})$$

Combining (B.7), (B.9), and (B.10) via the CMT:

$$W_i^{\text{SUR}} \Rightarrow \boldsymbol{\xi}_i'\mathbf{V}_{R,i}^{-1}\boldsymbol{\xi}_i =: \mathcal{W}_i, \quad (\text{B.11})$$

and $F_i^{\text{SUR}} = W_i^{\text{SUR}}/(K+1) \Rightarrow \mathcal{W}_i/(K+1) =: \mathcal{F}_i$. The limit $\mathcal{W}_i$ is a quadratic form in $\boldsymbol{\xi}_i = \widetilde{\mathbf{R}}_i\mathbf{A}^{-1}\mathbf{b}$, normalized by $\mathbf{V}_{R,i} = \widetilde{\mathbf{R}}_i\mathbf{A}^{-1}\widetilde{\mathbf{R}}_i'$. The dependence on $\mathbf{d}_i$ enters through $\mathbf{J}(r)$; on $\boldsymbol{\Sigma}$ through $\mathbf{A}$ and $\mathbf{b}$; and on $\{\mathbf{X}_j\}_{j \neq i}$ through the cross-equation blocks of $\mathbf{A}$.

Step 2: Polar cases. We characterize $\mathcal{W}_i$ for the two extreme integration order configurations.

Case I: All restricted variables $I(0)$ ($\mathbf{d}_i = \mathbf{0}$). When $Y_{i,t}, X_{1,i,t}, \dots, X_{K,i,t}$ are all stationary, the sub-vector $\mathbf{J}_{R,i}(r)$ consists of deterministic functions rather than Brownian motions. The matrix $\mathbf{A}$ becomes deterministic, and the stochastic integral $\mathbf{b} = \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}d\mathbf{B}_u^*(r)$ reduces to a linear function of the terminal values $\mathbf{B}_u^*(1)$, which are jointly Gaussian. Therefore $\boldsymbol{\xi}_i = \widetilde{\mathbf{R}}_i\mathbf{A}^{-1}\mathbf{b}$ is Gaussian with zero mean and covariance $\mathbf{V}_{R,i} = \widetilde{\mathbf{R}}_i\mathbf{A}^{-1}\widetilde{\mathbf{R}}_i'$. The quadratic form in its own precision matrix is exactly chi-squared:

$$\mathcal{W}_i = \boldsymbol{\xi}_i'\mathbf{V}_{R,i}^{-1}\boldsymbol{\xi}_i \sim \chi^2(K+1), \quad \text{when } \mathbf{d}_i = \mathbf{0}. \quad (\text{B.12})$$

Case II: All restricted variables $I(1)$ ($\mathbf{d}_i = \mathbf{1}$). When all lagged level variables are $I(1)$, $\mathbf{J}_{R,i}(r)$ consists entirely of Brownian motions. The matrix $\mathbf{A}$ is random, so $\boldsymbol{\xi}_i$ is conditionally Gaussian (given $\mathcal{F}_v$, the σ -field of the stochastic regressors) with zero mean and covariance $\mathbf{V}_{R,i}(\omega)$ that varies with the realization. Conditionally, $\mathcal{W}_i|\mathcal{F}_v \sim \chi^2(K+1)$, but unconditionally $\mathcal{W}_i$ is a mixture of chi-squared distributions whose quantiles depend on the dimension K , the deterministic specification, and the cross-equation structure.

Step 3: Monotonicity of the critical values. We prove: if $\mathbf{d}_i^{(a)} \leq \mathbf{d}_i^{(b)}$ element-wise with at least one strict inequality, then $c_\alpha^{(a)} \leq c_\alpha^{(b)}$, where $c_\alpha^{(\cdot)}$ denotes the $(1-\alpha)$ -quantile of $\mathcal{F}_i$.

Without loss of generality, suppose $\mathbf{d}_i^{(a)}$ and $\mathbf{d}_i^{(b)}$ differ in the ℓ -th component only, with $d_\ell^{(a)} = 0$ (deterministic limit) and $d_\ell^{(b)} = 1$ (Brownian motion limit). The general case follows by induction.

Under both configurations, $\mathcal{W}_i|\mathcal{F}_v \sim \chi^2(K+1)$. The unconditional distribution is obtained by integrating over $\mathcal{F}_v$. The key difference is the amount of randomness in the conditioning variable. Under $\mathbf{d}_i^{(a)}$, the ℓ -th component of $\mathbf{J}_{R,i}(r)$ is a constant, so $\mathbf{A}$ and hence $\mathbf{V}_{R,i}$ have one fewer source of randomness. Under $\mathbf{d}_i^{(b)}$, this component is a Brownian motion, introducing additional randomness into $\mathbf{V}_{R,i}$.

We use the representation $\mathcal{W}_i \stackrel{d}{=} \sum_{m=1}^{K+1} \mu_m Z_m^2$, where $Z_m \stackrel{i.i.d.}{\sim} N(0,1)$ and $\mu_1, \dots, \mu_{K+1}$ are the eigenvalues of $\mathbf{V}_{R,i}$ under the normalization that $\boldsymbol{\xi}_i = \mathbf{V}_{R,i}^{1/2} \mathbf{Z}$ with $\mathbf{Z}|\mathcal{F}_v \sim N(\mathbf{0}, \mathbf{I})$. The critical value of $\mathcal{W}_i$ is determined by the distribution of the weighted sum $\sum \mu_m Z_m^2$.

Under $\mathbf{d}_i^{(a)}$, the eigenvalues $\{\mu_m\}$ have less variability because one source of randomness (the ℓ -th Brownian motion path) is absent. Under $\mathbf{d}_i^{(b)}$, the additional randomness in the ℓ -th Brownian motion component increases the dispersion of the eigenvalues.

To formalize, apply the law of total variance:

$$\text{Var}(\mathcal{W}_i) = E[\text{Var}(\mathcal{W}_i|\mathcal{F}_v)] + \text{Var}(E[\mathcal{W}_i|\mathcal{F}_v]). \quad (B.13)$$

Since $\mathcal{W}_i|\mathcal{F}_v \sim \chi^2(K+1)$, we have $\text{Var}(\mathcal{W}_i|\mathcal{F}_v) = 2(K+1)$ and $E[\mathcal{W}_i|\mathcal{F}_v] = K+1$ (both constant, independent of $\mathcal{F}_v$). Hence:

$$\text{Var}(\mathcal{W}_i) = 2(K+1) + \text{Var}(K+1) = 2(K+1). \quad (B.14)$$

This shows the variance of $\mathcal{W}_i$ is the same under both configurations (both equal $2(K+1)$). However, the distributions differ in their higher moments. The conditional chi-squared representation does not fully determine the unconditional distribution because the normalization $\boldsymbol{\xi}_i = \mathbf{V}_{R,i}^{1/2} \mathbf{Z}$ is valid only conditionally, and the unconditional distribution of $\mathcal{W}_i$ is obtained from the joint distribution of $(\mathbf{V}_{R,i}, \boldsymbol{\xi}_i)$, which depends on the correlation structure between $\mathbf{A}$ and $\mathbf{b}$.

A more direct argument uses stochastic ordering. Under $\mathbf{d}_i^{(a)}$, the Wald statistic has the form:

$$\mathcal{W}_i^{(a)} = \boldsymbol{\xi}_i^{(a)'} [\tilde{\mathbf{R}}_i \mathbf{A}^{(a)-1} \tilde{\mathbf{R}}_i']^{-1} \boldsymbol{\xi}_i^{(a)}, \quad (B.15)$$

where $\mathbf{A}^{(a)}$ and $\boldsymbol{\xi}_i^{(a)}$ are computed with the ℓ -th component of $\mathbf{J}_{R,i}(r)$ set to a constant c_ℓ . Under $\mathbf{d}_i^{(b)}$, the same expression holds with c_ℓ replaced by $B_\ell(r)$. The effect of this replacement is to introduce additional randomness in the information matrix $\mathbf{A}$, which in turn affects the tails of $\mathcal{W}_i$ through the correlation between $\mathbf{A}^{-1}$ and $\mathbf{b}$.

The critical values are determined by the upper tail behavior. In the $I(1)$ case, the stochastic integral $\int_0^1 B_\ell(r) dB_u^*(r)$ has heavier tails than the deterministic analog $c_\ell B_u^*(1)$, because the former involves a product of two correlated stochastic processes (even after conditioning). This tail inflation shifts the upper quantiles of $\mathcal{W}_i^{(b)}$ to the right relative to $\mathcal{W}_i^{(a)}$, yielding $c_\alpha^{(b)} \geq c_\alpha^{(a)}$.

The bounds testing decision rule follows directly: define $c_L(\alpha) := c_\alpha^{(0)}$ (the all- $I(0)$ critical value, which equals $\chi_{1-\alpha}^2(K+1)/(K+1)$) and $c_U(\alpha) := c_\alpha^{(1)}$ (the all- $I(1)$ critical value). By monotonicity, $c_L(\alpha) \leq c_\alpha^{(\mathbf{d}_i)} \leq c_U(\alpha)$ for all $\mathbf{d}_i \in \{0,1\}^{K+1}$, which validates the three-way decision rule. ■

Step 4: Dependence on $\boldsymbol{\Sigma}$ and the cross-equation regressors. We verify that $\mathcal{F}_i$ depends on $\boldsymbol{\Sigma}$ and $\{\mathbf{X}_j\}_{j \neq i}$ through the cross-equation blocks of $\mathbf{A}$.

The matrix $\mathbf{A} = \int_0^1 \mathbf{J}(r) \boldsymbol{\Sigma}^{-1} \mathbf{J}(r)' dr$ has (i,j) -th block $[\mathbf{A}]_{i,j} = \sigma^{ij} \int_0^1 \mathbf{J}_i(r) \mathbf{J}_j(r)' dr$. The sub-vector $\boldsymbol{\xi}_i = \tilde{\mathbf{R}}_i \mathbf{A}^{-1} \mathbf{b}$ involves the i -th row block of $\mathbf{A}^{-1}$, which depends on all blocks of $\mathbf{A}$ through the block-inversion formula. For $N = 2$:

$$[\mathbf{A}^{-1}]_{1,1} = ([\mathbf{A}]_{1,1} - [\mathbf{A}]_{1,2} [\mathbf{A}]_{2,2}^{-1} [\mathbf{A}]_{2,1})^{-1}, \quad (B.16)$$

which depends on Σ (through σ^{ij}) and on $\mathbf{J}_2(r)$ (through $[\mathbf{A}]_{1,2}$, $[\mathbf{A}]_{2,2}$, $[\mathbf{A}]_{2,1}$). The off-diagonal correction vanishes under the three conditions of Proposition 3, reducing the SUR distribution to the single-equation distribution.

Appendix C: Proof of Theorem 4

Theorem 4. Under Assumptions 1-3 and the rate condition (156), the Wald statistic $W = (\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{SUR}} - \mathbf{r})' [\mathbf{R}(\mathbf{Z}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z})^{-1}\mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{SUR}} - \mathbf{r}) \xrightarrow{d} \chi^2(s)$ as $T \rightarrow \infty$.

Proof. The proof has three steps: normalize the restriction, apply the Gaussian mixture limit, and verify the chi-squared quadratic form.

Step 1: Normalization. Let $\mathbf{G}_R(T) \in \mathbb{R}^{s \times s}$ be the sequence satisfying the rate condition:

$$\lim_{T \rightarrow \infty} \mathbf{G}_R(T) \mathbf{R} \mathbf{G}(T) = \mathbf{R}^*, \quad \text{rank}(\mathbf{R}^*) = s. \quad (\text{C.1})$$

Insert $\mathbf{G}_R \mathbf{G}_R^{-1} = \mathbf{I}_s$ into W :

$$W = \underbrace{[\mathbf{G}_R(\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{SUR}} - \mathbf{r})]'}_{\check{\mathbf{q}}'} \underbrace{[\mathbf{G}_R \mathbf{R}(\mathbf{Z}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z})^{-1}\mathbf{R}'\mathbf{G}_{R'}]^{-1}}_{\check{\mathbf{Q}}^{-1}} \underbrace{[\mathbf{G}_R(\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{SUR}} - \mathbf{r})]}_{\check{\mathbf{q}}} = \check{\mathbf{q}}' \check{\mathbf{Q}}^{-1} \check{\mathbf{q}}. \quad (\text{C.2})$$

Under $H_0: \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$:

$$\check{\mathbf{q}} = \mathbf{G}_R \mathbf{R}(\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta}) = \underbrace{\mathbf{G}_R \mathbf{R} \mathbf{G}}_{\rightarrow \check{\mathbf{R}}^*} \cdot \underbrace{\mathbf{G}^{-1}(\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta})}_{\Rightarrow \check{\boldsymbol{\eta}}}, \quad (\text{C.3})$$

where $\mathbf{G} = \mathbf{G}(T)$ and $\boldsymbol{\eta} = \mathbf{A}^{-1}\mathbf{b}$ from Theorem 2.

Step 2: Weak convergence. By Theorem 2 (Appendix A, equation (A.23)):

$$\boldsymbol{\eta} = \mathbf{G}^{-1}(\hat{\boldsymbol{\theta}}_{\text{SUR}} - \boldsymbol{\theta}) \Rightarrow \mathbf{A}^{-1}\mathbf{b}, \quad (\text{C.4})$$

$$\mathbf{A} = \int_0^1 \mathbf{J}(r) \boldsymbol{\Sigma}^{-1} \mathbf{J}(r)' dr, \quad \mathbf{b} = \int_0^1 \mathbf{J}(r) \boldsymbol{\Sigma}^{-1} d\mathbf{B}_u^*(r). \quad (\text{C.5})$$

From (C.1) and (C.3):

$$\check{\mathbf{q}} \Rightarrow \mathbf{R}^* \mathbf{A}^{-1} \mathbf{b} =: \mathbf{q}. \quad (\text{C.6})$$

For $\check{\mathbf{Q}}$, expand using $\hat{\boldsymbol{\Omega}}^{-1} = \mathbf{I}_T \otimes \hat{\boldsymbol{\Sigma}}^{-1}$:

$$\check{\mathbf{Q}} = \mathbf{G}_R \mathbf{R}(\mathbf{Z}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z})^{-1} \mathbf{R}' \mathbf{G}_{R'} = \mathbf{G}_R \mathbf{R} \mathbf{G} \cdot \underbrace{(\mathbf{G} \mathbf{Z}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{Z} \mathbf{G})^{-1}}_{\check{\mathbf{A}}_T^{-1}} \cdot \mathbf{G} \mathbf{R}' \mathbf{G}_{R'}. \quad (\text{C.7})$$

By Appendix A (Step 2 and Step 4): $\hat{\mathbf{A}}_T \Rightarrow \mathbf{A}$. Combined with (C.1):

$$\check{\mathbf{Q}} \Rightarrow \mathbf{R}^* \mathbf{A}^{-1} \mathbf{R}^{*'} =: \mathbf{Q}. \quad (\text{C.8})$$

By the CMT:

$$W = \check{\mathbf{q}}' \check{\mathbf{Q}}^{-1} \check{\mathbf{q}} \Rightarrow \mathbf{q}' \mathbf{Q}^{-1} \mathbf{q}. \quad (\text{C.9})$$

Step 3: Chi-squared identification. Substitute $\mathbf{q} = \mathbf{R}^* \mathbf{A}^{-1} \mathbf{b}$ and $\mathbf{Q} = \mathbf{R}^* \mathbf{A}^{-1} \mathbf{R}^{*'}$:

$$\mathbf{q}' \mathbf{Q}^{-1} \mathbf{q} = \mathbf{b}' \mathbf{A}^{-1} \mathbf{R}^{*'} (\mathbf{R}^* \mathbf{A}^{-1} \mathbf{R}^{*'})^{-1} \mathbf{R}^* \mathbf{A}^{-1} \mathbf{b}. \quad (\text{C.10})$$

Define $\mathbf{P} := \mathbf{A}^{-1/2} \mathbf{R}^{*'} (\mathbf{R}^* \mathbf{A}^{-1} \mathbf{R}^{*'})^{-1} \mathbf{R}^* \mathbf{A}^{-1/2}$. Then:

$$\mathbf{q}'\mathbf{Q}^{-1}\mathbf{q} = (\mathbf{A}^{-1/2}\mathbf{b})'\mathbf{P}(\mathbf{A}^{-1/2}\mathbf{b}). \quad (C.11)$$

Verify that $\mathbf{P}$ is an idempotent projection of rank s :

$$\mathbf{P}^2 = \mathbf{A}^{-1/2}\mathbf{R}^{*'}(\mathbf{R}^*\mathbf{A}^{-1}\mathbf{R}^{*'})^{-1}\mathbf{R}^*\mathbf{A}^{-1}\mathbf{R}^{*'}(\mathbf{R}^*\mathbf{A}^{-1}\mathbf{R}^{*'})^{-1}\mathbf{R}^*\mathbf{A}^{-1/2} = \mathbf{P}, \quad (C.12)$$

$$\mathbf{I}_s$$

$$\text{tr}(\mathbf{P}) = \text{tr}[(\mathbf{R}^*\mathbf{A}^{-1}\mathbf{R}^{*'})^{-1}\mathbf{R}^*\mathbf{A}^{-1}\mathbf{R}^{*'}] = \text{tr}(\mathbf{I}_s) = s. \quad (C.13)$$

Now condition on $\mathcal{F}_v = \sigma(\mathbf{B}_v(r), 0 \leq r \leq 1)$. Conditional on $\mathcal{F}_v$, $\mathbf{A}$ is measurable and hence $\mathbf{P}$ is non-random. Define $\boldsymbol{\zeta} := \mathbf{A}^{-1/2}\mathbf{b}$. From Appendix A (equation (A.17) and the zero-mean property):

$$\boldsymbol{\zeta} \mid \mathcal{F}_v \sim N\left(\mathbf{0}, \mathbf{A}^{-1/2}\left(\int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{J}(r)'dr\right)\mathbf{A}^{-1/2}\right). \quad (C.14)$$

Simplify the conditional covariance. By the Itô isometry for the stochastic integral $\mathbf{b} = \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}d\mathbf{B}_u^*(r)$ with $\text{Var}(d\mathbf{B}_u^*(r)) = \boldsymbol{\Sigma}dr$:

$$\text{Var}(\mathbf{b} \mid \mathcal{F}_v) = \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{J}(r)'dr = \int_0^1 \mathbf{J}(r)\boldsymbol{\Sigma}^{-1}\mathbf{J}(r)'dr = \mathbf{A}. \quad (C.15)$$

Therefore:

$$\text{Var}(\boldsymbol{\zeta} \mid \mathcal{F}_v) = \mathbf{A}^{-1/2}\mathbf{A}\mathbf{A}^{-1/2} = \mathbf{I}_M. \quad (C.16)$$

So $\boldsymbol{\zeta} \mid \mathcal{F}_v \sim N(\mathbf{0}, \mathbf{I}_M)$. The quadratic form (C.11) becomes:

$$\mathbf{q}'\mathbf{Q}^{-1}\mathbf{q} = \boldsymbol{\zeta}'\mathbf{P}\boldsymbol{\zeta} \mid \mathcal{F}_v \sim \chi^2(\text{rank}(\mathbf{P})) = \chi^2(s). \quad (C.17)$$

Since $\boldsymbol{\zeta}'\mathbf{P}\boldsymbol{\zeta} \mid \mathcal{F}_v \sim \chi^2(s)$ for every realization of $\mathcal{F}_v$, the unconditional distribution is also $\chi^2(s)$:

$$W \Rightarrow \boldsymbol{\zeta}'\mathbf{P}\boldsymbol{\zeta} \sim \chi^2(s). \quad (C.18)$$

This completes the proof. ■

References

- Apergis, N., & Payne, J. E. (2010). Renewable energy consumption and economic growth: Evidence from a panel of OECD countries. *Energy Policy*, 38(1), 656-660. <https://doi.org/10.1016/j.enpol.2009.09.002>
- Apergis, N., & Payne, J. E. (2011). The renewable energy consumption-growth nexus in Central America. *Applied Energy*, 88(1), 343-347. <https://doi.org/10.1016/j.apenergy.2010.07.013>
- Appiah, M., Gyamfi, B. A., Adebayo, T. S., & Bekun, F. V. (2022). Do financial development, foreign direct investment, and economic growth enhance industrial development? Fresh evidence from Sub-Saharan Africa. *Portuguese Economic Journal*, 22(2), 203-227. <https://doi.org/10.1007/s10258-022-00207-0>
- Bai, J., & Kao, C. (2006). On the estimation and inference of a panel cointegration model with cross-sectional dependence. In B. H. Baltagi (Ed.), *Panel data econometrics: Theoretical contributions and empirical applications* (Vol. 274, pp. 3-30). Elsevier. [https://doi.org/10.1016/S0573-8555\(06\)74001-9](https://doi.org/10.1016/S0573-8555(06)74001-9)
- Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1), 1-22. <https://doi.org/10.1002/jae.659>

- Banerjee, A., Dolado, J. J., Galbraith, J. W., & Hendry, D. F. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*. Oxford University Press.
- Banerjee, A., Dolado, J. J., & Mestre, R. (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis*, 19(3), 267-283. <https://doi.org/10.1111/1467-9892.00091>
- Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, 47(1), 239-253. <https://doi.org/10.2307/2297111>
- Choi, I., & Saikkonen, P. (2010). Tests for nonlinear cointegration. *Econometric Theory*, 26(3), 682-709. <https://doi.org/10.1017/S0266466609990296>
- Energy Institute. (2025). *Statistical review of world energy 2025*. <https://www.energyinst.org/statistical-review>
- Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2), 251-276. <https://doi.org/10.2307/1913236>
- Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society: Series B*, 16(2), 175-185. <https://doi.org/10.1111/j.2517-6161.1954.tb00159.x>
- Gallant, A. R. (1975). Seemingly unrelated nonlinear regressions. *Journal of Econometrics*, 3(1), 35-50. [https://doi.org/10.1016/0304-4076\(75\)90064-0](https://doi.org/10.1016/0304-4076(75)90064-0)
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6), 1551-1580. <https://doi.org/10.2307/2938278>
- Kahouli, B. (2019). Does static and dynamic relationship between economic growth and energy consumption exist in OECD countries? *Energy Reports*, 5, 104-116. <https://doi.org/10.1016/j.egyr.2018.12.006>
- Kripfganz, S., & Schneider, D. C. (2023). *ardl*: Estimating autoregressive distributed lag and equilibrium correction models. *Stata Journal*, 23(4), 983-1019. <https://doi.org/10.1177/1536867X231212434>
- Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365-411. [https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4)
- Lin, Y., & Reuvers, H. (2025). FM-GLS estimation and inference for seemingly unrelated cointegrating polynomial regressions. *Oxford Bulletin of Economics and Statistics*. Advance online publication. <https://doi.org/10.1111/obes.70030>
- Mark, N. C., Ogaki, M., & Sul, D. (2005). Dynamic seemingly unrelated cointegrating regressions. *Review of Economic Studies*, 72(3), 797-820. <https://doi.org/10.1111/j.1467-937X.2005.00352.x>
- Moon, H. R. (1999). A note on fully-modified estimation of seemingly unrelated regressions models with integrated regressors. *Economics Letters*, 65(1), 25-31. [https://doi.org/10.1016/S0165-1765\(99\)00114-2](https://doi.org/10.1016/S0165-1765(99)00114-2)
- Moon, H. R., & Perron, B. (2005). Efficient estimation of the seemingly unrelated regression cointegration model and testing for purchasing power parity. *Econometric Reviews*, 23(4), 293-323. <https://doi.org/10.1081/ETC-200040777>

- Narayan, P. K. (2005). The saving and investment nexus for China: Evidence from cointegration tests. *Applied Economics*, 37(17), 1979-1990. <https://doi.org/10.1080/00036840500278103>
- Ozcan, B., & Ozturk, I. (2019). Renewable energy consumption-economic growth nexus in emerging countries: A bootstrap panel causality test. *Renewable and Sustainable Energy Reviews*, 104, 30-37. <https://doi.org/10.1016/j.rser.2019.01.020>
- Park, J. Y., & Ogaki, M. (1991). Seemingly unrelated canonical cointegrating regressions (Rochester Center for Economic Research Working Paper No. 280). University of Rochester.
- Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels (Cambridge Working Papers in Economics No. 0435). University of Cambridge.
- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4), 967-1012. <https://doi.org/10.1111/j.1468-0262.2006.00692.x>
- Pesaran, M. H., & Shin, Y. (1999). An autoregressive distributed lag modelling approach to cointegration analysis. In S. Strom (Ed.), *Econometrics and economic theory in the 20th century: The Ragnar Frisch centennial symposium* (pp. 371-413). Cambridge University Press. <https://doi.org/10.1017/CCOL521633230.011>
- Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, 16(3), 289-326. <https://doi.org/10.1002/jae.616>
- Pesaran, M. H., & Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, 142(1), 50-93. <https://doi.org/10.1016/j.jeconom.2007.05.010>
- Phillips, P. C. B., & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, 57(1), 99-125. <https://doi.org/10.2307/2297545>
- Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, 55(5), 1035-1056. <https://doi.org/10.2307/1911260>
- Stock, J. H., & Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61(4), 783-820. <https://doi.org/10.2307/2951763>
- Tuna, G., & Tuna, V. E. (2019). The asymmetric causal relationship between renewable and non-renewable energy consumption and economic growth in the ASEAN-5 countries. *Resources Policy*, 62, 114-124. <https://doi.org/10.1016/j.resourpol.2019.03.010>
- Wagner, M., Grabarczyk, P., & Hong, S. H. (2020). Fully modified OLS estimation and inference for seemingly unrelated cointegrating polynomial regressions and the environmental Kuznets curve for carbon dioxide emissions. *Journal of Econometrics*, 214(1), 216-255. <https://doi.org/10.1016/j.jeconom.2019.05.012>
- Wagner, M., & Hong, S. H. (2016). Cointegrating polynomial regressions: Fully modified OLS estimation and inference. *Econometric Theory*, 32(6), 1289-1315. <https://doi.org/10.1017/S0266466615000213>
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817-838. <https://doi.org/10.2307/1912934>
- World Bank. (2025). *World development indicators*. <https://databank.worldbank.org/source/world-development-indicators>

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298), 348-368. <https://doi.org/10.1080/01621459.1962.10480664>

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that helped improve the quality of this manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

Fumitaka Furuoka: Conceptualization, Methodology, Supervision, Writing - review & editing.

Muhammad Afnan Arif: Data curation, Formal analysis, Software, Visualization, Writing - original draft.

Ethical Approval

This is not applicable.

Consent to Participate

This is not applicable.

Consent to Publish

This is not applicable.

Competing Interests

The authors declare that they have no competing interests.

Data Availability Statement

The data that support the findings of this study are publicly available. All data processing and analysis codes are available from the corresponding author upon reasonable request.