Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Rethinking the agricultural cost of water scarcity: national water stress, irrigated cropland and total factor productivity

**Muhammad Afnan Arif**

muhdafnanarif@gmail.com

University of Malaya

**Research Article**

**Keywords:** Water stress, Water scarcity, Agricultural productivity, Total factor productivity, Irrigation, Panel data

**Posted Date:** November 11th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10839180/v1

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

**Additional Declarations:** No competing interests reported.

# Rethinking the agricultural cost of water scarcity: national water stress, irrigated cropland and total factor productivity

**Muhammad Afnan Arif** [1]

**Abstract**

Water scarcity is widely treated as a constraint on agricultural production, yet water does not enter the accounts from which agricultural productivity indices are built, so the scale of this constraint has never been directly measured. This study estimates that relationship using a long-run cross-country panel, controlling for irrigation, income, rural population and shocks shared across countries. The estimated relationship is negative but too small to distinguish from no effect, and a power analysis confirms the design is capable of detecting a meaningfully smaller effect than the one observed, so the finding sets an upper bound on the constraint rather than merely failing to find one. The result holds regardless of how heavily a country depends on irrigation, and does not vary systematically with income, aridity or observed water stress across countries. It is also robust to excluding unreliable records, to focusing on the countries with the most variation in scarcity, and to an alternative measure of agricultural output. Water scarcity may well constrain farming at the local level, but that constraint does not show up in the national data used to track and compare water stress across countries, which has implications for how such indicators are applied in policy analysis.

**Keywords:** Water stress; Water scarcity; Agricultural productivity; Total factor productivity; Irrigation; Panel data

## 1 Introduction

Global water stress, measured as annual freshwater withdrawals relative to renewable freshwater resources, has remained close to 18% since 2015, while several regions, including Northern Africa, Western Asia, and Central and Southern Asia, already exceed the 75% threshold classified by the United Nations as critical. This persistence matters directly for agriculture because productivity growth depends not only on land, labor, capital, and technology, but also on the reliability of water availability. Although SDG 6.4 calls for countries to reduce water scarcity substantially by 2030, current progress remains off track. Longstanding debates over how water scarcity should be defined and measured also show that indicator choice matters for interpreting cross-country stress comparisons (Falkenmark et al., 1989; Rijsberman, 2006; Damkjaer & Taylor, 2017). A central empirical question therefore remains unresolved: are agricultural productivity gains systematically more difficult under water-stress conditions, and does this relationship differ across countries once the econometric structure of global panel data is properly addressed?

The presumed effect has not, however, been estimated. Doeffinger and Hall (2020) provide the closest econometric treatment of stress and productivity jointly, but their dependent variable is water stress and their regressor is water productivity, output per unit of water withdrawn; their question is whether efficiency gains relieve or intensify extraction. Nyiwul et al. (2025) likewise take water productivity as the outcome, estimating the effect of water-sector patenting on it within a Cobb-Douglas framework. Neither estimates the effect of water scarcity on agricultural total factor productivity, and no cross-country study appears to do so. The gap is therefore not that two econometric procedures have gone uncombined; it is that a relationship assumed in policy design has not been measured. A second problem follows. Studies routinely divide countries by income,

[1] Corresponding author. Asia-Europe Institute, Universiti Malaya, Kuala Lumpur, Malaysia. Email: muhdafnanarif@gmail.com

aridity or observed stress before estimating, yet no framework in this literature specifies what would make the water-stress slope differ across such divisions, so the classifications are imposed rather than derived.

This study estimates that effect for a global panel of 127 countries observed annually from 1992 to 2021, combining internationally comparable measures of agricultural total factor productivity, water stress and standard productivity controls. The pairwise adaptive group fused lasso is used to recover any latent slope groups directly from the data, and a defactored variant of the same procedure removes unobserved common shocks. The estimated water-stress elasticity is small, negative and not distinguishable from zero, with a confidence interval bounding the response of agricultural total factor productivity to a 10% rise in water stress between a fall of roughly 1% and a rise of roughly one third of 1%. The group-selection procedure selects a single group across the full sample, two non-overlapping sub-periods and a cereal-yield specification, though a simulation reported in the Supplementary Material shows the procedure has limited power to detect group differences at this sample size, and the residual cross-sectional dependence left by the standard estimator is removed by defactoring without altering that selection.

The contribution is threefold. First, the paper supplies a theoretical basis for a relationship that applied work has specified by precedent, showing why water scarcity enters a productivity residual as a measurement wedge rather than as an omitted input, and why a mechanism operating within basins and growing seasons should attenuate at national annual aggregation. Second, it supplies the first cross-country estimate of that effect and bounds its magnitude, converting an assumption embedded in SDG 6.4 monitoring into a measured quantity. Third, it converts the pooling decision from a convention into a tested restriction, confirming a common slope against three conventional classifications that would each have imposed heterogeneity. The analysis is conducted entirely on observed panel data; its principal limitation is that a bound estimated at national annual aggregation cannot separate a negligible mechanism from a substantial one that the data cannot resolve.

The remainder of the paper is organized as follows. Section 2 reviews the literature on water stress, agricultural productivity, cross-sectional dependence and grouped-panel estimators, and develops the theoretical framework and hypotheses. Section 3 describes the data and empirical strategy. Section 4 presents and discusses the results. Section 5 concludes with implications and directions for future research.

## 2 Literature review

### 2.1 Water stress and agricultural productivity

A closer reading of the econometric literature reinforces this gap. Doeffinger and Hall (2020) provide the closest econometric precedent, though they estimate a different relationship. Their dependent variable is water stress and their regressor is water productivity, defined as output per unit of water withdrawn, and they report that productivity gains are associated with lower stress within countries over time. Their motivating question comes from resource economics, namely whether efficiency improvements relieve extraction pressure or induce rebound. Because water enters both of their variables, part of the association is definitional. The direction estimated here is the reverse, and the productivity concept is agricultural total factor productivity rather than water productivity, so their result neither anticipates nor contradicts the present one.

Agricultural productivity growth across countries has been traced primarily to technology and input use rather than to resource endowments. Gollin et al. (2021) attribute large output gains in developing countries to the diffusion of high-yielding crop varieties, and McArthur and McCord (2017) find comparable effects from fertiliser adoption. Sheahan and Barrett (2017) document

how uneven that adoption remains in Sub-Saharan Africa, while Fuglie (2018) reports that global productivity growth has slowed. Gollin et al. (2014) show that measured productivity gaps between agriculture and other sectors are smaller than raw data suggest once measurement is handled carefully, a caution that applies to cross-country productivity comparisons generally. Environmental constraints enter this literature mainly through climate rather than water: Schlenker and Roberts (2009) establish sharply nonlinear temperature effects on yields, Hsiang (2016) reviews the econometric issues such estimates raise, and Gouel and Laborde (2021) show that market-mediated adaptation offsets part of the physical damage. Where water enters directly it is usually as irrigation infrastructure rather than scarcity: Gatti et al. (2021) and Schuenemann et al. (2018) evaluate irrigation investment in Indonesia and Malawi respectively. Scarcity itself, measured as a national withdrawal ratio, has not been tested against productivity in this literature.

Evidence at finer scales points the other way. Wang et al. (2026) exploit the timing of government irrigation investment in Northern China and find that raising the irrigated share of farmland increased agricultural total factor productivity by roughly 14%, with drought mitigation among the operating mechanisms. Nyiwul et al. (2025) examine water-related innovation in 81 countries between 2000 and 2019 using fixed-effects instrumental-variable and system GMM estimators. Contrary to their own theoretical prediction, they find that water-sector patenting reduces water productivity, a result they interpret as a productivity paradox. Their outcome is again water productivity rather than agricultural productivity, and cross-sectional dependence is not addressed directly. Fuglie et al. (2021) provide a single-country productivity-accounting study of water-constrained agricultural growth in Egypt. Although that setting does not raise a multi-country dependence problem, it is relevant to the productivity measure used here.

Regional studies broaden the evidence base but leave the dependence and grouping questions unresolved. Abdou et al. (2024) link water scarcity to agricultural output in Gulf Cooperation Council countries. Jemmali et al. (2021) examine climate variables, agricultural sustainability and food-water poverty in selected MENA countries. Sun et al. (2021) split 41 Sub-Saharan African countries into income terciles and find income-specific effects on water resources. That income split is the type of conventional grouping this study tests directly.

This evidence shows that water stress and agricultural productivity are empirically connected, and that the association is robust across regions and specifications. What remains unresolved is why the association appears in a productivity measure that excludes water from its inputs, and what, if anything, would make its strength differ across countries.

### 2.2 Econometric approaches for environmental panel data

Environmental and agricultural panels often combine many countries over time, making inference vulnerable to common shocks. Pesaran (2006) proposes the common correlated effects estimator, which controls for unobserved common factors through cross-sectional averages. Pesaran (2021) develops the diagnostic test used here to detect residual cross-sectional dependence. Stauskas and De Vos (2025) extend this framework to cases in which outcomes and regressors respond to distinct but correlated factors, a relevant setting because the standard estimator in this paper leaves strong residual dependence.

The estimator used here sits within a literature that has developed rapidly since Pesaran (2006). Everaert and De Groote (2016) extend common correlated effects estimation to dynamic panels, and De Vos and Everaert (2021) correct the small-sample bias that arises when the time dimension is short. Subsequent work has concentrated on the conditions under which the approach is valid: Karabiyik et al. (2017) and De Vos et al. (2024) examine the rank condition, Margaritella and Westerlund (2023) show how information criteria can select the number of cross-sectional averages, and Juodis (2022) proposes a regularization approach when that number is

uncertain. Inference has been treated separately, through the incidental parameters problem in tests for remaining dependence (Juodis and Reese, 2022; Juodis and Sarafidis, 2022) and through bootstrap methods (De Vos and Stauskas, 2024). Juodis and Reese (2026) review two decades of applied practice. Grouped structure can also be placed in the factor rather than the slope (Ando and Bai, 2016), and inference on group membership itself is now available (Dzemski and Okui, 2024). A parallel strand asks whether unmodelled structure is better captured as breaks than as groups (Ditzen et al., 2025). Karabiyik et al. (2019) survey the multifactor error-structure literature, and Kapetanios et al. (2011) extend it to non-stationary factors. Implementations are available in standard software (Eberhardt, 2012; Ditzen, 2018; Ditzen and Reese, 2023), which matters for replication but does not settle the prior question of whether groups exist in a given application.

Applied environmental panels increasingly use dependence diagnostics to guide estimator choice. Olaniyi and Odhiambo (2025) apply several dependence tests before estimating models for resource-rich African economies. Shah et al. (2023) use the common correlated effects mean group estimator in a land-use and biodiversity study. Eberhardt et al. (2013) show that residual dependence can remain even after common-factor adjustment. These studies highlight the need for residual testing, but they do not combine dependence correction with data-driven group recovery in water or agricultural productivity panels.

Grouped-panel estimators address whether countries should be pooled, split into groups or treated separately. Mehrabani (2023) proposes the pairwise adaptive group fused lasso (PAGFL), which recovers both the number of groups and country membership from the data. Related approaches include the classifier-lasso of Su et al. (2016) and high-dimensional panel methods using non-concave penalties in Wang and Zhu (2022). These methods avoid imposing country groups in advance, but applications rarely ask whether recovered groups remain valid after residual cross-sectional dependence is removed.

Recent applications show the relevance of grouped-panel methods for environmental-economic questions. Hussain (2025) recovers three latent groups in a study of circular economy indicators and energy transition outcomes in European Union countries. Huang et al. (2024) document C-Lasso applications in cross-country saving and democracy-growth panels, each recovering two groups. Nibbering and Paap (2024) show that full pooling can outperform grouped alternatives in carbon-emissions forecasting when the signal is weak relative to estimation noise. This result is consistent with the complete pooling selected in the present water-stress-productivity application.

This literature leaves a specific gap, and the gap is conceptual before it is econometric. Grouped-panel estimators can recover country groups from the data, but whether groups should exist at all is a question about the economics of the relationship rather than about the estimator. Water-stress applications have not asked what mechanism would generate cross-country differences in the water-productivity slope, and in the absence of such a mechanism the income, aridity and stress classifications used in this literature are conventions rather than hypotheses. Section 2.3 derives a channel that answers this question and yields a prediction about slope structure, which the estimator is then used to test. Table 1 summarizes the studies reviewed in this section.

**Table 1. Summary of studies on water stress and agricultural productivity.**

| Study | Method(s) | DV | IV(s) | Country/Period | Key Findings |
|---|---|---|---|---|---|
| Doeffinger & Hall (2020) | FE panel | Water productivity | WS | Large cross-country panel | Productivity ↑ as WS ↓; CD infeasible (missing data) |

| | | | | | |
|---|---|---|---|---|---|
| Nyiwul et al. (2025) | FE-IV, sys-GMM | Water productivity | Water-sector patents | Global | Patenting ↑ productivity; CD not addressed |
| Fuglie et al. (2021) | Productivity accounting | TFP, TRP | Irrigation | Egypt (N=1) | Irrigation-driven growth; no CD concern |
| Abdou et al. (2024) | Regression | Agri. output | WS, fertilizer, CH4 | GCC, cross-section | WS ↓ agri. output in GCC states |
| Jemmali et al. (2021) | Panel cointegration, FE/RE | Food/water poverty | Climate, agri. productivity | MENA, 1990-2016 | Low agri. productivity and higher food poverty |
| Sun et al. (2021) | PMG/ARDL, income terciles | Water resources | Agri., CO2, growth | SSA (41 countries), 2000-2016 | Agri./CO2 effects on water vary by income tercile |
| Olaniyi & Odhiambo (2025) | Four CD tests | Resource-curse outcomes | Institutions | Resource-rich Africa | CD tests guide estimator choice |
| Eberhardt et al. (2013) | CCE + CD test | Output (mfg) | R&D, labor, capital | 10 countries, 1981-2005 | CD test rejects independence post-CCE |
| Hussain (2025) | C-Lasso, PSTR | CE-ET nexus | CCUS, fin. structure | 27 EU countries, 2000-2022 | 3 latent groups recovered |
| Huang et al. (2024) | C-Lasso | Savings; democracy-growth | Inflation; institutions | 56 & 98 countries | K=2 groups in both real applied panels |

Notes: WS = water stress, TFP = total factor productivity, TRP = total resource productivity, FE = fixed effects, IV = instrumental variable, GMM = generalized method of moments, FE-SEM = fixed-effects spatial error model, PMG/ARDL = pooled mean group autoregressive distributed lag, CD = cross-sectional dependence, CSD = cross-sectional dependence, C-Lasso = classifier-lasso, PSTR = panel smooth transition regression, CCUS = carbon capture, utilization, and storage, CE-ET = circular economy-energy transition nexus, GCC = Gulf Cooperation Council, MENA = Middle East and North Africa, SSA = Sub-Saharan Africa.

### 2.3 Theoretical framework

Water is unusual among agricultural inputs: rarely priced at scarcity value, rarely metered, and almost never recorded in the accounts from which national productivity statistics are built. A country near its renewable limit faces a constraint that does not enter its measured input bundle.

Let country $i$ in year $t$ produce output $Y_{it}$ with measured inputs $X_{it}$ and technology $A_{it}$. Water stress does not remove measured inputs; it degrades the service flow they deliver. Irrigation that cannot be supplied at capacity is still recorded as irrigated land. Effective input services $\tilde{X}_{it}$ therefore fall below measured inputs as stress rises

$$Y_{it} = A_{it}F(\tilde{X}_{it}), \qquad \tilde{X}_{it} = \varphi(s_{it})X_{it}, \qquad \varphi'(\cdot) < 0, \;\; \varphi(0) = 1$$

where $s_{it}$ denotes water stress and the utilization function $\varphi(\cdot)$ maps stress into the share of measured capacity delivered.

Measured productivity divides output by an index of measured, not effective, inputs. With $F$ Cobb-Douglas and $\theta_i$ the combined elasticity of water-dependent inputs, the recorded index is

$$TFP_{it} = \frac{Y_{it}}{F(X_{it})} = A_{it}\frac{F(\varphi(s_{it})X_{it})}{F(X_{it})} = A_{it}\varphi(s_{it})^{\theta_i}$$

so measured productivity falls with water stress at unchanged technology $A_{it}$. Taking logarithms and approximating the utilization function around low stress, $\ln\varphi(s_{it}) \approx -\delta \ln s_{it}$, gives

$$\ln TFP_{it} = \ln A_{it} - \theta_i \delta \ln s_{it}$$

The slope is the product of two parameters: a utilization elasticity set by crop physiology and water-delivery engineering, and a composition share measuring how far production depends on water-mediated inputs.

With country fixed effects and a one-period lag, the estimating equation is

$$\ln TFP_{it} = \eta_i + \beta \ln s_{i,t-1} + \boldsymbol{\gamma}' \mathbf{z}_{it} + u_{it}, \qquad \beta = -\theta\delta$$

Controls enter through the model. Irrigated land share is the empirical counterpart of the composition parameter; fertiliser intensity is a second channel, since nutrient response is moisture-conditional; income per capita blocks the path from income to technology; and rural population share captures labour-composition shifts independent of water. Average precipitation is a fixed climatological normal and is not identified alongside country fixed effects.

Three hypotheses follow, each conditional on the scale at which it can be tested.

**H1.** The water-stress coefficient is negative but small at national annual aggregation. The test is not significance but how tightly the magnitude can be bounded.

**H2.** Any surviving effect is stronger where production depends more on water-mediated inputs, giving a negative interaction with irrigated land share.

**H3.** The slope is common across countries, so classifications based on income or aridity impose heterogeneity in a dimension the mechanism does not operate through.

## 3 Data and methods

### 3.1 Data

The estimation panel covers 127 countries observed annually from 1992 to 2021. The balanced core is selected from a larger merged dataset based on complete coverage for the seven variables required by the specification. The dataset combines agricultural total factor productivity (TFP), normalized to 2015 = 100 for every country; cereal yield in kilograms per hectare as an alternative outcome; and water stress, defined as freshwater withdrawal as a share of available freshwater resources. The water-stress series is reported in the World Development Indicators and compiled from the national withdrawal and renewable-resource records held in the FAO AQUASTAT database (FAO, 2024). Agricultural productivity measurement relies on internationally comparable TFP accounting, for which recent USDA-ERS documentation and methodological reviews provide important context (Fuglie et al., 2024; Kryszak et al., 2023; Wang et al., 2024).

Europe and Central Asia account for the largest share of the sample, followed by Sub-Saharan Africa and the Middle East, North Africa, Afghanistan and Pakistan grouping. The income distribution ranges from high-income to low-income economies. Two data-quality issues are addressed before estimation. First, irrigated land data are incomplete and sometimes exceed the arable-land denominator because of source differences, so both capped and uncapped versions are retained, with the capped version used in the main specification. Second, 32 countries show unchanged water-stress values after the mid-2000s. These countries remain in the full-sample estimation but are excluded from the second sub-period, where the lack of variation would be mechanical rather than substantive.

Average annual precipitation was initially considered as a regressor but was excluded from the estimating equation because it is reported as a fixed long-run climatological average rather than

an annually varying series. It is retained only as a time-invariant classification variable for aridity groupings. High water-stress values, including observations above 500% in several Gulf economies, are retained because the indicator compares withdrawals with renewable resources only and can therefore exceed 100% where fossil aquifers or desalination are heavily used. The main specification regresses agricultural TFP on lagged water stress, irrigated share, fertilizer use, GDP per capita and rural population share, with country fixed effects. Cereal yield is used as a robustness outcome under the same control structure. Figures 1 and 2 provide descriptive regional context by showing variation in agricultural TFP and water stress over the sample period. Table 2 lists the data sources and descriptive statistics. Figures 1 and 2 show the regional distribution of both series.

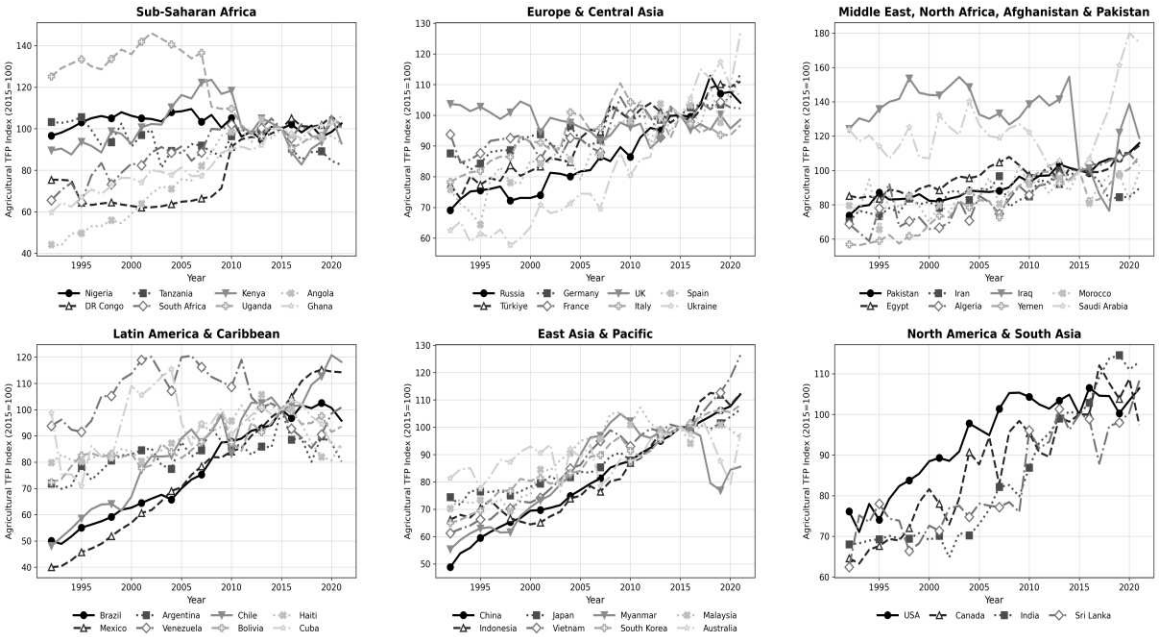



**Figure 1: Agricultural total factor productivity by region.**

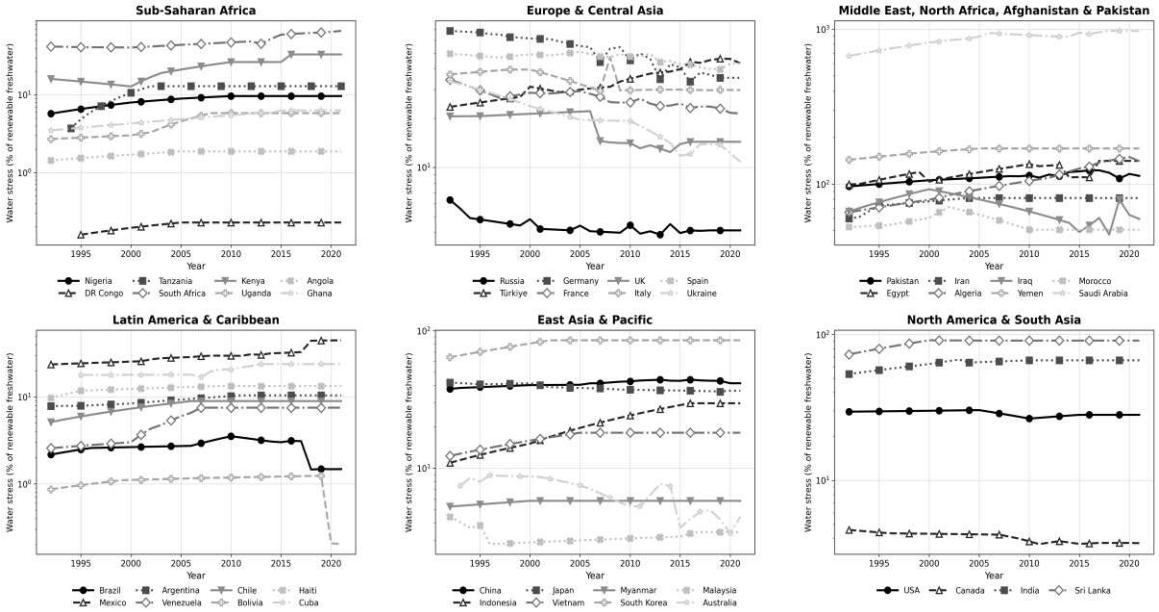



**Figure 2: Water stress by region.**

**Table 2. List of data sources and descriptive statistics.**

| **Panel A: Data sources** | | | | |
|---|---|---|---|---|
| **Acronym** | **Variable** | **Definition** | **Unit** | **Source** |
| **TFP** | Agricultural total factor productivity | Total factor productivity index, 2015 = 100 | Index | USDA-ERS |
| **YIELD** | Cereal yield | Cereal production per hectare harvested | kg/ha | WDI |
| **WS** | Water stress | Freshwater withdrawal as a share of available freshwater resources | % | WDI |
| **IRR** | Irrigated share | Irrigated land as a share of arable land, composite construction | % | WDI, FAOSTAT, USDA-ERS |
| **FERT** | Fertilizer use | Fertilizer consumption per hectare of arable land | kg/ha | WDI |
| **GDPPC** | GDP per capita | Gross domestic product per capita, constant prices | US dollars | WDI |
| **RURAL** | Rural population | Share of population classified as rural | % | WDI |

| **Panel B: Descriptive statistics** | | | | | |
|---|---|---|---|---|---|
| **Variable** | **Mean** | **Std. dev.** | **Min.** | **Max.** | **N** |
| **TFP** | 4.523 | 0.223 | 3.362 | 5.486 | 3,808 |
| **WS (lagged)** | 2.565 | 1.852 | -3.696 | 8.256 | 3,733 |
| **IRR (%)** | 22.499 | 29.330 | 0.001 | 100.000 | 3,808 |
| **FERT (kg/ha)** | 139.214 | 254.945 | 0.000 | 2335.096 | 3,783 |
| **GDPPC** | 8.408 | 1.472 | 5.195 | 11.418 | 3,807 |
| **RURAL (%)** | 42.388 | 22.196 | 0.000 | 93.351 | 3,810 |
| **YIELD** | 7.747 | 0.779 | 5.069 | 10.512 | 3,774 |

Notes: WDI = World Bank World Development Indicators, USDA-ERS = United States Department of Agriculture Economic Research Service, FAOSTAT = Food and Agriculture Organization Corporate Statistical Database.

### 3.2 Estimation methods

#### Panel model with latent group structure

The empirical framework follows the latent-group panel model in Mehrabani (2023). Slope coefficients may differ across countries, but countries within the same unknown group share identical slopes. Let $y_{it}$ denote the outcome, $\mathbf{x}_{it}$ the regressor vector, and $\eta_i$ the country fixed effect. The model, latent-group restriction and within-country transformation used to remove fixed effects are

$$
\begin{aligned}
y_{it} &= \boldsymbol{\beta}_i{}'\mathbf{x}_{it} + \eta_i + u_{it}, \qquad i = 1, \dots, N,\ t = 1, \dots, T, \\
\boldsymbol{\beta}_i^0 &= \begin{cases} \boldsymbol{\alpha}_1^0, & \text{if } i \in G_1^0 \\ \vdots & \\ \boldsymbol{\alpha}_{K^0}^0, & \text{if } i \in G_{K^0}^0 \end{cases}, \\
\tilde{y}_{it} &= \boldsymbol{\beta}_i{}'\tilde{\mathbf{x}}_{it} + \tilde{u}_{it}, \qquad \tilde{x}_{it} = x_{it} - T^{-1}\textstyle\sum_{t=1}^{T} x_{it}\,,
\end{aligned}
$$

where $\boldsymbol{\alpha}_k^0 \neq \boldsymbol{\alpha}_l^0$ for $k \neq l$ and $\{G_1^0, \dots, G_{K^0}^0\}$ forms an unknown partition of $\{1, \dots, N\}$.

#### Pairwise adaptive group fused lasso (PAGFL) estimator

The number of groups K° and country membership are not fixed in advance. They are recovered from the data using the pairwise adaptive group fused lasso (PAGFL) of Mehrabani (2023). PAGFL combines the adaptive group fused lasso of Qian and Su (2016) with the pairwise fusion

approach of Ma and Huang (2017), yielding a panel estimator that can identify latent slope groups without imposing the number of groups beforehand.

**Classifier-lasso (C-Lasso) and grouped fixed-effects (GFE) benchmarks**

Two related estimators clarify the role of PAGFL. Su et al. (2016) propose the classifier-lasso (C-Lasso), which jointly classifies units and estimates group-specific coefficients through a penalized likelihood framework, with a PGMM variant for endogenous regressors. Let$\varphi(\cdot)$ denote the log-likelihood contribution and $\hat{\mu}_i(\gamma_i^*)$ the profiled nuisance parameter. The penalized profile likelihood criterion and group-fusion penalty are

$$\begin{aligned} Q(\boldsymbol{\gamma}^*) &= \tfrac{1}{nT}\textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\phi\left(w_{it};\boldsymbol{\gamma}_i^*,\hat{\mu}_i(\boldsymbol{\gamma}_i^*)\right), \\ Q_{G,\lambda_{PPL}} &= Q(\boldsymbol{\gamma}^*) + \tfrac{\lambda_{PPL}}{N}\textstyle\sum_{i=1}^{N}\prod_{g=1}^{G_0}\parallel\boldsymbol{\beta}_i-\boldsymbol{\alpha}_g\parallel_2, \end{aligned}$$

where $\lambda_{PPL}$ is the tuning parameter and the product-form penalty over the $G_0$ candidate group centers $\boldsymbol{\alpha}_g$ is what distinguishes the C-Lasso from the additive pairwise penalty PAGFL uses instead. Bonhomme and Manresa (2015) propose a grouped fixed-effects (GFE) estimator based on a K-means-style least-squares search over possible country partitions. With group membership$g_i \in \{1,\dots,G\}$ and group-time intercept $c_{g_i,t}$, the estimator solves

$$\left(\widehat{\boldsymbol{\beta}},\hat{\mathbf{c}},\hat{\boldsymbol{\gamma}}\right) = \underset{(\boldsymbol{\beta},\mathbf{c},\boldsymbol{\gamma})}{\arg\min}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}-c_{g_i,t}\right)^2,$$

where the minimization is taken jointly over the slope coefficients, the group-time intercepts, and every candidate assignment of countries to groups.

Both benchmarks have practical limitations here. C-Lasso requires separate tuning choices for the number of groups and penalty strength. It can also leave some countries unclassified, requiring an external assignment rule. The GFE estimator requires the number of groups to be fixed before estimation and relies on a non-convex optimization problem that may converge to a local minimum. PAGFL addresses these issues by using a single tuning parameter and a convex penalized least-squares objective. The number of groups is recovered from the data rather than imposed by the researcher. Convexity also ensures a unique global minimizer, avoiding the local-minimum concern associated with K-means-style grouping.

**Penalized least-squares objective and alternating direction method of multipliers (ADMM) estimation**

PAGFL minimizes the following penalized least-squares objective, where the adaptive weight is constructed from preliminary ordinary least-squares estimates

$$\begin{aligned} Q_{1,NT}(\boldsymbol{\beta},\lambda_1) &= \tfrac{1}{T}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{y}_{it}-\boldsymbol{\beta}_i'\tilde{\mathbf{x}}_{it})^2 + \tfrac{\lambda_1}{N}\sum\sum_{1\le i<j\le N}\dot{w}_{ij}\parallel\boldsymbol{\beta}_i-\boldsymbol{\beta}_j\parallel, \\ \dot{w}_{ij} &= \parallel\dot{\boldsymbol{\beta}}_i-\dot{\boldsymbol{\beta}}_j\parallel^{-\kappa}, \\ \dot{\boldsymbol{\beta}}_i &= \left(\textstyle\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\,\tilde{\mathbf{x}}_{it}'\right)^{-1}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\,\tilde{y}_{it}, \end{aligned}$$

The parameter $\lambda_1$ controls the strength of pairwise fusion. The adaptive weight $\ddot{w}_{ij}$ increases the penalty for countries with similar preliminary slopes, allowing some pairwise coefficient differences to shrink exactly to zero while leaving surviving group-specific coefficients unpenalized after post-lasso re-estimation. Because the objective is not separable across countries, Mehrabani (2023) estimates it using the alternating direction method of multipliers (ADMM) developed by Boyd et al. (2011). ADMM rewrites the convex problem into smaller subproblems by introducing an auxiliary variable for each pairwise coefficient difference and minimizing the augmented Lagrangian. The update steps are set out in the Supplementary Material.

$$\widehat{\boldsymbol{\alpha}}_k^p = \left(\sum_{i\in\widehat{G}_k}\sum_{t=1}^T \tilde{\mathbf{x}}_{it}\,\tilde{\mathbf{x}}_{it}{}'\right)^{-1}\sum_{i\in\widehat{G}_k}\sum_{t=1}^T \tilde{\mathbf{x}}_{it}\,\tilde{y}_{it}.$$

The tuning parameter $\lambda_1$, and therefore the implied group count $\widehat{K}_{\lambda_1}$, is selected by minimizing the Bayesian-information-criterion-type rule proposed by Mehrabani (2023). Fixing the number of groups beforehand would assume the answer to the central empirical question.

The criterion and the within-group sum of squared residuals it penalizes are

$$\begin{aligned} IC_1(\lambda_1) &= \hat{\sigma}^2_{\widehat{G}_{\widehat{K}_{\lambda_1}}} + \rho_{1,NT}\, p\widehat{K}_{\lambda_1}, \\ \hat{\sigma}^2_{\widehat{G}_{\widehat{K}_{\lambda_1}}} &= \tfrac{1}{NT}\sum_{k=1}^{\widehat{K}_{\lambda_1}}\sum_{i\in\widehat{G}_k(\lambda_1)}\sum_{t=1}^T\left(\tilde{y}_{it} - \widehat{\boldsymbol{\alpha}}_k^{p\prime}\tilde{\mathbf{x}}_{it}\right)^2, \end{aligned}$$

The coefficient $\rho_{1,NT}$ grows with sample size to limit over-fitting. Across the candidate grid for $\lambda_1$, the value minimizing $IC_1(\lambda_1)$ determines the reported group count and coefficients.

**Defactored pairwise adaptive group fused lasso (DF-PAGFL) estimator**

The second estimator, denoted DF-PAGFL, applies PAGFL after a defactoring step based on the common correlated effects approach of Pesaran (2006). DF-PAGFL is therefore not a new estimator with new asymptotic theory. It is a two-step application of established procedures: first remove common cross-sectional components, then recover latent groups using PAGFL. Following Pesaran (2006), each variable, including the outcome and every regressor, is first replaced by its deviation from the cross-sectional average in that year. The defactored estimating equation is then

$$\begin{aligned} \tilde{v}_{it} &= v_{it} - \bar{v}_t, \qquad \bar{v}_t = \tfrac{1}{N}\textstyle\sum_{i=1}^N v_{it}\,, \\ \log\left(\widetilde{TFP}_{it}\right) &= \alpha_i + \boldsymbol{\beta}_{g_i}{}'\left(\tilde{x}_{1,it}, \tilde{x}_{2,it}, \tilde{x}_{3,it}, \tilde{x}_{4,it}, \tilde{x}_{5,it}\right)' + \tilde{u}_{it}, \end{aligned}$$

where $\bar{v}_t$ is the cross-sectional average of variable $v$ in year $t$, and $\tilde{x}_{1,it}, \ldots, \tilde{x}_{5,it}$ denote the defactored lagged water stress, irrigated share, fertilizer use, GDP per capita, and rural population share, respectively. This adjustment removes factors common to all countries in a given year, such as global commodity-price shocks or shared climate patterns, before PAGFL is applied to the adjusted variables. The label DF-PAGFL is used for brevity, but it denotes a sequence of two established procedures rather than a newly derived estimator.

A more integrated alternative to this two-step composition exists in the recent literature. De Vos and Westerlund (2026) propose a sequential group-estimation procedure that builds the CCE adjustment directly into the group-detection objective itself, rather than applying it as a separate first step, and report substantial computational gains from doing so in large panels. This study uses the two-step approach for two reasons. First, it keeps the published properties of each component estimator separately attributable. Second, it makes the comparison between standard and defactored residual dependence transparent, which is central to the empirical design.

The trade-off is computational. The two-step approach does not inherit the efficiency gains of a fully integrated estimator such as De Vos and Westerlund (2026). This is not material here because the sample contains 127 countries, well below the scale where those gains would drive the empirical choice.

$$\begin{aligned} CD &= \sqrt{\frac{2T}{N(N-1)}}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{ij}\,, \\ \hat{\rho}_{ij} &= \frac{\sum_{t=1}^T \hat{u}_{it}\hat{u}_{jt}}{\left(\sum_{t=1}^T \hat{u}_{it}^2\right)^{1/2}\left(\sum_{t=1}^T \hat{u}_{jt}^2\right)^{1/2}}, \end{aligned}$$

where $\hat{u}_{it}$ is the residual for country $i$ in year $t$. Under the null hypothesis of cross-sectional independence, $CD$ is asymptotically distributed as standard normal.

This version of the CD test is suitable when the cross-sectional dimension is large relative to the time dimension, matching the structure of the present panel. The test is applied separately to residuals from the standard PAGFL and DF-PAGFL estimators on the full estimation sample.

**Dynamic common correlated effects (DCCE) robustness check**

As a robustness check, the analysis also considers the dynamic common correlated effects (DCCE) extension of Chudik and Pesaran (2015). DCCE augments the CCE approach by adding lagged cross-sectional averages, which can be needed when common factors affect a dynamic panel with persistence. The defactored equation becomes

$$\log\left(\widetilde{TFP}_{it}\right) = \alpha_i + \boldsymbol{\beta}_{g_i}{}'\tilde{\mathbf{x}}_{it} + \sum_{l=0}^{p^*} \bar{\mathbf{z}}_{t-l}{}'\boldsymbol{\gamma}_{i,l} + \tilde{u}_{it}, \qquad \bar{\mathbf{z}}_t = \frac{1}{N}\sum_{i=1}^{N}[y_{it}, \mathbf{x}_{it}{}']',$$

where $\bar{\mathbf{z}}_t$ stacks the cross-sectional averages of the outcome and every regressor in year $t$, and $p^*$ lags of these averages are added to approximate the unobserved common factors in a dynamic setting, following the augmentation proposed by Chudik and Pesaran (2015). This augmentation checks whether contemporaneous-only defactoring is sufficient. In this application, the DCCE residual CD statistic is numerically identical to the simpler specification, so lagged cross-sectional averages do not change the main inference.

**Penalized generalized method of moments (PGMM) limitation**

The least-squares PAGFL estimator used here assumes weak exogeneity: regressors are uncorrelated with the contemporaneous idiosyncratic error after removing country fixed effects. This assumption is more plausible for lagged water stress than for GDP per capita or irrigated land share, which may respond to unobserved productivity shocks within the same country-year. Su et al. (2016) develop a PGMM variant of C-Lasso for precisely this situation, in which regressors are treated as endogenous and instrumented within the group-detection procedure itself. Writing $z_{it}$ for a $q \times 1$ vector of instruments, $\Delta y_{it}$ and $\Delta x_{it}$ for the first-differenced outcome and regressors, and $W_{i,NT}$ for a country-specific weight matrix, their PGMM objective is: Where that assumption is doubtful, two-stage instrumental-variable estimation under interactive effects offers an alternative (Cui et al., 2022).

$$Q_{2,NT}(\boldsymbol{\beta}, \lambda_2) = \sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T} z_{it}\,(\Delta y_{it} - \boldsymbol{\beta}_i{}'\Delta x_{it})\right]' W_{i,NT}\left[\frac{1}{T}\sum_{t=1}^{T} z_{it}\,(\Delta y_{it} - \boldsymbol{\beta}_i{}'\Delta x_{it})\right] + \frac{\lambda_2}{N}\sum\sum_{1\le i<j\le N} \ddot{w}_{ij} \parallel \boldsymbol{\beta}_i - \boldsymbol{\beta}_j \parallel,$$

where $\lambda_2$ plays the same role as $\lambda_1$ in the exogenous case, and $\ddot{w}_{ij}$ is the corresponding adaptive weight constructed from preliminary GMM estimates. Mehrabani (2023) extends this same PGMM logic to the pairwise adaptive fused-lasso framework used in this research, in place of the additive-multiplicative penalty of Su et al. (2016). The results should therefore be read as conditional associations under a weak-exogeneity assumption rather than causal estimates. This limitation is shared with close empirical precedents such as Doeffinger and Hall (2020) and Nyiwul et al. (2025), neither of which directly resolves regressor endogeneity. Figure 3 summarizes the analytical pipeline.

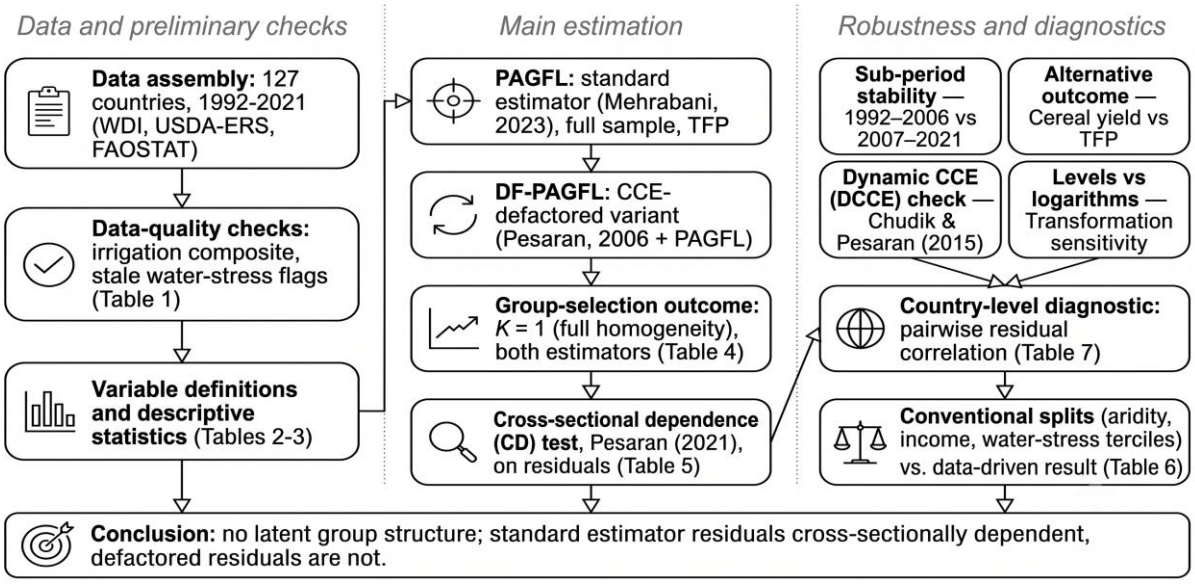



**Figure 3: Analytical flowchart.**

## 4 Results and discussion

### 4.1 Results

The group-selection results select a single group across all specifications tested (Table 4, Panel A). In the full sample, the standard estimator moves from the fully disaggregated solution at lower tuning values to a single group at the selected value, with no stable intermediate grouping. The defactored estimator reaches the same selection after common cross-sectional components are removed. The result also holds across two non-overlapping sub-periods and in the expanded cereal-yield specification. Neither the outcome measure, sample window, nor estimator variant produces a different selected group count. Section 4.2 addresses what this selection can and cannot support as evidence about the underlying relationship.

The pooled slope recovered by the post-lasso step provides the central quantitative result. Under the defactored specification the water-stress elasticity is -0.031, with a standard error of 0.034 clustered by country, so the effect is negative in sign but not distinguishable from zero at conventional levels. The associated 95% confidence interval runs from -0.099 to 0.037, which bounds the response of agricultural total factor productivity to a 10% increase in water stress between a fall of approximately 0.93% and a rise of approximately 0.35%. Estimating the same specification without the defactoring step returns a similar coefficient. The interaction between water stress and irrigated land share, which tests H2, is positive and not significant. Neither the level effect nor its dependence on composition is detectable at national annual aggregation, which is what Section 2.3 predicts, and the interval places an upper bound on how large the attenuated wedge can be in these data. Table 3 reports this estimate alongside the robustness set on which the bound rests.

**Table 3. Elasticity of agricultural total factor productivity with respect to lagged water stress.**

| **Panel A: Main estimate** | | | | |
|---|---|---|---|---|
| **Specification** | **Countries** | **Obs.** | **Elasticity** | **95% CI** |
| Defactored, country fixed effects | 127 | 3,706 | -0.031 (0.034) | [-0.099, +0.037] |
| **Panel B: Robustness** | | | | |

| | | | | |
|---|---|---|---|---|
| No defactoring | 127 | 3,706 | -0.013<br>(0.036) | [-0.084,<br>+0.058] |
| Excluding stale water-stress records | 95 | 2,769 | -0.031<br>(0.038) | [-0.106,<br>+0.044] |
| Highest within-country stress variation | 64 | 1,867 | -0.003<br>(0.042) | [-0.084,<br>+0.078] |
| Sub-period 1992-2006 | 127 | 1,806 | -0.015<br>(0.031) | [-0.076,<br>+0.045] |
| Sub-period 2007-2021 | 127 | 1,900 | -0.031<br>(0.040) | [-0.109,<br>+0.047] |
| Cereal yield as alternative outcome | 126 | 3,672 | +0.014<br>(0.036) | [-0.057,<br>+0.085] |
| **Panel C: Hypothesis 2** | | | | |
| Water stress x irrigated land share | 127 | 3,706 | +0.0002<br>(0.0003) | [-0.0003,<br>+0.0007] |

Notes: SE in parentheses below each coefficient, No coefficient is significant at conventional levels, Dependent variable is log agricultural total factor productivity except in the cereal-yield row, Standard errors are clustered by country, CI = confidence interval.

Two checks establish that this bound is not an artefact of limited treatment variation. Excluding the thirty-two countries whose water-stress series is unchanged after the mid-2000s leaves the estimate essentially unaltered. Restricting the sample to countries with above-median within-country variation in water stress moves the estimate closer to zero rather than away from it. The median within-country standard deviation of log water stress is 0.16, so the panel contains substantial variation in the treatment; the absence of a detectable response is a feature of the data rather than of the sample construction.

Residual diagnostics qualify how the group-selection result should be read. If unmodelled common factors were driving the selected group structure, the standard estimator would leave dependence in the residuals. The Pesaran (2021) CD test confirms this concern for the standard estimator: its residuals show large and statistically significant cross-sectional dependence (Table 5, Panel A). By contrast, residuals from the defactored estimator show no significant dependence. This pattern indicates that defactoring removes dependence operating mainly through the error structure rather than through the regressors individually. It also supports interpreting the single-group selection as unaffected by residual dependence, rather than as an artifact of it. Section 4.2 addresses the separate question of what the selection procedure was capable of detecting.

Table 4, Panel C compares the data-driven result with conventional country classifications. Splits based on aridity, income and observed water-stress levels all imply multiple groups before estimation, yet PAGFL selects a single group. These conventional classifications therefore impose heterogeneity that the data-driven procedure does not recover. The result suggests that country groupings commonly used in this literature should be tested rather than assumed.

The country-level diagnostic in Table 5, Panel B further clarifies the source of residual dependence. If dependence were concentrated within a particular region, income class or water-stress category, a more refined conventional split might address it. Instead, countries with the highest positive average residual correlations, including Yemen, Italy, Ghana, Angola and South Africa, do not form a coherent cluster. The same is true for countries with the lowest correlations, including Botswana, Suriname, Uruguay and Fiji. This pattern is consistent with broad common shocks, such as global agricultural commodity prices or correlated weather conditions, rather than omitted regional or income-based group effects.

The dynamic CCE robustness check addresses whether the contemporaneous defactoring step is sufficient when common shocks persist over time. Adding lagged cross-sectional averages

following Chudik and Pesaran (2015) produces only a temporary intermediate grouping along the tuning path. The information criterion remains lower at the homogeneous solution and the residual CD statistic is numerically identical to the contemporaneous-only specification. The simpler defactoring approach is therefore adequate for this application.

The log transformation is also checked because it may compress cross-country differences in productivity and water stress. Re-estimating the model with standardized levels rather than logarithms produces a different tuning path: no fusion occurs across the tested range until a much higher threshold. This sensitivity does not overturn the main result because the logarithmic specification is theoretically motivated and consistent with close empirical precedents. It does show, however, that transformation choices matter when grouped-panel methods are applied to macroeconomic variables with wide cross-country dispersion. Figure 4 compares the two series across the aridity, income and water-stress terciles used in this literature, and Figures 5 and 6 show their regional trends and cross-country dispersion.

### 4.2 What the group-selection result can support

The single-group selection reported above should not be read as direct evidence that all 127 countries share one water-stress elasticity. A simulation exercise, reported in the Supplementary Material, establishes that the group-selection procedure has essentially no power to recover two groups at this sample size and with this regressor set unless the true difference in elasticities between groups exceeds roughly eighty times the magnitude of the coefficient estimated here. No comparably large heterogeneity is suggested by any theoretical or empirical account of water-stress response in the literature reviewed in Section 2. The finding is therefore best read as an absence of detectable heterogeneity within the resolution of this design, rather than as confirmation that the response is uniform. This qualification does not affect the residual-dependence result above, which concerns the error structure rather than the slope parameters, and does not affect the headline elasticity reported in Section 4.1.

**Table 4. Group-selection results.**

| **Panel A: Main specifications** | | | | |
|---|---|---|---|---|
| Specification | Countries | $\lambda$ | Groups (K) | Information criterion |
| TFP, standard estimator, full sample | 127 | 0.152 | 1 | 0.069 |
| TFP, defactored estimator, full sample | 127 | 0.066 | 1 | 0.068 |
| TFP, standard estimator, 1992 to 2006 | 121 | 0.464 | 1 | 0.070 |
| TFP, standard estimator, 2007 to 2021 | 85 | 0.060 | 1 | 0.080 |
| YIELD, standard estimator, full sample | 151 | 0.351 | 1 | 0.105 |

| **Panel B: Data-quality robustness** | | | |
|---|---|---|---|
| Sample restriction | Countries | Groups (K), standard | Groups (K), defactored |
| Full sample, all 127 countries | 127 | 1 | 1 |
| Exclude 32 countries with stale water-stress values | 95 | 1 | 1 |
| Irrigated share, raw uncapped composite | 127 | 1 | - |
| Irrigated share, capped at 100% | 127 | 1 | 1 |

| **Panel C: Conventional splits** | | | | |
|---|---|---|---|---|
| Classification rule | Group 1 | Group 2 | Group 3 | Groups found by group fused lasso |

| | | | | |
|---|---|---|---|---|
| Aridity tercile (by average precipitation) | 43 arid | 42 moderate | 42 humid | 1 (homogeneous) |
| Income tercile (by average GDPPC) | 43 low income | 42 mid income | 42 high income | 1 (homogeneous) |
| Water-stress tercile (by average observed stress) | 43 low stress | 42 mid stress | 42 high stress | 1 (homogeneous) |

Notes: TFP = agricultural total factor productivity, YIELD = cereal yield, K denotes the number of selected groups, K is chosen by minimizing the Bayesian-information-criterion-type objective over a log-spaced $\lambda$ grid, K = 1 remains stable across consecutive $\lambda$ values and under a stricter convergence tolerance in the full-sample standard specification, the defactored estimator is not re-estimated for the uncapped irrigated-share specification, hence the corresponding cell is not applicable, terciles are based on country-level full-sample averages.

**Table 5. Cross-sectional dependence test and country-level diagnostic.**

| **Panel A: CD test** | | |
|---|---|---|
| Residuals from | CD statistic | Conclusion |
| Standard estimator | 8.17*** | Reject cross-sectional independence |
| Defactored estimator | -0.57 | Cannot reject cross-sectional independence |

| **Panel B: Country diagnostic** | | |
|---|---|---|
| Rank | Country | Average pairwise correlation |
| 1 (highest) | Yemen | 0.134 |
| 2 | Italy | 0.133 |
| 3 | Ghana | 0.129 |
| 4 | Angola | 0.128 |
| 5 | South Africa | 0.125 |
| 123 | Albania | -0.113 |
| 124 | Fiji | -0.117 |
| 125 | Uruguay | -0.120 |
| 126 | Suriname | -0.130 |
| 127 (lowest) | Botswana | -0.147 |

Notes: ***, **, * significant at 1%, 5%, 10%, CD = cross-sectional dependence, the CD test is applied to full-sample residuals from each estimator, Panel B reports average pairwise residual correlations from the standard estimator, rank 1 denotes the highest average correlation and rank 127 the lowest, only the five highest and five lowest countries are reported.

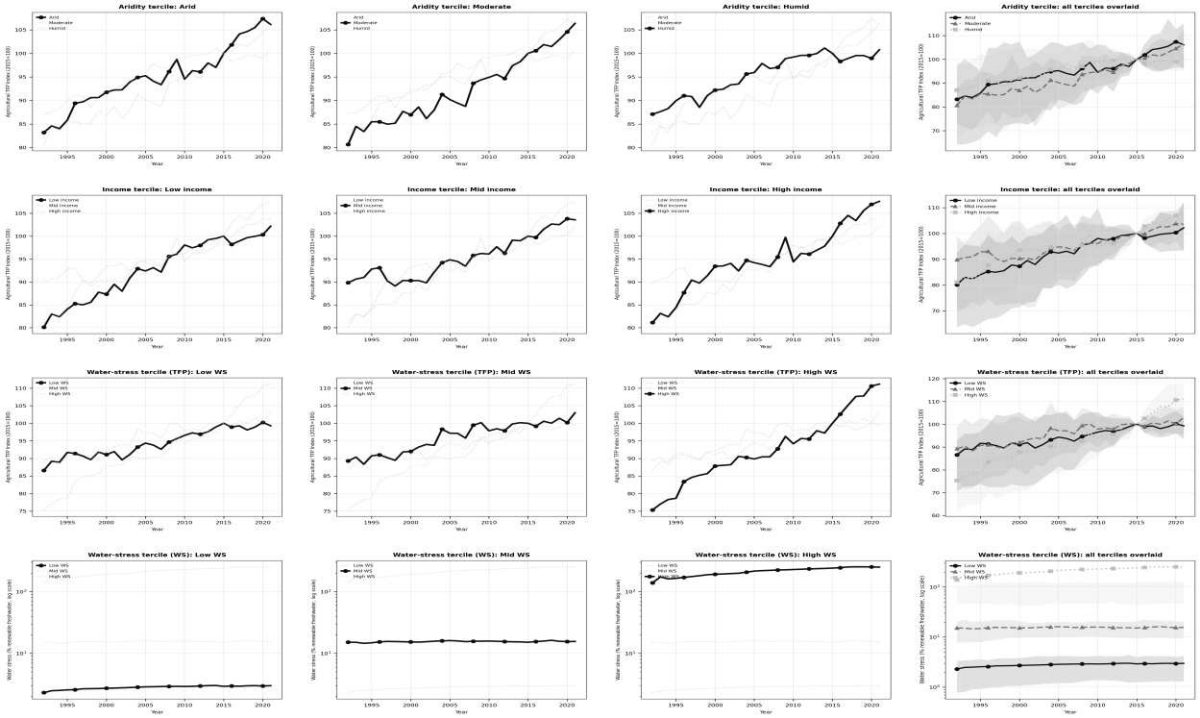



**Figure 4: Agricultural TFP and water stress by conventional classification tercile.**

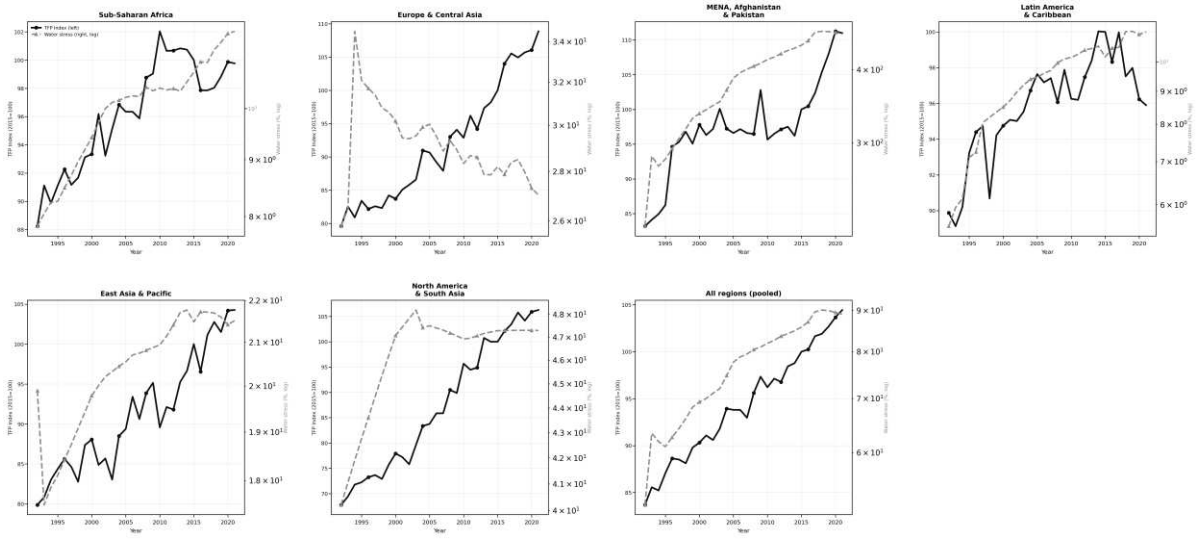



**Figure 5: Agricultural TFP and water stress trends by region.**

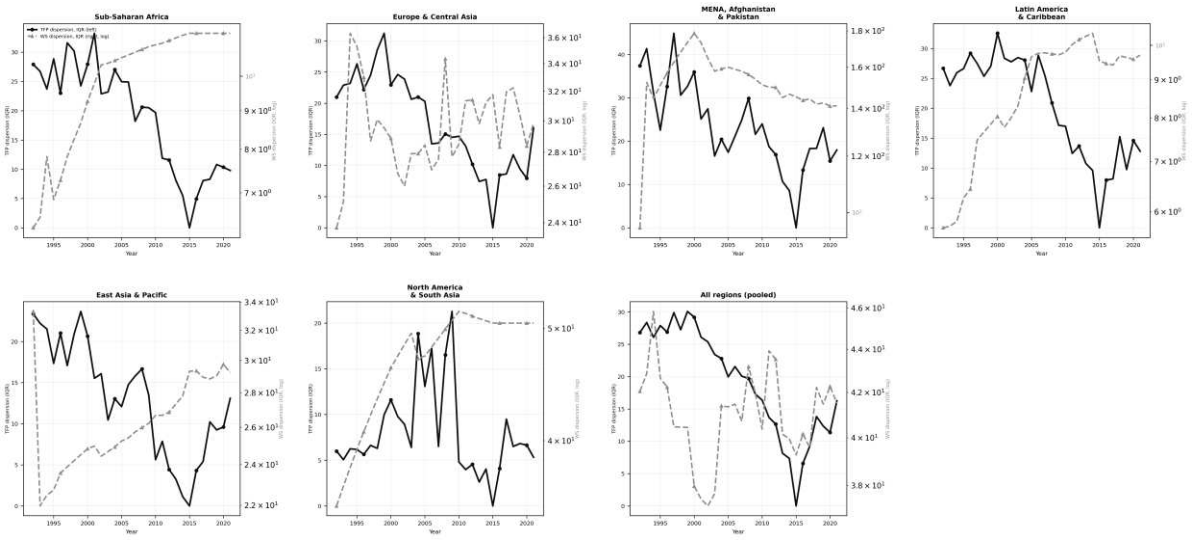



**Figure 6: Cross-country dispersion of agricultural TFP and water stress by region.**

### 4.3 Discussion

The principal empirical result has two related implications. The analysis selects a single group across every specification tested, while also demonstrating the importance of defactoring for valid inference. Across five specifications, covering two outcome variables, the full estimation sample, and two non-overlapping sub-periods, the selected model consistently returns one group rather than distinct water-stress regimes. This selection is stable: neither an alternative productivity measure nor a shorter temporal window produces a different result. As Section 4.2 discusses, the procedure's power to detect group differences at this sample size limits how far this stability can be read as evidence of a genuinely uniform response, rather than as an absence of heterogeneity large enough for the design to resolve. This finding contrasts with empirical approaches that motivate country splits by income, aridity, or observed water-stress levels, as in Doeffinger and Hall (2020).

These results align with the prediction of Section 2.3, and the alignment is the point rather than a convenience. The framework locates the utilization wedge within basins and growing seasons, and therefore predicts that national annual aggregation attenuates it. The estimated coefficient is duly small, negative in sign, and bounded within a narrow interval around zero. The framework further predicts a common slope, because a technical elasticity has no reason to vary with national income or aridity class once composition is controlled and once attenuation has removed most of the variation composition would generate. The group-selection procedure recovers exactly that. What the analysis cannot do is separate a negligible mechanism from a real one that national data cannot resolve, and the bound should be read with that limitation in view.

This single-group selection, however, must be interpreted alongside both the residual diagnostic evidence and the power limitation discussed in Section 4.2. The economic interpretation is that the main econometric threat in this application arises from residual cross-sectional dependence rather than from unobserved slope heterogeneity. Residuals from the standard estimator remain strongly correlated across countries, which undermines both conventional inference and the information-criterion comparisons used for group selection. By contrast, applying the defactored estimator to the same panel removes this dependence while preserving the single-group selection. In the sense of Pesaran (2006), the results therefore indicate that unobserved common factors, rather than an omitted group structure, constitute the relevant latent source of bias in this setting.

This distinction between slope homogeneity and residual dependence provides a basis for interpreting the findings relative to adjacent panel applications. The selection of complete pooling after accounting for cross-sectional dependence is directionally consistent with, although not identical to, evidence from adjacent cross-country panel applications, though the comparison is qualified by the power limitation in Section 4.2 in both this paper and, potentially, in related work using similar sample sizes. Nibbering and Paap (2024), for example, show that full pooling can dominate grouped alternatives in carbon-emissions forecasting when the underlying signal is weak relative to estimation noise, although their framework does not incorporate an explicit dependence correction. The country-level diagnostic reported here further clarifies the source of dependence. Countries with the strongest positive or negative average residual correlations do not cluster systematically by region, income, or water-stress exposure. This pattern weakens the interpretation that a more refined conventional country classification would absorb the dependence and instead supports the presence of broad common shocks, such as global commodity-price movements or contemporaneously correlated climatic conditions. The dynamic defactoring check reinforces this reading because adding lagged cross-sectional averages does not alter the residual-dependence result.

These empirical patterns suggest broader methodological implications for cross-country agricultural and environmental panels. Whether grouped-panel methods detect heterogeneity or select a single group, the credibility of that conclusion depends both on the residual dependence structure left by the estimator and on whether the design has adequate power to detect the alternative, a question Section 4.2 suggests deserves more routine attention in this literature. Diagnostics focused only on the regressors may therefore be insufficient, since the dependence in this application operates primarily through the error term. Country classifications based on income, aridity, or observed water stress, such as those motivated in related work by Doeffinger and Hall (2020) and Nyiwul et al. (2025), should consequently be treated as hypotheses to be tested rather than as maintained assumptions about the underlying parameter structure. Figure 7 shows this relationship for the twelve highest-stress countries, and Figure 8 repeats it in annual changes.

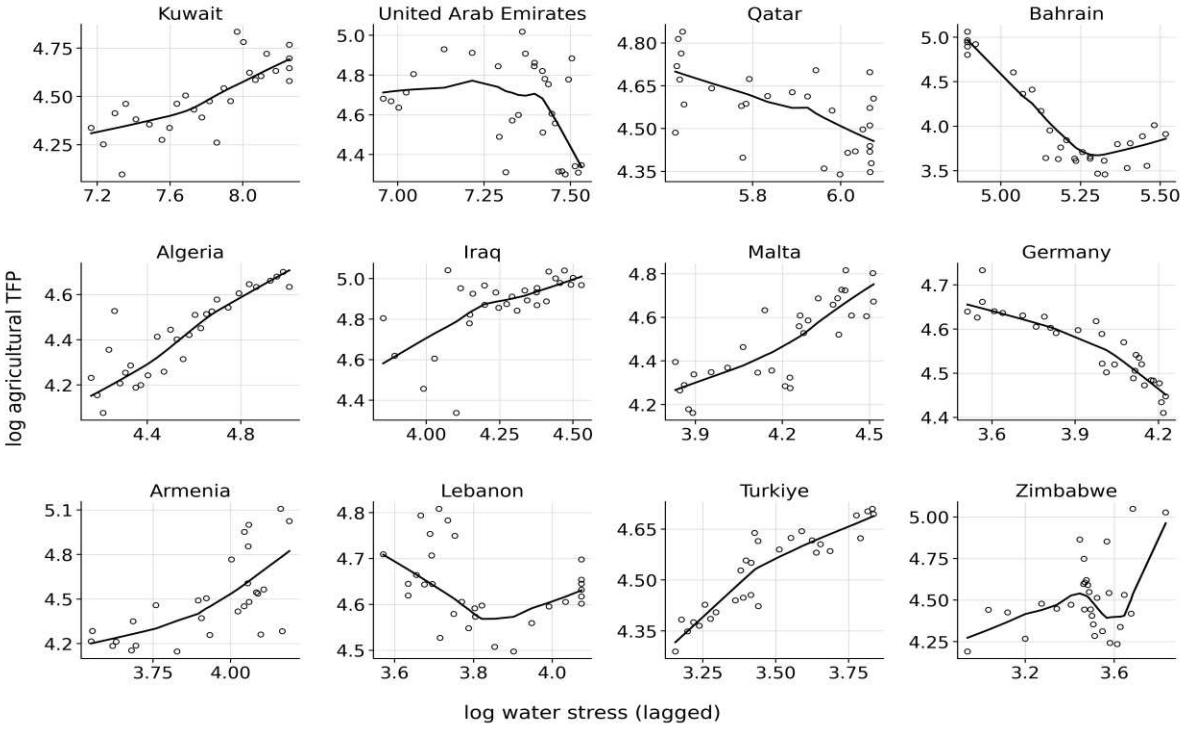


**Figure 7: Agricultural total factor productivity and water stress, twelve highest-stress countries.**

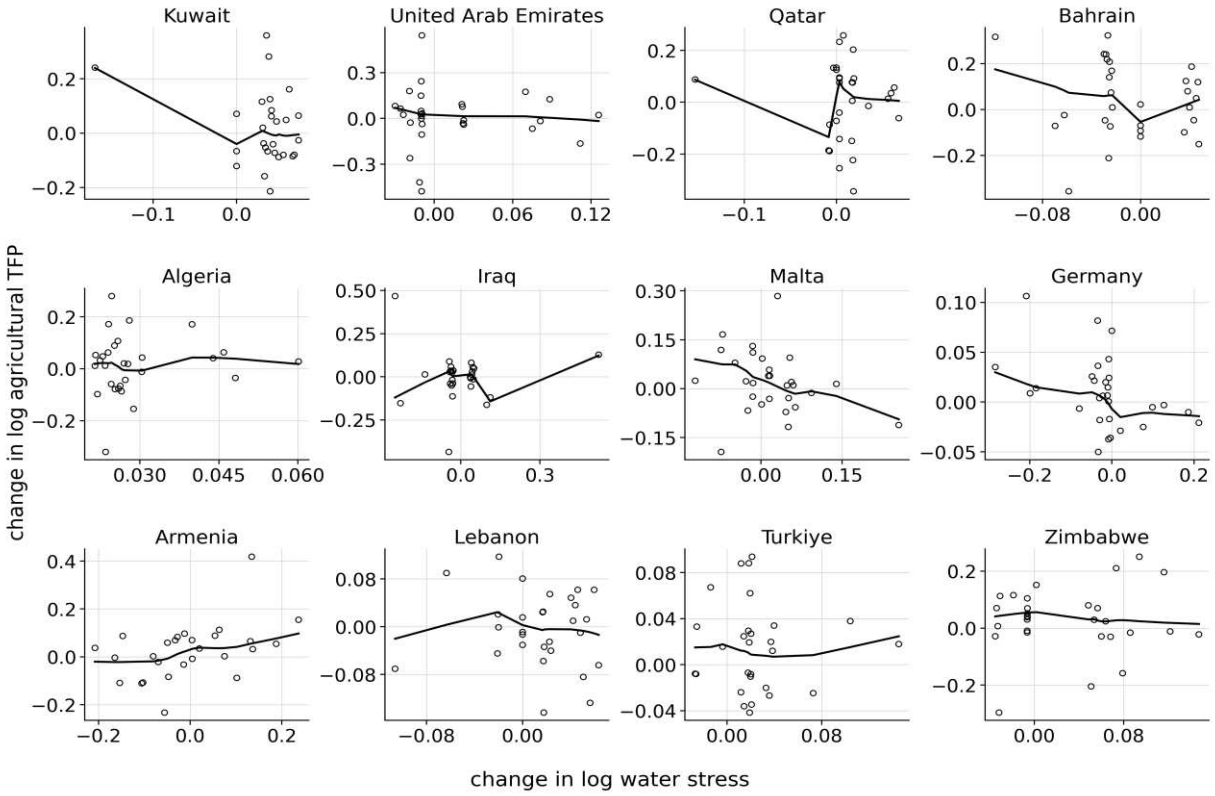



**Figure 8: Annual changes in agricultural total factor productivity and water stress, twelve highest-stress countries.**

## 5 Conclusion

This research asked whether national water stress reduces agricultural productivity, and estimated the effect for 127 countries observed from 1992 to 2021 using the pairwise adaptive group fused lasso and its defactored variant. A utilization-wedge framework motivated the specification and predicted that the mechanism, which operates at basin and seasonal scale, would be substantially averaged away in national annual data. The estimated elasticity is small, negative, and not statistically distinguishable from zero, and the confidence interval bounds the response of agricultural total factor productivity to a 10% rise in water stress between roughly minus one and plus one third of 1%. Because the design detects elasticities marginally larger than the one estimated, this is a bound on the constraint rather than a failure to observe it. The group-selection procedure finds no latent country groups across five specifications, contrary to the heterogeneous water-stress regimes assumed in related empirical work, and the defactored estimator removes strong residual cross-sectional dependence while preserving that conclusion.

Three implications follow for how water scarcity is measured and used in policy analysis. The first concerns the indicator itself. The ratio of freshwater withdrawals to renewable resources is the headline scarcity measure in national reporting and underpins the water-stress indicator monitored under Sustainable Development Goal target 6.4.2. The results establish that this indicator, at annual national resolution, carries no detectable signal about agricultural productivity. That is not an argument against the indicator, which was designed to describe pressure on the resource rather than to predict production outcomes. It is an argument against the secondary use

that has grown around it, in which national water-stress values are read as proxies for agricultural performance risk. Where the question is productivity, the indicator should be treated as uninformative unless it is disaggregated to basin and season.

The second concerns how scarcity effects enter economy-wide models. Computable general equilibrium and integrated assessment exercises that assign a water constraint to agriculture typically calibrate it from basin-level or farm-level studies and then apply it at national scale. The bound estimated here disciplines that step. A nationally detectable elasticity larger than the interval reported is inconsistent with three decades of cross-country evidence, so calibrations implying one require justification that the aggregation itself does not destroy the effect. Modellers requiring a large national water constraint should show why national data fail to reveal it, rather than assuming the transfer is innocuous.

The third concerns targeting. National water-stress classifications, along with income and aridity groupings, are widely used to allocate adaptation finance and to select countries for water-sector intervention. The group-selection results give no support for treating any of these classifications as a partition of the underlying response. Countries that share a stress category do not share an estimated elasticity in any recoverable sense. Allocation rules built on these categories are distributional conventions rather than empirical findings, and should be defended on equity or feasibility grounds rather than on the claim that they identify where water most constrains production.

Four limitations qualify these conclusions. First, the least-squares version of the estimator relies on a weak-exogeneity assumption, more defensible for lagged water stress than for potentially endogenous regressors such as gross domestic product per capita and irrigated land share. Second, the defactoring strategy approximates unobserved common factors using cross-sectional averages; richer factor structures remain a useful extension, although the dynamic augmentation considered here suggests the simpler contemporaneous adjustment is adequate. Third, a bound estimated at national annual aggregation cannot distinguish a negligible mechanism from a substantial one that the data cannot resolve. Fourth, the utilization wedge presumes that measured inputs remain in place while their service flow is compressed, so the framework does not apply where land is retired or irrigation abandoned.

## References


Abdou, D. S., Othman, A. K., & Moussa, M. N. (2024). Examining the impact of water scarcity on agricultural output in the Gulf Cooperation Council (GCC) region. World Water Policy, 10(4), 1243-1269. https://doi.org/10.1002/wwp2.12228

Ando, T., & Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. Journal of Applied Econometrics, 31(1), 163-191. https://doi.org/10.1002/jae.2467

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. Econometrica, 83(3), 1147-1184. https://doi.org/10.3982/ECTA11319

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning, 3(1), 1-122. https://doi.org/10.1561/2200000016

Chudik, A., & Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics, 188(2), 393-420. https://doi.org/10.1016/j.jeconom.2015.03.007

Cui, G., Norkuté, M., Sarafidis, V., & Yamagata, T. (2022). Two-stage instrumental variable estimation of linear panel data models with interactive effects. Econometrics Journal, 25(2), 340-361. https://doi.org/10.1093/ectj/utab029

Damkjaer, S., & Taylor, R. (2017). The measurement of water scarcity: Defining a meaningful indicator. Ambio, 46(5), 513-531. https://doi.org/10.1007/s13280-017-0912-z

De Vos, I., & Everaert, G. (2021). Bias-corrected common correlated effects pooled estimation in dynamic panels. Journal of Business & Economic Statistics, 39(1), 294-306. https://doi.org/10.1080/07350015.2019.1654879

De Vos, I., & Stauskas, O. (2024). Cross-section bootstrap for CCE regressions. Journal of Econometrics, 240(1), 105648. https://doi.org/10.1016/j.jeconom.2023.105648

De Vos, I., & Westerlund, J. (2026). Panel sequential group estimation of interactive effects models. Oxford Bulletin of Economics and Statistics, 1-19. https://doi.org/10.1111/obes.70055

De Vos, I., Everaert, G., & Sarafidis, V. (2024). A method to evaluate the rank condition for CCE estimators. Econometric Reviews, 43(2-4), 123-155. https://doi.org/10.1080/07474938.2023.2292383

Ditzen, J. (2018). Estimating dynamic common-correlated effects in Stata. Stata Journal, 18(3), 585-617. https://doi.org/10.1177/1536867X1801800306

Ditzen, J., & Reese, S. (2023). xtnumfac: A battery of estimators for the number of common factors in time series and panel-data models. Stata Journal, 23(2), 438-454. https://doi.org/10.1177/1536867X231175305

Ditzen, J., Karavias, Y., & Westerlund, J. (2025). Multiple structural breaks in interactive effects panel data models. Journal of Applied Econometrics, 40(1), 74-88. https://doi.org/10.1002/jae.3097

Doeffinger, T., & Hall, J. W. (2020). Water stress and productivity: An empirical analysis of trends and drivers. Water Resources Research, 56(3). https://doi.org/10.1029/2019WR025925

Dzemski, A., & Okui, R. (2024). Confidence set for group membership. Quantitative Economics, 15(2), 245-277. https://doi.org/10.3982/QE1975

Eberhardt, M. (2012). Estimating panel time-series models with heterogeneous slopes. Stata Journal, 12(1), 61-71. https://doi.org/10.1177/1536867X1201200105

Eberhardt, M., Helmers, C., & Strauss, H. (2013). Do spillovers matter when estimating private returns to R&D? The Review of Economics and Statistics, 95(2), 436-448. https://doi.org/10.1162/REST_a_00272

Everaert, G., & De Groote, T. (2016). Common correlated effects estimation of dynamic panels with cross-sectional dependence. Econometric Reviews, 35(3), 428-463. https://doi.org/10.1080/07474938.2014.966635

Falkenmark, M., Lundqvist, J., & Widstrand, C. (1989). Macro-scale water scarcity requires micro-scale approaches: Aspects of vulnerability in semi-arid development. Natural Resources Forum, 13(4), 258-267. https://doi.org/10.1111/j.1477-8947.1989.tb00348.x

FAO. (2018). Guidelines for the measurement of productivity and efficiency in agriculture. Food and Agriculture Organization of the United Nations. https://openknowledge.fao.org/handle/20.500.14283/ca6395en

FAO. (2024). AQUASTAT database. Food and Agriculture Organization of the United Nations. https://www.fao.org/aquastat/en/

Fuglie, K. O. (2018). Is agricultural productivity slowing? Global Food Security, 17, 73-83. https://doi.org/10.1016/j.gfs.2018.05.001

Fuglie, K., & Rada, N. (2013). Resources, Policies, and Agricultural Productivity in Sub-Saharan Africa (Economic Research Report No. 145). United States Department of Agriculture, Economic Research Service. https://ers.usda.gov/publications/pub-details?pubid=45047

Fuglie, K., Dhehibi, B., El Shahat, A. A., & Aw-Hassan, A. (2021). Water, policy, and productivity in Egyptian agriculture. American Journal of Agricultural Economics, 103(4), 1378-1397. https://doi.org/10.1111/ajae.12148

Fuglie, K., Morgan, S., & Jelliffe, J. (2024). World agricultural production, resource use, and productivity, 1961-2020 (Report No. EIB-268). United States Department of Agriculture, Economic Research Service. https://doi.org/10.32747/2024.8327789.ers

Gatti, N., Baylis, K., & Crost, B. (2021). Can irrigation infrastructure mitigate the effect of rainfall shocks on conflict? Evidence from Indonesia. American Journal of Agricultural Economics, 103(1), 211-231. https://doi.org/10.1002/ajae.12092

Gollin, D., Hansen, C. W., & Wingender, A. M. (2021). Two blades of grass: The impact of the Green Revolution. Journal of Political Economy, 129(8), 2344-2384. https://doi.org/10.1086/714444

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). The agricultural productivity gap. Quarterly Journal of Economics, 129(2), 939-993. https://doi.org/10.1093/qje/qjt056

Gouel, C., & Laborde, D. (2021). The crucial role of domestic and international market-mediated adaptation to climate change. Journal of Environmental Economics and Management, 106, 102408. https://doi.org/10.1016/j.jeem.2020.102408

Hsiang, S. (2016). Climate econometrics. Annual Review of Resource Economics, 8(1), 43-75. https://doi.org/10.1146/annurev-resource-100815-095343

Huang, W., Wang, Y., & Zhou, L. (2024). Identify latent group structures in panel data: The classifylasso command. The Stata Journal, 24(2), 1-25. https://doi.org/10.1177/1536867X241233642

Hussain, K., Jian, Z., & Khan, A. (2025). Circular economy and EU's energy transition: The moderating and transitioning effects of financial structure and circular carbon technology innovation: Evidence from C-Lasso and PSTR approaches. Journal of Cleaner Production. Advance online publication. https://doi.org/10.1016/j.jclepro.2025.145434

Jemmali, H., Morrar, R., & Ben Aissa, M. S. (2021). The dynamic nexus between climate changes, agricultural sustainability and food-water poverty in a panel of selected MENA countries. Journal of Water and Climate Change, 12(1), 1-17. https://doi.org/10.2166/wcc.2019.309

Juodis, A. (2022). A regularization approach to common correlated effects estimation. Journal of Applied Econometrics, 37(4), 788-810. https://doi.org/10.1002/jae.2899

Juodis, A., & Reese, S. (2022). The incidental parameters problem in testing for remaining cross-section correlation. Journal of Business & Economic Statistics, 40(3), 1191-1203. https://doi.org/10.1080/07350015.2021.1906687

Juodis, A., & Reese, S. (2026). Five lessons for applied researchers from twenty years of common correlated effects estimation. Journal of Econometrics, 253, 106120. https://doi.org/10.1016/j.jeconom.2025.106120

Juodis, A., & Sarafidis, V. (2022). An incidental parameters free inference approach for panels with common shocks. Journal of Econometrics, 229(1), 19-54. https://doi.org/10.1016/j.jeconom.2021.03.011

Kapetanios, G., Pesaran, M. H., & Yamagata, T. (2011). Panels with non-stationary multifactor error structures. Journal of Econometrics, 160(2), 326-348. https://doi.org/10.1016/j.jeconom.2010.10.001

Karabiyik, H., Palm, F. C., & Urbain, J.-P. (2019). Econometric analysis of panel data models with multifactor error structures. Annual Review of Economics, 11(1), 495-522. https://doi.org/10.1146/annurev-economics-063016-104338

Karabiyik, H., Reese, S., & Westerlund, J. (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. Journal of Econometrics, 197(1), 60-64. https://doi.org/10.1016/j.jeconom.2016.10.006

Kryszak, Ł., Świerczyńska, K., & Staniszewski, J. (2023). Measuring total factor productivity in agriculture: A bibliometric review. International Journal of Emerging Markets, 18(1), 148-172. https://doi.org/10.1108/IJOEM-04-2020-0428

Ma, S., & Huang, J. (2017). A concave pairwise fusion approach to subgroup analysis. Journal of the American Statistical Association, 112(517), 410-423. https://doi.org/10.1080/01621459.2016.1148039

Margaritella, L., & Westerlund, J. (2023). Using information criteria to select averages in CCE. Econometrics Journal, 26(3), 405-421. https://doi.org/10.1093/ectj/utad009

McArthur, J. W., & McCord, G. C. (2017). Fertilizing growth: Agricultural inputs and their effects in economic development. Journal of Development Economics, 127, 133-152. https://doi.org/10.1016/j.jdeveco.2017.02.007

Mehrabani, A. (2023). Estimation and identification of latent group structures in panel data. Journal of Econometrics, 235(2), 1464-1482. https://doi.org/10.1016/j.jeconom.2022.12.002

Nibbering, D., & Paap, R. (2024). Forecasting carbon emissions using asymmetric grouping. Journal of Forecasting, 43(6), 2228-2256. https://doi.org/10.1002/for.3124

Nyiwul, L., Hu, Z., & Koirala, N. P. (2025). Innovation and water productivity: Empirical evidence from water-related patents. Applied Economic Perspectives and Policy, 47(2), 515-555. https://doi.org/10.1002/aepp.13498

Olaniyi, C. O., & Odhiambo, N. M. (2025). Does economic complexity provide antidotal pathways to evade the resource curse syndrome? A novel role for institutions. Sustainable Development, 33(2), 3118-3142. https://doi.org/10.1002/sd.3266

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. (2021). General diagnostic tests for cross-sectional dependence in panels. Empirical Economics, 60(1), 13-50. https://doi.org/10.1007/s00181-020-01875-7

Qian, J., & Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused Lasso. Journal of Econometrics, 191(1), 86-109. https://doi.org/10.1016/j.jeconom.2015.09.004

Rijsberman, F. R. (2006). Water scarcity: Fact or fiction? Agricultural Water Management, 80(1-3), 5-22. https://doi.org/10.1016/j.agwat.2005.07.001

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proceedings of the National Academy of Sciences, 106(37), 15594-15598. https://doi.org/10.1073/pnas.0906865106

Schuenemann, F., Thurlow, J., Meyer, S., Robertson, R., & Rodrigues, J. (2018). Evaluating irrigation investments in Malawi: Economy-wide impacts under uncertainty and labor constraints. Agricultural Economics, 49(2), 237-250. https://doi.org/10.1111/agec.12412

Shah, M. I., Abbas, S., Olohunlana, A. O., & Sinha, A. (2023). The impacts of land use change on biodiversity and ecosystem services: An empirical investigation from highly fragile countries. Sustainable Development, 31(3), 1384-1400. https://doi.org/10.1002/sd.2454

Sheahan, M., & Barrett, C. B. (2017). Ten striking facts about agricultural input use in Sub-Saharan Africa. Food Policy, 67, 12-25. https://doi.org/10.1016/j.foodpol.2016.09.010

Stauskas, O., & De Vos, I. (2025). Handling distinct correlated effects with CCE. Oxford Bulletin of Economics and Statistics, 87(2), 448-475. https://doi.org/10.1111/obes.12650

Su, L., Shi, Z., & Phillips, P. C. B. (2016). Identifying latent structures in panel data. Econometrica, 84(6), 2215-2264. https://doi.org/10.3982/ECTA12560

Sun, D., Addae, E. A., Jemmali, H., Mensah, I. A., Musah, M., Mensah, C. N., & Appiah-Twum, F. (2021). Examining the determinants of water resources availability in sub-Sahara Africa: A panel-based econometrics analysis. Environmental Science and Pollution Research, 28, 21212-21230. https://doi.org/10.1007/s11356-020-12256-z

United States Department of Agriculture, Economic Research Service. (2024). International agricultural productivity [Data product]. https://www.ers.usda.gov/data-products/international-agricultural-productivity

Wang, S. L., Nehring, R., Mosheim, R., & Njuki, E. (2024). Measurement of output, inputs, and total factor productivity in U.S. agricultural productivity accounts (Report No. TB-1966). United States Department of Agriculture, Economic Research Service. https://doi.org/10.32747/2024.8755132.ers

Wang, W., & Zhu, Z. (2022). Group structure detection for a high-dimensional panel data model. Canadian Journal of Statistics, 50(3), 852-866. https://doi.org/10.1002/cjs.11646

Wang, X., Huang, J., Rozelle, S., & Zhang, L. (2026). Irrigation investments and agricultural productivity in Northern China. American Journal of Agricultural Economics. https://doi.org/10.1002/ajae.70072