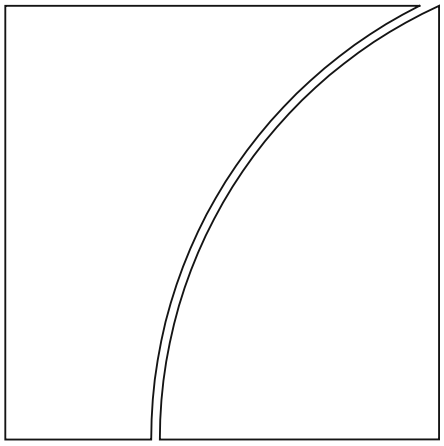


BIS Working Papers
No 1396

"Eppur si muove": Convergence Within and Between Countries

by Giovanni Lombardo

Monetary and Economic Department

November 2026

JEL classification: O47, C23, C15

Keywords: convergence, economic growth, cross-country income distribution, between-country and within-country variation, correlated random effects, dynamic panel

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in this publication are those of the authors and do not necessarily reflect the views of the BIS or its member central banks.

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2026. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

"Eppur si muove": Convergence Within and Between Countries*

Giovanni Lombardo
BIS

October 22, 2026

Abstract

Kremer, Willis, and You (2022) report that over six decades richer countries lost their growth advantage. Acemoglu and Molina (2022) and Pesaran and Smith (2026) reply that such regressions cannot measure catch-up when countries head toward different moving destinations. We nest both in one model. The movement survives across countries: the between-country income gradient of growth fell abruptly in the decade from 2003. Nulls with fixed destinations, or built to Pesaran and Smith's specification, do not reproduce it. Only a lower bound on convergence speed survives: the within-country bias cannot be separated from the estimate.

JEL classification: O47, C23, C15

Keywords: convergence, economic growth, cross-country income distribution, between-country and within-country variation, correlated random effects, dynamic panel

*The views expressed in this publication are those of the author and do not necessarily reflect the views of the BIS or its member central banks. Claude (Anthropic) was used for literature search, for writing and running the estimation and simulation code, and for editing the text. The author directed the research, checked all results and citations, and is solely responsible for the content.

1 Introduction

A few recent papers have revived a decades-old literature on income convergence. Up to the mid 1990s the literature reached a twin verdict: unconditional divergence over the long run (Pritchett, 1997) together with conditional convergence at roughly 2% per year (Barro and Sala-i-Martin, 1992).[1] Yet, Kremer et al. (2022, KWY henceforth) document that in pooled cross-country regressions convergence re-emerges after the 1990s.[2]

Acemoglu and Molina (2022, AM henceforth) point out that a pooled slope is not a convergence rate when countries approach different steady states, since the country effect is correlated with initial income; country fixed effects are therefore needed. Pesaran and Smith (2026, PS henceforth) point out that Barro cross-country regressions are biased toward zero unless the heterogeneity of the intercepts can be modeled exactly. The two-way fixed-effects estimator, which is AM's remedy, is biased in turn once countries are allowed trends that are not parallel, and the speed it reports tends to zero as that departure grows. Their own estimates rise once those trends are allowed to differ across countries: 0.3 to 0.5 percent a year from Barro regressions, about 3 percent from two-way fixed effects, and about 14 percent from a dynamic common-correlated-effects estimator that gives each country its own path. Both criticisms aim at the estimated convergence levels.

This paper takes those criticisms as given. Nothing below defends the level of a convergence coefficient, and the only statement about the convergence speed is a lower bound (Section 4.4). Yet, we note that the movement in convergence rate could survive those criticisms. This is the focus of our paper.

The time variation of the convergence rate pointed out by KWY is not challenged by AM and PS directly and deserves to be reviewed in light of the biases that the literature has identified. Is convergence an artifact of those biases too? Indeed, the bias AM and PS identify depends on the cross-country dispersion of steady states and on the loadings of common factors, and neither is fixed over sixty years. The question is also out of reach from where the critics stand. The estimators PS recommend need a long time-series dimension and deliver a single coefficient for the whole sample. On the contrary, this paper aims at disentangling the movement in the bias factors from changes in convergence rates.

To tackle this issue we specify a model that encompasses the two sides of the debate. Following Mundlak (1978) and Wooldridge (2019) we let the country effect be linear in the within-country means of the time-varying regressors; following Wooldridge (2005) we condition on the country's first observation. Setting these projection coefficients and the country-effect variance to zero collapses the model to KWY's pooled regression exactly. Holding the loading on the country mean common across decades reproduces AM's two-way fixed-effects slopes exactly, without estimating country-specific intercepts. The decade profile of the

[1] Johnson and Papageorgiou (2020) survey the half-century of evidence and find "a broad consensus of no evidence supporting absolute convergence".

[2] Patel et al. (2021) document the same turn independently, on a design that fixes the terminal year and varies the start.

convergence coefficient then separates into a within-country slope (how fast each country's income catches up with its—possibly varying—steady state) and a between-country gradient linking a country's average income level to its subsequent growth, and the question becomes whether the movement survives in the between gradient. The device is not new to growth econometrics — Islam (1995) projects the country effect on the regressors of every period by the minimum-distance method of Chamberlain (1982), and Moral-Benito (2012) estimates a cross-country growth panel by a correlated-random-effects likelihood that conditions on initial income, and concludes that the data cannot precisely identify the rate of convergence — but it is absent from the surveys that take stock of the field's econometric approaches (Durlauf et al., 2005; Smith, 2024), which set the case for fixed effects against the case for pooling.

Three findings emerge. First, the movement survives the criticisms that dispose of the level. The between-country gradient holds between 0.9 and 1.4 through the first four windows (1.42 in 1963–73) and steps down to 0.12 in 2003–13, where it stays; a linear trend is rejected at the 5 percent level and a single break at that window is not. The unconditional gradient KWY discuss moves with it. Yet, no point estimate of the speed can be recovered: the within-country slope carries a Nickell bias (Nickell, 1981) that cannot be separated from the estimate. The bias is governed by the calendar span of the sample rather than by the sampling interval (Barro, 2015, p. 914), as the power of unit-root tests is (Shiller and Perron, 1985; Perron, 1991), and by how far countries started from their destinations, which is unknown. The destinations nonetheless move. Let ρ_{10} denote the decade persistence of income. With destinations held constant, the posterior of Lancaster (2002), whose local mode is consistent at fixed T and which conditions on each country's first observation, places ρ_{10} between 0.92 and 1.00 (90 percent interval). At $\rho_{10} = 1$ the fixed-effects estimate $\hat{\rho}_{10}$ does not depend on the start: on this panel its mean is 0.55, with a simulation standard deviation of 0.04, and none of 2,000 replications reaches the observed $\hat{\rho}_{10} = 0.71$. A stationary start would settle the matter on the span alone, since the Nickell bias would then put a floor of $3/S$ a year under the speed a demeaned estimator reports on a span of S years: 4.29 percent on the seventy years of PS and 5.00 on our sixty. PS report 3 percent from two-way fixed effects; our two-way fixed-effects slope gives 3.48 percent (s.e. 0.46) in the same units, so both lie below that floor.

Second, the movement is not manufactured by the biases the critics identify. A null in which steady states are constant and no parameter varies by decade produces a decline of 0.12 (simulation s.d. 0.22) against the 1.30 observed when the between slopes are estimated as in the simulations. A null built to reflect PS's design — a trended latent factor, loadings that differ across countries and correlate with income, and country-specific dynamics — produces a profile that rises rather than falls, and never reaches the observed value in the replications.

Third, the mechanism is compositional. Durlauf et al. (2005, p. 589) observe that a negative cross-sectional slope is what one would expect if low-income countries tend to start below their steady states while high-income countries start above them, and add that they do not claim this is the case empirically. The decade profile of that covariance is what the between margin measures, and it is what moves: today's poor countries hold the deeper shortfalls from steady states. Over the same decades the growth correlates themselves converged: in

21 of 26 correlate-decade regressions, countries starting lower improved significantly faster (Section A.8).

The first finding implies that the speed at which a country closes its own gap is not identified at the frequency at which the debate is conducted. What can be stated is that the trend that KWY document is consistent with the evidence on the cross-country distribution of steady states.

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 lays out the encompassing specification and the restrictions that deliver each side of the debate. Section 4 reports the decade profiles, shows that the movement survives in the between gradient, while its split between the within slope and the wedge is not identified, and prices that movement against the nulls. Section 5 concludes.

2 Data

We use the annual panel from Penn World Table 11.0 (Feenstra et al., 2015), spanning 1960–2023, for 185 economies. Following KWY, growth rates are computed from real GDP at constant national prices (`rgdpna`) per capita and income levels from output-side real GDP at chained PPPs (`rgdpo`) per capita.[3] The sample drops countries whose population never exceeds 200,000 and countries for which natural-resource rents exceed 75% of GDP at any point in 1960–2017 (World Development Indicators); on the current WDI vintage the latter are Equatorial Guinea, Oman, Qatar, Saudi Arabia, Timor-Leste, and Turkmenistan (Timor-Leste does not appear in the PWT sample, so the two rules remove 21 economies). The panel is unbalanced, with poorer countries entering later.

The exact reproduction of KWY uses their published replication package (PWT 10.0 income measures, sample restrictions applied by the authors). The correlate analysis takes the institutional and policy variables from the same package, in the authors' normalization — each correlate divided by its standard deviation in 1985 or, following their replication code, the earliest year with at least forty observations — covering Polity 2, Freedom House political rights and civil liberties, the Worldwide Governance Indicators, the Heritage economic-freedom score, investment, and Barro–Lee educational attainment, matched to the PWT 11.0 panel by country and year.

[3] The split follows their footnote 5 and PWT's own guidance — `rgdpna` for growth over time, the chained-PPP series for level comparisons across countries (Feenstra et al., 2015). Section A.11 shows that it matters for the within slope.

3 Econometric framework

3.1 The two sides of the debate

KWY estimate, on a pooled country–year panel,

$$g_{i,t} \;=\; \sum_{d=1}^{6} \beta_d \, \mathbf{1}\{t \in d\} \, \log y_{i,t} \;+\; \mu_t \;+\; \varepsilon_{i,t}, \qquad g_{i,t} \equiv \frac{100}{10} \left(\log y_{i,t+10} - \log y_{i,t}\right), \tag{1}$$

where $y_{i,t}$ is GDP per capita of country i at time t, μ_t is a year effect, and standard errors are clustered by country. Timing: an observation dated t pairs annualized growth over the decade *following* t with income *at* t, so $t = 1960$ covers 1960–70 and $t = 2007$ covers 2007–17.[4] The index d refers to the decade windows of a grid of start years — the panel's two dimensions are countries and time, and throughout, *time* means the decade window — and the profile β_d is the object of interest. Their Table 1 uses the grid 1960, 1970, …, 2000, 2007 — six windows, non-overlapping except for the final pair, 2007 replacing 2010 because their data end in 2017 — and its remaining columns are the restrictions of Equation 1 to a common slope and to a slope linear in the start year. PWT 11.0's span allows a cleaner design:[5] our grid shifts three years to 1963, 1973, …, 2013, six non-overlapping, equally spaced decade windows covering 1963–2023 exactly and ending with the pandemic window 2013–23. Since no parameter links the windows, the point estimates are numerically identical to running one cross-sectional regression per window, and only the country-clustered covariance ties them together. $\beta_d < 0$ is absolute β-convergence; KWY's headline is the downward trend in the profile.

AM take as data-generating process the conditional-convergence model

$$g_{i,t} \;=\; \beta \log y_{i,t} + a_i + \mu_t + \varepsilon_{i,t}, \tag{2}$$

with country effects a_i capturing heterogeneous steady states. The steady states are detrended: common trend growth lives in μ_t, and a_i fixes the level of the country's balanced growth path relative to it. Their Proposition 1 shows that when Equation 2 generates the data, the pooled estimator of β in Equation 1 is inconsistent, biased toward zero by the covariance between a_i and income, time varying even under stable parameters, and degenerate at zero as countries approach their steady states. Estimating Equation 2 with fixed effects they find β strongly negative throughout, with no change over time. The rejoinder of KWY is that the within transformation discards the cross-country variation that the convergence question is about — a different exercise, and one with econometric issues of its own (Durlauf et al., 2005) — while accepting that the coefficients change substantially once country effects enter. Their reply contains three further elements that we build on. First, they decline to prejudge whether

[4] This reverses the design of the classical literature, which fixed the base year and stretched the horizon; here the horizon is fixed at ten years and the start year slides.

[5] A yearly dataset of overlapping decade growth rates would not reduce the bias of the within speed, which is governed by the calendar span of the sample (Section 4.3; Shiller and Perron, 1985; Perron, 1991), and it would make the errors serially correlated by construction.

convergence to fixed steady states or the evolution of steady-state income itself matters more for growth ("we do not want to assume away the latter", Kremer et al., 2022, p. 351), and they frame their correlates analysis as measuring that evolution. Second, they estimate a variant with country-decade fixed effects, which treat moving targets as unrestricted step functions. The variant is identified off the annual start-year panel, whose windows overlap nine years in ten, and it confirms that the within coefficient is stable. The dispersion of the estimated effects narrows from 1990 onward (their fig. A3), which they read as converging steady states. Third, citing Bernard and Durlauf (1996), they caution that time-series identification of convergence misleads when countries sit far from their steady states.

The country-decade effects and the between slope of the next section are related but answer different questions. A country-decade effect is what a window's cross section leaves unexplained by income, so its dispersion measures how unequally countries grow. The between slope β_d^B of the next section measures how income levels map into growth, which is the economic question the debate is about, and it comes with a standard error and a testable profile.

To this exchange we add an econometric fact the debate leaves untouched: with T short, the fixed-effects estimator of a dynamic panel is biased toward finding convergence (Nickell, 1981; Barro, 2015, p. 914), and at six decade transitions the bias is as large as the estimates themselves (Section 4.3). The democracy literature AM draw on (Acemoglu et al., 2019) uses the estimators built for this problem on annual data, where by its own simulations the bias they correct is small.

3.2 An encompassing correlated-random-effects specification

Both estimators are restrictions of one mixed-effects model: income enters split into the country mean and the deviation from it, each with its own decade slope,

$$\begin{aligned} g_{i,t} \;=\; & \sum_{d=1}^{6} \mathbf{1}\{t \in d\} \left[\beta_d^W \left(\log y_{i,t} - \overline{\log y_i} \right) \;+\; \varpi_d \, \overline{\log y_i} \right] \\ & +\; \mathbf{m}_i' \boldsymbol{\omega} \;+\; \delta \log y_{i,0_i} \;+\; u_i \;+\; \mu_t \;+\; \varepsilon_{i,t}, \end{aligned} \tag{3}$$

where $\overline{\log y_i}$ is the within-country mean of $\log y_{i,t}$ over country i's observed sample, $y_{i,0_i}$ its first observed income, $\mathbf{m}_i$ the vector of within-country means, over the same windows, of the time-varying regressors (the six products $\mathbf{1}\{t \in d\} \log y_{i,t}$, the six products $\mathbf{1}\{t \in d\} \overline{\log y_i}$, and the start-year dummies), and $u_i \sim N(0, \sigma_u^2)$ the residual country effect. The residual is Equation 2's a_i net of its projection on $(\mathbf{m}_i, \log y_{i,0_i})$, in the manner of Mundlak (1978), Chamberlain (1982), Wooldridge (2005), and Wooldridge (2019) (see the decomposition of Sections A.1 and A.2). It is uncorrelated with $\mathbf{m}_i$ and $\log y_{i,0_i}$ by construction, and the random-effects likelihood treats it as independent of the regressors given them. μ_t are time effects, entered as dummies given the six start years. The indicator $\mathbf{1}\{t \in d\}$ selects the decade window of start year t (1963–73, …, 2013–23), so each slope is an interaction with decade dummies, exactly as in Equation 1: within a window each is a single cross-sectional slope,

changing only over time.[6] Equation 3 is one regression estimated on the full panel, delivering six pairs of decade slopes, not six separate regressions. The distinction is immaterial in the pooled case — with no shared parameters, interacted slopes reproduce decade-by-decade cross sections exactly — but it is substantive here: $\boldsymbol{\omega}$, δ, σ_u, and u_i are common over time, so every decade's slopes are estimated holding the same country heterogeneity fixed, and the flatness of either profile is a testable hypothesis within one covariance matrix.

Three coefficients answer three different convergence questions. The *within* slope β_d^W, on the deviation, asks whether a country grows faster when its income is below its own path — convergence to the country's own steady state, the object of the fixed-effects literature; it is identified by comparing, within a decade, countries sitting below their own historical path with countries sitting above it. The *between* slope β_d^B, the response of growth to a country's income level with the initial condition held fixed, asks whether, among countries that started at the same income, the permanently richer one grows faster — the ordering of steady states; it is identified from the cross section of country means. The country mean contains the future income that defines growth, so β_d^B also carries a mechanical component, which changes little across decades when steady states are constant, declining by 0.12 between the first and fifth windows in simulation (Section 4.2.2); the data speak through its changes. The *unconditional* gradient $\beta_d^U \equiv \beta_d^B + \delta$, which lets the initial condition move with the level, asks the question of KWY: do poorer countries grow faster? The pooled coefficient is a variance-weighted combination of the within and between slopes; because the weights depend on the relative variances and on the panel's composition, the pooled slope can move without either margin changing — a restatement of the time-varying-bias result of AM.

The debate is then a sequence of nested restrictions on the wedge $\beta_d^B - \beta_d^W$, the gap between the two slopes. Set the wedge to zero in every decade and the projection with it, $\boldsymbol{\omega} = \mathbf{0}$ and $\delta = 0$, with $\sigma_u^2 = 0$, and the two slopes become one. Equation 3 is then Equation 1 estimated on the same six windows: the pooled specification. Section 4.2 tests the projection coefficients by likelihood ratio, holding the country effect free, and states what a rejection does and does not establish. Hold the wedge constant across decades instead, and the country effect is constant: this is Equation 2's a_i expressed through observables, and the within slopes equal the two-way fixed-effects estimates exactly (Wooldridge, 2019, Proposition 2.1) (Section A.1). The initial-condition term handles a separate problem. In a dynamic model the country effect is correlated with the first observed income by construction, so the random-effects likelihood is coherent only conditional on it (Wooldridge, 2005). Leaving the wedge free gives the object KWY absorb as a nuisance in their reply to AM. Section 4.2 tests both restrictions. No projection coefficient corrects the finite-T bias of the within slope, which is the subject of Section 4.3.

The debate turns on one quantity, the wedge, so it helps to write the model in coordinates that carry it as a single coefficient. Using the algebraically identical Mundlak coordinates of

[6] Unpacked, the term in the country mean reads $\varpi_1 D_{i,t}^1 \overline{\log y}_i + \cdots + \varpi_6 D_{i,t}^6 \overline{\log y}_i$ with $D_{i,t}^d = \mathbf{1}\{t \in d\}$: six regressors, of which exactly one is nonzero for any observation, so the slope on $\overline{\log y}_i$ switches with the observation's decade; the within term unpacks identically on the deviation.

Equation 3 — the ones we estimate —

$$g_{i,t} \;=\; \sum_{d=1}^{6} \mathbf{1}\{t \in d\} \left[\beta_d \log y_{i,t} \;+\; \gamma_d \overline{\log y_i} \right] \;+\; \mathbf{m}_i'\boldsymbol{\omega} \;+\; \delta \log y_{i,0_i} \;+\; u_i \;+\; \mu_t \;+\; \varepsilon_{i,t}, \quad (4)$$

with $\beta_d^W = \beta_d$ and $\varpi_d = \beta_d + \gamma_d$. The between slope β_d^B is the response of growth in window d to a unit rise of log income in every window with the initial condition held fixed. For a country observed in all six windows each element of $\mathbf{m}_i$ that averages an income term then rises by one sixth, so $\beta_d^B = \beta_d + \gamma_d + \bar{\omega}$, where $\bar{\omega}$ is one sixth of the sum of the twelve elements of $\boldsymbol{\omega}$ that load on those means. The wedge is $\gamma_d + \bar{\omega}$, and its changes across decades are those of γ_d, whose own level is not separately identified. The two coordinate systems are a nonsingular reparameterization of each other, so the fit is identical (this is the within–between parameterization of the multilevel literature; Section A.1). The constant-wedge model is Equation 4 with $\gamma_d \equiv \bar{\gamma}$, while the pooled specification sets γ_d, $\boldsymbol{\omega}$, δ, and σ_u^2 to zero.

This paper's conventions, then, are these.[7] *Country effect* refers to the object u_i (or a_i), with no commitment in the name. *Fixed effects* and *random effects*, used alone, refer to the panel-econometrics estimators and their identifying assumptions. *Mixed effects* refers to the estimation machinery that treats u_i as a draw from a distribution. In this vocabulary Equation 4 is a random-effects model in the panel sense whose Mundlak terms let the country effect correlate with income, delivering the fixed-effects estimand. The same structure carries over to estimators for which dummy-variable fixed effects are unavailable or inconsistent (Section A.10).

One feature of the specification drives every bias issue below. The country mean $\overline{\log y_i}$ is an average of the dependent variable's own path, so it is correlated with each disturbance $\varepsilon_{i,t}$ — the more strongly, the more of the country's subsequent history the mean contains. This is the mechanism of the bias of Nickell (1981), and its consequences differ only by which coefficient it reaches. In the within coefficient the correlation enters negatively and biases the slope toward convergence; its size at $T \leq 6$, and the failure of the standard remedies, are the subject of Section 4.3. In the between coefficient the same correlation enters positively and inflates the level. In a panel with constant steady states and no decade variation the between slope averages between 2.0 and 2.2 in every window, above every observed value, which is why the levels of split coefficients are treated as projection quantities throughout (Section 4.2). And because the correlation is stronger for early decades — their future occupies more of the mean — freed decade-varying coefficients absorb part of it as a spurious time pattern, which the rejections below partly reflect. The methodological literature is consistent with this account: the within–between model is developed for strictly exogenous covariates and does not treat the dynamic case (Allison, 2009; Bell and Jones, 2015; Bell et al., 2019); the correlated-random-effects device for dynamic models projects the heterogeneity on the initial condition and the exogenous covariates, not on the mean of the dependent variable (Wooldridge, 2005; Rabe-Hesketh and Skrondal, 2013); and mean-centering a lagged outcome

[7] The words "fixed" and "random" name different objects in panel econometrics (Wooldridge, 2010) and in the mixed-effects literature (Gelman and Hill, 2007).

is known to reproduce the Nickell bias in multilevel time-series models (Hamaker and Grasman, 2015), with latent-mean centering the proposed repair (Asparouhov and Muthén, 2019). The mean of income enters here regardless, because the between margin is the object of interest. Their movement is retained only where simulation can show that the correlation between the country mean and the disturbance does not produce it (Section 4.2). A lower bound on the speed of adjustment is estimated in Section 4.4 from predetermined regressors — current income and a target built from the correlates — which contain no country mean.

Two things would make γ_d vary across decades. The correlation just described is one: it is stronger for early windows, so a freed coefficient absorbs part of it. Factor structure is the other — common shocks loading on a country's permanent level, which no time-invariant effect absorbs (Pesaran, 2006; Bai, 2009).[8] If the wedge does vary, the within–between split is unstable from decade to decade, which is why the profiles below are read off the sums, β_d^B and β_d^U, estimated decade by decade under no wedge restriction. The within column of Table 1 is then the debate's own estimator, $\beta_{d,\bar{\gamma}}^W$: the within slope under $\gamma_d \equiv \bar{\gamma}$. Country dummies absorb only a time-invariant loading on $\overline{\log y_i}$, so fixed effects impose that same restriction. With the means of every time-varying regressor in the projection the two columns coincide exactly (Wooldridge, 2019, Proposition 2.1). Its level carries the finite-T bias of short dynamic panels (Nickell, 1981) — derived in Section A.4, quantified in Section 4.3 — so it is reported descriptively.

The difference $\beta_d^U - \beta_d^B$ is the coefficient on $\log y_{i,0_i}$, the country's income at the start of the sample. It is the regressor of the classical cross-sectional growth regression (Barro and Sala-i-Martin, 1992). There it carried the convergence hypothesis; here it is a control, and the question it answers is different — whether growth in each later decade still depends on where a country stood in the 1960s, conditional on where it stands now and on its own average. The first window is structurally excluded, since there the initial condition *is* current income and no separate δ exists. Freeing δ by decade is imprecise for the reason the wedge split is: $\overline{\log y_i}$ and $\log y_{i,0_i}$ are collinear across countries, as income and its mean are within them, while the total gradient β_d^U is well determined either way (Section A.6). For a country observed in all six windows the initial condition equals six times the element of $\mathbf{m}_i$ that averages first-window income, so the level of β_d^B is separated from δ only by the 53 countries that enter the panel after 1963, while β_d^U and the changes of β_d^B across decades are identified even on the 111 countries observed throughout. The same structure therefore governs both levels of the decomposition. β_d^U separates into β_d^B and δ, and β_d^B into β_d^W and the wedge, only under a restriction on how the components vary over time, and Section 4.2 and Section A.6 test each restriction in turn. The conclusions below rest on the sums, which need no such restriction; the components are reported with the restriction that defines them stated.

[8] A factor of this kind would also correlate residuals across countries, which the country-clustered standard errors used below do not allow for. Net of the window effects the excess dependence is negligible. An average pair of countries shares 5 decade windows, and at that length independent series already produce a mean absolute pairwise correlation of 0.46; the residuals give 0.46. Section A.9 reports this, and why the CD statistic of Pesaran (2004) cannot be read against a standard normal on this design.

4 Results

4.1 The moving unconditional convergence

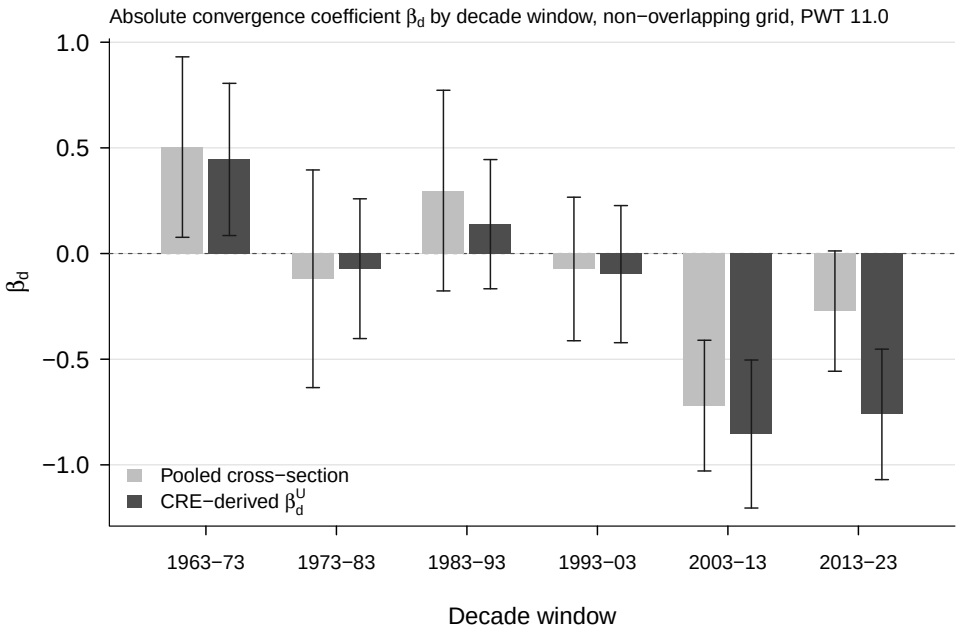


Figure 1: Absolute convergence coefficient β_d across the non-overlapping decade windows 1963–2023 (PWT 11.0). Light bars: the pooled cross-sectional gradient, standard errors clustered by country. Dark bars: the unconditional gradient β_d^U implied by the encompassing model below, CR2 cluster-robust standard errors. Whiskers: 95% confidence intervals.

PWT 11.0 adds six years of data,[9] extending the last full ten-year growth window to 2013–23 (on the non-overlapping grid of Section 3.1). The annual coefficient series replicates the published trajectory — no clear pattern in the 1960s–1970s, a drift toward divergence in the 1980s, a slide into convergence from around 1990 — and then adds a new phase: the series bottoms at -0.85 in 2006 and retraces. Figure 1 reports the same arc on the non-overlapping grid, where adjacent windows share no years: the final window gives a pooled coefficient of -0.272 (s.e. 0.145) for 2013–23, against -0.720 (s.e. 0.158) for 2003–13: absolute convergence, at its strongest in the windows ending around 2017, is no longer significant in the most recent window. The figure's second series is the unconditional gradient β_d^U implied by the correlated-random-effects model of Section 4.2: it tracks the pooled gradient window by window and parts company only in 2013–23, where the pooled gradient retreats to insignificance while β_d^U remains significantly negative. The 2013–23 window contains the pandemic collapse and recovery, so this retracement conflates any change in the convergence process with pandemic-window noise; the encompassing model below shows where it loads.

[9] On KWY's own replication data, their Table 1 is reproduced exactly; re-estimated on PWT 11.0, the decade profile matches the published one throughout. Section A.12 reports both.

Everything that follows is a β-object, a slope linking income levels to subsequent growth. So are KWY and AM: KWY report σ-convergence alongside their main results, and AM do not treat it. A negative β (mean reversion) does not imply a narrowing distribution; the inference from one to the other is Galton's fallacy (Quah, 1993; Durlauf et al., 2005). The unweighted cross-country standard deviation of log income per capita widens over the sample. On the unbalanced panel it runs from 0.94 in 1960 to 1.25 in 2000, but the country count grows from 109 to 163 over those years, so that comparison mixes a change in dispersion with a change in sample; the balanced figures below are the ones to read. It then falls to 1.20 by 2019 — the narrowing KWY date to the early 2000s — before the pandemic windows return it to 1.24. Only the widening is measured with any precision. On the 109 countries observed at all four dates, a country bootstrap puts the 1960–2000 rise at 0.37 (s.e. 0.05, $p < 0.001$) and the 2000–2019 fall at 0.04 (s.e. 0.05, $p = 0.34$); the post-pandemic retrace is 0.04 (s.e. 0.03, $p = 0.34$). The narrowing is therefore visible in these data but not distinguishable from no change, and it is an eighth of the earlier widening. Over the sixty years as a whole dispersion is wider than it began, on the balanced set as on the full one. A flat aggregate conceals movement in both directions: of the 108 countries other than the United States observed throughout, 60 narrowed their gap to the United States and 48 widened it. Ranks move little even so: the correlation of income ranks across consecutive decade nodes never falls below 0.96.

4.2 The encompassing model: convergence

Table 1: Decade profile of the convergence coefficient across nested estimators

Decade	Pooled OLS	Country FE	Within $\beta^W_{d,\bar{\gamma}}$	Between β^B_d	Unconditional β^U_d
1963–73	0.504 (0.218)	-1.944 (0.442)	-1.944 (0.450)	1.421 (0.627)	0.445 (0.184)
1973–83	-0.119 (0.263)	-2.378 (0.406)	-2.378 (0.412)	0.904 (0.676)	-0.072 (0.169)
1983–93	0.298 (0.242)	-2.221 (0.310)	-2.221 (0.313)	1.114 (0.704)	0.139 (0.156)
1993–03	-0.073 (0.173)	-2.612 (0.349)	-2.612 (0.353)	0.878 (0.735)	-0.098 (0.166)
2003–13	-0.720 (0.158)	-3.321 (0.321)	-3.321 (0.324)	0.121 (0.716)	-0.854 (0.179)
2013–23	-0.272 (0.145)	-2.977 (0.317)	-2.977 (0.321)	0.214 (0.698)	-0.761 (0.158)

Notes. Extended grid to 2013–23; 881 country–years and 164 countries (880 and 163 under country FE). Pooled OLS and country FE estimate Equation 1 without and with country intercepts. Within: $\beta^W_{d,\bar{\gamma}}$ under $\gamma_d \equiv \bar{\gamma}$ (Equation 4), identical to the country-FE estimates. Between: $\beta^B_d = \beta_d + \gamma_d + \bar{\omega}$, for a country observed in all six windows. Unconditional: $\beta^U_d = \beta^B_d + \delta$, common δ. All models include time dummies. Standard errors clustered by country, CR2 in the last three columns. Section A.6 gives the column definitions in full.

4.2.1 The decade profile

The mixed-effects estimates of Equation 4 (restricted maximum likelihood, REML; estimation diagnostics for all models are collected in Section A.13) bear on the debate as follows.[10]

First, the pooling restriction is rejected: the likelihood-ratio test that every projection term is zero — the country means of the decade-interacted income and of the start-year dummies, and the initial condition — gives $\chi^2(10) = 323$ ($p < 10^{-15}$), with $\hat{\delta} = -1.96$ (s.e. 0.30) in the constant-wedge fit. In both fits the country-effect variance is estimated at zero, so the projection absorbs the whole country effect and REML coincides with least squares on the same regressors. The between and unconditional columns of Table 1 and Figure 2 come from the decade-varying-wedge fit, where $\hat{\delta} = -0.98$ (s.e. 0.59); that is the value differencing the table's between and unconditional columns. Yet, more is needed to measure differences in steady states, as shown below.[11] With log income detrended by window, the between-country variance of country means is 1.19 against a within-country variance of 0.12, a ratio of 10 to one. Under a single common steady state every country would be a transitory deviation from the same path, and the country means would differ only by sampling: at the estimated decade persistence and this panel's length that mechanism delivers a between variance of 0.063, smaller than the observed by a factor of 19. That calibration needs a persistence which Section 4.3 shows is not identified, so the factor is reported across its range: 55 with no persistence at all, 19 at the $\rho_{10} = 0.65$ used here, and 10 at a unit root, which is its floor. Steady states differ, by an amount no transitory process on this panel could generate, and the pooled slope of Equation 1 therefore cannot be read as a convergence rate, consistent with the analysis of AM. Consistently, the within slopes, in slope units throughout, are strongly negative in every decade, between -1.9 and -3.3 percent per year (Table 1). As a specification check, the equivalence of Section 3.2 holds in the estimates: the correlated-random-effects within column and the fixed-effects column coincide in every decade. With the country means of every time-varying regressor in the projection, the GLS within slopes are the two-way fixed-effects estimates exactly, on an unbalanced panel as on a balanced one (Wooldridge, 2019, Proposition 2.1). The two estimators therefore also share their finite-T bias; Section 4.3 establishes what the within column can measure. The constant-wedge restriction is itself

[10] REML estimates the variance components from residual contrasts — the data projected orthogonally to the fixed effects — removing the downward bias that maximum likelihood incurs by ignoring the degrees of freedom the fixed effects absorb; given the variance components, the fixed-effect estimates are generalized least squares. Because REML likelihoods are not comparable across different fixed-effects structures, all likelihood-ratio tests we report use maximum-likelihood refits.

[11] The likelihood-ratio statistic cannot itself measure how much steady states differ. Its regressors are country averages that contain the future income defining growth, so they are correlated with the outcome whether or not steady states differ. Simulating a panel with one common steady state on this panel's own window geometry, where the restriction is true and pooled least squares consistent, the test rejects at the one percent level in every replication. Nor is this a short-panel artifact that longer histories would cure. In a balanced panel with one common slope, where the projection reduces to the country mean and the initial condition, $\hat{\delta}$ does fade with T, but the loading on the country mean, 7.6 at $T = 6$, settles near 2.7 and the rejection rate stays at one, because with a common steady state the country mean has no permanent component, so its variance shrinks at the same rate as the mechanical covariance and their ratio does not. The test serves here as a specification statement only.

rejected: a likelihood-ratio test of $\gamma_d \equiv \bar{\gamma}$, jointly with the country means that the decade-varying loading adds to the projection, gives $\chi^2(7) = 48.6$. Freeing the wedge lowers the between column by 0.84 to 1.14 in every window, while $\hat{\delta}$ moves from -1.96 to -0.98, so the unconditional column moves by at most 0.16, and the fall in the between column across the panel is -1.03 under the restriction against -1.21 without it. The within profile that the same fit delivers runs from -0.17 to -5.47, a range of 5.31 against 1.38 under the restriction, with standard errors of 0.50–0.90 against 0.24–0.31. The movements of the two columns that carry the argument are insensitive to the restriction; the within column is not, which is the identification result of Section 4.3 in another form and the reason the within slope is reported under the restriction the debate itself imposes.

The estimates also say how the model's components sit in the variation of decade growth. Regroup Equation 3, folding the initial condition into the level term:

$$g_{i,t} = \underbrace{\beta_d^W \left(\log y_{i,t} - \overline{\log y_i}\right)}_{\text{gap}} + \underbrace{\varpi_d \overline{\log y_i} + \mathbf{m}_i'\boldsymbol{\omega} + \delta \log y_{i,0_i}}_{\text{level}} + \underbrace{\mu_t}_{\text{common time}} + \underbrace{u_i}_{\text{country}} + \varepsilon_{i,t}, \tag{5}$$

where the *gap* term carries the within slope applied to the country's distance from its own average, and the *level* term is a function of the country's income level and history. One implication of these components matters for interpreting the results (Section A.6 adds the anatomy of the residual country effect).

At the fitted within slopes the *gap* term moves decade growth by $|\beta^W| \times 0.49 \approx 1.2$ percentage points a year, some 13 percent of income compounded over a decade. The idiosyncratic shock is larger, $\hat{\sigma}_\varepsilon = 2.12$, and the common time effects μ_t larger still. The pull therefore shows only in accumulation. That is why the annual β_t series is noisy, why the full panel is needed, and why a single pandemic window moves the picture. Seen this way, the sign reversal in β_d^U was not a change in what dominates growth outcomes — decade growth remains mostly shocks — but the fading of a systematic tilt: since $\beta_d^U = \beta_d^B + \delta$, with δ held common across decades the reversal is exactly the level-aligned divergence gradient β_d^B falling below $-\delta$ — a movement built entirely from cross-sectional projections.

Second, the convergence coefficients vary over time, which is the substance of the KWY headline. The between gradient sits near 1 through the first four windows and drops to near zero in 2003–13, where it stays. Cluster-robust Wald tests on the six coefficients (CR2 by country, referred to the small-sample F of Pustejovsky and Tipton, 2018) reject a flat profile ($F(5, 60) = 8.71$, $p = 3.0 \times 10^{-6}$) and, at the 5 percent level, a linear trend ($F(4, 61) = 2.75$, $p = 0.036$), while a single break at 2003–13 is not rejected ($F(4, 61) = 1.22$, $p = 0.31$). What the data show is one level shift, not a gradual decline. The between slope β_d^B falls from 1.42 (s.e. 0.63) in 1963–73 to 0.12 (s.e. 0.72) in 2003–13. The correlation between the country mean and the disturbance shifts its level (Section 3.2): the null with constant steady states of Section 4.2.2, in which income does not predict growth, puts it at 2.0–2.2 in every window, above every observed value. Only its movement is read. The movement has an antecedent inside the fixed-effects coordinates: the narrowing dispersion of country-decade fixed effects

that KWY report from 1990 onward (their fig. A3).

The unconditional gradient assembles these pieces into the object KWY plot. β_d^U is positive in 1963–73 (0.45, s.e. 0.18), hovers around zero through the 1990s, and is significantly negative from the 2000s (-0.85, s.e. 0.18 in 2003–13). It tracks the pooled column of the table, so absolute convergence is reproduced as a derived quantity.

The decomposition shows where that convergence comes from. The initial-condition term attributes the faster growth of poor countries to their steady-state shortfalls: today's poor countries are disproportionately those with the deeper shortfalls. That term is imprecisely measured: $\hat{\delta} = -0.98$ (s.e. 0.59, $p = 0.10$) is not separated from zero at the 5 percent level. Over the same decades their growth correlates caught up (Section A.8). Growth closes those gaps while they stay open, at a speed the decade panel cannot pin down (Section 4.3) and that Section 4.4 bounds from below. The positive β_d^U of the early decades is the same mechanism with the gaps held elsewhere. The fast growers of the early 1960s sat around or above the middle of the income distribution, in the economies of postwar reconstruction and late industrialization, so the pull toward steady states then tilted the cross-section toward divergence. The sensitivity of convergence findings to who populates the early cross-section is a classical point (DeLong, 1988).

Table 2 documents the passenger list. The decade's ten fastest growers sat around or above the middle of the start-of-window income distribution through the 1980s (median percentiles 44 to 62) and in its bottom third from 1993–2003 onward (30 to 32). The ten slowest sat in the bottom third in 1963–73 and in the upper half of the two most recent windows. The cross-section changed sign because the fast growers moved down the income distribution. Measured against the fitted targets of Section 4.4, the fastest ten sit at or below the median of the gap distribution in every window (median gap percentiles 19 to 50). Fast growers hold the deeper shortfalls from target throughout; what changed is the incomes at which those shortfalls sit.

Figure 2 sets the identity out window by window: the light bars are β_d^B, the dark bars β_d^U, and the arrow between them is δ. With δ held common across decades, absolute convergence appears exactly where β_d^B falls below $-\delta$, the dotted rule. Freed by decade (Section A.6), the unconditional gradient barely changes, and the 2003–13 step divides between the two components: the between gradient falls by 0.33 and δ_d by 0.41.

Within convergence contributes no catch-up by itself — trend reversion around dispersed steady states closes no cross-country gaps — but it sets the speed at which movements in the steady-state configuration become realized income; a rising steady state widens a country's gap and *raises* its growth, which then decays exponentially as the gap closes.

Third, the extended grid shows what the retracement of the pooled coefficient after its 2006 trough is made of. From 2003–13 to 2013–23 the between slope moves from 0.12 to 0.21 and the fitted within profile from -3.32 to -2.98: under the common-δ device a modest re-steepening of the cross-sectional income–growth gradient and a modest retreat of the within profile, both small against their decade-scale movements since the 1960s — an allocation

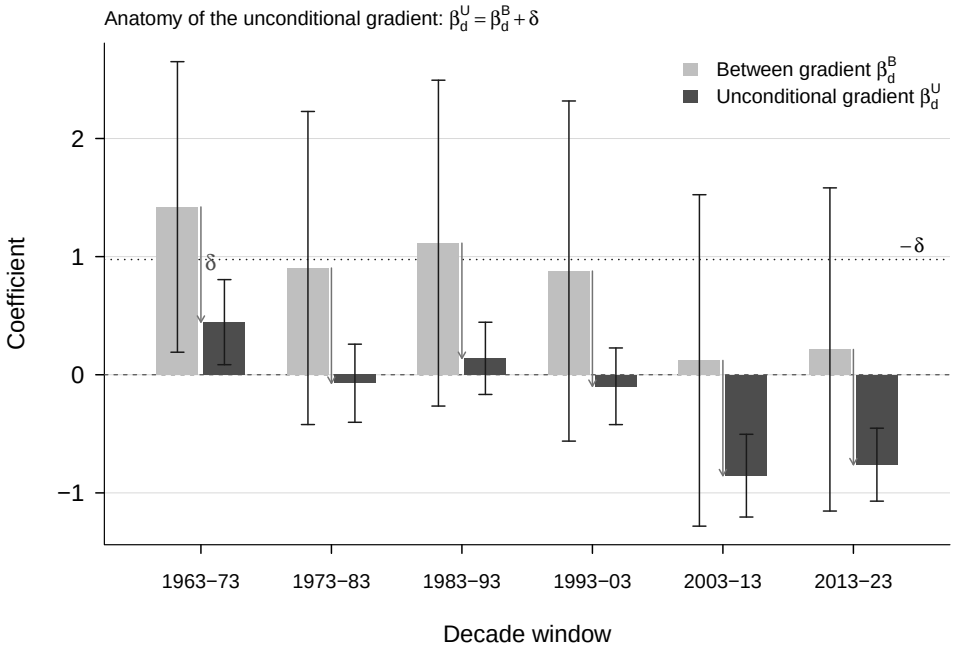


Figure 2: Anatomy of the unconditional gradient. Light bars: the between-country income gradient β_d^B, which holds the initial condition fixed. Dark bars: the unconditional gradient $\beta_d^U = \beta_d^B + \delta$, the object comparable to the pooled cross-section. The arrow is the initial-condition coefficient δ, common across decades; the dotted rule marks $-\delta$, the level at which β_d^U changes sign. Whiskers are 95% confidence intervals from CR2 cluster-robust standard errors. Decade-varying-Mundlak variant, non-overlapping grid, PWT 11.0.

Table 2: The passenger list: median start-of-window income percentile of each decade's ten fastest- and ten slowest-growing countries (percentiles within the decade's own cross-section, extended grid, PWT 11.0), the fastest ten's median percentile in the distribution of gaps to the fitted targets of Section 4.4 (composite sample; low percentile = far below target), and the five fastest growers by ISO code.

Decade	Ten fastest	Ten slowest	Fastest, gap pctile	Five fastest growers
1963–73	44	30	33	BWA, TWN, SGP, JPN, GRC
1973–83	62	61	50	MLT, KOR, MAC, TWN, ROU
1983–93	50	48	35	CHN, KOR, MDV, BLZ, TWN
1993–03	31	41	19	BIH, MLI, ARM, CHN, MMR
2003–13	30	83	22	AZE, MAC, CHN, MMR, ETH
2013–23	32	62	36	GUY, IRL, BGD, CHN, VNM

the freed initial-condition projection reverses (Section A.6). The division of the retracement between the two channels is therefore not robust, though the retreat of the unconditional gradient toward zero is. The convergence regime of the 2000s weakens in the pandemic window and remains significant in the unconditional gradient; whether the weakening is cyclical will only be visible in windows that clear the pandemic years. The timing, moreover, splits the retracement into a pre-pandemic and a pandemic part: the annual series bottoms at -0.85 in 2006 and has already retraced to -0.45 (s.e. 0.16) by the 2009–19 window — the last that clears the pandemic entirely — so roughly two-thirds of the move off the trough predates COVID, with the remainder, to -0.27 in 2013–23, spanning it (the annual series overlaps adjacent windows and the 2009 start sits at the trough of the financial crisis, so the split is indicative rather than sharp).

4.2.2 Alternatives to a change in the world

The decline is a movement in an estimated coefficient, so Section A.7 asks what else could produce it. Each alternative is calibrated to this panel and set against the observed fall, 1.30 when the between slopes are estimated by least squares on the regressors of Equation 4 as in the simulations, the same value as in the REML fit of Table 1: the mechanical endogeneity of the country mean, which declines over time by construction; the changing composition of the panel; the cyclical position of the single-year income anchor; steady states that drift; the influence of single countries and country groups; more than one convergence regime; a common factor with income-correlated loadings; and the design PS recommend, with a trending factor, heterogeneous loadings and heterogeneous dynamics. Apart from regimes, which the data cannot settle, none of them produces the fall. A null with constant steady states manufactures 0.12 (simulation s.d. 0.22), and the null built to PS's design moves the profile the other way. A common factor reaches the observed decline in a tenth of replications only when the common shock is larger than the entire cross-country spread of growth rates it would explain. Within regimes the decline is imprecise: for either regime, with the threshold

fixed at the median or chosen by the data, it does not differ detectably from the pooled decline (country-bootstrap p between 0.33 and 0.58).

Taken together, the estimates connect the two positions. The level of the pooled coefficient conflates within and between slopes that are large and of opposite sign, and carries no structural interpretation, which is the specification concern of AM. Its movement survives that concern. The decline of the between margin exceeds what the mechanical endogeneity of the country mean manufactures, and what any of these alternatives manufactures, and the unconditional gradient built on it reproduces the trend documented by KWY. Heterogeneity that moves over time, such as factor structures and heterogeneously trending steady states (Pesaran, 2006; Bai, 2009), is absorbed by no time-invariant country effect; Section 4.3 takes up what that leaves the within margin able to measure.

4.2.3 What the between gradient can and cannot say

The object carrying the movement is a cross-sectional projection, and cross-sectional convergence regressions are the ones this literature has most often rejected. Three objections stand. Quah (1993) shows that a negative slope does not imply a narrowing distribution. Bernard and Durlauf (1996) and Durlauf et al. (2005, §4.2.2) show that under multiple steady states the slope is governed by a within-sample covariance between initial and steady-state income, so it moves with composition while no structural parameter changes. PS add that the cross-country regression is biased toward zero unless the heterogeneity of the intercepts is modeled exactly.

Each of these concerns the level of β_d^B, and nothing here uses that level. Projection arithmetic alone puts it at 2.0–2.2 under the null with constant steady states (Section A.7), which is why it is treated above as a projection rather than as a structural gradient.

The claim is about the profile, and a bias that is the same in every window leaves a profile unchanged. The objections therefore bite to the extent that the bias itself moves across windows, which is a measurable quantity and the one these nulls measure. The mechanical endogeneity of the country mean drifts by 0.12 against the 1.30 observed; country-specific target drift correlated with income manufactures between -0.34 and 0.23; a trended factor with heterogeneous loadings and heterogeneous dynamics manufactures between -0.54 and -0.34, which is to say a rise.

One question the panel cannot answer is where the movement comes from. The between gradient is the sum of the within slope and the wedge, and either could have moved. The unconditional gradient cannot separate them: with a common initial-condition coefficient it follows the between gradient one for one, whatever the within slope does. Nor can the pooled slope. Simulating growth on the observed income data with three different within paths — constant, and the two estimated ones — reproduces the observed pooled decline of 1.22 in every case: 1.09, 1.22 and 1.08, against simulation standard deviations of 0.23 to 0.25. That the gradient fell is measured. Which margin moved is not.

On the distributional question this paper takes no position beyond what Section 4.1 says about

β-objects. Johnson and Papageorgiou (2020) conclude that there is no evidence supporting absolute convergence in the cross-country distribution, and the dispersion evidence here agrees with them: the unweighted standard deviation of log income is wider in 2023 (1.24) than in 1963 (0.98), and ranks are close to immobile: their correlation across consecutive decade nodes never falls below 0.96. That is consistent with the gradient KWY document, because the two are different objects: a conditional-mean slope can turn negative while the distribution it is fitted to fails to contract. This paper's finding concerns the first and leaves the second where Johnson and Papageorgiou left it.

4.3 What the within margin measures at decade frequency

The within column of Table 1 is the object both fixed effects and the correlated-random-effects model deliver, and this subsection asks what it measures. That short dynamic panels distort the answer has long been known: Islam (1995) chose his estimators by simulating it, Caselli et al. (1996) built a GMM estimator around it, Hauk and Wacziarg (2009) measured it by simulation, and Barro (2015) invokes it against panel convergence rates of the size reported here; what follows quantifies the hazard at this debate's frequency and tests the remedies on this grid. Section A.4 derives the finite-T bias the within column inherits: demeaning over a country's short history manufactures mean reversion, the more so the shorter the history. At the estimated persistence, the exact bias formula gives -3.2 to -6.0 in the units of β_d^W across the panel lengths in our data — the same order as the estimates themselves. A within slope of -3 at six decade transitions is, on its face, compatible with no within convergence at all; whether it reflects convergence, bias, or a mixture cannot be settled from the size of the coefficient. Nor is the problem one of too few observations. Sampling the same sixty years more often shrinks the bias in the coefficient the regression estimates, but that coefficient is a per-period object; converting it to a rate per year multiplies by the reciprocal of the sampling interval, and to leading order the two effects cancel. In terms of the annual convergence rate r, the leading term of the bias is $2/S$ a year whatever the interval, where S is the calendar span of the sample in years. The exact bias does fall as the grid is refined, but only toward a limit the span fixes: at the benchmark persistence of Section A.4 it is 6.56 points a year at decade spacing, 4.42 at annual spacing and 4.29 in the limit of continuous sampling. In that limit the bias is a function of rS alone, the expression Barro (2015, eq. 1, p. 914) derives from Nickell's formula, whence his conclusion that observing a fixed span more often does not reduce it. Consistency therefore requires a longer sample. That span-not-frequency point is long established for unit-root testing (Shiller and Perron, 1985; Perron, 1991): consistency requires the span to grow, and sampling a fixed span more finely does not deliver it. The $2/S$ term is the bias at a given true speed; as the true speed goes to zero, the speed the estimator reports tends to $3/S$ a year, the floor used below. The prescription can be taken to data. The Maddison Project reaches past 1900, and on the 38 countries observed at every decade node since 1902 the same within estimator gives -2.50 (s.e. 0.66) over 1962–2022 and -1.67 (s.e. 0.58) over 1902–2022. With composition held fixed, doubling the span moves the slope by $+0.83$, where the exact bias formula predicts $+1.65$ to $+1.91$ under constant targets, a premise the floor argument below rejects. The contrast has the predicted sign and about half the

predicted size. At $S = 120$ the remaining bias, 1.6 to 2.0 in the same units, is still the size of the estimate itself. A true within speed near the bound of Section 4.4 is therefore consistent with the long-span estimate, although the remaining bias is large enough that speeds well above it are not excluded.

The standard remedies fail at this frequency. Difference GMM (Arellano and Bond, 1991) instruments the decade change in income with lagged levels; on the non-overlapping grid the instruments are irrelevant (first-stage partial $F = 3.8$), the point estimate lands above the unit root ($\hat{\rho}_{10} = 1.09$, where $\rho_{10} = 1 + \beta^W/10$ is the decade autoregressive root; Section A.4), and the estimate is uninformative. The failure comes from the frequency: in the annual democracy panels of the literature AM draw on (Acemoglu et al., 2019), countries contribute close to forty transitions — the mechanical bias is an order of magnitude smaller and lagged levels retain predictive power for annual changes — while at decade horizon income is close to a random walk, and four usable transitions leave the instruments nothing to bite on. Annual instruments do not repair them while the equation stays at decade horizon: annual lags dated before the earlier window — the closest valid dates — raise the first-stage partial F to 7 and no further, below conventional thresholds, with point estimates where weak-instrument bias places them.[12] Inverting the exact bias formula fails likewise: under a stationary start, no persistence at $T = 4$–6 maps into the fixed-effects estimate ($\hat{\rho}_{10} = 0.71$). A start further from the targets can raise the estimate beyond that range (Section A.4), so the failure speaks against a stationary start as much as against constant targets.

System GMM (Blundell and Bond, 1998) — the standard prescription for this weakness in growth panels (Bond et al., 2001) — adds level equations instrumented by lagged differences. The instrument it adds is strong (first-stage partial $F = 41$), but its power comes from moments whose validity the data reject: system GMM requires initial deviations that are mean-stationary around constant or commonly trending steady states. The Sargan–Hansen overidentification test asks whether the instruments are jointly orthogonal to the errors, so that a rejection indicts the instruments or the specification they are asked to satisfy; it rejects in every system variant and never in the difference variants. The incremental statistic — the difference between the system and difference Sargan statistics, distributed χ^2 with as many degrees of freedom as added moments — isolates the level conditions and rejects them at $C = 7.1$ ($p = 0.008$), or $C = 8.5$ ($p = 0.004$) under one-step common weighting, so the rejection is not an artifact of two-step weight matrices. The fragility of these level moments in cross-country growth applications has its own literature (Roodman, 2009; Bazzi and Clemens, 2013).

When both routes are re-run at five-year spacing over the same span, and at five-year spacing over 1963–88 — the frequency and roughly the period of Caselli et al. (1996) — the incremental statistic weakens from $p = 0.008$ on the decade grid to 0.034 at five years and 0.46 on the short early span, where it does not reject at all. The point estimate does not move with it:

[12] Shortening the window does not help much either: halving it doubles the number of transitions but also doubles the annualization factor, so the bias in annualized units falls only toward the limit the span fixes, and the shorter horizon brings its own measurement-error problem (Section A.4).

system GMM returns a decade persistence of 1.08, 1.09 and 1.36 in the three designs, at or above unity in every one, and the orthogonalized-effects posterior of Section A.4, which uses no instruments at all, places its mode at 1.00. Three designs, two frequencies, two spans and two unrelated corrections for the finite-T bias give one answer: under constant targets every estimator places the root at or above unity.

The statistical claim is best carried without those standard errors. Inverting the difference-GMM moments gives an Anderson–Rubin confidence set (Anderson and Rubin, 1949) that requires no assumption of strong identification: clustered by country, it rejects the interior interval $(-0.65, 0.94)$ and nothing else on the grid $[-2, 3]$, so the set that survives is disconnected and covers 68 percent of it. A valid confidence set that is the complement of an interval is what weak identification produces (Dufour, 1997, Theorem 3.3 and Sect. 5), and it agrees with the first-stage F. Restricting to monotone adjustment, $0 < \rho_{10} < 1$, what survives is $\rho_{10} \geq 0.95$, with the fixed-effects value itself rejected at $p = 0.008$ and the differenced-error structure never rejected (AR(2) $p \geq 0.27$ across all variants). A country with a reported speed of s closes s percent of the gap to its own steady state each year, so the fixed-effects slope of -2.94 is a speed of 2.94 read linearly, or 3.48 in the log convention of Section A.4. In those terms, under a constant-target model the data reject within convergence faster than about half a percent a year, and reject the fixed-effects reading specifically.

This agrees with the verdict PS reach analytically: once countries are allowed trends that are not parallel, the speed a demeaned estimator reports is not the speed of anything. It also bounds what can be said about the earlier literature. Caselli et al.'s convergence estimate of roughly ten percent a year comes from difference GMM on five-year intervals, where the instruments are weak in the same way: the difference route returns persistences of 1.1 and 2.4 in the same designs, and on the decade grid its estimates swing from $\hat{\rho}_{10} = 0.40$ to 1.70 as instrument choices vary. Bond et al. (2001) reach the same verdict from the other side: a consistent estimate should lie between the within-groups and pooled least-squares bounds, and on the data and model of Caselli et al. it does not (Durlauf et al., 2005, p. 633). On a sample like theirs the overidentifying restrictions are not rejected, so the objection concerns identification.

The remaining consistent route is likelihood-based: the orthogonalized-effects posterior of Lancaster (2002) (Section A.4), which under constant country effects is free of the Nickell bias, and it conditions on each country's first observation, so on simulated constant-target panels it recovers the true persistence whatever the start, strong convergence included (Section A.4). Every estimator so far maintains constant steady states; the Lancaster posterior estimates that model without the bias, so its answer is the constant-target model's own. On the balanced decade panel, fixed effects give $\hat{\rho}_{10} = 0.62$ — in growth units a within slope of -3.8; the balanced-subsample counterpart of the $\hat{\rho}_{10} = 0.71$ above — while the Lancaster posterior concentrates at the unit root: its log-posterior rises monotonically over the entire stationary region — no interior mode exists (Figure A.1) — so the mass piles up at $\rho_{10} \to 1$, a within slope of zero.[13] Had the fixed-effects reading been right — constant targets with

[13]The correction factor grows without bound beyond the stationary region and the posterior is not integrable

strong convergence — the bias-free posterior would have found it. Its verdict therefore leaves two worlds standing: constant targets with no within convergence at all, or steady states that move. A comparison that does not depend on the start excludes the first. At a unit root the initial level is absorbed by the country effect, so the fixed-effects estimate has the same distribution whatever the start. Simulated on this panel's structure, it averages 0.55 (s.d. 0.04), and none of 2,000 replications reaches the observed 0.71.

Bernard and Durlauf (1996) define convergence for *pairs* of countries, in two senses. The weaker is *catching up*: the expected gap between a pair narrows over a given horizon, which is what Equation 1 tests. The stronger is equality of their long-run forecasts, which Pesaran (2007) relaxes to stationarity of each pairwise output gap around a constant and tests across all pairs of countries. Reversion of a country toward its own detrended path appears in neither. By itself it is a persistence property: with country-specific trends, reversion can be fast while the cross-country distribution of income is set by the trends alone (Lee et al., 1997, 1998; Islam, 1998). The same holds for any within rate, PS's included, which is why the distributional question is carried above by the between margin and by the dispersion evidence.

PS reject the two-way fixed-effects estimator, and under a stationary start the span alone reaches the same verdict. They report what that estimator gives on annual data for 1950–2019: a speed of about 3 percent a year. It is the estimator Section A.4 describes, and their span is 70 years. With a stationary start, no covariates and serially uncorrelated errors, the slowest speed it can report under constant targets is 4.29 percent in rate units. Heterogeneous persistence across countries leaves that floor in place (Section A.4). Their figure lies below it. So does the fixed-effects slope here, 3.48 percent (s.e. 0.46) in the same rate units, or 2.94 read as a slope, against the 5.00 a sixty-year span allows.[14]

Under a stationary start, two designs, one spanning sixty years and one seventy, report speeds their own spans forbid under constant destinations. Without that assumption, the Lancaster posterior and the unit-root comparison reach the same conclusion: the destinations move.

None of this touches the profiles of the cross-sectional objects. β_d^B, β_d^U, δ, the correlate cross-sections and the accounting decomposition of Section 4.5 are projections in which no lagged regressor is demeaned. The correlation between the country mean and the disturbance shifts the level of β_d^B (Section 3.2), but under the null with constant steady states it moves the profile by 0.12, against the 1.30 observed (Section 4.2.2). The objections of AM to the pooled level and of PS to demeaned estimators, whatever their force against the pooled coefficient and the within column, leave the movement standing. The within column is therefore reported descriptively; its level is not a speed. Whether countries adjust toward their targets at all,

there (Figure A.1), so the corner is the estimator's answer.

[14] Their panel is unbalanced, averaging sixty-three annual observations per country, so $S = 70$ is the most conservative choice: the floor rises as the span shortens, to 4.76 at $S = 63$. Their figure is read from their abstract rather than from a table entry. The derivation covers the two-way estimator and does not extend to the estimators PS recommend in its place, which are consistent only as the time dimension grows and, as PS note, retain a small-T bias analogous to the Nickell bias whether or not a latent factor is present.

and at what minimum rate, is taken up in Section 4.4, which bounds the speed from below with an estimator that involves no demeaning.

4.4 A bound on the within speed

The within column of Table 1 carries no identified level, so the debate's second question — how fast a country closes its own gap — is not answered by it. A bound on that speed can be estimated without demeaning. Build a target for each country from the three growth correlates, in the KWY normalization, and regress decade growth on the gap between income and that target:

$$g_{i,t} \;=\; \beta\left(\log y_{i,t} - \hat{y}^{*}_{i,t}\right) + \mu_t + \varepsilon_{i,t}. \tag{6}$$

Unrestricted, this is the conditional convergence regression in panel form, and it involves no demeaning anywhere: if the observables span the steady states, the gap slope is the speed.

The gap slope, in slope units, is −0.95 (s.e. 0.18), a one-sided 95 percent bound of 0.66 percent a year. Three qualifications are in order. The observables do not span the targets. Put the country-effect machinery back into the gap regression — the country means of income, of each correlate and of the year dummies, the initial condition and the country random effect — and test the null that all of them are zero, which is the null that the constructed targets already capture the steady state. It is rejected decisively ($\chi^2(11) = 163$), so the regression conditions on a proxy for the steady state rather than on the steady state. What identifies it is largely cross-sectional — purged of year effects and the correlates, 81 percent of the variation in the gap regressor lies between countries, and the country-means regression alone gives −0.77 (s.e. 0.19) — so the objection of Section 4.2.3 to reading a cross-sectional slope as a convergence speed reaches it too. And its direction rests on an economic premise rather than on an estimate: the steady-state determinants the three correlates fail to measure stand higher in richer countries, in which case the omitted-variable formula fixes the direction of the bias: it pushes β toward zero, so the estimate understates the speed.

The bound survives the conditioning choices that would ordinarily be suspected. Investment is the classic worry (Bond et al., 2001): dropping it gives 0.88, replacing it by its previous-window value 0.96, lagging all three correlates 0.95, and anchoring income one year before the growth window 0.97. The point estimate stays between 0.88 and 0.99 across the six variants and the one-sided bound never falls below 0.61. Correcting the three correlates jointly for classical measurement error moves the estimate away from zero rather than toward it:[15] 1.14 at an assumed reliability of 0.90 and 1.58 at 0.80, against 0.95 uncorrected.

[15]The error variances are not identified here, so the reliability is imposed rather than estimated. For an assumed reliability ρ — the share of a correlate's variance that is signal — the quantity $(1-\rho)$ times its observed variance is subtracted from the diagonal of the moment matrix, on the within-year variation the estimator uses, and the equation is re-estimated; the diagnostics are those of Klepper and Leamer (1984). Income is treated as measured without error. On the annual panel, writing the deviation of log income from its country and year means as a first-order autoregression plus white noise, the variance and the first two autocovariances leave no room for the white-noise (classical) component: its estimated variance is below zero, and it stays below zero in every country-bootstrap resample.

The speed is not the same in every decade. Letting the gap slope vary by window rejects a constant value ($\chi^2(5) = 30.9$, $p < 10^{-4}$), and across the six windows it runs -0.75, -1.27, -0.31, -0.80, -1.39, -0.81. It has no trend. A speed that wanders like this cannot produce the one-time step in the between gradient of Section 4.2.

The bound holds under the premise just stated, and on that premise it is a finding in its own right. Countries do close the gap to their own moving target, at no less than 0.66 percent a year in slope units, estimated without demeaning (i.e. thus avoiding the Nickell bias). What the decade panel cannot supply is the level of that speed, and the within column of Table 1 cannot supply it either (Section 4.3).

4.5 Correlates and growth

The second front of the debate concerns the growth correlates. KWY find that the growth payoff to institutions and policies has flattened, and read that flattening as the reason absolute convergence caught up with conditional convergence; AM reply that the flattening is an artifact of the same omitted heterogeneity. For a correlate $x_{i,t}$ the encompassing model becomes

$$\begin{aligned} g_{i,t} \;=\; \beta \log y_{i,t} \;+\; \sum_d \lambda_d \, \mathbf{1}\{t \in d\} \, x_{i,t} \;+\; \sum_d \xi_d \, \bar{x}_i^d \;+\; \gamma \, \overline{\log y}_i \\ +\; \sum_t \phi_t \, \bar{D}_i^t \;+\; \delta \log y_{i,0_i} \;+\; u_i \;+\; \mu_t \;+\; \varepsilon_{i,t}, \end{aligned} \tag{7}$$

where $\bar{x}_i^d$ is the within-country mean of $\mathbf{1}\{t \in d\} \, x_{i,t}$ and $\bar{D}_i^t$ that of the dummy for start year t, both over the country's observed windows. With $\xi_d = \gamma = \phi_t = \delta = \sigma_u^2 = 0$ this is KWY's income-controlled growth regression; with the Mundlak terms free, λ_d is the within-country slope, equal to the two-way fixed-effects estimate (Wooldridge, 2019, Proposition 2.1), the quantity AM argue remains stable for democracy. Only the correlate slope carries the decade index, so that movements in λ_d are not absorbed into a co-moving income slope. Section A.8 reports the decade profiles and the decomposition below in full.

The pooled profiles reproduce KWY's flattening, and a flat profile is rejected for all eight correlates (Table 3). The Mundlak split adds two results. First, it sorts the correlates by margin. For education the association is between-country heterogeneity: its within slope is negative in every window and its Mundlak term strongly positive ($\xi = 1.01$, s.e. 0.40), while investment is weak on both margins. For the political-institutions measures, Polity 2 and the two Freedom House scores, the Mundlak term is small and insignificant, so the cross-country association adds nothing detectable to the within-country one, and the within slope moves: flatness is rejected for Freedom House political rights and civil liberties at $p < 0.001$ and for Polity 2 at $p = 0.006$, unadjusted, and $p = 0.046$ after a Bonferroni correction across the eight correlates (Section A.8). Heritage economic freedom is the exception among the institutional measures, with a significant Mundlak term ($\xi = -1.04$, s.e. 0.43). The two Worldwide Governance Indicators, observed in two windows only, do not reject flatness. Second, on the extended data the flattening is a trough centered on the 2000s, and the institutional slopes rebound in 2013–23.

Table 3: Correlate–growth slopes: sample sizes, the Mundlak coefficient ξ, the average of ξ_d over the correlate's windows (SE in parentheses), and p-values of Wald tests of a flat λ_d profile in the pooled and CRE specifications.

Correlate	N	ξ (Mundlak)	flat λ_d: p, pooled	flat λ_d: p, CRE
Polity 2	769	-0.15 (0.22)	0.002	0.006
FH political rights	718	-0.37 (0.22)	< 0.001	< 0.001
FH civil liberties	718	-0.21 (0.24)	< 0.001	< 0.001
WGI rule of law	319	-0.19 (0.70)	< 0.001	0.296
WGI gov. effectiveness	318	-0.06 (0.74)	< 0.001	0.360
Heritage econ. freedom	299	-1.04 (0.43)	< 0.001	0.012
Investment (% GDP)	698	-0.04 (0.23)	0.033	0.591
Barro–Lee education	733	1.01 (0.40)	< 0.001	0.078

The same window carries the retracement of absolute convergence after 2006. KWY link absolute and conditional convergence by the omitted-variable formula. In each window, cross-sectional regressions on a common sample give

$$b_d \;=\; b_d^* \;+\; \pi_d\,\theta_d, \tag{8}$$

where b_d is the absolute convergence slope, b_d^* the income slope conditional on the correlate, θ_d the cross-sectional correlate slope and π_d the correlate–income gradient. The identity is an accounting decomposition and attributes no causality. The conditioning gap $\pi_d\theta_d$ closes by 2003–13, as KWY find, and reopens to 0.10–0.96 for the institutional block in 2013–23, while conditional convergence stays strong (b_d^* between -0.31 and -1.23 across correlates). In this decomposition the reopening gap carries the whole weakening of absolute convergence in the pandemic window for the Heritage score and the two Worldwide Governance Indicators, 67 and 70 percent of it for the two Freedom House scores, and 26 percent for Polity 2, whose conditional slope itself weakens. Repeated on within variation, with country and year fixed effects, the same identity shows that adding the correlate moves the convergence slope by at most 0.34 across the eight correlates. The correlates channel therefore works through the cross-country association between institutions and income. Within a country, conditioning on its institutions barely moves the convergence slope.

5 Conclusion

KWY document a movement: across six decades the pooled income–growth slope turns from positive to negative. AM attribute that slope's level to omitted country effects; PS show that neither the pooled regression nor the fixed-effects estimator that replaces it recovers a convergence rate once countries are allowed trends that are not parallel. This paper accepts both criticisms. What it adds is that the movement survives them.

Nesting the two specifications as restrictions on one dynamic correlated-random-effects model separates the decade profile into a within-country slope and a between-country gradient. The between gradient holds near 1 through the first four windows, forty years, and then steps down to 0.12 in 2003–13. It is one level shift, not a gradual decline: the data reject a linear trend and do not reject a single break. No speed can be read off the within slope: at six decade transitions the finite-sample bias in the within slope is of the order of the estimate, and it is governed by the calendar span of the sample rather than by the sampling interval, so a finer grid over the same sixty years reduces it only toward a limit the span fixes. That is the same verdict PS reach analytically, by a different route and on a different design.

The movement is not an artifact of the biases that make the level untrustworthy. A null with constant steady states, in which no parameter varies by decade, manufactures 0.12 of the observed 1.30, both by least squares. A null built to PS's own design — a trended latent factor, loadings that differ across countries and correlate with income, and country-specific dynamics — produces a profile that rises rather than falls, and its dynamic heterogeneity makes the null less able to manufacture the decline. The configurations that reach the observed value at all describe samples whose cross-country income dispersion is flat or narrowing, and these data widened.

The mechanism the evidence supports is compositional, and it is the one Durlauf et al. (2005, p. 589) conjectured without claiming that it holds empirically. Poor countries grow faster because today's poor countries are disproportionately those carrying the deeper shortfalls from their steady states, and over the same decades they caught up on the growth correlates. In the early decades the open gaps sat at middle and high incomes, and the cross-section showed divergence; since the 1990s they belong to the poor, and it shows convergence. The same decomposition sorts the debate's second front: the growth association of education is between-country and investment is weak on both margins, while the Freedom House and Polity profiles move within country, dipping in the 2000s and recovering in the latest windows.

For anyone using a convergence rate, this narrows what the rate can claim. The speed at which a country closes its own gap is not identified at the frequency at which this debate is conducted. The only number offered for it is a lower bound: under the premise of Section 4.4, countries close the gap to their own moving target at no less than 0.66 percent a year. The trend KWY document is a statement about the changing cross-country distribution of steady states, and it survives the objections that dispose of the coefficient's level.

References

Acemoglu, D. and Molina, C. (2022). Comment on "Converging to Convergence". *NBER Macroeconomics Annual*, 36:425–442.

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, 127(1):47–100.

Allison, P. D. (2009). *Fixed Effects Regression Models*. Number 160 in Quantitative Applications in the Social Sciences. Sage, Thousand Oaks.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46–63.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2):277–297.

Asparouhov, T. and Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(1):119–142.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Barro, R. J. (2015). Convergence and modernisation. *The Economic Journal*, 125(585):911–942.

Barro, R. J. and Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy*, 100(2):223–251.

Bazzi, S. and Clemens, M. A. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics*, 5(2):152–186.

Bell, A., Fairbrother, M., and Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality & Quantity*, 53(2):1051–1074.

Bell, A. and Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, 3(1):133–153.

Bernard, A. B. and Durlauf, S. N. (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 71(1–2):161–173.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Bond, S., Hoeffler, A., and Temple, J. (2001). GMM estimation of empirical growth models. Discussion Paper 3048, CEPR.

Caselli, F., Esquivel, G., and Lefort, F. (1996). Reopening the convergence debate: A new look at cross-country growth empirics. *Journal of Economic Growth*, 1(3):363–389.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5–46.

DeLong, J. B. (1988). Productivity growth, convergence, and welfare: Comment. *American Economic Review*, 78(5):1138–1154.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65(6):1365–1387.

Durlauf, S. N., Johnson, P. A., and Temple, J. R. W. (2005). Growth econometrics. In Aghion, P. and Durlauf, S. N., editors, *Handbook of Economic Growth*, volume 1A, pages 555–677. Elsevier, Amsterdam.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–3182.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.

Hamaker, E. L. and Grasman, R. P. P. P. (2015). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, 5:1492.

Hauk, W. R. and Wacziarg, R. (2009). A Monte Carlo study of growth regressions. *Journal of Economic Growth*, 14(2):103–147.

Islam, N. (1995). Growth empirics: A panel data approach. *The Quarterly Journal of Economics*, 110(4):1127–1170.

Islam, N. (1998). Growth empirics: A panel data approach—a reply. *The Quarterly Journal of Economics*, 113(1):325–329.

Johnson, P. A. and Papageorgiou, C. (2020). What remains of cross-country convergence? *Journal of Economic Literature*, 58(1):129–175.

Klepper, S. and Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica*, 52(1):163–183.

Kremer, M., Willis, J., and You, Y. (2022). Converging to convergence. *NBER Macroeconomics Annual*, 36:337–412.

Lancaster, T. (2002). Orthogonal parameters and panel data. *Review of Economic Studies*, 69(3):647–666.

Lee, K., Pesaran, M. H., and Smith, R. P. (1997). Growth and convergence in a multi-country empirical stochastic Solow model. *Journal of Applied Econometrics*, 12(4):357–392.

Lee, K., Pesaran, M. H., and Smith, R. P. (1998). Growth empirics: A panel data approach—a comment. *The Quarterly Journal of Economics*, 113(1):319–323.

Moral-Benito, E. (2012). Determinants of economic growth: A Bayesian panel data approach. *Review of Economics and Statistics*, 94(2):566–579.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Patel, D., Sandefur, J., and Subramanian, A. (2021). The new era of unconditional convergence. *Journal of Development Economics*, 152:102687.

Perron, P. (1991). Test consistency with varying sampling frequency. *Econometric Theory*, 7(3):341–368.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. Cambridge Working Papers in Economics 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.

Pesaran, M. H. (2007). A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics*, 138(1):312–355.

Pesaran, M. H. and Smith, R. P. (2026). The output convergence debate revisited: Lessons from recent developments in the analysis of panel data models. arXiv preprint 2602.04060.

Pritchett, L. (1997). Divergence, big time. *Journal of Economic Perspectives*, 11(3):3–17.

Pustejovsky, J. E. and Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4):672–683.

Quah, D. (1993). Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics*, 95(4):427–443.

Rabe-Hesketh, S. and Skrondal, A. (2013). Avoiding biased versions of Wooldridge's simple solution to the initial conditions problem. *Economics Letters*, 120(2):346–349.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158.

Shiller, R. J. and Perron, P. (1985). Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters*, 18(4):381–386.

Smith, R. P. (2024). Econometric aspects of convergence: A survey. *Open Economies Review*, 35(4):701–721.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1):39–54.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2nd edition.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1):137–150.