Methodological Details of the Error Correction Model with MARS

José Mauricio Gómez Julián

Table of Contents

1 Econometric Framework for Cointegration and Error Correction Analysis

This document details the methodology implemented to evaluate long-run equilibrium relationships and short-run adjustment dynamics between production and circulation variables through a hybrid approach that combines classical econometric cointegration techniques with modern non-parametric statistical learning methods. This methodology integrates the rigor of cointegration analysis with the flexibility of Multivariate Adaptive Regression Splines (MARS) to capture non-linearities in the error correction mechanism.

1.1 Determination of Integration Order I(1)

1.1.1 Theoretical Foundation

Cointegration analysis requires that the involved time series be integrated of order one, denoted as $I(1)$. An $I(1)$ series is non-stationary in levels but becomes stationary after first differencing. This property is fundamental because only $I(1)$ processes can maintain long-run equilibrium relationships without diverging indefinitely.

1.1.2 Verification Protocol

Let $z_t \in \{Y_t, X_t\}$ be a time series. We determine that $z_t \sim I(1)$ through the following two-stage protocol:

Stage 1: Non-stationarity in levels

We apply the Augmented Dickey-Fuller (ADF) test with deterministic components:

$$\Delta z_t = \mu + \beta t + \rho z_{t-1} + \sum_{i=1}^{p} \psi_i \, \Delta z_{t-i} + \varepsilon_t$$

where we test $H_0: \rho = 0$ (unit root). We require failure to reject H_0 at 10% significance for both drift and trend specifications, indicating robust non-stationarity across deterministic specifications.

Stage 2: Stationarity in first differences

We test the first difference without deterministic components:

$$\Delta^2 z_t = \rho' \Delta z_{t-1} + \sum_{i=1}^{p} \psi'_i \Delta^2 z_{t-i} + \eta_t$$

We require rejection of $H_0: \rho' = 0$ at 10%, confirming stationarity after differencing.

1.1.3 Justification of Significance Level

We use 10% in $I(1)$ tests as a conservative pre-filter to reduce the False Negative Rate (FNR)[^1]. This more permissive threshold at the initial stage is compensated by stricter filters in subsequent stages (cointegration at 5%, unidirectional ECM at 5%, out-of-sample validation).

1.2 Cointegration Analysis

1.2.1 Dual Approach: Engle-Granger and Johansen

We implement two complementary cointegration methodologies, recognizing that each has specific strengths:

1.2.1.1 Engle-Granger Procedure with Phillips-Ouliaris

Step 1: Cointegration regression

We estimate the long-run relationship:

$$Y_t = \alpha + \beta X_t + u_t$$

where α and β are the parameters of the cointegrating vector $(1, -\alpha, -\beta)'$.

Step 2: Residual stationarity test

We apply two tests on $\hat{u}_t$:

- **ADF without deterministics**: Tests unit root in residuals at level $p < 0.05$
- **Phillips-Ouliaris (PO)**: Test robust to endogeneity and serial correlation

Step 3: “Either” decision rule

We accept cointegration if **either** test (ADF or PO) validates at 5%. This rule reduces FNR in the presence of structural breaks, compensated by stricter subsequent validation.

1.2.1.2 ***Johansen Trace Test***

We specify a VAR of order K (selected by BIC criterion via `VARselect`) and transform it to its VECM representation:

$$\Delta Z_t = \Pi Z_{t-1} + \sum_{i=1}^{K-1} \Gamma_i \Delta Z_{t-i} + \Psi D_t + \varepsilon_t$$

where: - $Z_t = (Y_t, X_t)'$ is the variable vector - $\Pi = \alpha\beta'$ is the long-run impact matrix - D_t contains deterministics (constant and/or trend)

We test $H_0: r = 0$ (no cointegrating vectors) using the trace statistic. We test with specifications {const,trend} and accept cointegration if the statistic exceeds the critical value at 5% for any specification.

1.2.2 Justification of the "Either" Rule

The "either" rule (EG **or** Johansen) instead of "both" (EG **and** Johansen) is justified by:

1. **Robustness to regime changes**: Different tests may be sensitive to different types of structural breaks
2. **Compensation with subsequent filters**: The ECM with $\lambda < 0$ and out-of-sample validation filter false positives
3. **Empirical evidence**: Monte Carlo tests show lower FNR without substantial FPR inflation when combined with predictive validation

1.3 Error Correction Model (ECM)

1.3.1 Linear ECM Specification

Given the cointegrating vector (α, β) from Engle-Granger, we construct:

$$\text{ECM1}_t = Y_{t-1} - \alpha - \beta X_{t-1}$$

This term captures the deviation from long-run equilibrium at $t-1$. The complete ECM model is:

$$\Delta Y_t = \lambda \cdot \text{ECM1}_t + \sum_{i=1}^{L} \phi_i \Delta Y_{t-i} + \sum_{i=1}^{L} \gamma_i \Delta X_{t-i} + \varepsilon_t$$

where: - $\lambda < 0$ is the speed of adjustment toward equilibrium - L is the number of lags in differences - ϕ_i, γ_i capture short-run dynamics

1.3.2 Optimal Lag Selection

We implement an automatic selection procedure:

1. **Search over** $L \in \{1,2,\ldots,L_{\max}\}$ with $L_{\max} = 4$ by default

2. **Primary criterion**: BIC for parsimony
3. **White noise constraint**: If the model with lowest BIC fails Ljung-Box ($p \leq 0.05$ with 12 lags), we select the model with lowest BIC that passes the test
4. **Optional guidance**: $L \approx \max(1, K-1)$ where K is the VAR order, though not binding

1.3.3 Unidirectional Test with HAC Errors

1.3.3.1 Hypothesis and Justification

We test: - $H_0: \lambda \geq 0$ (no correction or divergence) - $H_1: \lambda < 0$ (correction toward equilibrium exists)

The unidirectional test is fundamental because $\lambda > 0$ would imply divergence from equilibrium, which is economically incoherent and would violate the system's stability condition.

1.3.3.2 Robust Inference

We employ HAC (Heteroskedasticity and Autocorrelation Consistent) standard errors using the Newey-West estimator:

$$\hat{V}_{NW} = \hat{\Omega}_0 + \sum_{j=1}^{m} w_j \left(\hat{\Omega}_j + \hat{\Omega}'_j\right)$$

where $w_j = 1 - j/(m+1)$ are Bartlett weights and m is automatically selected. The robust t statistic is:

$$t = \frac{\hat{\lambda}}{\sqrt{\hat{V}_{NW,\lambda\lambda}}}$$

with one-sided p-value $p = P(T \leq t | H_0)$. We reject if $p < 0.05$ **and** $\hat{\lambda} < 0$.

1.4 Non-Linear Extension with MARS

1.4.1 Economic Motivation

Economic relationships frequently exhibit non-linearities: - **Threshold effects**: Different responses depending on variable levels - **Asymmetries**: Different adjustments for positive vs. negative deviations - **Regime changes**: Parameters varying with economic context

MARS captures these characteristics through adaptive basis functions without requiring *a priori* specification of the functional form.

1.4.2 MARS-ECM Model Specification

The non-linear model is specified as:

$$\Delta Y_t = f(\text{ECM1}_t, \Delta X_t, \Delta Y_{t-1}, \Delta X_{t-1}, \Delta Y_{t-2}) + \eta_t$$

where $f(\cdot)$ is approximated by MARS as:

$$f(\mathbf{x}) = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{k=1}^{K_m} h_{km}\left(x_{v(k,m)}\right)$$

with: - h_{km} are hinge functions: $\max(0, x - c)$ or $\max(0, c - x)$ - M is the number of basis functions (controlled by nk) - K_m is the interaction degree (controlled by degree)

1.4.3 Hyperparameter Configuration

The search grid specifies: - **degree** $\in \{1,2\}$: Controls interactions (1 = additive, 2 = allows interactions) - **nk** $\in \{15,25,35,50,65\}$: Maximum number of terms before pruning

This grid balances flexibility with overfitting risk, expandable with more historical data.

1.5 Temporal Cross-Validation

1.5.1 Rolling-Origin with Sliding Window

We implement temporal cross-validation respecting causality through rolling-origin with sliding window:

1.5.1.1 Configuration Parameters

- **Initial size**: $\max(40, 0.80 \times n)$ observations
- **Test horizon**: 12 months (annual forecast)
- **Step between origins**: 12 months (avoids overlap)
- **Window type**: Sliding (constant size) vs Expanding (cumulative)

1.5.1.2 Sliding Window Justification

The sliding window maintains “comparable memory” between folds and is more sensitive to regime changes than the expanding window. This is crucial for economic series with non-constant parameters in a broad sense. Empirical evidence shows that sliding:

1. Preserves global stability (same number of robust models)
2. Increases local sensitivity (detects more regime-dependent relationships)
3. Improves adaptation to recent dynamics

1.5.2 Nested Cross-Validation

For hyperparameter selection without contaminating evaluation:

1. **Outer level**: Rolling-origin for performance evaluation
2. **Inner level**: Within each outer train, additional rolling-origin with:
   - Initial: 60% of outer train
   - Inner test: 6 months
   - Inner step: 3 months

This structure avoids *data snooping*[^2] and provides unbiased estimates of generalization error.

1.6 Evaluation Metrics

1.6.1 Scale-Dependent Metrics

- **RMSE**: $\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$
- **MAE**: $\frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$

1.6.2 Relative Metrics

- **MAPE**: $\frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right|$ (protected for $|Y_t| > \epsilon$)
- **sMAPE**: $\frac{100}{n}\sum_{t=1}^{n}\frac{2|Y_t-\hat{Y}_t|}{|Y_t|+|\hat{Y}_t|}$
- **Theil's U**: $\frac{\sqrt{\overline{(Y_t-\hat{Y}_t)^2}}}{\sqrt{\overline{Y_t^2}}}$

1.6.3 Explanatory Metric

- **Protected R^2**:

$$R^2 = \begin{cases} 1 - \dfrac{\sum\left(Y_t - \hat{Y}_t\right)^2}{\sum(Y_t - \bar{Y})^2} & \text{if } SST > \epsilon \\ \text{NA} & \text{if } SST \leq \epsilon \end{cases}$$

1.6.4 Theil Decomposition

The MSE decomposes into interpretable components:

$$\text{MSE} = \underbrace{\left(\bar{\hat{Y}} - \bar{Y}\right)^2}_{\text{Bias}^2} + \underbrace{(\sigma_{\hat{Y}} - \sigma_Y)^2}_{\text{Var. differential}} + \underbrace{2\sigma_{\hat{Y}}\sigma_Y\left(1 - \rho_{\hat{Y},Y}\right)}_{\text{Imperfect covariance}}$$

The proportions (bias_prop, var_prop, cov_prop) diagnose the primary source of predictive error.

1.7 Temporal Stability Criteria

1.7.1 Support Metric

We define support as:

$$\text{support} = \frac{\text{folds\_proceed}}{\text{folds}}$$

where `folds_proceed` counts folds that pass all econometric filters and have acceptable predictive performance.

1.7.2 Validation Thresholds

- **Strict threshold**: support ≥ 0.75 **and** folds_proceed ≥ 5
- **Moderate threshold**: support ≥ 0.60 **and** folds_proceed ≥ 3

The absolute minimum requirement prevents "false robustness" from small denominators.

1.7.3 Stability-Adjusted Metrics

- **Stable R^2**: $R^2_{\text{stab}} = R^2 \times \text{support}$
- **Stable U**: $U_{\text{stab}} = \frac{U}{\text{support}}$ (penalizes instability)

These metrics integrate predictive performance with temporal consistency, favoring "good and constant" models over "sometimes excellent" ones.

1.8 Computational Implementation

1.8.1 Multi-Level Parallelization

The parallelization architecture operates at two levels:

1. **Pair level**: Each combination $(X \to Y)$ is processed in an independent worker
2. **BLAS control**: `blas_set_num_threads(1)` is set to avoid CPU over-subscription

Workers are independent R processes (not threads) coordinated by `future::multisession`, each with its own memory. The seed `future.seed=TRUE` ensures reproducibility in parallel.

1.8.2 Progress Management

The `progressr` package provides real-time feedback without interfering with parallelization, crucial for long executions (84 pairs × multiple folds × nested validation).

1.8.3 Computational Complexity

Total complexity is:

$$\mathcal{O}\left(N_{\text{pairs}} \times F_{\text{outer}} \times F_{\text{inner}} \times G \times C_{\text{model}}\right)$$

where: - $N_{\text{pairs}} = 84$ (6 circulation × 7 production × 2 directions) - F_{outer} = number of outer folds (typically 8-15) - F_{inner} = inner folds per outer fold (typically 3-5) - G = grid size